# Biased Inference due to Prior Beliefs: Evidence from the Field

Martin Kapons and Peter Kelly*

October 31, 2022

**Abstract**

Prior-biased inference is a subset of confirmation bias - it suggests that agents update from observed signals in a way that favors their current beliefs. There is experimental evidence of prior-biased inference, but field evidence is much more limited. We provide evidence of prior-biased inference in three high-stakes field settings: the earnings forecasts of sell-side analysts, the macroeconomic forecasts of professional forecasters, and the pitch calls of Major League Baseball umpires. Our evidence is consistent with the idea that priors bias beliefs by influencing the constructs through which agents interpret the world.

*Kelly is at the University of Notre Dame; Kapons is at the University of Amsterdam. We thank Nicholas Barberis, Tony Cookson, Benjamin Golez, Ben Matthies, Sebastian Park, and seminar participants at the University of Notre Dame for helpful comments and feedback.

# 1 Introduction

Beliefs and preferences guide economic decisions. Therefore, it is critical for economists to understand how beliefs are formed and how beliefs deviate, if at all, from Bayesian updating. A large literature in psychology and economics examines deviations from correct reasoning (see Benjamin [2019] for a review). One important subset of this research area concerns how priors influence belief updating. Prior-biased inference captures the idea that people update from observed signals in a way that favors their current beliefs (Benjamin [2019]). It is a subset of confirmation bias, which captures the empirical tendency of people to interpret and seek out information that confirms their priors.[1] There is significant experimental evidence that priors bias how agents update beliefs, but there is very limited field evidence (Benjamin [2019]). Due to significant concerns about the generalizability of experimental results[2], it is critical to examine evidence from the field to determine real-world applicability. In this paper we provide evidence of prior-biased inference from three distinct high-stakes field settings.

We first consider sell-side analysts' earnings forecasts. To see whether prior-biased inference is important for earnings forecasts, we need to see if people's priors affect how they interpret earnings signals. The IBES (Institutional Brokers' Estimate System) data set is well suited for this analysis. IBES contains sell-side analysts' firm-level earnings forecasts and recommendations. We use the most recent recommendation of the analyst as a prior: those with a buy (sell) recommendation presumably expect the stock price to increase (decrease). We consider positive (negative) value-relevant news to be prior consistent if the analyst had a buy (sell) recommendation. We proxy for the direction of the news with forecast revisions - we consider the news to be good if the analyst updates his forecast in a positive direction and we consider the news to be bad if the analyst updates his forecast in

[1]Thus, how people update their beliefs is only one part of confirmation bias. Another example of confirmation bias would concern the *tendency* to update beliefs.

[2]Winkler and Murphy [1973] note that problem structures can differ significantly between the lab and field. There are also concerns that problems are framed differently in the field and in the lab (Tversky and Kahneman [1983]). Finally, there are concerns that cognitive processes like attention may be different in the field and the lab (Benjamin [2019]).

a negative direction. Prior-biased inference suggests that sell-side analysts will react more strongly to prior-consistent news compared to prior-inconsistent news. Therefore, we hypothesize that we will see a greater degree of overreaction to prior-consistent news compared to prior-inconsistent news.

We test this using Coibion and Gorodnichenko [2015] style regressions. Coibion and Gorodnichenko [2015] consider regressions of forecasts errors, or the actual minus the forecast, on the forecast revision. Negative coefficient estimates then suggest overreaction - the forecast revision was too large - and positive coefficient estimates suggest underreaction - the forecast revision was too small. We first follow Bordalo et al. [2019], which documents evidence of diagnostic expectations in these forecasts, and consider consensus forecasts. We construct a buy-consensus based on those analysts with buy recommendations and a sell-consensus based on those analysts with sell-recommendations. We find economically and statistically significant evidence that the consensus reacts more strongly to prior-consistent information. For example, when we consider forecast errors at a 4-year horizon (as in Bordalo et al. [2019]), and consider prior-consistent signals, we estimate the Coibion and Gorodnichenko [2015] coefficient to be $-0.600$. In contrast, we estimate the coefficient to be $-0.237$ when considering prior-inconsistent signals. The difference between the two coefficients is statistically significant at the one percent level.

In work about macroeconomic expectations, Bordalo et al. [2020b] suggest that individual-level Coibion and Gorodnichenko [2015] style regressions are a better measure of how forecasters actually react to signals. This motivates our analysis of analyst-level Coibion and Gorodnichenko [2015] style regressions. We consider pooled OLS regressions and analyst-specific regressions. In our analyst-specific regressions, we consider prior-consistent and prior-inconsistent samples. Again, for both estimation methods, we find evidence of greater overreaction to prior-consistent signals. For example, when we consider forecast errors at the 4-year horizon and prior-consistent signals, we find that the median Coibion and Gorodnichenko [2015] coefficient across all analysts is $-0.654$. In contrast, we find that the median coefficient equals $-0.285$ when considering prior-inconsistent signals. Again, the difference between the two coefficients is statistically significant at the one percent level.

These empirical results suggest that sell-side analysts react more strongly to prior-consistent signals relative to prior-inconsistent signals. There is evidence that sell-side analysts face incentives to give biased forecasts to curry favor with management (Lin and McNichols [1998]). This may lead them to overreact to positive information and underreact to negative information. Because most recommendations are positive, our results may be capturing these incentives. To address this concern, we consider two subsamples: (1) A sample of unaffiliated analysts, (2) a sample of only negative recommendations. In these samples, the incentives to curry favor with management should be very limited. However, we find similar, or stronger, results in these subsamples.

We do not know exactly what information analysts are digesting. However, it seems likely that revisions shortly after earnings announcements are in response to the information embedded in earnings reports. Therefore, to isolate how agents respond to specific information signals, we consider a subsample of earnings forecasts that are made shortly before the earnings announcement and revised shortly after the earnings announcement. In this sample, we again find that the degree of overreaction is much greater for prior-consistent forecast revisions relative to prior-inconsistent forecast revisions.

These results provide strong evidence that prior-biased inference exists in the field. We next provide evidence that these biases are strong enough to predict asset prices. When analysts observe an information signal about a firm's expected cash flows, they will revise their earnings forecast. However, given that their inference is biased by priors, the information needed to generate a revision of a given size should be stronger, on average, for prior-inconsistent signals compared to prior-consistent signals. This logic is similar to the logic presented in Kelly [2018], which argues that the information content of an insider sale at a loss is stronger than the information content of an insider sale at a gain since the psychological hurdle needed to cross to sell at a loss is greater than the psychological hurdle needed to cross to sell at a gain. To the extent that analyst forecast revisions contain new information, and the market is slow to understand these differential effects, we should find that prior-consistent revisions have less return predictability than prior-inconsistent revisions. We find evidence in line with this hypothesis. For example, at the six-month horizon, we find that prior-inconsistent negative forecast revisions predict a six-month return 316 basis points lower than prior-consistent negative forecast revisions (F-stat: 9.90). Similarly, we find that prior-inconsistent positive forecast revisions predict a six-month return 263 basis points higher than prior-consistent positive forecast revisions (F-stat: 4.94). These results suggest that prior-biased inference is important for our understanding of expectation formation in capital markets.

Our second setting concerns the macroeconomic forecasts of professional forecasters. We consider forecasts from the Survey of Professional Forecasters run by the Federal Reserve Bank of Philadelphia. One disadvantage of this setting is that macroeconomic forecasters do not have recommendations that we can use as a prior. We proxy for their priors by comparing those with past forecasts above the median value to those with past forecasts below the median value in any given quarter. For example, we assume that forecasters with above-median inflation forecasts in the last quarter are expecting high inflation and those with below median forecasts in the last quarter are expecting low inflation. This assumes that macroeconomic forecasters think in terms of categories and have taken a stance on inflation. We consider a revision to be prior-consistent if the forecast revision is positive (negative) and the forecaster had an above-median (below-median) forecast before the forecast revision. We find strong evidence of prior-biased inference. For example, across the 15 macroeconomic variables we consider, we find evidence of prior-biased inference in all 15 of them using pooled OLS regressions of individual-level forecasts. Additionally, the evidence is statistically significant at the 5-percent level in 12 of them. Another possible prior is that macroeconomic forecasters are biased by the specific number of their most recent forecast. This would likely lead to sticky expectations in both directions and no asymmetry. Our results suggest that prior-biased inference and categorical thinking are important considerations in how these professional forecasters form macroeconomic expectations.

Our final setting considers Major League Baseball (MLB) umpires' decisions.[3] A key part of an umpire's job is judging whether a pitch is thrown in the strike zone. Pitches deemed to be thrown inside (outside) the strike zone are called strikes (balls) and are considered a good (bad) outcome for the pitcher. We can judge the accuracy of the umpire's call by comparing the umpire's call with data from a system, PITCHf/x, that tracks the trajectory of pitched baseballs. A non-negligible fraction, over 10 percent, of pitches are called incorrectly by umpires. To address our research question, we examine whether umpire errors are related to their priors regarding the likelihood a pitcher will throw a strike. Unlike in the sell-side analyst setting, priors are not explicitly disclosed - we can only proxy for an umpire's prior. We assume that these priors emerge from the pitcher's historical tendency to throw strikes. Specifically, we consider *Percent Walks,* which proxies for the pitcher's tendency to throw balls in the previous season. We find that umpires are more likely to make incorrect calls that are favorable to the pitcher when the pitcher was successful in the previous season. For example, when we look at the set of all pitched balls, control for the distance from the strike zone, and compare pitchers within the same game, we find that a one-standard deviation decrease in *Percent Walks,* is associated with greater than a 0.3 percentage points increase in the fraction of balls that are called strikes (t=-8.17). When we restrict our sample to pitches that are close to the strike zone, we find even stronger results - we find that a one-standard deviation decrease in *Percent Walks,* is associated with about a 1.3 percentage points increase in the fraction of balls that are called strikes (t=-10.71). We consider a number of tests to show that our results are not driven by statistical discrimination or the Matthew effect (Kim and King [2014]), which captures the idea that those pitchers with high-status earn favorable treatment.

Lastly, we discuss the underlying frameworks that could drive these results. We consider

[3]This setting has also been considered for field studies by Parsons et al. [2011], Kim and King [2014] and Chen et al. [2016] for other judgment biases.

a number of frameworks related to motivated beliefs, bounded rationality, and local memory. We argue that motivated beliefs can explain our sell-side analyst and macroeconomic forecaster results, but are unlikely to explain our baseball results. Bounded rationality could explain our baseball results, but is unlikely to explain our sell-side analyst or macroeconomic forecaster findings. An explanation rooted in local memory can explain the results in all three settings. We argue that all frameworks are likely relevant for prior-biased inference, and we speculate on the conditions that could lead to greater importance for the different frameworks.

Most generally, we contribute to the literature on field evidence for judgment biases. Benjamin [2019] notes concerns about the generalizability of laboratory evidence and advocates field studies as a high priority. The paper notes there is limited field evidence for judgment biases outside of gambler's fallacy, the hot hand effect and base-rate neglect (e.g. Chen et al. [2016], Gilovich et al. [1985] and Green and Zwiebel [2018]). We contribute to this literature by offering field evidence of prior-biased inference. The field evidence for prior-biased inference that we are aware of relates to political priors. Analyzing data from an investor social platform, Cookson et al. [2020] show that likely Republicans' outlook on equities changed little during the COVID-19 crisis, while others became more pessimistic. Meeuwis et al. [2022] also offers evidence of prior-biased inference by examining investor portfolios: Likely Republicans increased the equity share and market beta of their portfolio after the 2016 presidential election while likely Democrats moved more into safe assets.[4] We provide evidence for prior-biased beliefs in three high-stakes field settings unrelated to political beliefs: sell-side analyst forecasts, macroeconomic forecasts, and Major League Baseball umpire calls. Our analysis differs in that (1) we show prior-biased inference in reaction to a large number of events, (2) we document evidence of prior-biased inference on both sides of beliefs - we show prior-biased inference for both those with negative recommendations (negative beliefs) and positive recommendations (positive beliefs), and (3) our results contribute

[4]Coibion et al. [2021] also offer support for this interpretation in their experimental/survey work on political polarization and expected economic outcomes.

to the discussion about the underlying mechanism by showing that prior-biased inference exists in a field setting where there is little reason to think that preferences drive the priors.

More specifically, our results contribute to macroeconomic expectations formation. There is a growing literature that documents deviations from full-information rational expectations (e.g. Coibion and Gorodnichenko [2015], Bordalo et al. [2020b], D'Acunto et al. [2021],Bianchi et al. [2022], Cassella et al. [2022b]) and studies the implications of these belief distortions in macroeconomic models (e.g. Maxted [2022]). This is important for our understanding of macroeconomic dynamics - we contribute to this literature by highlighting the importance of priors in macroeconomic expectation formation. Our results also contribute to a growing literature on expectations formation in financial markets. This setting is particularly important given the importance of investor expectations for financial markets (Klaus and Nagel [2022]). Traditionally, researchers in asset pricing assume rational expectations. While this assumption is analytically convenient, it is not supported by research on expectations formation (Benjamin [2019]). This motivates a move towards psychologically more realistic asset pricing frameworks. In order to develop these frameworks, it is critical to understand how financial market participants form expectations. A number of empirical papers use surveys to analyze how financial market participants form expectations. These papers highlight, among other things, the importance of personal experiences (e.g. Malmendier and Nagel [2011]) and extrapolation (e.g. Da et al. [2021]) in return expectations. Given the importance of cash flow expectations for asset prices (Chen et al. [2013], Bordalo et al. [2020a], Cassella et al. [2022a]), it is also important to understand how cash flow expectations are formed. Recent work shows that sticky expectations (Bouchaud et al. [2020]) and diagnostic expectations (Bordalo et al. [2019]) successfully explain patterns in earnings forecasts. We contribute to this growing literature by documenting the importance of an individual's priors on how they interpret earnings information. The financial research that most relates to our work concerns how agents seek out information, not how they interpret information. For example, Cookson et al. [2022] find that bulls and bears put themselves in echo chambers where they expose themselves to information that coincides with their prior. Additionally, Pouget et al. [2017] shows that analysts that hold positive (negative) views do not incorporate subsequent negative (positive) news in their forecasts. Hence, analysts fail to update to prior-inconsistent signals. In contrast, we examine how analysts update their forecasts *conditional* on updating. Hence, our work serves as a complement to Pouget et al. [2017] - analysts are not only biased in what information they react to, but also in how they react conditional on reacting.

## 2 Psychological review

In this section, we review the existing evidence on prior-biased beliefs. In their seminal experiment, Lord et al. [1979] recruited experimental subjects in favor of the death penalty and experimental subjects opposed to the death penalty. After reading a detailed account of a study on the death penalty, both opponents and proponents of the death penalty reported more extreme positions. The result has been replicated in a number of experimental settings (e.g. Fryer et al. [2019]) and is often interpreted as a bias because people updated their beliefs in opposite directions. However, it is possible that the behavior is rational. For example, if people have private information that does not relate to their initial assessment, but relates to their interpretation of evidence then we can see belief polarization among Bayesians (Benoit and Dubra [2019]). However, in some experimental designs, priors are randomly assigned (e.g. Darley and Gross [1983]), which eliminates concerns about non-common priors driving belief polarization.

There is also experimental evidence on the underlying mechanism behind prior-biased inference. Charness and Dave [2017] consider a sequential updating experiment. Subjects are told initial conditions that two states are equally likely. They are then shown signals that have a seventy percent chance of matching the correct state. After each signal, they record their subjective probability of each state. In this experiment, where subjects were incentivized for accuracy, subjects responded more strongly to prior-consistent information (confirming signals) than to prior-inconsistent information (disconfirming signals). Interestingly, when subjects were incentivized to favor a particular state - experimentally adjusting their preferences - the degree of prior-biased inference decreased. This is consistent with the idea that preferences do not drive prior-biased inference. Rather, it may be that people assess how consistent the signal is with their prior and do not consider inconsistent information (Fischhoff and Beyth-Marom [1983]). Relatedly, it is consistent with the idea that we exhibit representativeness based on what enters our local memory (Bordalo et al. [2019]), and ideas related to our priors are more likely to appear in our local memory. However, there is also evidence that preferences could drive prior-biased inference. Eil and Rao [2011] find that individuals respond more strongly to information that gives them pleasure (higher IQ or more beauty) even if this information does not conform with their prior. If agents have priors that coincide with their preferences, as the average subject did in this experiment, this may manifest as prior-biased inference. After presenting evidence of prior-biased inference in the field, we will discuss potential underlying mechanisms for our field evidence.

# 3 Data

## 3.1 Data sources and variable construction

Our first setting concerns sell-side analyst forecasts. We collect earnings forecasts and actuals from I/B/E/S (Institutional Brokers' Estimate System) unadjusted detail. We only consider long-term growth forecasts because (1) there is evidence that sell-side analysts may adjust their shorter-term forecasts in order to manipulate the consensus (e.g. Call et al. [2022]), (2) we are using recommendations to proxy for the prior. Recommendations are based on stock price expectations and long-term growth forecasts are more important for stock prices than short-term growth forecasts (Bordalo et al. [2020a]). We collect sell-side analyst recommendations from the I/B/E/S recommendations file. We assume analysts have a positive prior if the most recent IBES recommendation code before a revision is less than 3 (buy or strong buy), and we assume analysts have a negative prior if the most recent IBES recommendation code is greater than 3 (sell or strong sell). To adjust for stock splits, we use the cumulative factor to adjust shares from the CRSP daily file. We merge I/B/E/S data with CRSP data using the I/B/E/S-CRSP linking table provided by WRDS. Following Bordalo et al. [2019], our primary analysis compares long-term growth forecasts with realized outcomes. We calculate 3-year, 4-year and 5-year growth rates to compare with long-term growth forecasts. Long-term growth forecasts are an annualized percentage number. In order to compare the forecast with the actual, one needs to be able to compute a realized annual growth rate. This is difficult to do when the base is a negative earnings number. Bordalo et al. [2019] only consider firms with positive earnings. A lot of firms, however, have negative earnings. To expand our coverage, we follow Da and Warachka [2011] and use the absolute value in the denominator when calculating the annualized n-year growth rates for earnings $E$

$$EG_t(n) = ((\frac{E_{t+n} - E_t}{|E_t|} + 1)^{1/n} - 1) * 100.$$

We follow Bordalo et al. [2019] in our calculation of different consensus forecasts. We first determine the last forecast for each analyst-firm-year combination. We then calculate an average long-term growth forecast for those with buy recommendations and those with sell recommendations. We determine the consensus revision based on year over year changes. We drop observations where the absolute value of the revision is greater than 200 percentage points due to likely data errors. We also run individual-level regressions (analyst-by-analyst), where we consider all forecast revisions.

We consider some analyses of unaffiliated analysts. We identify affiliated analysts using a number of different data sets. We thank John Loudis for sharing a data set that connects broker identities (as listed in IBES) with banks from the SDC database since 1999. Broker identities in IBES are no longer reliable for data sets downloaded from WRDS. We thank

Jessie Watkins for providing an IBES data set downloaded in 2015 and John Loudis for providing a recommendation data set that runs through 2017. These data sets allow us to tie analysts to specific banks over the period 1999-2015. We identify an analyst as affiliated with a firm if their associated bank was connected to the firm's IPO in the past 5 years, the firm's SEO within the past 2 years, or was the lead underwriter on bond issuance in the past year.

We also examine some asset pricing implications. We draw return information from CRSP. For the return analysis, we consider firms that have a sell-side analyst revision in our sample. We consider ordinary common shares listed on the AMEX, NASDAQ, or NYSE and do not consider REITs, closed-end funds, ETFs, or Americus Trust Components. Following Shumway [1997], we replace missing delisting returns with a return equal to -0.3 for performance-related delistings. We construct a book-to-market control equal to the log value of common equity divided by market capitalization, where market capitalization is equal to the quarterly closing price times the number of common shares outstanding. We only consider observations that have a pre-log book-to-market ratio greater than 0 and less than or equal to 100. Size is calculated as the log value of market capitalization. We control for momentum by calculating the previous six-month return, excluding the most recent month.

Our second setting concerns macroeconomic forecasts. We collect individual forecast data from the survey of professional forecasters (SPF) which is conducted by the Federal Reserve Bank of Philadelphia. The forecast variables are quarterly macroeconomic outcomes (Nominal GDP, Real GDP, GDP price index, consumer price index (CPI), real consumption, industrial production, real nonresidential investment, real residential investment, real federal government consumption, real state and local government consumption, housing start, and unemployment), and financial information (three-month Treasury rate, ten-year Treasury rate, and AAA corporate bond rates). Hence, we look at the variables analyzed by Bordalo et al. [2020b] and follow their variable definition and forecast horizon. The forecast horizon is one year. For variables in levels, such as GDP, we transform the variables to growth rates and look at quarters t-1 to t+3. For relative variables, such as unmployment rates and interest rates, we look at the level in quarter t+3. Specific variable constructions can be found in Bordalo et al. [2020b]. For actual values of the macroeconomic outcomes and financial information, we use initial releases from the Philadelphia Fed's Real-Time Dataset for Macroeconomists (as in Coibion and Gorodnichenko [2015]). The results are similar when using the most recent release of actual outcomes.

The last quarter that we consider is the second quarter of 2022. The first quarter that we consider depends on when the respective variable has been included in the SPF. Nominal GDP, Real GDP, industrial production, GDP price index, housing start, and unemployment start in the last quarter of 1968. CPI, real consumption, real nonresidential investment, real residential investment, real federal government consumption, real state and local government consumption, three-month Treasury rate, and AAA corporate bond rates start in the third quarter of 1981. The ten-year Treasury rate starts in the first quarter of 1992.

Unlike for the sell-side analysts setting, we do not have a separate variable that measures the prior belief of the macroeconomic forecasters. We hence use the rank of the forecaster's forecast at quarter t-1 for quarter t+3 compared to all other forecasters that make a forecast in quarter t-1 for quarter t+3 as a proxy for the forecaster's prior. That is, the prior is measured with the forecast that is made right before the forecast revision. We consider forecasters that have a forecast at quarter t-1 that is above (below) the median forecast at quarter t-1 to have a positive (negative) prior.

Our third setting concerns Major League Baseball (MLB). We use the baseballr package in R in order to scrape baseball data from the website Baseball Savant (baseballsavant.mlb.com). In baseball, the pitcher's objective is to get outs or retire batters. Batters take turns standing at home plate and (potentially) swinging at pitches from the pitcher (i.e. swinging the bat to attempt hitting the pitched ball). A plate appearance (PA), or a completed turn batting for a batter, can end without a batter hitting the ball. If the batter gets 4 balls, he will walk (a good outcome for the batter) and if he gets 3 strikes, he will strike out (a bad outcome for the batter). When the batter does not swing, the umpire has to judge whether the pitch crossed the plate in the strike zone - the zone above home plate between the batter's knees and the midpoint of their torso. If the umpire judges the ball to have crossed the plate in the strike zone, it is called a strike; otherwise, it is called a ball. Unsurprisingly, umpires occasionally make incorrect assessments. The PITCHf/x system tracks the trajectory and location of every major league pitch. We use this data, which had its first full season in 2008, to determine the right call. We compare umpire's calls with this data to determine when umpires make mistakes. The data spans from 2008 to 2021.

We are, of course, interested in how priors affect umpire's decision making. We proxy for an umpire's prior about a pitcher's likelihood of throwing a strike based on the pitcher's tendency to walk people. If a pitcher had a tendency to walk batters in the previous year, we assume that the umpire will think it's more likely that the pitcher will throw a ball rather than a strike. Specifically, we consider the percentage of plate appearances that result in a walk for MLB pitchers with at least 100 PAs (plate appearances) in the previous year.

## 3.2 Summary statistics

The summary statistics for the different data used in the respective analyses can be found in Table 1. The summary statistics for analyst data are reported for the 4-year forecasting horizon and are pooled summary statistics. The magnitudes are comparable for the 3-year and 5-year horizon, respectively. For the analyst consensus data, the median forecast error is negative for all groups (prior-consistent/prior-inconsistent and the groups split by recommendation type and revision sign). The revision for prior-consistent observations is positive (median of 2.613) and negative for prior-inconsistent observations (median of -3.300). The split of the data by recommendation type and the sign of the revision shows that there are far fewer observations for sell recommendations (728 + 863 = 1,591) than for buy recommendations (8,714+ 11,654 = 20,368). The data on individual analyst observations shows that each analyst revises their forecast on average 2.382 times in each firm year.

The statistics for revisions and forecast errors by prior-consistent and prior-inconsistent observations are similar to those reported for the consensus data.

The summary statistics for the macroeconomic forecasters are average statistics across quarters. Forecast errors do not follow a systematic pattern and are most of the times not statistically different from zero. Mean (median) forecast revisions are on average negative for twelve (eleven) of the fifteen variables.

The baseball data has 3,503,950 observations; pooled summary statistics are reported. 11.7 percent of observations are incorrectly called and 8 percent are favorable to the pitcher. The Percent Walks of the prior year amounts to 8.54 percent, on average. The Distance from the Strike Zone amounts to 0.612, on average.

# 4 Empirical analysis

In this section, we test how priors impact the interpretation of information. We first study analyst forecasts - we measure the level of overreaction by analyst prior and signal type. Using Coibion and Gorodnichenko [2015] style regressions, we find evidence of more overreaction to signals that confirm priors than to signals that do not confirm priors. We also use Coibion and Gorodnichenko [2015] style regressions to document prior-biased inference in macroeconomic forecasts. Overreaction is stronger for signals that are in line with priors. We then present results related to MLB umpires. We provide evidence that umpires are likely influenced by their priors - they are more likely to give favorable calls to pitchers with a successful track record.

## 4.1 Sell-side analysts

### 4.1.1 Prior-consistent consensus reactions

In this subsection, we analyze how analysts update long-term growth forecasts in response to recommendation-consistent, or prior-consistent, signals. We consider a signal prior-consistent if the analyst increases (decreases) his long-term growth forecast for a firm and his most recent stock recommendation for that firm was positive (negative). We hypothesize that we will find a greater degree of overreaction to prior-consistent signals, or information. We measure the degree of overreaction in two subsamples: a prior-consistent subsample and a prior-inconsistent subsample. We first consider forecast revisions at the consensus level. We follow Bordalo et al. [2019] in our calculation of different consensus forecasts. We first determine the last forecast for each analyst-firm-year combination. We then calculate an average long-term growth forecast for those with buy recommendations and those with sell recommendations. The revision is based on year over year changes in the buy consensus or year over year changes in the sell consensus. We measure how analysts react using Coibion and Gorodnichenko [2015] style regressions. Specifically, we estimate regressions of the following form, and cluster standard errors by year:

$$FE = \alpha * Revision + \gamma_t * Year_t. \tag{1}$$

Forecast errors, $FE$, are defined as equal to the actual minus the forecast. The $\alpha$ coefficient captures how agents react to information. Full-information rational expectations suggest $\alpha = 0$. If $\alpha > 0$, this suggests that the forecaster underreacts to the information because positive and negative forecast revisions are insufficient. Similar logic implies that $\alpha < 0$ indicates overreaction. We present the results in Table 2. The first column presents prior-consistent results (e.g. when the buy consensus forecast revision is positive) and the second column presents prior-inconsistent results (e.g. when the buy consensus forecast revision is negative). We find much higher levels of overreaction when the signal is consistent with the prior. For example, when we consider forecast errors at a 4-year horizon (as in Bordalo et al. [2019]), and consider prior-consistent signals, we estimate $\alpha = -0.600$. In contrast, we estimate $\alpha = -0.237$ when considering prior-inconsistent signals. The difference between these two coefficient estimates is statistically significant at the 1-percent level. This result suggests that priors influence how agents interpret financial market information.

We also consider a unified regression framework to test the importance of prior-biased inference. Specifically, we estimate an equation of the following form for firm i in year t:

$$FE_{i,t+x} = \alpha_1 * PC_{i,t} + \alpha_2 * Revision_{i,t} + \alpha_3 * (Revision_{i,t} * PC_{i,t}) + \gamma_t * Year_t + \epsilon_{i,t+x}. \quad (2)$$

*PC* is a dummy that equals one if the forecast revision is prior consistent. We cluster standard errors by year. We predict that $\alpha_3 < 0$, since this would suggest greater overreaction for prior-consistent news. We present the results in Table 3 for a 4-year horizon.[5] We estimate $\alpha_3 = -0.385$ (t=-5.72). This again suggests greater overreaction to prior-consistent news. We also assess how our results change with uncertainty. We consider three proxies for uncertainty. The first is size - in column 2 (3) of Table 3, we present the results from estimating equation 2 for the bottom (top) size quintile. Size is proxied by market capitalization. In the bottom size quintile (the one that proxies for high uncertainty), we estimate $\alpha_3 = -0.577$ (t=-4.07). For the top quintile, the point estimate is positive and not statistically different from zero. Our second proxy is age - measured as the time since the firm entered the CRSP data set. Again, we find evidence that the results are greater with greater uncertainty. In the bottom age quintile, we estimate $\alpha_3 = -0.409$ (t=-3.14). For the top quintile, the point estimate is positive and not statistically different from zero. Our final proxy for uncertainty is analyst forecast dispersion - measured by the standard deviation of long-term growth forecasts. For this analysis, we ignore observations where there is only one analyst forecast. When we estimate equation 2 for the top quintile, $\alpha_3 = -0.411$ (t=-5.34). For the bottom quintile, the point estimate is positive and not statistically different from zero. To summarize, this subsection documents strong evidence of prior-biased inference among sell-side analysts that increases with uncertainty.

[5] The results are stronger at the 5-year horizon and weaker at the 3-year horizon.

### 4.1.2 Individual-level analysis

In the previous subsection, we presented results from using changes in the consensus forecast. Bordalo et al. [2020b] suggest that individual-level Coibion and Gorodnichenko [2015] data is more informative about how forecasters actually react to signals. Hence, we use individual-level data in the two ways suggested by Bordalo et al. [2020b]. First, we use pooled OLS regressions for individual j and firm i:

$$FE_{i,j,t+x} = \delta_1 * PC_{i,j,t} + \delta_2 * Revision_{i,j,t} + \delta_3 * (Revision_{i,j,t} * PC_{i,j,t}) + \gamma_i * Firm_i + \gamma_t * Year_t + \epsilon_{i,t+x}. \tag{3}$$

Standard errors are clustered by analyst. Second, we run regressions of equation 1 analyst-by-analyst. For each individual, we consider prior-consistent samples and prior-inconsistent samples. We require at least 10 observations for each individual to estimate the regression. We estimate Coibion and Gorodnichenko [2015] style regressions for each individual on a prior-consistent sample and a prior-inconsistent sample.

We present the coefficient estimates for the different groups in Table 4. For the pooled OLS regressions, the column for prior-consistent estimates reports $\delta_2 + \delta_3$. Again, we see evidence of more overreaction when the signal is prior consistent than when the signal is prior inconsistent. For example, for pooled OLS regressions, when we consider forecast errors at the 4-year horizon and prior-consistent signals, $\delta_2 + \delta_3 = -0.823$, while $\delta_2 = -0.151$. The difference is statistically significant at the 1-percent level. Also, for analyst-by-analyst regressions, when we consider forecast errors at the 4-year horizon and prior-consistent signals, we find that the median $\alpha = -0.654$. In contrast, we find that the median $\alpha = -0.286$ when considering prior-inconsistent signals. The difference is statistically significant at the 1-percent level when using median regressions with robust standard errors to test the difference. These results suggest that sell-side analysts react more strongly to prior-consistent information.

### 4.1.3 Updating around earnings announcements

We have provided evidence that analysts update differently depending on whether the information is consistent with their prior. One interesting question concerns whether analysts react differently because of the information that captures their attention, or because the same information is interpreted differently based on their priors. Given that this is not an experimental setting, it is difficult to completely control what draws analysts' attention.[6] However, we can comment on this distinction by examining forecast revisions after earnings announcements. In these windows, it seems reasonable to assume that the information from the earnings announcement is largely responsible for most analysts' forecast revisions. Additionally, we can proxy for whether the information is consistent with the forecast revision by only considering forecast revisions that are consistent with the analyst-specific earnings surprise. To test this, we consider a sample of forecast revisions that are consistent with the prior and have an earnings surprise consistent with the prior and compare them to forecasts revisions that are inconsistent with the prior and also have an earnings surprise that is inconsistent with the prior. We estimate a regression of the following form:

$$FE_{i,t+x} = \delta_1 * BothPC_{i,t} + \delta_2 * Revision_{i,t} + \delta_3 * (Revision_{i,t} * BothPC_{i,t}) + \gamma_i * Firm_i + \gamma_t * Year_t + \epsilon_{i,t+x}, \tag{4}$$

where $BothEC$ is a dummy that equals one if the forecast revision and the earnings surprise were both prior-consistent. Standard errors are clustered by analyst. We present the results in Table 5. When we consider forecast errors at the 4-year horizon, we estimate $\delta_3 = -0.629$ (t=-1.93) and the results are even stronger at the 3-year and 5-year horizons. This suggests that analysts react much more strongly in response to earnings announcements that are consistent with the recommendation compared to earnings announcements that are inconsistent with the recommendation. Our results suggest that, holding the information

[6] In the baseball setting, what the agent (the umpire) pays attention to is largely controlled by the nature of the job.

constant, analysts respond differently depending on their priors.

### 4.1.4 Unaffiliated analysts

A number of papers highlight that sell-side analysts may adjust their forecasts to curry favor with management (e.g. Lin and McNichols [1998] note that analysts may bias their forecasts and recommendations in order for their employer to receive more underwriting business). It is possible that those with buy recommendations overreact (underreact) to positive (negative) signals to curry favor with management. We address this concern in two ways. First, in this subsection, we will show that our results from the last subsection hold for the subset of unaffiliated analysts. In the next subsection, we will show that those with negative recommendations - who are presumably not biased by these incentives - overreact more to negative news than to positive news.

We present our results of pooled OLS estimates, as specified in equation 3, for unaffiliated analysts in Table 6. We find very similar results compared to when considering all analysts. For example, when we consider forecast errors at the 4-year horizon and prior-consistent signals, we find that$\delta_2 + \delta_3 = -0.851$, while $\delta_2 = -0.222$. The difference is statistically significant at the 1-percent level.

### 4.1.5 Recommendation-specific analysis

In prior sections, we provided evidence that analysts react more strongly to prior-consistent information. We now test whether this holds in both positive (buy) recommendation and negative (sell) recommendation subsamples. This is important because analysts with buy recommendations may be currying favor with management. Hence, if our results hold in the subsample with sell recommendations, they are unlikely driven by analyst incentives to curry favor with management. In total, we consider four different subsamples: positive signals with positive recommendations, negative signals with positive recommendations, positive signals with negative recommendations, and negative signals with negative recommendations. We hypothesize that we will find a greater degree of overreaction to positive signals than negative signals when the most recent recommendation is positive and a greater degree of overreaction to negative signals than positive signals when the most recent recommendation is negative. Again, we measure the degree of overreaction following Equation 1.

We present the results for the consensus forecasts in Table 7 and the results for the individual level data in Table 8. We find evidence that our results hold in both the positive recommendation sample and the negative recommendation sample. The first column of Table 7 presents the results when the most recent recommendation was positive and the second column presents results when the most recent recommendation was negative. For buy recommendations, the results suggest that analysts overreact more to positive signals than negative signals, consistent with prior-biased beliefs. Specifically, when considering 4-year and 5-year forecast errors, we find much higher levels of overreaction to positive signals than negative signals. And, when considering forecast errors at the 3-year horizon, we see similar levels of overreaction for positive signals and negative signals. For sell recommendations, the results are especially striking. For negative signals, we see strong levels of overreaction at all horizons. For positive signals, we even see evidence of *underreaction*. Table 8 reports the results for pooled OLS regressions using the individual level data. The results are similar to those reported in Table 7. The primary difference is that there is no underreaction to positive signals for sell recommendations. These results suggest that prior-biased beliefs are a robust feature of sell-side analyst data, and are likely not driven by analyst incentives.

## 4.2 Asset pricing implications

We provided evidence that sell-side analysts overreact more to prior-consistent signals than to prior-inconsistent signals. This suggests that for a revision of a given positive size, it is more informative if it comes from an analyst, whose most recent recommendation was negative. Similarly, a revision of a given negative size should be more informative if it comes from an analyst, whose most recent recommendation was positive. The logic is similar to the logic in Kelly [2018], which shows that company insider sales at a loss are a more negative signal for future returns than company insider sales at a gain. This is attributed to the fact that the information shock, on average, should be greater to overcome the psychological hurdle needed to realize a loss. In our setting, we hypothesize that the information shock needed to overcome prior biases must be greater if the shock is prior-inconsistent than if the shock is prior-consistent. To the extent that sell-side analyst revisions capture information shocks, this should manifest in returns.

We test this conjecture with future monthly (and future six month) returns after the forecast revision. We run pooled OLS regressions with month fixed effects to test this hypothesis. Specifically, we estimate regressions of the following form for firm i in month t:

$$Return_{i,t\rightarrow t+x} = \beta_1 BuyNeg_{i,t} + \beta_2 SellNeg_{i,t} + \beta_3 BuyPos_{i,t} + \beta_4 SellPos_{i,t} + \Omega Controls + \gamma Month_t + \epsilon_{i,t}, \quad (5)$$

where $BuyNeg_{i,t}$ is a dummy that equals one if there was a sell-side analyst, whose most recent recommendation for firm i was positive and who decreased his long-term growth forecast for firm i by at least five percent in month $t$, $SellNeg_{i,t}$ is a dummy that equals one if there was a sell-side analyst, whose most recent recommendation for firm i was negative and who decreased his long-term growth forecast for firm i by at least five percent in month $t$, $BuyPos_{i,t}$ is a dummy that equals one if there was a sell-side analyst, whose most recent recommendation for firm i was positive and who increased his long-term growth forecast for firm i by at least five percent in month $t$, and $SellPos_{i,t}$ is a dummy that equals one if there was a sell-side analyst, whose most recent recommendation for firm i was negative and who increased his long-term growth forecast for firm i by at least five percent in month $t$. In all regressions, we include controls for the mean forecast revision in the month (zero if there was no forecast revision) and a dummy that equals one if there was no forecast revisions. We cluster standard errors in two dimensions by month and by firm. We predict that $\beta_1 < \beta_2$

and $\beta_3 < \beta_4$. These orderings suggest that the information content of a revision is stronger if it is a prior-inconsistent signal than a prior consistent signal. We consider long-term horizons because the forecast revisions concern long-term growth, it may take time for these information signals to manifest, and the market may underreact to the initial signal.

We present the results at a one-month and a six-month horizon in Table 9. At the one-month horizon, the difference between $\beta_1$and $\beta_2$ is about -42 basis points (F-stat: 0.92) and the difference between $\beta_3$ and $\beta_4$ is about -61 basis points (F-stat: 3.47). This difference is consistent with our hypothesis, but the difference is either marginally significant or not statistically significant. When we add controls for short-term reversals, momentum, value and size, the difference shrinks, which suggests that analysts may be picking up information from some well-known return predictors. At the six-month horizon, the difference between $\beta_1$ and $\beta_2$ is about -316 basis points (F-stat: 9.90) and the difference between $\beta_3$ and $\beta_4$ is about -263 basis points (F-stat: 4.94). These differences are economically and statistically significant. Again, when we add controls for short-term reversals, momentum, value and size, these differences shrink. These results suggest that analysts require stronger information signals to make prior-inconsistent revisions and this manifests in future returns.

## 4.3 Macroeconomic forecasts

In this subsection, we examine whether macroeconomic forecasters are influenced by their priors. We consider a signal to be prior-consistent if the forecast revision is positive (negative) and the last quarter's individual level forecast of quarter t+3 is above (below) the median of all forecasts in the last quarter. As for the sell-side analysts, we consider Coibion and Gorodnichenko [2015] style regressions for individual forecaster data. We run pooled OLS regressions of equation 3 and cluster standard errors by time and by forecaster. We predict that the coefficient on *PC*Revision* is negative. We also run regressions of equation 1 forecaster-by-forecaster, on a subsample with prior-consistent observations and on a subsample with prior-inconsistent observations. Both estimation methods follow Bordalo et al.

[2020b]. We require at least 10 observations for each individual forecaster to estimate the forecaster-by-forecaster regression. We expect the median coefficient of the prior-consistent subsample to be negative and smaller than the median coefficient of the prior-inconsistent subsample.

Table 10 reports the regression estimates; Figure 1 shows the regression estimates graphically. For pooled OLS regressions, the coefficient on *PC*Revision* is negative for all fifteen variables that the macroeconomic forecasters provide predictions for and statistically significant for twelve out of the fifteen variables at the five percent significance level.[7] The median coefficient estimates of the forecast-by-forecaster regressions provide similar inferences. Median coefficients of the prior-consistent sub-sample are negative and smaller than the median coefficients of the prior-inconsistent sub-sample for thirteen of the fifteen variables. Eight of those thirteen median coefficient differences are statistically significant at the five percent significance level when testing for significance using median regressions with robust standard errors. These results provide evidence that macroeconomic forecasters are biased by their priors. Our work builds on the findings from Bordalo et al. [2020b], which documents that macroeconomic forecasters overreact to signals. We show that overreaction is concentrated in forecasters whose forecast revision is consistent with their prior.

## 4.4 Umpire's calls

In this subsection, we examine whether umpires exhibit prior-biased inference. As they are constantly around the game, we assume that umpires are aware of pitchers' successes. Therefore, we proxy for an umpire's prior about a pitcher with the pitcher's tendency to throw balls in the previous season. Specifically, we consider *Percent Walks*, a variable that equals the fraction of plate appearances that resulted in a walk - or 4 balls - for the pitcher in the previous year. We identify incorrect calls by comparing the umpire's call - a ball or a

[7] The high coefficient on *Revision* for the variable GDP price index is consistent with Bordalo et al. [2020b]. They report a high coefficient for Coibion and Gorodnichenko [2015] style regressions - without splitting the sample into prior-consistent and prior-inconsistent sub-samples.

strike - with what a machine - PITCHf/x - would have called the pitch. If the two disagree, we deem the umpire call incorrect. If the machine would've called the pitch a ball, but the umpire called the pitch a strike, we refer to this umpire call as *Favorable to Pitcher.* We first estimate regressions of the following form for all *incorrect* calls from 2008 to 2021:

$$\text{Favorable to Pitcher}=\beta_0 + \beta_1\text{Percent Walks}+ \gamma\text{Game}+\epsilon. \tag{6}$$

We include game fixed effects to distinguish between how historically good pitchers' pitches are evaluated relative to their less successful counterparts' pitches within the same game. This is important if umpire tendencies change from game to game. We present the results in Table 11. We cluster standard errors by pitcher to account for correlated errors within pitcher across time, and we also cluster standard errors by game. We estimate $\beta_1 = -0.67$ (t=-11.00). This means that the pitcher's historical tendency to throw walks, which we use as a proxy for the umpire's priors, is negatively correlated with how often he'll receive favorable calls. Specifically, a one-standard deviation decrease in *Percent Walks* is associated with almost a 2 percentage points increase in the fraction of incorrect calls that will be favorable to the pitcher. We interpret this to mean that if the umpire has positive (negative) priors about the pitcher, he is more likely to give the pitcher favorable (unfavorable) calls.[8] It is possible that this result is driven by the way the catcher - or the person catching the balls that the pitcher throws - frames the pitch. Namely, it is possible that the umpire isn't biased by his priors, but rather he is influenced by the way the catcher catches the pitch. In the second column of Table 11, we include catcher fixed effects to account for a catcher's framing ability. In this specification, we estimate $\beta_1 = -0.63$ (t=-10.49). This suggests that catchers' framing ability is not the driver of our results.

Another possibility is that pitchers with a successful track record are more likely to pitch balls on the edge of the strike zone. To address this concern, we consider another specification

[8]We can make a similar inference if we replace *Percent Walks* with the pitcher's ERA, a commonly used measure of pitcher ability but less direct for our purposes, from the previous year.

that allows us to control for the distance from the strike zone. Specifically, we consider a sample of all pitches that the machine thinks should've been called a ball. We then estimate the following regression for all such pitches from 2008 to 2021:

$$\text{Favorable to Pitcher}{=}\beta_0 + \beta_1\text{Percent Walks}{+}\beta_2\text{Distance to Strike Zone}{+}\ \gamma\text{Game}{+}\epsilon. \quad (7)$$

Again, we present the results in Table 11 and cluster standard errors by pitcher and game. Again, we find results consistent with the idea that umpires' priors bias their calls. Specifically, we estimate $\beta_1 = -0.13$ (t=-8.48). This suggests that a one standard deviation increase in *Percent Walks* is associated with over a 0.3 percentage point increase in the fraction of balls that will be incorrectly called strikes.[9] We include catcher fixed effects in column 4 and find a very similar point estimate.

One possible explanation for this result is that umpires receive a signal, and then rationally update in a way that maximizing their chances of calling the pitch correctly. Consider a parallel from a traffic cop. He rationally knows that men are more likely to speed than women. He is supposed to pull over any car driving over 75 miles per hour. His radar gun shows him 75 if the speed is measured between 74.6 and 75.4. He thinks that, since men are more likely to speed, they are more likely to be speeding when clocked at 75. Therefore, he maximizes his chances of giving appropriate tickets by being more likely to pull over men clocked at 75 miles per hour than women clocked at 75 miles per hour. Now, suppose we're considering a baseball pitcher. Pitchers with successful track records are more likely to throw strikes than throw balls. The umpire then considers it rational to call a truly uncertain pitch a strike rather than a ball if it's thrown by a pitcher with a good track record. This argument would rely on the assumption that a pitch just inside the strike zone is indistinguishable from a pitch just outside the strike zone. We view this as unlikely. Even within a narrow range of the strike zone border, umpires' pitch calls are much more accurate than chance. Therefore,

[9] This is about a 3 percent increase of the fraction of all pitches outside the strike zone called incorrectly.

umpires seem to pick up an important signal about the balls location even close to the border. Additionally, we show in column 1 of Table 12 that those pitchers with worse pitching records are statistically no less likely to throw strikes within 0.25 units of the border of the strike zone[10]. This suggests that even if these pitches were completely uncertain, favoring successful pitchers would not be a dominant strategy. However, as we show in columns 2 and 3 of Table 12, even within 0.25 units of the strike zone, pitches are much more likely to be called favorably for those with better pitching records. Specifically, a one standard deviation increase in *Percent Walks* is associated with over a 1 percentage point increase in the fraction of balls that will incorrectly be called strikes. In our final column, we add dummy variables for the pitch count since the umpire may favor the pitcher when the count is in the batter's favor (Moskowitz and Wertheim [2011]). After adding these dummy variables, we find even stronger results.

Our findings are very related to Kim and King [2014], which also finds evidence that successful pitchers get more favorable calls. However, their interpretation of this result is different. They argue that individuals are biased to positively evaluate high-status individuals irrespective of quality (the Matthew effect). In Table 13, we provide evidence against this conjecture. Kim and King [2014] use All-star designations as a measure of status. This designation is likely connected to umpire priors and, therefore, not a good statistic for distinguishing between prior-biased inference and the Matthew effect. We, however, believe that batting average is a good statistic for this purpose.

If status offers players a more favorable strike zone, then we should see that higher-status batters get more favorable calls. We consider batters with at least 100 plate appearances in the previous year. We use *Batting average* as a statistic for our setting as it is likely not strongly related to priors, but is related to status. Namely, those with higher batting averages in the previous year likely have higher status, but it is not immediately obvious how

[10]This is for pitches where the batter does not swing. Outside of this narrow zone, historical success is correlated with the tendency to throw pitches in the strike zone. We consider a narrow range of pitches within 0.25 units of the border of the strike zone. For context, this is the 25th percentile for distance from the strike zone for pitches recorded outside the strike zone by Pitchf/x.

batting average would impact umpire priors. In column 1 of Table 13 we show that batting averages *positively* predicts the likelihood of a favorable call for the pitcher when the pitch is a narrow range outside of the strike zone. Of course, All-star status is an elite designation while batting average is a continuous variable. Therefore, we also consider a dummy *Strong Batter* that equals one if the batter had a batting average greater than or equal to .300 in the previous season, an elite benchmark. In column 2 of Table 13, we show that there is little evidence of a relationship between this dummy and the likelihood of a favorable call for the pitcher.

It is likely that MLB umpires were made aware of the results from Kim and King [2014] - there was an article about this study in the New York Times in 2014. Therefore, it is interesting to see whether umpires changed the way they treated all-stars after becoming aware of this study. In column 3 of of Table 13, we consider data from before 2014. Like Kim and King [2014], we find evidence that all-star pitchers receive favorable treatment after controlling for their walk percentage. However, if we restrict the sample to pitches after 2014, we find evidence that all-star pitchers receive *unfavorable* treatment after controlling for walk percentage.[11] This is consistent with the idea that umpires can change their tendencies in the presence of feedback.

# 5 Discussion of the Underlying Mechanism

The primary contribution of our paper is to show that priors bias inference in high-stakes field settings. However, the underlying mechanism behind *why* priors bias inference is also interesting. In this section, we discuss three potential drivers of prior-biased inference and how well they fit with our evidence.

We first consider motivated reasoning. Under motivated reasoning, people deviate from rational expectations to make themselves happier. For example, a future college graduate

[11] Without this control, the point estimate for all-star pitcher is still negative, but is no longer statistically significant.

may derive happiness from thinking it will not rain on his graduation day (anticipation utility) - this may lead him to react differently to information depending on whether it suggests rain or no rain on his graduation day. For sell-side analysts, it seems plausible that they would gain happiness from their recommendation proving right. This may lead to biased updating based on new information in a way that reflects their recommendation. A similar argument would also apply to macroeconomic forecasters - to the extent that they give categorical recommendations as part of their work. However, motivated reasoning seems less likely to influence umpire calls. First, it is not obvious why umpires would derive utility from successful pitchers throwing more strikes. Major League Baseball cares deeply about the integrity of the game. Therefore, umpires should not have intrinisic preferences for certain players to do better. Furthermore, umpires are graded on the accuracy of their calls - they are highly incentivized to provide accurate calls. This suggests that any happiness would be tied to the accuracy of their calls. Finally, umpires make a judgment about information that has already transpired and are given near immediate feedback - fans and players may indicate displeasure immediately and umpires receive reports about their performance shortly after the game ends. For these reasons, anticipation utility should not play a major role in their decisions. This suggests that, since we find evidence of prior-biased decision-making among umpires, preferences are unlikely to be the *sole* driver of prior-biased inference.

Second, we consider an explanation rooted in bounded rationality. Umpires will - rationally - think that the likelihood a successful pitcher throws a strike is higher than the likelihood an unsuccessful pitcher throws a strike. They will then view a pitch in a location X. They view the pitch location with noise - they know that there is a chance it is a ball and a chance it is a strike. They do not recognize that - conditional on the perceived pitch location - the likelihood of a strike is independent of the pitcher's track record. Therefore, they decide to combine the two pieces of information to determine the likelihood of a strike. This leads them to favor pitchers with a successful track record. This explanation, however, seems unlikely to explain analysts behavior or professional forecaster's behavior. We find evidence that they overreact to prior-consistent information. For example, analysts with a buy recommendation may expect long-term growth of 8 percent and see a signal which suggests a long-term growth of 10-percent. If analysts combined these two signals, we would expect a long-term growth forecast somewhere between 8 and 10 percent. However, we find evidence of overreaction to prior-consistent signals.

Lastly, we consider an explanation rooted in local memory, or the memories that help us perform cognitive tasks. It seems likely that our priors influence what is in our local memory - we are more likely to have available memories that are consistent with our priors than available memories that are inconsistent with our priors. Therefore, we will be biased towards finding information to be representative of our priors. This will lead us to overreact to information that is prior-consistent and underreact to information that is prior-inconsistent. Specifically, our results are consistent with a diagnostic expectations framework where the degree of representativeness depends on whether the information is prior consistent or prior inconsistent (see also Charness and Dave [2017]). Consider the example of diagnostic expectations in earnings from Bordalo et al. [2019]: The paper notes that Google had a much higher frequency of positive earnings reports than other firms. Therefore, if a firm reports very good earnings, it is likely that the firm is more representative of being the next Google than if a firm reports subpar earnings. According to Bordalo et al. [2019], this suggests that when a firm reports good earnings, Google's history is more likely to come to mind (i.e. to the local memory). As such, people will think the firm is more likely to be the next Google than is rational. Our earnings forecast results suggest that Google is even more likely to come to mind if the sell-side analyst had positive priors, and is less likely to come to mind if the analyst had negative priors. A nearly identical logic applies to macroeconomic forecasts and a similar logic can be applied in our baseball setting. Suppose that the umpire perceives the ball to be pitched in a location X. He views the true location with noise. Conditional on viewing pitch location X, without diagnostic expectations, he knows that the true location is in the strike zone 60 percent of the time. That is, 60 percent of the time he perceives pitch location X, it will be because the underlying conditions (e.g. the sun, exactly where he's standing behind home plate) lead him to view a true strike in that location. The remainder of the time he perceives pitch location X, the underlying conditions lead him to view this pitch location as a ball. With diagnostic expectations, the umpire will overstate the probability that the pitch is a strike given that the perceived pitch location is highly diagnostic of being a strike. Our results suggest that the umpire's reaction will depend on his priors. If he thinks the pitcher is likely to throw strikes, it is more likely that the umpire will overreact to the likelihood of conditions that would lead true strikes to have a perceived pitch location X and will underreact to the likelihood of conditions that would lead true balls to have a perceived pitch location X. The intuition behind this result is that if an umpire has strong positive priors about the pitchers tendency to throw strikes, then the good pitcher's pitches are more likely to remind him of strikes.

Our results are most consistent with a diagnostic expectations framework where the degree of representativeness depends on whether the information is prior consistent or prior inconsistent. This suggests that priors bias our inferences by influencing the constructs through which we view the world. Differences in local memory can lead to different constructs through which we interpret the world. For example, suppose that we have a prior that it will be sunny on our wedding day. This will make it more likely that our local memory will be more full of sunny days rather than rainy days. Then, it would be likely that a few clouds will not remind us of a day that turned to be full of clouds and rainy, but will rather remind us of a day that continued to have limited cloud coverage and lots of sunlight. We view this explanation as very plausible, but we do not think it captures the entire picture of why priors bias beliefs. We also believe that motivated beliefs and bounded rationality will play a role. We conjecture that the importance of motivated reasoning in prior-biased inference increases as the happiness you get from your prior increases. Priors will align strongly with preferences in many settings, and we encourage researchers to not discount the influence of motivated reasoning in those settings. We also think that bounded rationality will be a driver behind prior-biased inference in a setting in which agents are strongly incentivized to make accurate decisions and have repeated feedback.

# 6 Conclusion

We heed the call from Benjamin [2019] for more field evidence of judgment biases: we provide evidence of prior-biased inference in multiple high-stakes field settings. We document evidence of prior-biased inference among sell-side analysts: sell-side analysts exhibit greater overreaction to prior-consistent signals compared to prior-inconsistent signals. This result has implications for how investors form cash flow expectations and should be considered in future asset pricing models. We also provide evidence of prior-biased inference in macroeconomic forecasts - this has implications for macroeconomic models and central bank policies. For example, our results suggest that it may be difficult to change agents' inflation expectations if information signals do not coincide with priors. Finally, we show that our results extend to MLB umpires - a setting where the information received is naturally controlled.

These results contribute to our understanding of how agents form beliefs outside of experimental laboratories. Our results suggest that prior-biased inference is a robust feature of human behavior - this has implications for most fields within economics.

## References


Daniel Benjamin. *Handbook of behavioral economics: Errors in probabilistic reasoning and judgment biases*. Elsevier Press, 2019.

Jean-Pierre Benoit and Juan Dubra. Apparent bias: What does attitude polarization show? *International Economic Review*, 60(4):1675–1703, 2019.

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

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

Pedro Bordalo, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. Expectations of fundamentals and stock market puzzles. *NBER Working Paper*, 2020a.

Pedro Bordalo, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–2782, 2020b.

Jean-Philippe Bouchaud, Philipp Krueger, Augustin Landier, and David Thesmar. Sticky expectations and the profitability anomaly. *Journal of Finance*, 74(2):639–674, 2020.

Andrew Call, Nathan Sharp, and Mason Snow. Bearish analysts and the issuance of difficult-to-beat earnings forecasts. *Working Paper*, 2022.

Stefano Cassella, Benjamin Golez, Huseyin Gulen, and Peter Kelly. Horizon bias and the term structure of equity returns. *The Review of Financial Studies*, 2022a.

Stefano Cassella, Benjamin Golez, Huseyin Gulen, and Peter Kelly. Motivated beliefs in macroeconomic expectations. *Working paper*, 2022b.

Gary Charness and Chetan Dave. Confirmation bias with motivated beliefs. *Games and Economic Behavior*, 104:1–23, 2017.

Daniel Chen, Tobias Moskowitz, and Kelly Shue. Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *The Quarterly Journal of Economics*, 131(3):1181–1242, 2016.

Long Chen, Zhi Da, and Xinlei Zhao. What drives stock price movements? *The Review of Financial Studies*, 26(4):841–876, 2013.

Olivier Coibion and Yuriy Gorodnichenko. Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8): 2644–2678, 2015.

Olivier Coibion, Yuriy Gorodnichenko, and Michael Weber. Political polarization and expected economic outcomes. *Working paper*, 2021.

Tony Cookson, Joey Engelberg, and William Mullins. Does partisanship shape investor beliefs? evidence from the covid-19 pandemic. *The Review of Asset Pricing Studies*, 10 (4):863–893, 2020.

Tony Cookson, Joey Engelberg, and William Mullins. Echo chambers. *The Review of Financial Studies*, 2022.

Zhi Da and Mitch Warachka. The disparity between long-term and short-term forecasted earnings growth. *Journal of Financial Economics*, 100(2):424–442, 2011.

Zhi Da, Xing Huang, and Lawrence J Jin. Extrapolative beliefs in the cross-section: What can we learn from the crowds? *Journal of Financial Economics*, 140(1):175–196, 2021.

Francesco D'Acunto, Ulrike Malmendier, Juan Ospina, and Michael Weber. Exposure to grocery prices and inflation expectations. *Journal of Political Economy*, 129(5):1615–1639, 2021.

John Darley and Paget Gross. A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1):20–33, 1983.

David Eil and Justin Rao. The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2): 114–138, 2011.

Baruch Fischhoff and Ruth Beyth-Marom. Subjective cash flow and discount rate expectations. *Psychological Review*, 90(3):239–260, 1983.

Roland Fryer, Philipp Harms, and Matthew Jackson. Updating beliefs when evidence is open to interpretation: Implications for bias and polarization. *Journal of the European Economic Association*, 17(5):1470–1501, 2019.

Thomas Gilovich, Robert Vallone, and Amos Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17(3):295–314, 1985.

Brett Green and Jeffrey Zwiebel. The hot-hand fallacy: Cognitive mistakes or equilibrium adjustments? evidence from major league baseball. *Management Science*, 64(11):5315–5348, 2018.

Peter Kelly. The information content of realized losses. *The Review of Financial Studies*, 31 (7):2468–2498, 2018.

Jerry Kim and Brayden King. Seeing stars: Matthew effects and status bias in major league baseball umpiring. *Management Science*, 60(11):2619–2644, 2014.

Adam Klaus and Stefan Nagel. Expectations data in asset pricing. *NBER Working Paper*, 2022.

Hsiou-wei Lin and Maureen McNichols. Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics*, 25(1): 101–127, 1998.

Charles Lord, Lee Ross, and Mark Lepper. Biased assimilation and attitude polarization:

The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11):2098–2109, 1979.

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

Peter Maxted. A macro-finance model with sentiment. *Review of Economic Studies*, 2022.

Maarten Meeuwis, Jonathan Parker, Antoinette Schoar, and Duncan Simester. Belief diagreement and portfolio choice. *Journal of Finance*, 2022.

Tobias Moskowitz and Jon Wertheim. *Scorecasting: The hidden influences behind how sports are played and games are won.* Crown Archetype, 2011.

Christopher Parsons, Johan Sulaeman, Michael C. Yates, and Daniel Hamermesh. Strike three: Discrimination, incentives, and evaluation. *American Economic Review*, 101(4): 1410–1435, 2011.

Sebastien Pouget, Julien Sauvagnat, and Stephane Villeneuve. A mind is a terrible thing to change: confirmatory bias in financial markets. *The Review of Financial Studies*, 30(6): 2066–2109, 2017.

Tyler Shumway. The delisting bias in crsp data. *The Journal of Finance*, 52(1):327–340, 1997.

Amos Tversky and Daniel Kahneman. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315, 1983.

Robert Winkler and Allan Murphy. Experiments in the laboratory and the real world. *Organizational Behavior and Human Performance*, 10(2):252–270, 1973.

Figure 1: Coibion-Gorodnichenko (2015) Regression Estimates for Macroeconomic variables
This figure presents evidence of prior-biased inference among macroeconomic forecasters. Figure 1A reports coefficient estimates from equation 3 where PC reports $\delta_2 + \delta_3$ and PIC reports $\delta_2$. Figure 1B reports the median coefficient estimate from forecaster-by-forecaster regressions for the prior-consistent subsample and prior-inconsistent subsample respectively. The dependent variable is the forecast-error for quarter t+3. We consider a signal to be prior-consistent (PC) if the forecast revision is positive (negative) and the last quarter's individual level forecast of quarter t+3 is above (below) the median of all forecasts in the last quarter. The macroeconomic variables are Nominal GDP (ngdp), Real GDP (rgdp), GDP price index (pgdp), consumer price index (cpi), real consumption (rconsum), industrial production (indprod), real nonresidential investment (rnresin), real residential investment (rresinv), real federal government consumption (rfedgov), real state and local government consumption (rslgov), housing start (housing), unemployment (unemp), three-month Treasury rate (tbill), ten-year Treasury rate (tbond), and AAA corporate bond rates (bond).

**Figure 1A: Pooled OLS**


**Figure 1B: Forecaster-by-forecaster regressions**

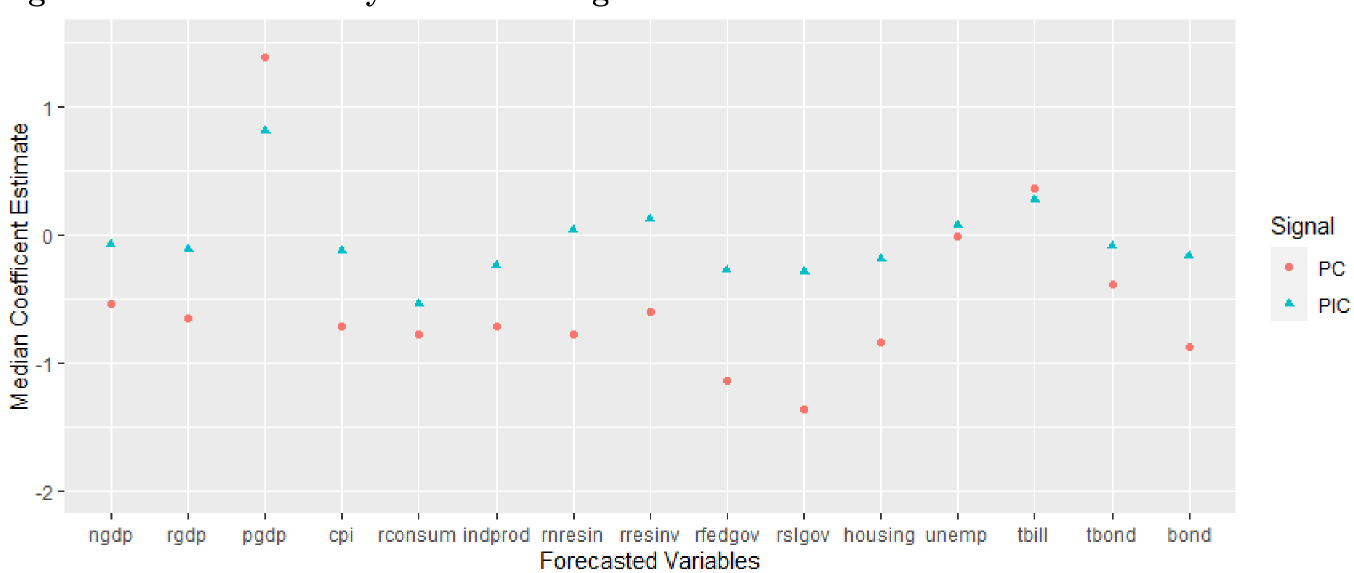

Table 1: Summary statistics

This table presents the mean, the median, the standard deviation, and the number of observations for variables we consider in our analysis. For the macroeconomic forecasts, we provide the mean of the quarterly statistics across quarters. For sell-side analyst variables: *FE4* is the forecast error for the 4-year horizon. The forecast error is the difference between the realized, actual earnings growth and the forecasted earnings growth. *Revision* is the difference of the current analyst earnings growth forecast and the last analyst earnings growth forecast. For macroeconomic forecast variables: Forecast error is the difference between the realized value and the forecasted value. Realized values are initial values provided by the real time data of the Survey of Professional Forecasters. Forecast revision is the difference between the forecast made for quarter t+3 at time quarter t and time quarter t-1 For Major League Baseball variables: *Percent Walks* is the percent of PA (plate appearances) for the pitcher that result in a walk. It is measured on a pitcher-year basis. *Incorrect Call* is a dummy variable that equals one if the pitch was called incorrectly. *Favorable to Pitcher* is a dummy variable that equals one if the pitch was incorrectly called a strike instead of a ball. *Distance from Strike Zone* is a measure of how far the pitch was from the strike zone. We include all pitches that the machine measured as being outside the strike zone in this summary.

| | Mean | Median | SD | N |
|---|---|---|---|---|
| **Analyst (consensus) variables** | | | | |
| Prior-consistent observations | | | | |
| *Revision (FE4 sample)* | 5.475 | 2.613 | 14.780 | 9,577 |
| *FE4* | -13.025 | -13.615 | 43.037 | 9,577 |
| Prior-inconsistent observations | | | | |
| *Revision (FE4 sample)* | -6.155 | -3.300 | 13.626 | 12,382 |
| *FE4* | -7.815 | -10.043 | 40.134 | 12,382 |
| Buy recommendation, positive revision | | | | |
| *Revision (FE4 sample)* | 6.900 | 3.000 | 13.818 | 8,714 |
| *FE4* | -14.284 | -14.728 | 43.172 | 8,714 |
| Sell recommendation, positive revision | | | | |
| *Revision (FE4 sample)* | 9.147 | 3.722 | 17.167 | 728 |
| *FE4* | -3.397 | -4.585 | 58.432 | 728 |
| Buy recommendation, negative revision | | | | |

| | | | | |
|---|---|---|---|---|
| *Revision (FE4 sample)* | -7.111 | -3.667 | 12.781 | 11,654 |
| *FE4* | -8.091 | -10.304 | 38.692 | 11,654 |
| Sell recommendation, negative revision | | | | |
| *Revision (FE4 sample)* | -8.916 | -4.000 | 16.392 | 863 |
| *FE4* | -0.313 | -3.578 | 39.481 | 863 |
| **Analyst (individual) variables** | | | | |
| *Number of revisions in analyst-firm-year* | 2.382 | 2.000 | 2.088 | 63,375 |
| Prior-consistent observations | | | | |
| *Revision (FE4 sample)* | 3.053 | 1.300 | 10.518 | 36,611 |
| *FE4* | -10.834 | -9.706 | 38.528 | 36,611 |
| Prior-inconsistent observations | | | | |
| *Revision (FE4 sample)* | -3.701 | -2.000 | 10.512 | 43,526 |
| *FE4* | -7.986 | -7.890 | 35.505 | 43,526 |
| **Macroeconomic forecasts** | | | | |
| Forecast errors | | | | |
| *Nominal GDP* | 0.003 | 0.002 | 0.012 | 6,153 |
| *Real GDP* | 0.048 | 0.047 | 0.010 | 6,167 |
| *GDP price index* | -0.029 | -0.029 | 0.008 | 6,138 |
| *CPI* | 0.048 | 0.047 | 0.010 | 6,165 |
| *Real consumption* | 0.053 | 0.053 | 0.008 | 4,407 |
| *Industrial production* | -0.037 | -0.038 | 0.018 | 5,802 |
| *Real nonresidential investment* | 0.044 | 0.044 | 0.027 | 4,293 |
| *Real residential investment* | 0.072 | 0.073 | 0.047 | 4,283 |
| *Real federal gov. consumption* | 0.066 | 0.067 | 0.024 | 4,137 |
| *Real state and local gov. consumption* | 0.059 | 0.059 | 0.012 | 4,159 |

| | | | | |
|---|---|---|---|---|
| *Housing start* | -0.023 | -0.020 | 0.093 | 5,897 |
| *Unemployment* | 0.027 | 0.036 | 0.348 | 6,280 |
| *Three-month Treasury rate* | -0.528 | -0.510 | 0.454 | 2,545 |
| *Ten-year Treasury rate* | -0.485 | -0.465 | 0.392 | 2,171 |
| *AAA coporate bond rate* | -0.524 | -0.501 | 0.534 | 3,692 |
| Forecast revisions | | | | |
| *Nominal GDP* | -0.001 | -0.001 | 0.011 | 6,153 |
| *Real GDP* | -0.002 | -0.002 | 0.009 | 6,167 |
| *GDP price index* | 0.000 | 0.000 | 0.007 | 6,138 |
| *CPI* | -0.002 | -0.002 | 0.009 | 6,165 |
| *Real consumption* | -0.001 | -0.001 | 0.008 | 4,407 |
| *Industrial production* | -0.003 | -0.003 | 0.017 | 5,802 |
| *Real nonresidential investment* | -0.003 | -0.003 | 0.023 | 4,293 |
| *Real residential investment* | -0.006 | -0.005 | 0.040 | 4,283 |
| *Real federal gov. consumption* | 0.001 | 0.000 | 0.020 | 4,137 |
| *Real state and local gov. consumption* | -0.001 | 0.000 | 0.011 | 4,159 |
| *Housing start* | -0.026 | -0.024 | 0.087 | 5,897 |
| *Unemployment* | 0.055 | 0.062 | 0.331 | 6,280 |
| *Three-month Treasury rate* | -0.198 | -0.176 | 0.437 | 2,545 |
| *Ten-year Treasury rate* | -0.134 | -0.126 | 0.356 | 2,171 |
| *AAA coporate bond rate* | -0.114 | -0.111 | 0.490 | 3,692 |

**Umpire variables**

| | | | | |
|---|---|---|---|---|
| *Incorrect Call* | 0.117 | 0 | 0.321 | 3,503,950 |
| *Favorable to Pitcher* | 0.080 | 0 | 0.272 | 3,503,950 |
| *Percent Walks* | 8.54 | 8.3 | 2.82 | 5,265 |

| | | | | |
|---|---|---|---|---|
| *Distance from Strike Zone* | 0.612 | 0.51 | 0.480 | 2,469,836 |

Table 2: Prior-consistent Analysis Consensus-level

This table presents coefficient estimates from Coibion and Gorodnichenko [2015] style regressions run at the consensus-level with year fixed effects. The column labels indicate whether the signal was prior consistent or prior inconsistent. We consider a signal prior-consistent if the buy (sell) consensus long-term growth forecast increases (decreases). The row labels indicate the horizon used for the actual in the forecast error. We cluster standard by year. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | Prior-consistent | Prior-inconsistent |
|---|---|---|
| 3 Years | -0.488*** | -0.361*** |
| 4 Years | -0.600*** | -0.237*** |
| 5 Years | -0.635*** | -0.208*** |

Table 3: Prior-consistent Analysis Consensus-level

This table presents evidence of prior-biased inference that seems to increase with uncertainty. The dependent variable is the forecast-error at the 4-year horizon. The column headers indicate the subsample considered. PC is a dummy that equals one if the signal was prior-consistent. We consider a signal prior-consistent if the buy (sell) consensus long-term growth forecast increases (decreases). Age is measured as the time since the firm first appeared in CRSP. Dispersion is the standard deviation of the forecasts. We ignore observations where there is only one analyst forecast. All regressions include year fixed effects. The row labels indicate the horizon used for the actual in the forecast error and indicate whether the revision was positive or negative. We cluster standard errors by year and put t-statistics in parentheses. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | Full Sample | Bot Quintile Size | Top Quintile Size | Bot Quintile Age | Top Quintile Age | Bot Quintile Dispersion | Top Quintile Dispersion |
|---|---|---|---|---|---|---|---|
| PC | 0.184 | -0.443 | 0.201 | -0.750 | 0.236 | 3.966** | -4.088* |
| | (0.19) | (-0.21) | (0.19) | (-0.34) | (0.19) | (2.29) | (-1.74) |
| Revision | -0.229*** | -0.296** | -0.333** | -0.229** | -0.479*** | -0.941*** | -0.103* |
| | (-4.09) | (-2.72) | (-2.23) | (-2.13) | (-3.02) | (-2.85) | (-1.92) |
| PC*Revision | -0.385*** | -0.577*** | 0.110 | -0.409*** | 0.120 | 0.323 | -0.411*** |
| | (-5.72) | (-4.07) | (0.36) | (-3.14) | (0.39) | (0.85) | (-5.34) |
| N | 22,519 | 2,879 | 4,763 | 3,728 | 3,825 | 4,061 | 3,228 |

Table 4: Prior-consistent Analysis Individual-level

This table presents coefficient estimates from a pooled OLS regression of forecast errors on forecast revisions intereacted with a dummy for prior-consistent observations and includes firm fixed effects and year fixed effects (Panel A), and median coefficient estimates from Coibion and Gorodnichenko [2015] style regressions run at the individual-level - analyst-by-analyst (Panel B). We present the median coefficient estimates across all individuals for each group. We require at least 10 observations for each individual regression. The column labels indicate whether the signal was prior consistent or prior inconsistent. We consider a signal prior-consistent if the analyst increases (decreases) his long-term growth forecast for a firm and his most recent recommendation for that firm was positive (decreases). The row labels indicate the horizon used for the actual in the forecast error. For the pooled OLS regression, we cluster standard errors by analyst. For the individual-level regressions, we test significance using median regressions and robust standard errors. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Standard errors for the pooled OLS regressions are clustered by analyst.

**Panel A: Pooled OLS regressions**

| | Prior-consistent | Prior-inconsistent |
|---|---|---|
| 3 Years | -0.827*** | -0.137*** |
| 4 Years | -0.823*** | -0.151*** |
| 5 Years | -0.902*** | -0.078 |

**Panel B: Individual-level regressions**

| | Prior-consistent | Prior-inconsistent |
|---|---|---|
| 3 Years | -0.704*** | -0.269*** |
| 4 Years | -0.654*** | -0.286*** |
| 5 Years | -0.742*** | -0.208*** |

Table 5: Prior-consistent Analysis around Earnings

This table presents evidence of prior-biased inference around earnings announcements. We consider revisions where a long-term growth forecast was made the quarter before an earnings announcement, and revised within 20-days after that earnings announcement. The dependent variable is the forecast-error, and the column-header indicates the horizon considered. BothPC is a dummy that equals one if the revision was consistent with the earnings surprise, and the analyst-specific earnings surprise was also consistent with the prior. We consider a revision prior-consistent if the analyst increases (decreases) his long-term growth forecast for a firm and his most recent recommendation for that firm was positive (negative). We consider an analyst-specific earnings surprise to be prior-consistent if the actual is greater (less) than the analyst's quarterly forecast and the most recent recommendation was positive (negative). All regressions include year fixed effects and firm fixed effects. We cluster standard errors by analyst and put t-statistics in parentheses. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | 3-year horizon | 4-year horizon | 5-year horizon |
|---|---|---|---|
| BothPC | -6.983** | -4.517 | -1.618 |
| | (-2.13) | (-2.28) | (-1.08) |
| Revision | -0.029 | -0.187 | 0.090 |
| | (-0.17) | (-1.44) | (0.34) |
| BothPC*Revision | -0.753** | -0.629* | -1.419*** |
| | (-2.59) | (-1.93) | (-4.09) |
| N | 4,577 | 3,966 | 3,369 |

Table 6: Unaffiliated Prior-consistent Samples Individual-level

This table only considers unaffiliated analysts. We present coefficient estimates from Coibion and Gorodnichenko [2015] style regressions estimated with pooled OLS regressions. The column labels indicate whether the signal was prior consistent or prior inconsistent. We consider a signal prior-consistent if the analyst increases (decreases) his long-term growth forecast for a firm and his most recent recommendation for that firm was positive (decreases). The row labels indicate the horizon used for the actual in the forecast error. Standard errors are clustered by analyst. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | Prior-consistent | Prior-inconsistent |
|---|---|---|
| 3 Years | -0.797*** | -0.300*** |
| 4 Years | -0.851*** | -0.222*** |
| 5 Years | -0.988*** | -0.110 |

Table 7: Coibion-Gorodnichenko (2015) Regressions for EPS

Each entry in the table corresponds to the estimated coefficient of regressing consensus forecast errors on consensus forecasts revisions with year fixed effects. The column labels indicate the type of consensus. Buy (Sell) indicates that it is the consensus from analysts with Buy (Sell) recommendations. The row labels indicate the horizon used for the actual in the forecast error and indicate whether the revision was positive or negative. Standard errors are clustered by year. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | Buy | Sell |
|---|---|---|
| 3 Years - Positive Signal | −0.446*** | 0.456 |
| 3 Years - Negative Signal | −0.492*** | −0.989*** |
| 4 Years - Positive Signal | −0.587*** | 0.235 |
| 4 Years - Negative Signal | −0.327*** | −0.988*** |
| 5 Years - Positive Signal | −0.643*** | 0.221 |
| 5 Years - Negative Signal | −0.281*** | −0.947*** |

Table 8: Coibion-Gorodnichenko (2015) Regressions for EPS

Each entry in the table corresponds to the estimated coefficient of regressing consensus forecast errors on consensus forecasts revisions with firm fixed effects. The column labels indicate the type of consensus. Buy (Sell) indicates that it is the consensus from analysts with Buy (Sell) recommendations. The row labels indicate the horizon used for the actual in the forecast error and indicate whether the revision was positive or negative. Standard errors are clustered by analyst. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | Buy | Sell |
|---|---|---|
| 3 Years - Positive Signal | $-0.749$*** | -0.006 |
| 3 Years - Negative Signal | $-0.132$** | $-0.773$*** |
| 4 Years - Positive Signal | $-0.755$*** | -0.149 |
| 4 Years - Negative Signal | $-0.162$*** | $-0.810$*** |
| 5 Years - Positive Signal | $-0.839$*** | $-0.037$ |
| 5 Years - Negative Signal | $-0.090$* | $-0.906$*** |

Table 9: Return Predictability

This table presents the relationship between returns and analyst behavior. We run pooled OLS regressions with month fixed effects for all columns. We consider all firms with a single sell-side analyst revision and we consider data starting in 1993. The column headers indicate the future return horizon being predicted. *Buy - Negative Revision* is a dummy that equals one if a sell-side analyst, whose most recent recommendation was positive, made a negative revision to his long-term growth forecasts of at least five percent. *Sell - Negative Revision* is a dummy that equals one if a sell-side analyst, whose most recent recommendation was negative, made a negative revision to his long-term growth forecasts of at least five percent. *Buy - Positive Revision* is a dummy that equals one if a sell-side analyst, whose most recent recommendation was positive, made a positive revision to his long-term growth forecasts of at least five percent. *Sell - Positive Revision* is a dummy that equals one if a sell-side analyst, whose most recent recommendation was negative, made a positive revision to his long-term growth forecasts of at least five percent. We include a control for the average forecast revision in the month, and a dummy control that equals one if there was no forecast revision. We include extra controls when noted. These controls are the previous month return, the previous six-month return skipping the most recent month, the log value of the book-to-market ratio, and the log value of the market capitalization. F-statistic 1 tests the difference between the coefficient estimates on the negative revision dummies. F-statistic 2 tests the difference between the coefficient estimates on the positive revision dummies. We multiply coefficient estimates by 100. We cluster standard errors in two dimensions, by permno and by month. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | One month | One month | Six months | Six months |
|---|---|---|---|---|
| Buy - Negative Revision | -0.27 | -0.37** | -1.54*** | -2.00*** |
| | (-1.52) | (-2.20) | (-3.34) | (-4.52) |
| Sell - Negative Revision | 0.15 | -0.03 | 1.62 | 0.79 |
| | (0.34) | (-0.06) | (1.65) | (0.80) |
| Buy - Positive Revision | -0.03 | -0.00 | 0.53 | -0.02 |
| | (-0.17) | (-0.01) | (0.91) | (-0.03) |
| Sell- Positive Revision | 0.58* | 0.52 | 3.16** | 1.73 |
| | (1.66) | (1.48) | (2.60) | (1.48) |
| Extra Controls | No | Yes | No | Yes |
| F-statistic 1 | 0.92 | 0.57 | 9.90*** | 7.66*** |
| F-statistic 2 | 3.47* | 2.56 | 4.94** | 2.29 |

Table 10: Coibion-Gorodnichenko (2015) Regressions for Macroeconomic variables

This table presents evidence of prior-biased inference in macroeconomic forecasters. The dependent variable is the forecast-error for quarter t+3. PC is a dummy that equals one if the signal was prior-consistent. We consider a signal to be prior-consistent if the forecast revision is positive (negative) and the last quarter's individual level forecast of quarter t+3 is above (below) the median of all forecasts in the last quarter. For the pooled OLS regressions, the standard errors are clustered by time and by forecaster. The forecaster-by-forecaster regressions report the median coefficient estimate. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Statistical significance between prior-inconsistent and prior-consistent median coefficient estimates is tested using median regressions with robust standard errors. Yes/no indicates statistical significance at the five percent significance level.

| | Pooled OLS regressions | | | Forecaster-by-forecaster regressions | | |
|---|---|---|---|---|---|---|
| | PC | Revision | PC*Revision | Prior-inconsistent | Prior-consistent | Statistically sign. |
| Nominal GDP | 0.000 | −0.095 | −0.603*** | −0.074 | −0.540 | Yes |
| Real GDP | 0.008* | 0.135 | −0.833** | −0.110 | −0.653 | No |
| GDP price index | −0.001 | 0.510 | −0.151 | 0.808 | 1.379 | No |
| CPI | 0.008* | 0.142 | −0.857** | −0.121 | −0.714 | No |
| Real consumption | −0.004*** | −0.309 | −0.573* | −0.542 | −0.774 | No |
| Industrial production | 0.001 | −0.328 | −0.804*** | −0.243 | −0.715 | Yes |
| Real nonresidential investment | −0.006 | −0.030 | −0.678*** | 0.033 | −0.780 | Yes |
| Real residential investment | −0.021 | 0.401* | −0.912*** | 0.125 | −0.597 | Yes |
| Real federal gov. consumption | 0.013*** | −0.605** | −1.278*** | −0.276 | −1.143 | No |
| Real state and local gov. consumption | 0.005 | −0.286 | −0.325 | −0.287 | −1.364 | Yes |
| Housing start | 0.003 | −0.085 | −0.834*** | −0.191 | −0.841 | Yes |
| Unemployment | 0.043 | −0.082 | −0.341*** | 0.071 | −0.017 | No |
| Three-month Treasury rate | −0.015 | 0.349*** | −0.354*** | 0.269 | 0.367 | No |
| Ten-year Treasury rate | 0.002 | −0.004 | −0.649*** | −0.084 | −0.393 | Yes |
| AAA corporate bond rate | 0.028 | −0.014 | −0.980*** | −0.164 | −0.877 | Yes |

Table 11: Umpires' Calls

This table documents how umpires' calls change with their priors. We approximate the extent to which the umpire expects a pitcher to throw strikes with *Percent Walks*, a variable that equals the fraction of plate appearances that resulted in a walk for the pitcher in the previous year. In the first two columns, the sample is the set of all pitches that resulted in a mistake by the umpire. In the third and fourth columns, the sample consists of all pitches outside of the strike zone. *Favorable to Pitcher* is the dependent variable - it is a dummy that equals one if the pitch should've been called a ball, but was called a strike. *Distance from Strike Zone* is a control that equals the distance from the strike zone. We use catcher fixed effects where noted. We multiply coefficient estimates by 100. Standard errors are clustered by game and by pitcher. We put t-statistics in parentheses. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | *Favorable to Pitcher* | *Favorable to Pitcher* | *Favorable to Pitcher* | *Favorable to Pitcher* |
|---|---|---|---|---|
| *Percent Walks* | -0.67*** | -0.63*** | −0.13*** | −0.12*** |
| | (-11.00) | (-10.49) | (-8.48) | (-8.17) |
| *Distance from Strike Zone* | | | −22.96*** | −22.98*** |
| | | | (-121.07) | (-120.85) |
| Game Fixed Effects | Yes | Yes | Yes | Yes |
| Catcher Fixed Effects | No | Yes | No | Yes |
| N | 408,055 | 408,028 | 2,469,827 | 2,469,825 |

Table 12: Umpires' Calls - Additional Tests

This table also documents how umpire's calls change with their priors. We approximate the extent to which the umpire expects a pitcher to throw strikes with *Percent Walks*, a variable that equals the fraction of plate appearances that resulted in a walk for the pitcher in the previous year. The first column consists of all pitches within 0.25 units of the strike zone border. The last three columns consist of pitches outside of the strike zone, but less than 0.25 units away. *Favorable to Pitcher* is the dependent variable in most regressions - it is a dummy that equals one if the pitch should've been called a ball, but was called a strike. *Strike* is a dummy that equals one if the pitch should've been called a strike. *Distance from Strike Zone* is a control that equals the distance from the strike zone. We use catcher fixed effects where noted. We multiply coefficient estimates by 100. Standard errors are clustered by game and by pitcher. We put t-statistics in parentheses. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | *Strike* | *Favorable to Pitcher* | *Favorable to Pitcher* | *Favorable to Pitcher* |
|---|---|---|---|---|
| *Percent Walks* | -0.006 | -0.49*** | −0.46*** | -0.52*** |
| | (-0.25) | (-11.69) | (-10.71) | (-12.04) |
| *Distance from Strike Zone* | | −181.92*** | −181.93*** | -176.01*** |
| | | (-105.73) | (-107.04) | (-103.65) |
| Game Fixed Effects | No | Yes | Yes | Yes |
| Catcher Fixed Effects | Yes | No | Yes | Yes |
| Pitch Count Dummies | No | No | No | Yes |
| N | 1,105,927 | 598,713 | 598,698 | 598,696 |

Table 13: Umpires' Calls - Additional Tests

This table provides evidence against the Matthew effect, and shows that umpires have favored all-star pitchers less over time. The sample consists of pitches outside of the strike zone, but less than 0.25 units away. *Favorable to Pitcher* is the dependent variable in most regressions - it is a dummy that equals one if the pitch should've been called a ball, but was called a strike. *Favorable to Pitcher* is the dependent variable in most regressions - it is a dummy that equals one if the pitch should've been called a ball, but was called a strike. *Strike* is a dummy that equals one if the pitch should've been called a strike.We include controls for the distance from the strike zone and the current pitch count. We multiply coefficient estimates by 100. Standard errors are clustered by game and by batter in the first two columns and by game and by pitcher in the last two columns. We put t-statistics in parentheses. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level.

| | *Favorable to Pitcher* | *Favorable to Pitcher* | *Favorable to Pitcher* | *Favorable to Pitcher* |
|---|---|---|---|---|
| *Batting Average* | 9.75** | | | |
| | (2.33) | | | |
| *Strong Batter* | | -0.01 | | |
| | | (-0.04) | | |
| *All-star Pitcher* | | | 1.07* | -1.06* |
| | | | (1.89) | (-1.77) |
| *Percent Walks* | | | -0.50*** | -0.55*** |
| | | | (-7.82) | (-8.09) |
| Game Fixed Effects | Yes | Yes | Yes | Yes |
| Catcher Fixed Effects | Yes | Yes | Yes | Yes |
| Pitch Count Dummies | Yes | Yes | Yes | Yes |
| N | 1,221,800 | 1,221,800 | 445,770 | 364,305 |