Forecaster Herding, Timing and Accuracy in the Age of ESG

by

Louis H. Ederington[1], Scott C. Linn[2*], Huiming (Minna) Zhang[2]

September 25, 2022

Abstract

Using a panel of 19,677 forecasts of the change in natural gas in storage in the U.S. for the period 2003-2021 we find that as the focus on Environmental, Social and Governance (ESG) increased forecast accuracy in this highly "E" sensitive sector improved, and, while forecasters predominantly anti-herd (rely on private information) anti-herding intensity declined. Greater anti-herding intensity and the early release of forecasts are associated with larger forecast error, together largely determine accuracy, and the relation is unaffected by the shift in ESG focus. The results suggest that the ESG focus had the complimentary effect of increasing attention on fundamentals.

Keywords: ESG; analyst forecasts; herding behavior; forecast timing; forecast accuracy; natural gas in storage

JEL: G1, D83, D84

[1]This study is dedicated to our dear friend and partner on this research Louis Ederington, late of the University of Oklahoma, who sadly passed away while the investigation was in progress.

[2]University of Oklahoma, Division of Finance, Price College of Business, 307 W. Brooks, Norman, OK 73019 USA. Linn: slinn@ou.edu; 405-595-7426; Zhang: huimingzhang@ou.edu; 405-325-5591.

Forecaster Herding, Timing and Accuracy in the Age of ESG

1. Introduction

The atmosphere created by the 'Environmental, Social, and Governance' (ESG) movement has had what many believe to be a profound impact on economic behavior and decisions (see for instance the website maintained by the World Bank, https://datatopics.worldbank.org/esg/). Little is known however about how the behavior of forecasters of economic variables has been influenced by the ESG movement even though those forecasts comprise an important piece of the information set used by investors, traders, and policy makers. Has the increased focus on ESG pushed analysts away from fundamental supply and demand issues, or has it had the complimentary effect of driving them to an even greater overall deepening of understanding and subsequently improved forecast accuracy, or simply, has the ESG emphasis been a neutral event? The commitment of resources to the greater ESG vision has been large. The consultancy Opimas reports that spending on ESG data hit $1 billion in 2021 and is on target to surpass $1.3 billion in 2022. Bloomberg reports that over $120 billion flowed into ESG-focused exchange-traded funds during 2021 and projects that more than one-third of all globally managed assets could carry explicit ESG labels by 2025, exceeding $50 trillion.[1] These statistics make clear that an important shift has occurred and that answers to the questions we have posed have and will in the future take on increased importance. We present answers to those questions through the study of a key segment of forecasters who cannot have escaped the exchange on ESG, forecasters who concentrate on what might be called ESG sensitive economic sectors. No sector has attracted more ESG attention, and in particular the 'E' of ESG, than the oil and natural gas sector.[2] The call for a more intense focus on ESG issues by analysts makes this an ideal focus group. The question we explore in this study is whether the behavior of analysts making forecasts of fundamental variables

[1] Opimas: https://www.opimas.com/research/742/detail/ ; Bloomberg: https://www.bloomberg.com/news/newsletters/2021-12-01/the-esg-market-is-controlled-by-a-few-big-investors, and, https://www.bloomberg.com/company/press/esg-assets-rising-to-50-trillion-will-reshape-140-5-trillion-of-global-aum-by-2025-finds-bloomberg-intelligence/ .

[2] Notwithstanding the coal mining sector and users of coal.

for the natural gas sector changed during the period 2003 – 2021, a period during which the ESG narrative has become an everyday topic of conversation for many and a touchstone for many major investors and traders, e.g., BlackRock, signatories of the United Nations Principles for Responsible Investment.[3] We explore several important dimensions of analysts' forecasting behavior before and after the ESG movement became a focal point in economic as well as political and cultural discussions. Specifically, we examine whether the propensity to herd, the timing of analyst forecast releases, forecast accuracy and the relations between them, changed over the course of 2003-2021, a period during which there has been increased intensity of focus on ESG. We study 19,677 forecasts of the change in natural gas in storage in the United States issued by 133 forecasters employed by 102 firms. The forecasts are compiled and reported by Bloomberg weekly. Forecast errors are computed using the actual changes in storage reported weekly by the U.S. Energy Information Administration.

We treat the end of 2012 as the point in time at which attention to ESG began to escalate which we elaborate on in the following section of the paper.[4] Further, we consider the ESG movement attention shift an external shock to the set of forecasters we study and not a phenomenon driven by those forecasters. We present evidence that the accuracy of forecasts of the changes in U.S. natural gas storage levels, which we measure using the seasonally adjusted mean absolute forecast error (RAFE) (1) deteriorates as anti-herding increases and, also, (2) deteriorates as the time between the forecast release and the date of the announcement of the actual change by the U.S. Energy Information Administration (EIA) increases. These two variables account for roughly 50% of the variation in forecast accuracy across the forecasts we study, and the result is not influenced by the increased focus on ESG. This distinguishes our study from others in another important way as this result is not driven by demographic characteristics of the forecasters such as an analyst's experience. Forecasters in our sample exhibit a strong tendency to not herd, that is to anti-herd. However, there is a subset of forecasters that neither herd nor anti-herd, consistent with

[3] BlackRock: https://www.blackrock.com/ch/individual/en/themes/sustainable-investing/esg-integration; UNPRI: https://www.unpri.org/about-us/about-the-pri .

[4] The results we report are not sensitive to a series of alternative specifications for time that we describe later.

that group optimally weighting their private information and the public ‘consensus’ forecasts available, that is not overweighting either information set. Our results indicate that the *cross-sectional* relations between accuracy and herding and forecast timing for the forecasts studied experienced no significant change between the period prior to the increase in ESG focus (pre-2013) and the period of intense focus (post-2012). On the other hand, we find that the level of accuracy for our sample of forecasters improved during the post-2012 period and there was a decrease in the standard deviation of forecast errors computed across forecasters. These results are broadly consistent with the hypothesis that the increased attention to ESG during the post-2012 era had the complimentary effect of increasing the overall attention on fundamental supply and demand conditions in the natural gas market, however this did not negate the result that anti-herding and early releases of forecasts are the primary factors associated with worse forecasts.

We also find that forecaster accuracy varies greatly across forecasting firms and is persistent but imperfect, the correlation of accuracy between the subperiods is 64%, consistent with some forecasters improving. The mean absolute forecast error of the least accurate forecasters is 1.9 times that of the most accurate during the pre-2013 period, but declines marginally during the post-2012 period, consistent with an overall reduction in forecast error following the increased focus on ESG issues. However, the less accurate forecasters rarely improve over time. This raises the question of how the least accurate forecasters continue to successfully sell their services? Our results indicate two potential reasons for the continued demand for the forecasts of less accurate forecasters. First, they tend to release their forecasts early, possibly providing a first-mover advantage to clients, and second, they tend to base their forecasts on their private intelligence rather than the consensus view which would be indicative of herding, thereby disseminating information that clients find useful, but which is specific to the forecaster. Such forecasters may survive if clients demand this private intelligence but nevertheless require that an ensemble of less accurate forecasts be unbiased. We show that the consensus of the less accurate forecasts in our sample is nevertheless an unbiased estimate of the actual, suggesting that there is demand for individual less accurate forecasts when users understand the consensus of the ensemble of these forecasts to be unbiased. The unbiasedness conclusion holds both before and after the increase in ESG attention, suggesting that no material change occurred following the end of 2012.

Studies of forecaster accuracy differences in other markets have focused on cross-sectional determinants such as forecaster experience, specialization, and available resources. Our evidence indicates the two primary determinants for the sample studied are the proclivity to anti-herd and the timing of when the forecast is released, both of which are choices made by an analyst. Little is known about how the impacts of these determinants are influenced by significant shifts in attention. Our evidence indicates that the increased attention to ESG issues, which forecasters in the energy sector have undoubtedly been confronted with yet has had no significant impact on the cross-sectional relations between herding, timing, and accuracy in the market studied. On the other hand, we find that accuracy increased between the 'pre-intensity ESG' period prior to 2013 and the 'post-intensity ESG' period consistent with the view that this increased attention had the complimentary effect of increasing overall attention to factors driving supply and demand. Finally, we find no relation between forecaster experience and accuracy after controlling for year fixed effects. Our results are robust to whether we focus on firm level behavior or individual forecaster level behavior as well as numerous model specifications.

We provide additional evidence on the behavior of anti-herding by the forecasters in our sample using individual forecaster panel models in which we control for firm fixed effects and separately forecaster fixed effects. These results also support the conclusion that anti-herding behavior is the norm. We find that anti-herding on average declined between the two subperiods, but that this can be attributed to two separate causes. First, the top 20 forecasters in terms of accuracy are on average associated with less anti-herding while the bottom 20 forecasters on average exhibit greater anti-herding, and that these relations did not change between the two subperiods. Finally, within a subperiod, anti-herding is most intense 3 days prior to the release of the EIA's report, declines 2 days out and increases the day before the EIA release. The relation does not change for the 2-day out cases between the subperiods, but anti-herding declines for the 1 day out cases during the second subperiod.

We also ask and answer several questions specifically regarding herding behavior. First, was anti-herding the predominant form of behavior? Our results show that the answer is yes. Second, did the level of anti-herding change following the increased focus on ESG? Anti-herding is dominant in both subperiods however anti-herding by individual analysts during the second period was less intense, Third, were better forecasters greater anti-herders or conversely, were poorer forecasters greater anti-herders? We find that better forecasters tend to neither herd nor anti-herd while poorer forecasters anti-herd. These results do not vary between the subperiods. We find that anti-herding on average declined between the two subperiods, but that this can be attributed to two separate causes. First, the top 20 forecasters in terms of accuracy are on average associated with less anti-herding while the bottom 20 forecasters on average exhibit greater anti-herding, and that these relations did not change between the two subperiods. Fourth, was anti-herding related to the timing of an analyst's forecast release? The results indicate the answer is yes. Anti-herding intensity decreases between the third day and the second day prior to the EIA report date but then increases. Within a subperiod, anti-herding is most intense 3 days prior to the release of the EIA's report, declines 2 days out and increases the day before the EIA release. The relation does not change for the 2-day and 3-day out cases between the subperiods, but anti-herding declines for the 1-day out cases during the second subperiod. We provide evidence on the behavior of anti-herding using individual forecaster panel models in which we control for firm fixed effects and separately forecaster fixed effects. These results also support the conclusion that anti-herding behavior is the norm.

The data also permit us to identify the date of release, and based upon information obtained from the source, Bloomberg, we know the order in which the forecasts are released for each date for a subset of the forecast dates studied. Based upon this subset of data, which represent within day releases, we find that our base results on herding are further confirmed. Several additional robustness tests reported also confirm our basic findings.

Section 2 provides a description of our research design and the sample of forecasts we study. We follow with a review of how the paper contributes to the extant literature in Section 3. We provide a first look and discussion of the sample forecast errors in Section 4. Section 5 provides an examination of the herding behavior of the forecasters in the sample. Section 6 examines the joint impact of herding and timing on forecast accuracy through the estimation of between-firm models as well as panel models focused on individual analysts. In section 7 we explore the behavior of anti-herding in more detail. Section 8 presents a discussion of our evidence on forecast dispersion across analysts and relates dispersion to our evidence on the lack of herding. Section 9 presents a discussion of why poor forecasters and anti-herding survive. Section 10 reviews several robustness experiments and Section 11 presents our conclusions.

2. Background, Research Design, and the Sample of Forecasts

December 2004 saw the publication of an influential report titled "Who Cares Wins", a joint initiative of a group of global-level leading financial institutions and the World Bank. Notably in that report the phrase 'Environmental, Social and Governance' (ESG) was coined.[5] ESG has since garnered world-wide attention gaining significant prominence in the United States as elsewhere. Some might even call the movement a significant cultural shift. Hirshleifer (2020) has recently proposed an economic paradigm he labels the "social approach to investor sentiment" (pg. 34) in which economic agents in essence 'feed' off each other's views. Indeed, the ESG *movement* could potentially fall within this class of agent behaviors. Hirshleifer (2020) points out that such social interaction can result in social transmission bias, which he describes as " the systematic directional modification of ideas or signals as they pass from person to person" (pg. 1).[6] He highlights two driving factors that we feel may have helped to propagate the ESG movement: 1) shocks to environmental cues (such as the publication of the "Who Cares Wins" document and later the Paris Climate Agreement, and subsequent press coverage, and, 2) the content of folk models (the spread of "folk" interpretations of complex scientific data, for instance motivated by a casual reading and reporting of the several publications of The Intergovernmental Panel on

[5] "Who Cares Wins", 2004. Source document: https://documents1.worldbank.org/curated/en/280911488968799581/pdf/113237-WP-WhoCaresWins-2004.pdf

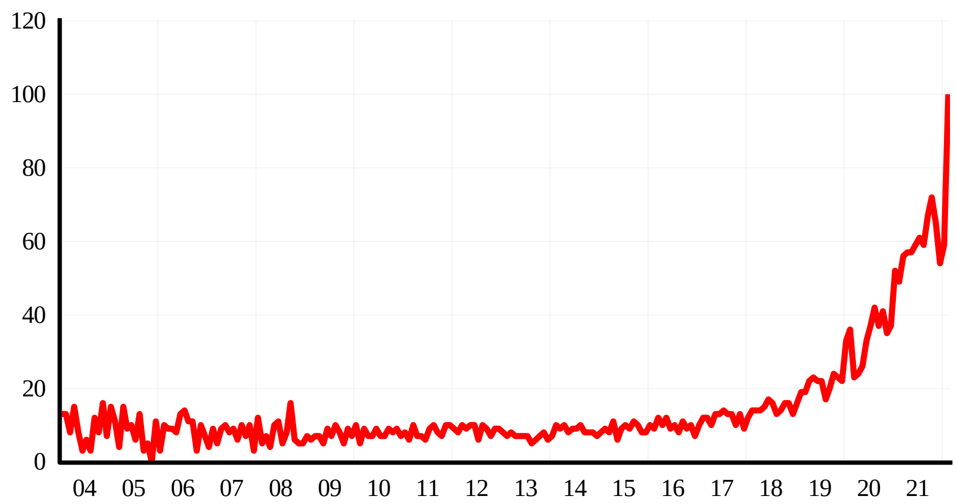
[6] The impact of social interaction and agent interaction in general has a long and rich history in sociology and psychology, too voluminous to mention here, refer to Hirshleifer (2020). Hirshleifer (2020, pg. 1) states: "Social economics and finance recognizes that people observe each other and talk to each other, where talking includes written text and social media."

Climate Change and the commentaries on the many complex simulation models that have been constructed to predict future climate change behavior).[7] The mechanism by which the increased attention to all things ESG may have created social transmission bias could come through signals and ideas being modified or distorted, or simply not conveyed in the transfer of information between "senders" and "receivers" (Hirshleifer, 2020, pg. 5).

There seems to be little doubt that sentiment runs high on both sides in the discussion of ESG and that the potential for bias to emerge in the spirit of Hirshleifer could be present. Evidence on the extent of the sharing of ideas regarding ESG is apparent from an inspection of Figure 1 which displays the Google trend index for 'interest' in ESG. The Google trend index for 'ESG' was essentially flat for numerous years following 2004 but began an upward trend following 2012 and escalated dramatically after 2016.[8] Notably, 2016 corresponds to the effective date of the Paris Climate Agreement.[9]

Environmental, Social and Governance
United States, Google Trends Index
Monthly 2004-2022

Figure 1



[7] See the website of the 'The Intergovernmental Panel on Climate Change' https://www.ipcc.ch/.
[8] Further details on the construction of a Google Trend index can be found at: https://support.google.com/trends/answer/4365533?hl=en
[9] See https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

Of particular interest for our study are the recommendations published in the "Who Cares Wins" report advocating that those analysts operating in the greater financial community should bring greater consideration of ESG issues to bear in their analyses. One recommendation of the "Who Cares Wins" report is the following:

> **"Analysts are asked to better incorporate environmental, social and governance (ESG) factors in their research where appropriate and to further develop the necessary investment know-how, models and tools in a creative and thoughtful way." (pg. iii)**

which is followed later by:

> **"We encourage analysts to prioritize ESG issues on the basis of their potential impact on financial value and on a sector-by-sector basis." (pg. 28) [10]**

These recommendations are proposals for a shift of attention, which could potentially lead to phenomena of a form like those hypothesized in Hirshleifer (2020). However, the impact of this shift in attention is an empirical question.

Three hypotheses emerge regarding predictions about the implications of the shift in attention on ESG. The first is that increased scrutiny of markets that have a broad impact on ESG may cause attention to be diverted away from other fundamental issues, resulting, in a forecasting context, in a degradation of forecast accuracy and increased dependence on the opinions of other forecasters, increased herding. This could potentially be the result of social transmission bias. Conversely, enhanced scrutiny could have the opposite beneficial consequence that there is a complementarity effect of a more intense focus on ESG resulting in a deeper dive into fundamentals overall and an improvement in both accuracy and less dependence on consensus views. Finally, it is possible that the shift in attention had no impact at all, a result which would suggest that while ESG may be on the minds of many, the impact on the behavior of analysts who operate in the energy sector has been minimal. Our investigation provides evidence on these competing hypotheses.

---

[10] Further illustrative of this emphasis is the position of the Chartered Financial Analyst Institute https://www.cfainstitute.org/en/research/esg-investing .

We treat the end of 2012 as the point in time at which the increased attention to ESG began and as an external shock to the set of forecasters we study. Analysts forecasting economic variables, whether of a micro or macro nature, face several decisions, each of which could affect the accuracy of their forecasts and which could be impacted by the type of societal attention shift brought on by the ESG movement. First is the decision whether to form and release their forecast before others, thereby satisfying a demand by users for early but potentially less accurate forecasts. Intertwined with this choice is the companion decision of whether to release a less accurate versus a more accurate forecast. An opposing alternative is whether to delay releasing a forecast until closer to the time when the actual data being forecasted are disclosed, thereby accumulating more information to reflect in the forecast and thereby possibly improving the forecast's accuracy. Last, the analyst must decide whether to herd—that is, whether to base the forecast predominantly or wholly on the prior forecasts of others rather than relying on private information. Users' preference for professional forecasts of economic variables, revealed by their demand for such, indicates that professional forecasts play an important informational role in users' decisions.[11] A critical element in the assessment of a forecast's value is, of course, its accuracy. The primary objective of this study is to shed light on the relation among the timing of forecast issuance, the propensity to herd, and forecast accuracy and whether these choices have changed because of the atmosphere created by the ESG movement, within the context of the commodity market for natural gas.

The setting we investigate is tailor-made for a study of forecast timing, the propensity to herd, and forecast accuracy and whether we observe changes during recent times coincident with the increased intensity of interest in all issues ESG. We view three characteristics as being important for the design of the study. First, to ensure that the forecasts are not slanted by a prior relationship between the analyst and the user of the forecast or any entity whose value might be influenced by the forecast, the forecasts should not be related to either estimates of the value or performance of such parties. Second, the focus should be on a variable that has been shown to have economic relevance, such as through surprises in the variable's value having an impact on market

[11] Indeed, Bloomberg as well as Haver Analytics supply extensive analyst-based forecasts of fundamental economic variables as part of their subscription services.

prices. Third, the forecasts should not have the potential to influence the actual value of the economic variable being forecasted. The setting we investigate meets these criteria. Finally, and importantly, the study should span a period during which the focus on ESG has changed dramatically.

We investigate professional analyst forecasts of changes in the weekly natural gas storage figures reported in the Weekly Natural Gas Storage Report (WNGSR) compiled and released by the Energy Information Administration's (EIA), an arm of the United States Department of Energy. The criteria mentioned above are satisfied in several ways. First, forecasts are prepared by professional analysts and predict a macro variable that is not tied directly to any user of the forecasts nor to the value of any specific entity/firm. Second, the observed impact of these forecasts on the futures price of natural gas indicates that market participants pay close attention to these forecasts and their accuracy.[12] Third, the timing and frequency of the forecasts' releases cannot influence the actual value of natural gas in storage. This is because the analysts issue their forecasts after the date when the EIA measures the actual amount of natural gas in storage (as of the Friday before the EIA release) but before the actual amount is released to the public (the following Thursday).

Before proceeding there are several facts about the forecasts we study that are worth highlighting. First, numerous studies have shown a strong natural gas futures price reaction to deviations between forecasts of the change in the amount of natural gas in storage and the change reflected in the EIA announcement, so there is a strong incentive for analysts to forecast the EIA figures early and accurately.[13] Ederington, Lin, Linn, and Yang (2019) find evidence of a price reaction to analysts' forecasts prior to the EIA announcement. Second, unlike the case with earnings forecasts, the variable being forecasted is an aggregate economy-wide variable and is not directly tied to the performance of a particular firm (see the survey Kothari, So, and Verdi, 2016). This eliminates the incentive for analysts to issue biased forecasts that cater to specific clients, as

---

[12] Linn and Zhu (2004); Gay, Simkins, and Turac (2009); Chiou-Wei, Linn, and Zhu (2014); Halova, Kurov and Kucher, 2014, Ederington, Lin, Linn and Yang, 2019; Prokopczuk, Wese Simen and Wichman, 2021).
[13] Refer to footnote 11.1

can also potentially be the case with earnings forecasts. Third, the EIA issues the WNGSR each Thursday morning, but the data being reported are measured as of the prior Friday. Thus, forecast users cannot take action to manipulate the actual storage numbers and thereby influence the accuracy of the forecasts. Fourth, many forecasters follow and make forecasts for this single variable, ostensibly competing against one another. Fifth, forecasts are issued weekly, which provides both relatively high frequency and a large sample size focused on a single economic variable.

Our data set consists of 19,677 individual forecasts by 102 different forecasters of the weekly EIA announcement of the change in total natural gas storage in the lower 48 U.S. states for the period from 3/4/2003 through 12/9/2021. Finally, forecasts each week are submitted to Bloomberg up to three days prior to the EIA release date but, as mentioned, after the Friday for which the actual change is compiled by the EIA. The dates on which forecasts are submitted are also posted on Bloomberg. The availability of the forecast submission dates provides an opportunity for evaluating the relation between forecast timing and accuracy, the prevalence of herding, the relation between herding and accuracy, and the relative joint contributions of timing and herding to forecast accuracy. Finally, the period we study spans a significant ramp-up in the focus on ESG as illustrated in Figure 1. Our emphasis on forecasting behavior before and after this ramp-up has not been examined in the extant literature but with the attention being paid to ESG our study adds a new feature to the evidence regarding the importance of the ESG emphasis.

3. Contributions to the literature

The impact of the ESG movement on economic outcomes, which arguably also captures the focus on climate change issues and climate risk exposure, has attracted considerable attention both within academic circles and without, notably amongst public policy makers. Recent contributions to the academic literature have explored the implications of ESG issues on aggregate economic welfare and real outcomes (Hong, Wang, and Yang, 2022; Naaraayanan, Sachdeva, and Sharma, 2022; Bolton and Kacperczyk, 2022), on the decision criteria of investors (Hart and Zingales, 2022), on the risks and valuation of equities (Goldstein, Kopytov, Shen, and Xiang, 2022; Lindsey, Pruitt, and Schiller, 2022; Brogger and Kronies, 2022; Giglio, Kelly, and Stroebel, 2022, Pastor, Stambaugh and Taylor, 2022), on the fixed income risks and pricing (Kyung Auh, Choi, Deryugina, and Park, 2022) and on the banking sector (Kacperczyk and Luis Peydro, 2022; Ivanov, Kruttli, and Watugala, 2022). The impact of ESG on the actual behavior of market participants has generally been inferred however from aggregate studies. We contribute to this expanding literature by focusing on whether the increased attention given to ESG has impacted the micro behavior of an important group of market participants, professional forecasters. Our specific focus is on whether the ESG movement has impacted forecaster herding behavior and the timing of forecast releases along with the joint impact of these behaviors on forecast accuracy. Our results are new to the literature and provide an additional perspective on how the ESG movement may have influenced economic behavior.

Several studies of earnings forecasters have concluded that there are factors that influence the attention of company analysts, in particular that analysts who are proximate to the firms being studied are better informed than non-local analysts (e.g., Coval and Moskowitz 2001; Ivkovic and Weisbenner 2005; Malloy 2005; Engelberg and Parsons 2011; Miller and Shanthikumar 2012). The conclusion drawn is that proximity to the company being studied increases the attention of analysts. These studies do not however examine whether there are triggers that increase attention. We add to the literature on analyst attention by examining whether widespread attention triggers (the increased focus on ESG issues) are associated with a change in forecast accuracy as well as a change in forecaster herding behavior through a study of forecasters of changes in natural gas in storage in the U.S. But the characteristics of the economic variable we focus on, and the forecasters involved, have some added benefits. First, our focus is on a commodity traded in a national market where the forecast variable is a quantity (the change in the commodity in storage nationwide) and is related to the aggregate supply and demand for that commodity. Second, unlike earnings forecasts for a company, proximity of analysts to the 'market' for natural gas is not relevant as natural gas is a commodity traded in a national market with information about supply and demand conditions widely available. And finally, the forecast is for an aggregate market quantity and not a company specific variable such as earnings.

We also contribute to the literature on forecasters' herding behavior by providing empirical evidence on the relation between herding and the advent of the ESG movement.[14] In particular, we study how the relation between a forecaster's propensity to herd and their forecasting accuracy was impacted following the ramp in the focus on ESG. Several authors have presented evidence that at least some corporate earnings forecasters and stock analysts herd, including, Trueman (1994), Hong, Kubik, and Solomon (2000), Welch (2000), Clement, and Tse (2005), Jegadeesh and Kim (2010), Clement, Hales, and Xue (2011), and Huang, Krishnan, Shon and Zhou (2017), andAltınkılıç, Balashov and Hansen (2019).Zitzewitz (2003), Bernhardt, Campello, and Kutsoati (2006) (BCK) and Chen and Jiang (2006) (CJ) argue based upon the evidence they present that most common stock analysts anti-herd, meaning that they tend to overweight their private information relative to what would be optimal behavior. Results presented in Ivković and Jegadeesh (2004) are also consistent with the 'no herding' thesis. These findings suggest that the question of whether forecasters herd, anti-herd or act optimally has not been fully resolved. We present evidence on this question through an examination of a sample of forecasters that has received scant attention even though the forecasts they produce are closely watched by traders in an important energy market, the market for natural gas. Further, our results provide new insight into how the behavior of these forecasters has changed following the increase in attention to ESG.

With several exceptions, such as Lamont (2002) who studies GDP growth estimates, few studies document or explain differences in herding behavior between forecasters of macro-level variables, which is another contribution we make to the literature. Gallo, Granger, and Jeon (2002) find evidence consistent with herding in GDP forecasts while Lamont (2002) finds that in forecasting GDP growth, older forecasters tend to issue forecasts further from the consensus than younger forecasters. Pierdzioch, Rulke, and Stadtmann present evidence of anti-herding among oil price forecasters (2010) and metals price forecasters (2013) based upon the BCK (2006) statistic. Their results are based upon forecasts compiled in the Survey of Professional Forecasters

[14] For reviews of the theoretical literature on herding including informational herding and the area of informational cascades see SHirshleifer and Teoh (2003, 2009) Hirshleifer (2020), Bikhchandani, S., Hirshleifer, Tam Welch (2021).

published by the European Central Bank. Bewley and Fiebig (2002) find evidence of herding in interest rate forecasts, but Ehrbeck and Waldmann (1996) and Pierdzioch and Rulke (2013) find that interest rate forecasters tend to anti-herd. Broughton and Lobo (2018), using a measure developed in BCK measure, find that forecasters of the Purchasing Managers' Index (PMI) issued by the Institute of Supply Management tend to anti-herd. We present evidence that most forecasters of natural gas in storage tend to anti-herd, but some – specifically, the most accurate forecasters – exhibit neither herding nor anti-herding behavior.

In a recent study, Fernandez-Perez, Garel, and Indriawan (2020) also present tests of whether natural gas storage forecasters engage in herding. However, their analysis uses data that would not have been available to forecasters before the EIA report release date. As explained below, the herding measures we compute involve comparing a forecaster's individual forecast with the consensus of prior forecasts when the individual's forecast is made. The dataset that Fernandez-Perez, Garel, and Indriawan (2020) use is obtained from Reuters and does not identify forecast dates, so those authors do not have data on prior forecasts. Thus, their herding measure involves a comparison of the individual's forecast with the final Reuters consensus average, which was not known to the forecasters until later. The measures of herding we compute use prior rather than subsequent consensus forecasts. The Bloomberg compiled estimates contain each forecast's exact date of release and detailed information on who made the forecasts and which firm hired the forecasters. The estimates are posted on Bloomberg and accessible to all forecasters and investors on the release date. We therefore know which forecasts are available to forecasters when they publish new estimates. The measures of herding we compute use prior rather than subsequent consensus forecasts. In addition, we present results for multiple herding measures, establishing the robustness of our conclusions. Another distinguishing characteristic of our study is the examination of whether the expansion of the ESG movement has impacted herding, a question those authors do not address.

Our results also provide new insights regarding professional forecaster accuracy. The primary studies that have focused on the determinants of forecaster accuracy have concentrated on the accuracy of earnings forecasts produced by professional analysts (see the surveys: Kothari, So, and Verdi, 2016; Ramnath, Rock, and Shane, 2008). Although some early studies found no significant differences in forecast accuracy among earnings forecasters, Stickel (1992) and Sinha, Brown, and Das (1997) find systematic and persistent differences in forecaster accuracy. Exploring possible reasons for these differences, Clement (1999) reports that more experienced analysts and analysts employed by large firms (with possibly more information sources) tend to be more accurate. He also finds that forecaster accuracy is negatively correlated with the number of firms followed by the analyst, suggesting a specialization advantage, and this finding is corroborated by Hirst, Hopkins, and Wahlen (2004) for bank analysts. Both Brown (2001) and Michaely et al. (2020) find that earnings analyst accuracy is persistent; that is, forecasters who accurately forecast earnings in the past tend to be accurate in the future, and vice versa. Clement (1999) and Sinha, Brown, and Das (1997) find that late earnings forecasts tend to be more accurate than early forecasts. Michaely et al. (2020) also find that a consensus earnings forecast based on just the most accurate forecasters has slightly lower forecast error than the widely available consensus based on all forecasters when there are four or more high-quality forecasters. Whereas Clement (1999) reports that forecast accuracy increases with forecaster experience for earnings forecasters, Lamont (2002) finds that as forecasters of macroeconomic variables became older and more established, they produce more radical and less accurate forecasts. The aforementioned studies have not explored the implications of the type of cultural shift we suggest may have accompanied escalation of the ESG movement. Therefore, one of the significant differences between our study and the extant literature is an examination of whether forecast accuracy differed between the period leading up to an increased emphasis on ESG and following and an exploration of cross-sectional relations between accuracy and herding and the timing of forecast release and whether they were influenced by this shift in attention.

Cooper, Day, and Lewis (2001) explore the importance of forecast release timing for earnings estimates and argue that earnings analysts have strong incentives to be among the first to release their earnings estimates. They find that stock prices respond significantly more strongly to earnings forecasts by forecasters who are generally among the first to release their forecasts than to forecasts by historically late forecasters. In contrast, they do not find a significant difference in the price response to forecasts by the historically most accurate forecasters versus the less accurate. This suggests that investors value earlier forecasts. We specifically provide evidence on the timing of forecast release and its impact on accuracy and whether this relation has changed with the increased focus on ESG.

Hirshleifer, Levi, Louriea, and Teoh (2019) study the impact of decision fatigue on the herding behavior of earnings forecasters and find that decision fatigue leads to increased herding. The authors measure decision fatigue using the number of forecasts an analyst issues during a particular day. Our setting is different because analysts issue only a single forecast regarding the commodity in question and so are not exposed to the type of 'fatigue' Hirshleifer et al. study.

The studies by Gay, Simkins, and Turac (2009) and Gu and Kurov (2018) explore the accuracy of forecaster predictions about the amount of natural gas in storage and find significant differences among natural gas forecasters' accuracy. However, these authors neither explore the reasons for forecast accuracy differences nor whether there have been changes over time. Our exploration of the relation between determinants of accuracy and whether those relations have changed over time with the shift in attention to ESG questions expands our understanding of these issues. BCK and CJ focus on the measurement and determinants of herding in general. Our study focuses on the relation between forecast accuracy, herding, and the timing of forecast release and how significant attention triggers influence these relations. To our knowledge, no research to date has directly explored the implications of the intense focus on ESG and the joint relations between herding, forecast timing and forecast accuracy.[15]

Finally, a body of literature exists that confronts the question of analyst forecast bias, but such studies have devoted only limited attention to why there is continued user demand for the forecasts of inaccurate analysts (see Kothari et al., 2016; Chen and Jiang, 2006). Ederington, Lin, Linn, and Yang (2019) and Chiou-Wei, Linn, and Zhu (2014) report that the Bloomberg median forecast is surprisingly accurate in that when the change in storage levels announced by the EIA is regressed on the median Bloomberg forecast change, the adjusted r-square is over 99%.

[15] See Ramnath, Rock, and Shane (2008) and Spyrou (2013).

Nonetheless, there is a strong market reaction to the EIA announcement. We contribute to this literature by presenting evidence consistent with a theoretical explanation first proposed by Laster et al. (1999) for why individual forecasts with varying degrees of accuracy are observed and whether this changed following 2012 and the increased attention given to ESG issues.

4. Forecast errors pre and post year-end 2012

Each Thursday at 10:30 AM ET, the EIA releases its survey of natural gas in storage in the United States measured as of the preceding Friday.[16] Although the report presents storage levels by U.S. geographic regions, the headline figure is the total of working gas in storage for the lower 48 U.S. states in billions of cubic feet (Bcf).[17] Bloomberg collects analyst forecasts of the change in total gas in storage by email or phone beginning several days prior to the EIA announcement. Bloomberg reports these individual forecasts as they are received, along with the average, median, standard deviation, and high and low values. Bloomberg publishes the date, but not the time, that it receives each individual forecast. We collect these individual forecasts along with the final summary measures for each week during the period 3/4/2003 through 12/9/2021 (978 weeks). Figure A1 in the Appendix presents a screen shot of the Bloomberg screen that displays the data. The data set includes 19,677 individual storage forecasts, or an average of roughly 20 forecasts per week by 102 different firms.[18] Based upon the change in interest in ESG (Figure 1), we divide the sample into two time periods: Subperiod 1 spans 3/4/2003 through 12/30/2012 and Subperiod 2 spans 1/1/2013 through 12/9/2021. "Forecast Error" (FEr) is measured as the difference between the individual forecast and the actual change subsequently announced by the EIA and the absolute forecast error is the absolute value of FEr.

[16] Details about the compilation of the Weekly Natural Gas Storage Report by the U.S. EIA are available at http://ir.eia.gov/ngs/ngs.html.

[17] Natural gas is generally stored in depleted oil or natural gas fields or in salt caverns. To maintain pressure, these facilities require the presence of a cushion, or an amount of "base" gas. Working gas in storage is the amount stored in addition to the base gas. https://www.eia.gov/tools/glossary/index.php?id=W#work_gas

[18] The total of 102 firms is after consolidating observations where the same firm was listed under two or more slightly different names on different dates. In most cases, the same firm name was recorded slightly differently on different weeks, such as an abbreviated name some weeks and a full name in other weeks. In other cases, firm names changed slightly over our data period. There were 146 names prior to consolidation. Economist names were much more consistent.

The data indicate that absolute forecast errors are highest in January, decline monotonically through May, are low through the summer and fall, and then rise in November and December. To control for this seasonality in our tests, we define the relative absolute forecast error (RAFE) for a forecaster as the forecaster's absolute forecast error divided by the monthly mean for the month in which the forecast is issued. In addition to controlling for seasonality of the absolute forecast error, the Brown-Forsythe Modified Levene test does not reject the null hypothesis of equality of variances of RAFE across the months-of-the-year (p-value .101). Therefore, our adjustment controls for both mean and variance effects due to seasonality.

Descriptive statistics are reported in Table 1 for the full sample period along with the two subperiods. Numerous studies have found evidence of bias in analysts' earnings forecasts.[19] In Panel A (full sample results) we find that the mean forecast error (FEr) is negative (-0.368, p-value for test that mean equals 0 is .000) and significantly different from zero, but the median FEr is zero. The results also indicate there is considerable cross-sectional variation in the forecast errors. The average absolute forecast error equals a positive 8.542 bcf and the median is 6.000 bcf when considering the full sample. For comparison purposes note that the median forecasted storage change equals 40. The standard deviation of the forecast error is large relative to the mean. Together these statistics indicate there is variation of forecast errors across forecasters. The subperiod statistics show that in the first period (2003-2012) and the second period (2013-2021), the mean FEr is negative but is less negative during the second period. Likewise, both the mean and median Absolute Forecast Error (AFE) are smaller during the second period, as is the standard deviation of this variable.

Relative absolute forecast error (RAFE) is our study's primary measure of forecasting error. As mentioned above, the mean forecast error varies across the months-of-the-year. The seasonal corresponds to a seasonal in total storage that is driven by a seasonal in natural gas consumption. Similar to AFE the mean RAFE is smaller during the second period as is the median. RAFE displays less skewness and kurtosis in the first period than in the second period.

[19] See the review by Kothari, So, and Verdi (2016) of the earnings forecast literature.

We test whether the means and medians of the three variables are equal between the first and second subperiods. The results indicate that there are statistically significant differences between the two periods for both AFE and RAFE, the variable we focus on henceforth. Here and in subsequent tests we employ a conservative 1% significance level criterion. Both the mean and median RAFE are smaller during the second period, indicating by this measure that accuracy improved. The take-away is that there was improvement in accuracy during the second subperiod following the increased focus on ESG. However, the standard deviation of RAFE relative to the mean indicates considerable variation across forecasters.[20] Mean RAFEs for the four least accurate forecasting firms are 1.9 times the four most accurate during the first subperiod, and 1.7 times greater in the second subperiod. The null that forecast accuracy does not differ across forecast issuing firms is rejected at the <1% level for both subperiods (Subperiod 1: $F = 14.46$, $p < .01$; Subperiod 2: $F = 19.47$, $p < .01$).[21] While these results do not control for other factors, they are consistent with an improvement in accuracy following the increased focus on ESG. We also investigate how persistent forecast accuracy was between the subperiods studied. Specifically, was future firm forecast accuracy correlated with past forecast accuracy despite the increased focus on ESG during the second period?[22] The correlation coefficient between the RAFEs in the two subperiods equals .64 and is significantly different from zero. We conclude that in a comparison of the two subperiods future forecast accuracy was positively correlated with past forecast accuracy, however, the relation is not conclusively perfect, suggesting the possibility that the second period of increased ESG attention was associated with changes in analyst behavior that improved accuracy.

[20] Gay, Simkins, and Turac (2009) and Gu and Kurov (2018) explore natural gas forecaster accuracy for the 1997–2005 and 2008–2016 periods, respectively, and reject the null that forecast accuracy does not differ among forecasters. However, those authors do not explore why some forecasters are more accurate than others nor whether there have been changes over time.
[21] The ANOVA test allows for unequal variances and includes all weekly observations for each firm.
[22] Brown (2001) and Michaely et al. (2020) find that the accuracy of earnings forecasters tends to be persistent.

5. Herding

5.1 Introduction

A forecaster that concludes a forecast based on his private information would differ substantially from those of earlier forecasters, might ignore or underweight his private information and adjust his forecast toward those released earlier, even though it might otherwise be optimal to incorporate his private information. That is, the later forecaster might herd by tilting his forecast towards the prior consensus. Conversely, if a forecaster tilts more towards his private information and ignores the consensus he is anti-herding. The remaining case is when the forecaster strikes the optimal weighting of public and private information and neither herds nor anti-herds relative to the optimal weighting scheme.

The shift of attention to ESG issues during the post-2012 era could have created distortions of the forms suggested by Hirshleifer (2020) leading to increased reliance on consensus views and increased herding. Conversely, if the shift of attention resulted in an enhanced focus on fundamentals overall and expanded private information, leading to less reliance on consensus views, then increased anti-herding could have resulted. Finally, if the focus on ESG issues was essentially a neutral activity, then neither a tilt towards more or less herding/anti-herding should be observed.

5.2 Measures of Herding Behavior

5.2.1 The Bernhardt, Campello, and Kutsoati (2006) (BCK) Measure of Herding

BCK assume that a rational forecaster seeking to maximize the accuracy of his forecast rationally incorporates into his forecast F_{jt} all information available at time (t), including previous forecasts that are available, which we summarize as the consensus forecast C_t.[23] That is, the rational forecaster seeking to maximize accuracy weights all relevant available information, including the forecasts of others, in an optimal fashion.[24] BCK define a herder as a forecaster who

[23] The BCK test has been used by several researchers to test for herding,including Pierdzioch, Rulke, and Stadtmann (2010), Pierdzioch and Rulke (2013), and Broughton and Lobo (2018).

[24] In BCK, rationality within the context of the development of their herding measure means optimal Bayesian updating.

overweights the information in previous forecasts issued by others in forming his forecast F_{jt}, and an anti-herder as a forecaster who underweights such information. In other words, the BCK herder classification includes not just those who simply mimic the prevailing consensus, but also those who base their forecast on both their private information and the information in existing forecasts but overweight the latter relative to the most accuracy-maximizing weighting. Similarly, in addition to those who ignore the information in existing forecasts in forming their own, the BCK anti-herder classification includes those who consider both the prevailing consensus and their private information in forming their forecast but underweight the prevailing consensus. We now lay out the details of the test.

The BCK test is based on the following reasoning. Let A_t be the actual storage level change based upon the EIA report, which, of course, is unknown at the time the analyst is preparing his forecast. Let E_{jt} be forecaster j's estimate of A_t based solely on his private information, that is, before seeing the forecasts of others. If j's private information forecast is unbiased, the probability of $E_{jt}>A_t$ and the probability of $E_{jt}<A_t$ should both be about .5.

Let C_t be the consensus forecast of other forecasters observable by j before publishing his forecast and let F_{jt} be the forecast released by j after observing the previous consensus forecast. If other forecasters are unbiased, $\Pr(C_t<A_t) \approx .5$ and $\Pr(C_t>A_t) \approx .5$. However, for any particular date t, the consensus C_t may deviate from the actual A_t and from the private information forecast E_{jt}, even if both E_{jt} and C_t are unbiased over many experimental trials. Suppose that the consensus, C_t, is below the private information forecast E_{jt} of forecaster j. If forecaster j issues unbiased forecasts, the probability that the ultimate forecast issued, F_{jt}, exceeds A_t should be independent of all conditioning information outside j's private information available at the time of his forecast, including C_t, and the probability should equal .5. The same holds for the probability that F_{jt} is less than A_t. This will be the case if the forecaster neither herds nor anti-herds.

On the other hand, if $E_{jt}>C_t$ and the analyst herds, the actual forecast issued, $F_{jt,}$ will be shifted toward C_t and away from E_{jt} by more than would be called for by the optimal weighting of private and public information. That is, if herding occurs, $F_{jt}>C_t$. As a result, if herding does occur when $E_{jt}>C_t$, the probability that $F_{jt}>A_t$ should be influenced by the fact that the forecaster has shifted away from E_{jt} towards C_t. Thus, the probability that $F_{jt}>A_t$ should be less than .5.

Similarly, suppose $E_{jt}<C_t$ - that is, that the prevailing consensus forecast exceeds j's forecast based on his private information. If $E_{jt}<C_t$ and j herds, and analyst j herds, then j will adjust his forecast toward C_t, but if all the weight is not placed on C_t, then $F_{jt}<C_t$. If F_{jt} was unbiased, the probability that $F_{jt}<A_t$ should equal about .5. However, if F_{jt} is being shifted away from E_{jt} upwards towards the observed C_t, this will no longer be the case, and the probability that $F_{jt}<A_t$ will be less than .5.

Hence, the BCK test proceeds as follows. First, we separate the forecasts into two groups where (1) $F_{jt}>C_t$ and (2) $F_{jt}<C_t$, and we test whether the proportion of times $F_{jt}>A_t$ is less than .5 in case 1 and the proportion of times $F_{jt}<A_t$ is less than .5 in case 2. If these proportions are both statistically significantly less than .5, it constitutes evidence of herding, according to BCK. Similarly, a finding that the proportions are greater than .5 in these two cases implies that forecasters anti-herd.

5.2.2 The Chen and Jiang (2006) Measures of Herding

We continue with the notation used to describe the BCK test for expository purposes but suppress the individual forecaster identifier j. Chen and Jiang (2006) posit a setting in which an analyst is privy to private as well as public information about the unknown value of a variable to be forecasted, which henceforth we define following our earlier convention as the change in storage A. A sufficient statistic for the private information is assumed to exist and is defined as $\mathrm{Y} = \mathrm{A} + \varepsilon_{\mathrm{y}}$ where $\varepsilon_{\mathrm{y}} \sim \mathrm{N}(0, \sigma_{\mathrm{y}}^2)$. Similarly, the public information is characterized as the consensus forecast $\mathrm{C} = \mathrm{A} + \varepsilon_{\mathrm{c}}$, $\varepsilon_{\mathrm{c}} \sim \mathrm{N}(0, \sigma_{\mathrm{c}}^2)$. Each information signal is characterized by a level of precision $\left(\frac{1}{\sigma_{\mathrm{i}}^2}\right)$ where i= (Y, C). The analyst's optimal Bayesian forecast of the unknown quantity A utilizes both private information as well as public information. Specifically, the conditional expectation of A is given by

$$\mathrm{E}\left(\mathrm{A} \middle| \mathrm{Y},\mathrm{C}\right) = \left(h \bullet \mathrm{Y}\right) + \left(1-h\right) \bullet \mathrm{C} \tag{1}$$

where h is the forecaster's optimal weight on the private information if the forecaster acts as a rational Bayesian.

Suppose analysts construct their forecasts of A conditional on Y and C using a weighting factor k on private information that is possibly different from h. Specifically, the general forecast is given by

$$\mathrm{F} = (k \bullet \mathrm{Y}) + (1 - k) \bullet \mathrm{C} \tag{2}$$

Combining the two expressions and incorporating the assumptions regarding Y and C, the conditional expectation of the *difference* between F and A (the forecast error) is given by

$$\mathrm{E}(\mathrm{F} - \mathrm{A} \mid \mathrm{Y}, \mathrm{C}) = \frac{k - h}{k}(\mathrm{F} - \mathrm{C}) \tag{3}$$

Herding occurs if the analyst underweights the private information relative to what would be optimal, or conversely, overweights the public information. If the forecaster employs the optimal weight, h, then k=h and $\frac{k-h}{k}$ will equal zero. If the forecaster overweights her private information when making forecasts, then $k > h$ and the multiplier $\frac{k-h}{k}$ will be positive, an indication of anti-herding. In contrast, if the forecaster underweights her private information, $k < h$ and the multiplier $\frac{k-h}{k}$ will be negative, an indication of herding. A regression of the actual forecast error, F–A, on the forecast deviation from the consensus, F–C, will provide an estimate of $\frac{k-h}{k}$ and permit a test of herding behavior, specifically if $\frac{k-h}{k} < 0$, the analyst herds, if $\frac{k-h}{k} = 0$ the analysts neither herds nor anti-herds, and if $\frac{k-h}{k} > 0$ the analysts anti-herds. We refer to the test statistic as CJ_Dev.

The dependent variable in the estimation, the forecast error F–A, is the forecaster's estimate of the natural gas storage change in week t minus the actual storage change for week t; the regressor is the forecaster's estimate of the storage change minus the consensus forecast (F–C). For forecasts issued on Day 0 (the day of the EIA release), the consensus is the median of the forecasts issued on Day 1, Day 2, and Day 3; for forecasts issued on Day 1, the consensus is the median of the forecasts issued on Day 2 and Day 3; and for forecasts issued on Day 2, the consensus is the median of forecasts issued on Day 3.

The intuition behind the second measure proposed by CJ runs as follows. If the analyst utilizes the efficient forecasting parameter *h*, her forecast is equally likely to deviate positively or negatively from the actual value of A. Because the consensus forecast is unbiased, it is also equally likely that an analyst's forecast, if constructed efficiently using *h*, will deviate positively or negatively from the consensus. Because of distributional symmetry, there is therefore a 50% chance that the sign of the forecast error will be the same as that of the deviation from the consensus, assuming that the forecasts are generated using the parameter *h*. This logic serves as the threshold for the null hypothesis: If the analyst uses the optimal factor *h*, the proportion of cases in which the sign of the forecast error equals the sign of the forecast–consensus deviation will equal .5. Conversely, if the actual forecast parameter *k* does not equal *h* (i.e., if $k > h$ or $k < h$), this will no longer be true. If $k > h$, the analyst overweights the private information, and the proportion of cases in which the deviation of the forecast from C matches the forecast error (the deviation of the forecast from the actual) will exceed .5, and if $k < h$, the proportion of cases with matching signs will be less than .5 (CJ), as the analyst underweights private information. We refer to this test as CJ_Prob.[25] CJ_Prob exceeds .5 if the analyst anti-herds and is less than .5 if the analyst herds. Table A1 of the Appendix summarizes the three measures and the interpretation of each regarding the presence of herding and anti-herding.

5.2.3 Herding by the Sample Forecasters

In Table 2, we present the herding measures in each subperiod for firms making at least 20 forecasts in each. The measures reported are a weighted average accounting for whether $F_{jt}>C_t$ or $F_{jt}<C_t$, where F_{jt} is the week t forecast released by forecaster j, and C_t is the consensus forecast at the time forecaster j releases her forecast, and the weights depend upon the instances of $F_{jt}>C_t$ and

[25] The BCK and CJ_Prob statistics are based upon similar assumptions and, as we will see, produce statistics that are similar in sign and magnitude. CJ show that the measures they propose are robust to alternative assumptions.

$F_{jt}<C_t$ over all weeks.[26] For forecasts released two days prior to the EIA announcement, we use the median forecast on the previous day (Day 3) for C_t. For forecasts released on Day 1, C_t is measured as the median of the forecasts on Days 2 and 3, and for forecasts released on Day 0 (that is, before the 10:30 EIA announcement), C_t designates the median of the forecasts on Days 1, 2, and 3. It should be noted that C_t cannot be measured for forecasts released on Day 3 or for forecasts released on Day 2 if there were no forecasts on Day 3.

As shown in Table 2, there is considerable variation in herding across the sample firms, and this is true for all three herding measures, however anti-herding predominates. The level of forecaster herding behavior is persistent. The correlation between the two subperiods overall is greater than 60%. Roughly 88% of the firms did not deviate in behavior between the two subperiods based upon the CJ_Dev measure. Of the firms that were anti-herders in the first period 12% shifted, becoming neither herders nor anti-herders. There is no evidence of herding behavior only anti-herding behavior, and the case where the firm neither herds nor anti-herds. We conclude that forecasting firms choose different herding strategies, but generally anti-herding, and tend to stick to those strategies over time. However, while highly correlated between the two subperiods, the level of anti-herding behavior is slightly lower during the second period. We test whether the mean level of the herding variable differs between the two subperiods for each herding measure. The test results are reported near the bottom of Table 2. The tests reject at the 1% level the null of equality of means between the two subperiods for two of the herding variables.[27] The results suggest that forecasters shifted their herding behavior slightly between the two periods, but generally engaged in anti-herding.

6. Forecaster Accuracy, Herding and Timing

6.1 Preliminaries

The relation between forecast accuracy and herding or anti-herding rests upon both the implications of how forecasters weight private versus public information when composing their

[26] In unreported results we present statistics showing that over all forecasting firms each of the three herding measures confirms that anti-herding is the dominant behavior. The results are available from the authors upon request.
[27] The p-values for tests of mean equality between subperiods: BCK (0.007), CJ_Dev (0.091), CJ_Prob (0.000).

forecasts as well as the choice of when to release a forecast. Consider the weighting scheme. Given an optimal weighting scheme, the prediction would be that forecasts based upon those weights would have a lower mean squared error than forecasts that tilt away from those weights in either direction. The expectation is therefore that accuracy will be worse for both anti-herders and herders relative to forecasters who neither herd nor anti-herd (those who form their forecasts optimally).

If the focus on ESG during the post 2012 period had the consequence of a greater and more in-depth focus on fundamentals we expect that overall accuracy should have improved regardless of which of the three behaviors, anti-herding, herding, or neither is dominant. Conversely, if such attention distracted from a focus on fundamentals, then accuracy may have degraded. Likewise, the relation between herding, anti-herding and accuracy after 2012 could have shifted if the increased attention to ESG brought more fundamental information to bear on forecasts and the weighting of public and private information but could have alternatively resulted in a shift away from optimal weighting with accuracy becoming more strongly linked to the choice to herd or anti-herd.

The timing of a forecast's release relative to when the actual is reported involves a trade-off between accuracy and factors that may influence supply or demand for forecasts other than accuracy. If fundamental information accumulates, allowing better forecasts to emerge as the EIA report date approaches, we should see that later forecasts are more accurate than early forecasts. This of course raises the question then of why early forecasts would be released. Rosen (1981) and Scharfstein and Stein (1990) have argued that a "superstar" (forecaster) effect may manifest if the best forecasters are rewarded monetarily but also with reputation which may include the reputation of being first or breaking away by being early. Ehrbeck and Waldmann (1996) define what they label "rational cheating" as a description of the situation in which forecasters, "…balance their aims of minimizing forecast errors and looking good before the outcome is observed" (p. 22) giving rise to biased but early forecasts. Laster, Bennett, and Geoum (1999) develop a model of competition amongst forecasters who are compensated for both the accuracy of their forecasts and the attraction (through publicity) that their forecasts bring to the firms that employ them, essentially a reputation effect. The upshot is that each forecaster faces a tradeoff between accuracy and publicity and will make the choice depending on how their employer weights, either explicitly or implicitly, the two factors in setting the forecaster's compensation. The relation between forecast timing and accuracy is of course an empirical question. We test whether later forecasts are more accurate than earlier forecasts

Likewise, if the attention shift was towards improved analysis of all fundamentals including ESG, then some analysts may have adjusted their behavior and begun issuing their forecasts early because of their belief that their forecasts were more accurate, responding as well to increased demand for their forecasts. Consequently, we could see a change in the relation between accuracy and the timing of forecast release during the period following 2012.

The impact of timing and herding on forecast accuracy may however depend also upon on the confluence of the two. That is, there may be an independent effect of each influence but also an interaction effect. For instance, as the EIA release date approaches and forecasters' information becomes more consistent and similar, the marginal benefit to the analyst whose forecast signals unique private information may be larger than for forecasts issued at earlier dates. However, the marginal cost to the analyst may be higher for issuing a less accurate forecast when other forecasts are quite consistent and closer to one another. Which scenario dominates is an empirical question. We account for these effects by incorporating an interaction variable that is the product of the herding and timing measures.

We conclude with a comment on the shift in forecast issuance timing that we observe at the firm-level. Overall, there was a shift during the second subperiod to a greater frequency of early forecasts. The mean number of days that forecasts were issued prior to the EIA report equaled the following: Subperiod 1: 1.29, Subperiod 2: 1.35 based upon a total sample size of 19,677 individual forecasts. The difference is statistically different from zero at the 1% level.

6.2 Estimation Results

We first present between firm regression results, which emphasize forecast accuracy variation across forecasting firms. We standardized all variables, and then computed the average values for each firm across the sample observations for that firm. The use of standardized variables enables us to assess the relative impact of the explanatory variables on forecast accuracy, as the raw variables are of different scales. We regress the average values for our standardized measure of RAFE (relative absolute forecast error), for each of the firms in the sample on: (1) the standardized average herding measure for that firm, (2) the average standardized number of days between the firm's forecast release and the EIA announcement and (3) an interaction variable equal to the product of the herding and timing variables. We account for heteroskedasticity in coefficient tests by using Huber-White standard errors. Using standardized variables allows us to emphasize the relative impacts of the explanatory variables on our measure of accuracy. The estimated coefficients equal the corresponding changes in the standard deviation of the dependent variable, RAFE, due to a one unit increase in the standard deviations of the explanatory variables. Table 3 presents the results for models in which we test the null hypotheses that there is no relation between a herding measure and accuracy, and likewise the null that there is no relation between the timing of the forecast release and accuracy. We also test whether there were any changes in the relations between accuracy and herding and timing following the increased emphasis on ESG at the end of 2012. We consider the ESG movement shift to have occurred external to the set of forecasters we study, being an economy-wide phenomenon. As such we consider this to be a shock and test for shifts in the coefficients of the herding and timing explanatory variables. All forecasters in our first set of results are considered to have been shocked by this event. We define a dummy variable D1 equal to 1 for observations during the period *1/1/2013-12/9/2021* and 0 otherwise.[28] The dummy variable is interacted with the other explanatory variables and is included as a constant effect as well. The estimated coefficients of the interaction variables allow us to test whether there was a statistically significant change in the impacts of timing and herding on accuracy between

[28] To address whether the results are sensitive to controlling for time using a dummy variable to distinguish observations after 2012 we estimated two alternative specifications where the base model is that estimated in Table 4. First, we replaced the dummy variable with a linear time trend variable which increases linearly over the sample years. The estimation results for the time trend specification showed no significant effects of the time trend on forecasters' tendency to herd, timing choice, or the interaction between herding and timing. Second, we replaced the period dummy variable with the Google Search Trend index for "ESG" (graphed in Figure 1.). The results of this specification also produced no significant differences from the original specification based upon the period dummy. These results are for brevity not reported but are available from the authors upon request.

the subperiods. Table 3 reports the results for a model which includes the interaction variables. We restrict the sample to those firms making 20 or more forecasts in either subperiod. Several results are noteworthy.

First, the estimated coefficients on each of the herding measures are positive and significantly different from zero at the 1% level. A more positive value for each of the herding measures implies anti-herding behavior, so we conclude that greater anti-herding intensity is associated with less accuracy. The estimated coefficients on the timing variable 'Days before EIA release' are also positive and generally significantly different from zero at the 1% level. Thus, early forecasters tend to be less accurate. Second, the estimated coefficients on the interaction variables equal to the period dummy times the herding measures are never significantly different from zero at the 1% level. These results suggest that there was no differential impact of herding (anti-herding) on accuracy between the two subperiods. Likewise, there was no differential impact of the relation between accuracy and the timing of the forecast release at the 1% level (variable 'Days before EIA release') between the two subperiods. Results based upon the full sample, irrespective of how many observations a firm had reported in either subperiod, are found to be similar to those reported in Table 3. Further, the average adjusted r-squared across the models presented in Table 3 is equal to roughly 50%, suggesting that timing and herding explain a large fraction of the variation in accuracy between firms.

The size of each estimated coefficient provides an indication of the relative importance of the variable. A one standard deviation change in the CJ_Dev herding measure (*Model 2* of Table 3) leads to roughly a .67 to .69 standard deviation change in the mean RAFE measure during both subperiods. Correspondingly a one standard deviation change in the timing variable (Days before EIA release) for the same model leads to a .25 to .24 standard deviation change in RAFE. We conclude that the forecaster's herding choice is more important to accuracy than the timing release choice, although both are individually important and statistically significant. The estimated coefficients on the interaction variable constructing as the product of a herding measure and the timing measure are not significantly different from zero for any of the estimation results, indicating there is no moderating effect of timing on herding. Likewise, this result does not shift between the subperiods.

Accuracy may also be influenced by forecaster ability. Ability should plausibly be positively related to experience. That is, if analysts who have made a large number of forecasts have used that experience to refine the models they use for forecasting, then we predict that more experienced analysts should on average be more accurate. *Model 2* includes a proxy for experience. The proxy variable we use is the number of forecasts produced by a firm which we label "Number of forecasts".[29] We add to the regression models the standardized number of forecasts issued by the firm over the sample period. The coefficient estimates for the variable measuring the number of forecasts are not significantly different from zero in either subperiod for two of the herding measure models but is positive and significant for the CJ_Dev model, however the overall results for herding and timing are consistent across all the models estimated under the *Model 2* structure.

The results presented in Table 3 as mentioned are between firm regressions utilizing individual firm averages. Those regressions do not capture any time-series variation within a single firm. We could potentially supplement those results by presenting panel regressions with fixed firm effects and weekly firm measures of timing and accuracy and an interaction variable for each subperiod. Table 2 suggests however that herding behavior is relatively invariant at the firm level. The expectation might therefore be that weekly measures of herding would likewise exhibit the time invariance. The herding measures however cannot be estimated on a week-by-week basis. But even if they could, to capture within firm effects of anti-herding we would need to include both the herding measures and fixed firm effects. But because herding (anti-herding) is relatively time invariant at the firm level, if we were to include weekly herding measures and firm fixed effects we would encounter a severe multicollinearity problem. Nevertheless, the results presented in Table 3 suggest that a large fraction of the accuracy variation is due to between-firm variation.

[29] We recognize that our measure of experience is a noisy proxy. A limitation of the measure is that some firms in the sample may have issued forecasts prior to the start date of our sample. This could potentially add noise and weaken any empirical relation between experience and accuracy in our data.

Our data do however allow us to estimate a disaggregated model in which we focus on individual analysts in contrast to firms. While not perfect, this allows us to estimate a panel data model that allows for within-firm variation of accuracy by estimating a model using forecaster-level forecast errors. We regress individual *i*'s relative absolute forecast error in week t on firm-level herding measures (BCK(*i*), CJ_Dev(*i*), or CJ_Prob(*i*)), analyst *i*'s choice of timing in week *t* (Days before EIA release(i,t)), the interaction of the two variables, and the analyst's experience. All variables are standardized. The primary weakness of the model is our use of the firm level herding measures as proxies for the individual forecaster herding measures and the assumption that these measures are both constant during each subperiod and are reflective of the individual analyst's behavior. We justify the use of the firm level herding measures in the regressions based upon the observation that in general each firm is represented by generally only one or two analysts and that the measures are relatively invariant between the two subperiods for most firms. Our assumption is that the invariance property holds generally week-by-week during each subperiod. We cannot however include firm fixed effects in the regressions due to the multicollinearity issue mentioned above. But the disaggregation implicitly captures some of the within-firm variation through the individual forecaster errors. However, as we are interested in time variation in accuracy we do include year fixed effects. Our sample ranges from 2003 to 2021, we include dummy variables for each year except year 2003 which serves as the base year. The year-fixed effects allow us to control for the influences of year-level omitted variables on individual analyst forecast accuracy. These could for instance include fundamental shocks to the gas industry as well as any time trend in forecast accuracy. The variable "Number of forecasts" which proxies for forecasting experience, is now time-varying. We use the number of forecasts before week t of the analyst i instead of the total number of estimates for a firm.

Table 4 presents the estimation results for the analyst-week specifications. In *Model 1*, we find results that are consistent with those presented in Table 3. Anti-herding and timing continue to have a positive and statistically significant influence on forecast accuracy. Earlier forecasts and those associated with greater anti-herders tend to be less accurate. The low r-squared values for the regressions however when compared to those for the between-firm regressions presented in

Table 3, suggest that it is the cross-sectional relation reflected in the between-firm regressions that is the primary explanation for accuracy variation. The *Model 2* results which include year fixed effects support this conclusion. Although the coefficient on the variable ”Number of forecasts” (our proxy for experience) is significant in *Model 1*, it is not significantly different from zero when year fixed effects are included in *Model 2*.

7. Additional Evidence on Anti-herding

The extent to which a forecaster anti-herds is a choice. The structure of the CJ_Dev measure offers an opportunity for us to explore anti-herding behavior at the level of the analyst. Our tests are designed to investigate several questions. First, was anti-herding the predominant form of behavior? Second, did the level of anti-herding change following the increased focus on ESG? Third, were better forecasters greater anti-herders or conversely, were poorer forecasters greater anti-herders? Fourth, was anti-herding related to the timing of an analyst’s forecast release? Equation (3) presents the basis for estimation of the measure, which involves a regression of the deviation of an analyst’s forecast from the actual at time t (F(i,t)-A(t)), on the deviation of the forecast from the consensus forecast, (F(i,t)-C(t)). Recall a positive coefficient is an indication of anti-herding, while a negative coefficient is an indication of herding. In this section we present results in which the focus is on the forecasts of the individual forecasters. We use the median of the prior forecasts as our estimate of the consensus. As this specification uses analyst-week level observations, we estimate two model variations, one that includes firm-fixed effects, where the firm is the company that employs the analyst, and a separate model that includes analyst (individual) fixed effects.

Was anti-herding the predominant form of behavior? The estimation results are presented in Table 5. All models use the same dependent variable, the individual forecaster errors (F(i,t)-A(t)). These are unbalanced panel models because forecasters might issue forecasts in different weeks over our sample years 2003-2021. *Model 1* is our baseline specification including as a right-hand side variable the forecaster’s deviation from the consensus F(i,t)-C(t)) and firm fixed effects. *Model 2* differs from Model 1 in that the specification includes analyst fixed effects in place of firm fixed effects. In both *Models 1* and *2* the estimate of the coefficient on the analyst forecast deviation from the consensus (F(i,t)-C(t)) is positive and statistically significant. These results support the conclusion drawn earlier that analysts in our sample tend towards anti-herding, that is, to present forecasts that overweight their private information. The fact that the estimated coefficient on (F(i,t)-C(t)) is positive indicates that generally forecasters anti-herd.

Did the level of anti-herding change following the increased focus on ESG? Given the results for *Models 1* and *2*, *Model 3* introduces a dummy variable, D1, which equals 1 for observations after the end of 2012 and an interaction variable equal to D1 times (F(i,t)-C(t)). The estimated coefficient on the interaction variable is an indicator of whether analysts tended to tilt more towards or away from anti-herding during the second subperiod. The coefficient on the dummy variable D1 has a negative sign but is not significantly different from zero. The coefficient on the interaction variable is both negative and statistically significantly different from zero. The net coefficient for the second period equals .698-.374 = .324 indicating that anti-herding by individual analysts during the second period was still present but less intense, similar to the firm-level results presented in Table 5.

Were better forecasters greater anti-herders or conversely, were poorer forecasters greater anti-herders? *Model 4* turns to the question of whether the top (more accurate) forecasters tended to anti-herd, herd or neither, and likewise for the bottom (least accurate) forecasters. The model is streamlined and includes as explanatory variables (F(i,t)-C(t)), firm fixed effects, and two interaction variables. If a forecaster was in the top 20 most accurate (forecasters who have the smallest average forecast error over all weeks), Top 20 equals 1 and otherwise 0. If a forecaster was in the bottom 20 least accurate then Bottom 20 (forecasters who have the largest average forecast errors over all weeks), equals 1 and otherwise 0. We interact both dummy variables with (F(i,t)-C(t)). The results indicate that the coefficient on (F(i,t)-C(t)) remains positive and significant and is equal to 0.547 indicating anti-herding is prevalent overall. However the coefficient on the interaction variable [Top 20 x (F(i,t)-C(t))] is negative and significant and equal to -.504. The sum of the two coefficients, 0.547+(-.504), is close to zero, although the sum is significantly different from 0, nevertheless indicating the top 20 forecasters tend more towards neither herding nor anti-herding but display choices consistent with acting optimally. Conversely the estimated coefficient on the interaction variable [Bottom 20 x (F(i,t)-C(t))] is positive and statistically significant, indicating that the worst forecasters display the greatest anti-herding behavior.

Was anti-herding related to the timing of an analyst's forecast release? Further did this relation change following the increased focus on ESG? *Model 5* extends the analysis by examining whether anti-herding differs across the days prior to the release of the EIA report. We define two dummy variables, the first, DA, takes the value 1 if the analyst's forecast is released 2 days prior to the EIA report and 0 otherwise. The second DB takes the value 1 if the forecast is released one day prior to the EIA report and 0 otherwise. We interact these dummies with (F(i,t)-C(t)). The forecasts released 3 days prior to the EIA report serves as the base case. The estimated coefficients on the two interaction variables provide estimates of whether the intensity of anti-herding changed as the EIA report date approached. We also interact the dummies DA and DB with [D1 x (F(i,t)-C(t))] to test whether the relations changed during the second subperiod. The estimation results for *Model 5* indicate that anti-herding predominates as found for *Models 1-4*. The estimated coefficient on the variable (F(i,t)-C(t)) is positive and statistically significantly different from 0 at the 1% level and equal to 1.047, indicating that forecasts issued 3 days prior to the EIA report are associated with intense anti-herding. Controlling for the timing of the forecast release and jointly for the sample subperiod, the coefficient on the base variable that accounts for a shift in anti-herding in general, the coefficient on [D1 x (F(i,t)-C(t))], while negative is not significantly different from zero at the 1% level. The result indicates that forecasts issued 3 days early during the second period show the same level of anti-herding as during the first period. However both coefficients on the variables that control for the day of the release are negative and statistically significant but are less in absolute value than the base coefficient on (F(i,t)-C(t)) which equals 1.047. Anti-herding intensity decreases between the third day and the second day prior to the EIA report date but then increases. The net anti-herding coefficient two days out equals 1.047-0.901 = 0.146, showing a decrease in intensity but nevertheless that anti-herding is present, and one day out equals 1.047-0.319 = 0.728. The increase in intensity of anti-herding as the EIA release date approaches potentially is in response to an effort by forecasters to distinguish themselves as well as potentially satisfy the demand for unique forecasts as discussed in Section 6.1. These results do however support the overall conclusion that anti-herding is always the norm.

Controlling for the subperiod there is an incremental statistically significant decrease in anti-herding intensity one day out during the second period, the net intensity during the second subperiod equals 1.047-0.319-0.344 = 0.384, as compared with the net intensity during the first period of 1.047-.319 = .728. This was not the case for the forecasts released 2 and 3 days ahead of the EIA release. Thus, while the intensity of anti-herding did change for forecasts issued one day out, our conclusion that anti-herding continued to be the dominant form of behavior.

Model 6 presents the results of the full model. Here we see that the conclusions drawn for the disaggregated *Models 1-5* continue to be supported when all explanatory variables are included. Summarizing the results based upon *individual forecaster behavior*, 1) anti-herding is the dominant form of behavior in both subperiods studied, 2) anti-herding intensity first decreases and then increases as the EIA report date approaches relative to the earliest forecast dates, 3) the top 20 forecasters in terms of accuracy neither herd nor anti-herd while the bottom 20 forecasters tend to anti-herd. The result suggests that the most accurate forecasters behave as if they weight their private information and the public consensus view optimally, 4) during the second subperiod, anti-herding by early forecasters, those issuing forecasts 2 or 3 days prior to the EIA release, exhibits the same relative behavior as during the first subperiod, but as mentioned above anti-herding intensity declined on the day before the release of the EIA report.

8. Herding and Forecast Dispersion

A related issue is forecast dispersion. In Table 6, we present evidence on how forecast dispersion across forecasters differs across days prior to EIA report release. We measure dispersion as the standard deviation of forecast error across forecasters. Again, we present results for both subperiods. We calculate the mean standard deviation across all days with observations for each day relative to the EIA release. Column 3 presents the results by day prior to the EIA release for the first subperiod and column 6 the results for the second subperiod. The standard deviations decline monotonically across the days leading up to the EIA announcement in both subperiods. As reported in the table each day prior to the EIA release does not have an identical number of observations. We therefore also present average standard deviations for Days 0, 1, and 2 for the 114 weeks (first subperiod) and 386 weeks (second subperiod) with enough forecasts to make the results over all three days comparable. Those results virtually mimic the results presented in columns 3 and 6. In the table's final row, we report Welch's *F*-statistics for a test of the null that the mean standard deviations do not differ for the four (or three) days. The null is rejected at the 1% level in both tests. Clearly, forecast dispersion declines over time as the EIA announcement day approaches. Overall dispersion is *smaller* during the second subperiod, and this is true for each day prior to the EIA release. This result is also consistent with overall scrutiny increasing during the second subperiod.

We see two possible reasons for the decline in forecast dispersion each day prior to the EIA release. One possibility is that forecasters herd. The forecast dispersion pattern is consistent with the proposition that analysts observe others' forecasts and adjust their forecasts towards the consensus before releasing their numbers. This behavior will lead to forecasts that become more similar as the EIA release date approaches. A second possibility is that, as the time until the EIA announcement declines, forecasters accumulate better and more similar information, leading to more similar forecasts. The presence of anti-herding behavior among the forecasters we study however suggests that the decline in forecast dispersion shown in Table 6 as the EIA report date approaches is likely caused by the arrival of additional fundamental information rather than forecasters herding by overweighting prior forecasts. As we have also suggested, the decrease in dispersion between the subperiods could potentially be the result of increasingly widespread attention on the part of analysts brought on by an increased focus on this market.

Separately, the smaller dispersion of forecast errors across forecasters in the second period could be due to more widespread herding behavior during that period. Such would not necessarily reduce the overall error rate but would produce similar error rates and thus a lower cross-sectional dispersion. However, the smaller dispersion could also be due to forecasters drilling down more on fundamentals during the second period so that observationally the errors are again similar. The fact that the results presented in Table 6 show that anti-herding or neither anti-herding nor herding, are the norms indicates the lower dispersion is not due to herding.

Notice, that as mentioned earlier, dispersion overall declines during the second subperiod. This is consistent with the hypothesis that greater focus on ESG also led to greater overall analysis of fundamentals which caused analyst forecasts to be more alike. Consistent with this conclusion is the result shown in Table 1 that overall accuracy improved during the second subperiod. Likewise, the results presented in Table 2 do not support the hypothesis that convergence was due to herding since the evidence is more supportive of analysts either anti-herding or neither herding nor anti-herding.

9. Why Does Demand for Poor Forecasts Persist?

The results presented in Tables 3 and 4 indicate that forecasters exhibiting strong anti-herder tendencies are the least accurate. Because anti-herders could improve the accuracy of their forecasts by placing more weight on the existing forecasts of others when forming their forecasts, two questions arise: (1) Why do anti-herders persist in anti-herding and in the issuance of less accurate forecasts? and (2) Why do less accurate forecasts remain in demand? We posit that forecast users may be willing to give up some accuracy in return for the accumulation of private information conditional on the consensus forecast built from an ensemble of individual inaccurate forecasts being unbiased. A pure herder who simply repeats others' forecasts brings no new information to the market. In contrast, anti-herders who base their forecasts solely or primarily on their private information may bring new and valuable information to the market. So anti-herders may provide their clients with additional perceived valuable forecasts the consensus of which is unbiased. In this section we present evidence that, despite being less accurate, the consensus of the less accurate forecasts is nevertheless unbiased. This could explain why less accurate forecasters persist in anti-herding when a move toward herding would improve their forecast accuracy and why there is continued demand for their persistently less accurate forecasts.

Laster, Bennett, and Geoum (1999) develop a model in which forecasters are compensated for the accuracy of their predictions as well as the attraction (publicity) that their forecasts bring to the firms employing them. A conclusion drawn from the equilibrium of the model is that individual forecasters may issue inaccurate forecasts, but the consensus forecast will "… be unbiased, or nearly so" (Laster et al., 1999, p. 17). Users who understand this and who demand an unbiased forecast will employ a consensus of the ensemble of forecasts, accepting some inaccuracy in return for an unbiased consensus. Scharfstein and Stein (1990) also make the case that reputation could be an important element of an analyst's compensation in addition to accuracy.

We test whether the consensus of the least accurate forecasts issued on days 2 and 3 is unbiased. Treating the top 10 most accurate firms on days 2 and 3 as the most accurate and the remainder as less accurate, we test the hypothesis that the less accurate group's consensus forecast of the change in storage is an unbiased predictor of the actual EIA-reported values using all weeks for which forecasts by these firms were issued. Our test is based upon parameter estimates for the model $A_t = \alpha + \beta \overline{F}_{p,c,t} + e_t$ for the less accurate forecasters. The variable A_t is the actual storage change at t, $\overline{F}_{p,c,t}$ represents the consensus (median) forecast of the change in storage at t for the less accurate group, and e_t is a mean zero constant variance error. If $\alpha = 0$ and $\beta = 1$, we can conclude that the consensus forecast of the less accurate group is unbiased.[30] We estimate the model separately for both subperiods. The F statistic for the test of the null hypothesis ($\alpha = 0$ and $\beta = 1$) is F= 3.48, p=0.031 for the first subperiod, and is F= 0.37, p=0.694, for the second subperiod. Based upon this result, we cannot reject the null hypothesis that the consensus forecast of the less accurate group is unbiased at the 1% level in both periods, though confidence in the inference improves during the second period. These results are consistent with two predictions of the Laster et al. model: 1) inaccurate forecasts exist in equilibrium and 2) the consensus of the ensemble of inaccurate forecasts is unbiased. The fact that the ensemble of inaccurate forecasts is nevertheless unbiased provides evidence for why such forecasts are demanded and because of the connection between inaccuracy and the level of anti-herding, why anti-herders are observed. The conclusion applies not only to the pre-2013 period but also to the second subperiod during which the intensity of focus on ESG accelerated.

[30] We follow Mincer and Zarnowitz (1969) and many subsequent researchers in the specification of the model as $A_t = \alpha + \beta P_t + u_t$, where A_t represents the actual realization and P_t is the prediction (forecast). We find that the test result is robust to accounting for autocorrelation and heteroskedasity of the error term.

10. Robustness Tests

In addition to the alternative controls for time described in footnote 30, we estimated several supplementary models. The base structure of the estimated models of accuracy is as shown in Table 4. First to test whether the basic results are sensitive to the choice of forecast we use when computing forecast errors, we repeat the estimation using only the one day ahead forecasts and find the basic results are unchanged. Second, to control for the potential that an information shock received prior to the forecast release might influence the results, we include the lagged standardized oil futures price change. Here the assumption we are making is that any information shock will reflect itself in a change in the futures price. The basic results are unaffected. Finally, based upon information obtained from the source for the forecasts in the sample, Bloomberg, we know the order of the forecast releases for each date for the year 2021. Based upon this subset of data, which represent within day releases, we find that our base results on herding and accuracy are further confirmed.

11. Conclusions

The atmosphere created by the 'Environmental, Social, and Governance' (ESG) movement has had what many believe to be a profound impact on economic behavior and decisions. We present evidence on one dimension of this potential impact, the accuracy of forecasts made by professional analysts that focus on the energy sector. The energy sector is arguably a sector that has received intense 'ESG' focus by both investors, industry participants, the media and public policy makers in recent years. Changes in natural gas in storage reported by the U.S. Energy Information Administration (EIA) are closely followed by energy market participants and are known to reflect changes in supply and demand conditions. A professional cadre engage in forecasting such changes which serve as the platform for our analysis. We present evidence that the accuracy of forecasts of the changes in natural gas storage levels is strongly correlated with and largely determined by (1) the forecaster's tendency to herd and (2) how long before the EIA announcement the forecast is released. Specifically, anti-herders' forecasts tend to be much less accurate than those who neither herd nor anti-herd, and early forecasts are less accurate than late forecasts. These two variables account for roughly 78% of the variation in forecast accuracy across forecasting firms (Table 3). Our results indicate that the influence of anti-herding on the accuracy of the forecasts studied experienced no significant change between the period prior to the increase in ESG focus (pre-2013) and the period of intense focus (post-2012) (Table 3). A similar result is found for forecast timing for two of the herding measures. On the other hand, we find that the mean absolute forecast error (AFE) and relative absolute forecast error (RAFE) for our sample of forecasters declined during the post-2012 period as did the standard deviation of forecast errors (Tables 1 and 6). These results are broadly consistent with the hypothesis that the increased ESG attention during the post-2012 era had the complimentary effect of increasing overall attention on fundamental supply and demand factors.

We find that forecaster accuracy varies greatly across forecasting firms and is highly persistent. The mean absolute forecast error of the least accurate forecasters is 1.9 times that of the most accurate during the pre-2013 period, but that ratio declines marginally during the post-2012 period, consistent with an overall reduction in forecast error. The less accurate forecasters rarely improve over time. This raises the question of how the least accurate forecasters continue to successfully sell their services; that is, what services do they provide to clients? Our results indicate two potential reasons for the continued viability of less accurate forecasters. First, they tend to release their forecasts early, possibly providing a first-mover advantage that clients demand. Second, they tend to base their forecasts on their private intelligence rather than herding, thereby disseminating information that clients find useful. As evidence in support of this thesis, we show that the consensus of the less accurate forecasts in our sample is nevertheless an unbiased estimate of the actual change in natural gas in storage. The result is consistent with the hypothesis that there is demand for individual less accurate forecasts when users understand the consensus of the ensemble of these forecasts to be unbiased. The unbiasedness conclusion holds both before and after the increase in ESG attention.

We find that anti-herding behavior is the norm, differs substantially across forecasters, and is strongly persistent (Tables 2-5). Specifically, anti-herders continue to anti-herd but there is little variation in this behavior between the pre-2013 and post-2012 periods, suggesting that this behavior was relatively stable. This is despite calls for an increase in ESG focus during the latter period. We do find evidence however that forecasters with the highest accuracy tend to neither herd nor anti-herd, while the worst forecasters tend to strongly anti-herd (Table 5). Anti-herding exhibits a tendency to increase just prior to the EIA release, following a decrease in intensity, but this does not change between the subperiods. The increase in intensity of anti-herding as the EIA release date approaches is consistent with forecasters attempting to distinguish themselves as well as potentially strategically satisfy the demand for unique forecasts as discussed in Section 6.1.

Studies of forecaster accuracy differences in other markets have focused on cross-sectional determinants such possible determinants as forecaster experience, specialization, and available resources. Our evidence indicates the two primary determinants for the sample studied here are the proclivity to anti-herd and the timing of when the forecast is released. Little is known about how the impacts of these determinants is influenced by significant shifts in attention. Our evidence indicates that the increased attention to ESG issues, which forecasters in the energy sector have undoubtedly been confronted with, has had no significant impact on the relations between herding, timing and accuracy. While these relations persist, recall that overall forecast accuracy did improve during the post 2012 period. We do find a tendency for forecasters to issue forecasts earlier during the post 2012 period. One explanation for this change in behavior is that if there was an increase in focus on the part of analysts during the second subperiod, this could have increased their beliefs that their private information had improved in quality, leading them to issue more forecasts based upon those beliefs that they have acquired 'better' private information. The observation that the average forecast error declined between the 'pre-intensity ESG' period prior to 2013 and the 'post-intensity ESG' period is consistent with the view that this increased attention had the complimentary effect of increasing overall attention on factors driving supply and demand. Additional specifications involving alternative controls for time, the choice of forecast to evaluate and shocks to information prior to forecast release do not alter our basic inferences.

Although our study has focused solely on forecasters of changes in natural gas storage levels, our results potentially have implications for understanding forecaster accuracy for other economic variables and the influence of the shift of attention to all things 'ESG' on forecasting behavior in other domains.

References

Auh, J.K., J. Choi, T. Deryugina, T. Park. 2022. Natural disasters and municipal bonds, Working Paper, University of Illinois Urbana-Champaign.

Bernhardt, D., M. Campello, and E. Kutsoati. 2006. Who herds? *Journal of Financial Economics* 80: 657–675.

Bewley, R., and D. G. Fiebig. 2002. On the herding instinct of interest rate forecasters. *Empirical Economics* 27: 403–425.

Bikhchandani, S., D. Hirshleifer, O. Tamuz, I. Welch. 2021. Information cascades and social learning. *Working Paper*. University of Southern California and NBER Wp. 28887.

Bolton P., and M. Kacperczyk , 2022. Firm commitments. *Working Paper*, Columbia University and Imperial College

Brogger, A., A. Kronies. 2022. Skills and sentiment in sustainable investing, Working Paper, Copenhagen Business School.

Broughton, J. B., and B.J. Lobo. 2018. Herding and anchoring in macroeconomic forecasts: The case of the PMI. *Empirical Economics* 55: 1337–1355.

Brown, L. D. 2001. A temporal analysis of earnings surprises: Profits versus losses. *Journal of Accounting Research* 39: 221–241.

Brown, M. B., and A. B. Forsythe. 1974. Robust tests for the equality of variances. *Journal of the American Statistical Association* 69: 364–367.

Chen, Q., and W. Jiang. 2006. Analysts' weighting of private and public information. *Review of Financial Studies* 19: 319–355.

Chiou-Wei, S. Z., S. C Linn, and Z. Zhu. 2014. The response of US natural gas futures and spot prices to storage change surprises: Fundamental information and the effect of escalating physical gas production. *Journal of International Money and Finance* 42: 156–173.

Clement, M. B. 1999. Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics* 27: 285–303.

Clement, M. B., and S. Y. Tse. 2005. Financial analyst characteristics and herding behavior in forecasting. *Journal of Finance* 60: 307–341.

Clement, M. B, J. Hales, and Y. Xue. 2011. Understanding analysts' use of stock returns and other analysts' revisions when forecasting earnings. *Journal of Accounting and Economics* 51: 279–299.

Cooper, R. A., T. E. Day, and C. M. Lewis. 2001. Following the leader: A study of individual analysts' earnings forecasts. *Journal of Financial Economics* 61: 383–416.

Dechow, P., A. Hutton, and R. Sloan. 2000. The relation between analysts' forecasts of long-term earnings growth and stock price performance following equity offerings. *Contemporary Accounting Research* 17: 1–32.

Ederington, L. H., F. Lin, S. C. Linn, and L. Yang. 2019. EIA storage announcements, analyst storage forecasts and energy prices. *The Energy Journal* 40: 121–142.

Ehrbeck, T. and R. Waldmann. 1996. Why are professional forecasters biased? Agency versus behavioral explanations. *Quarterly Journal of Economics* 111: 21-40.

Fernandez-Perez, A., A. Garel, and I. Indriawan. 2020. Natural gas storage forecasts: Is the crowd wiser? *The Energy Journal* 41: 213–238.

Gallo, G., C. W. Granger, and Y. Jeon. 2002. Copycats and common swings: The impact and use of forecasts in information sets. *IMF Staff Papers* 49: 1–18.

Gay, G. D., B. J. Simkins, and M. Turac. 2009. Analyst forecasts and price discovery in futures markets: The case of natural gas storage. *Journal of Futures Markets* 29: 451–477.

Giglio, Kelly, and Stroebel, 2022 "Climate Finance", with Bryan Kelly and Johannes Stroebel, *Annual Review of Financial Economics*, forthcoming

Goldstein, I., A. Kopytov, L. Shen, H. Xiang. 2022. On ESG investing: Heterogeneous preferences, information, and asset prices, Working Paper, University of Pennsylvania.

Gu, C., and A. Kurov. 2018. What drives informed trading before public releases? Evidence from natural gas inventory announcements. *Journal of Futures Markets* 38: 1079–1096.

Halova, M.W., Kurov, A., Kucher, O., 2014. Noisy inventory announcements and energy prices. *Journal of Futures Markets* 34, 911–933.

Hart, O., and L. Zingales. 2022. The new corporate governance. *University of Chicago Business Law Review*, forthcoming.

Hirshleifer, D., Y. Levi, B. Lourie,, S.H. Teoh. 2019. Decision fatigue and heuristic analyst forecasts. *Journal of Financial Economics*, 133, 83-98.

Hirshleifer, D., 2020. Presidential Address: Social transmission bias in economics and finance. *Journal of Finance* 75, 1779-1831.

Hirshleifer, D., S.H. Teoh. 2003. Herd behaviour and cascading in capital markets: A review and synthesis. *European Financial Management* 9, No. 1, 2003, 25 – 66.

Hirshleifer, D., S.H. Teoh. 2009. "Thought and behavior contagion in capital markets", in *Handbook of Financial Markets: Dynamics and Evolution. Handbooks in Finance*, eds. Thorsten Hens and Klaus Schenk-Hoppe, Chapter 1, 1–46. Amsterdam, The Netherlands: North-Holland.

Hirst, D. E., P. E. Hopkins, and J. M. Wahlen. 2004. Fair values, income measurement, and bank analysts' risk and valuation judgments. *The Accounting Review* 79: 453–472.

Hong, H., J. D. Kubik, and A. Solomon. 2000. Security analysts' career concerns and herding of earnings forecasts. *Rand Journal of Economics* 31: 121–144.

Hong, H. , N. Wang, J. Yang. 2022. Welfare consequences of sustainable finance, Working Paper, Columbia University.

Huang, R., M. Krishnan, J. Shon, and P, Zhou. 2017. Who herds? Who Doesn't? Estimates of Analysts' Herding Propensity in Forecasting Earnings. *Contemporary Accounting Research* 34: 374-399.

Ivanov, I. (United States Federal Reserve Board of Governors), M. S. Kruttli, S. W. Watugala. 2022. Banking on carbon: Corporate lending and cap-and-trade policy, Working Paper, United States Federal Reserve Board of Governors.

Ivković, C. Z., and N. Jegadeesh. 2004. The timing and value of forecast and recommendation revisions. *Journal of Financial Economics* 73: 433–463.

Jegadeesh, N., and W. Kim. 2010. Do analysts herd? An analysis of recommendations and market reactions. *Review of Financial Studies* 23: 901–937.

Kacperczyk, M., J.L. Peydro. 2022. Carbon emissions and the bank-lending channel, Working Paper, Imperial College.

Ke, B., and Y. Yu. 2006. The effect of issuing biased earnings forecasts on analysts' access to management and survival. *Journal of Accounting Research* 44: 965–999

Kothari, S. P., E. So, and R. Verdi. 2016. Analysts' forecasts and asset pricing: a survey. *Annual Review of Financial Economics* 8: 197–219.

Lamont, O. 2002. Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behavior and Organization* 48: 265–280.

Laster, D., P. Bennett, and I. S. Geoum. 1999. Rational bias in macroeconomic forecasts. *Quarterly Journal of Economics* 114: 293–318.

Shangchen Li, S., R. Hongxun, S. Titman, H. Xiang. 2022. ESG Spillovers. Peking University and the University of Texas at Austin.

Lin, H. W., and M. F. McNichols. 1998. Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics* 25: 101–127.

Lindsey,L., S. Pruitt, C. Schiller. 2022. Cost of ESG investing, Working Paper, Arizona State University.

Linn, S. C., and Z. Zhu. 2004. Natural gas prices and the gas storage report: Public news and volatility in energy futures markets. *Journal of Futures Markets* 24: 283–313.

Michaely, R., A. Rubin, D. Segal, and A. Vedrashko. 2020. Do differences in analyst quality matter for investors relying on consensus information? Working Paper, Swiss Finance Institute.

Mincer, J., and V. Zarnowitz. 1969. The evaluation of economic forecasts. In J. Mincer (ed.), *Economic Forecasts and Expectations*, New York: National Bureau of Economic Research.

Naaraayanan, L., K. Sachdeva , V. Sharma. 2022. The real effects of environmental activist investing, Working Paper, London Business School.

Pastor, L., R. F. Stambaugh, L. A. Taylor, 2022. Dissecting green returns, Working Paper, University of Chicago.

Pierdzioch, C., J. C. Rulke, and G. Stadtmann. 2010. New evidence of anti-herding of oil-price forecasters. *Energy Economics* 32: 1456–1459.

Pierdzioch, C., J. C. Rulke, and G. Stadtmann. 2013a. Forecasting metal prices: Do forecasters herd?, *Journal of Banking and Finance* 37: 150–158.

Pierdzioch, C., and J. C. Rulke. 2013b. A note on the anti-herding instinct of interest rate forecasters. *Empirical Economics* 45: 665–673.

Prokopczuk, M., Wese Simen, C., and R. Wichman, R., 2021. The natural gas announcement day puzzle. *Energy Journal* 42: 91-112.

Ramnath, S., S. Rock, and P. Shane. 2008. The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting* 24: 34–75.

Richardson, S. A., S. H. Teoh, and P. D. Wysocki. 2004. The walk-down to beatable analyst forecasts: The role of equity issuance and insider trading incentives. *Contemporary Accounting Research* 21: 885–924.

Rosen, S., 1981. The economics of superstars. *American Economic Review* 7: 845–858.

Scharfstein, D., and J. Stein. 1990. Herd behavior and investment. *American Economic Review* 80: 465–479.

Sinha, P., L. D. Brown, and S. Das. 1997. A re-examination of financial analysts' differential earnings forecast accuracy. *Contemporary Accounting Research* 14: 1–42.

Spyrou, S. 2013. Herding in financial markets: A review of the literature. *Review of Behavioral Finance* 5: 175–194.

Stickel, S. E. 1992. Reputation and performance among security analysts. *Journal of Finance* 47: 1811–1836.

Trueman, B. 1994. Analyst forecasts and herding behavior. *Review of Financial Studie*s 7: 97–124.

Welch, I. 2000. Herding among security analysts. *Journal of Financial Economics* 58: 369–396.

Woolridge, J. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press.

Zitzewitz, E. 2003. Measuring herding and exaggeration by equity analysts and other opinion sellers.

Table 1
Descriptive statistics for forecasts and forecast errors for all individual forecasters

	Mean (1)	Median (2)	Standard Deviation (3)	Skewness (4)	Kurtosis (5)	Mininum (6)	Maximum (7)
A. Full sample (2003-2021): 19,677 obs							
Forecast storage change (Bcf)	1.620	40.000	95.876	-0.996	2.971	-350.000	135.000
Forecast error (FEr)	-0.368	0.000	12.149	-0.076	9.658	-145.000	97.000
Absolute forecast error (AFE)	8.542	6.000	8.646	2.904	19.421	0.000	145.000
Relative absolute forecast error (RAFE)	1.000	0.757	0.930	2.253	11.798	0.000	10.744
B. First period (2003-2012): 11,200 obs							
Forecast storage change (Bcf)	3.224	38.000	91.222	-0.956	2.789	-279.000	131.000
Forecast error (FEr)	-0.483	0.000	13.119	-0.220	9.587	-145.000	97.000
Absolute forecast error (AFE)	9.235	7.000	9.330	2.857	19.483	0.000	145.000
Relative absolute forecast error (RAFE)	1.078	0.823	0.992	2.043	9.978	0.000	10.744
C. Second period (2013-2021): 8,477 obs							
Forecast storage change (Bcf)	-0.500	42.000	101.665	-1.014	3.035	-350.000	135.000
Forecast error (FEr)	-0.216	0.000	10.739	0.293	8.501	-82.000	81.000
Absolute forecast error (AFE)	7.631	6.000	7.558	2.806	16.068	0.000	82.000
Relative absolute forecast error (RAFE)	0.898	0.688	0.829	2.587	15.607	0.000	8.626

Equality tests (First period=Second period)	FEr	AFE	RAFE
p-value: Mean equality (Median equality)	0.127 (0.095*)	0.000***(0.000***)	0.000*** (0.000***)

(continued on next page)

Table 1 (continued)
Descriptive statistics for forecasts and forecast errors for all individual forecasters

The table reports statistics for 19,677 individual forecasts by 102 different forecasters of the weekly EIA announcement of the change in total natural gas in storage in the lower 48 U.S. states in billions cubic feet (Bcf). The forecasts are compiled by Bloomberg and span 3/4/2003 through 12/9/2021. Forecast Error (FEr) is measured as the difference between the individual forecasted change and the actual change subsequently announced by the EIA. Absolute Forecast error is the absolute value of FEr. Relative absolute forecast error (RAFE) is the absolute forecast error standardized by monthly means. In Panel A, we report full sample statistics, in Panel B and Panel C, respectively, we summarize forecasts and errors from 2003 to 2012 and from 2013 to 2021. In the Equality tests panel, we test if the mean errors are equal between the two periods and we report p-values of the statistics. In the brackets we report p-values for equality of median tests. *** indicates rejection of the null that the means are equal between subperiods at the 1% level.

Table 2
Forecasting firm herding behavior by subperiod

	BCK		CJ_Dev		CJ_Prob	
Forecasting Firm	3/4/2003-12/31/2012	1/1/2013-12/9/2021	3/4/2003-12/31/2012	1/1/2013-12/9/2021	3/4/2003-12/31/2012	1/1/2013-12/9/2021
C.H. Guernsey	0.5296	0.5359	0.7726	0.7093	0.7763	0.7778
Capital One Securities Inc	0.6863	0.5354	1.0980	0.7005	0.7500	0.6667
Citi Futures Perspective	0.6366	0.5690	0.8234	0.8045	0.8387	0.7419
Citigroup Global Markets	0.5946	0.5173	0.2124	0.1487	0.6000	0.5289
ConocoPhillips	0.5780	0.5540	0.4456	0.1162	0.6189	0.5561
Cormark Securities Inc	0.7419	0.6342	0.4562	0.7261	0.7576	0.6429
Credit Suisse	0.5514	0.5517	0.2701	0.2116	0.5955	0.5741
Energy Overview	0.7753	0.7105	0.8262	0.7360	0.8000	0.7778
Fellon-McCord	0.7581	0.7067	0.8263	1.4690	0.7541	0.7027
FirstEnergy Capital	0.5509	0.5466	0.3144	0.2259	0.6220	0.6505
Gelber & Associates	0.6274	0.6721	0.6331	0.6190	0.6827	0.6895
Hencorp	0.7428	0.8019	1.0327	1.2347	0.7506	0.8100
ICAP Energy	0.6286	0.5385	0.5282	0.1228	0.6815	0.5517
ION Energy	0.5000	0.4708	0.1152	-0.0238	0.5444	0.4645
J.P. Morgan	0.5517	0.4781	0.1152	-0.0428	0.6091	0.4755
KLR Group LLC	0.5781	0.5194	1.1378	0.2965	0.6667	0.5161
Macquarie	0.4945	0.5065	0.2824	-0.0278	0.5636	0.5024
Noble Americas Corp.	0.5211	0.5556	0.1897	0.2063	0.5082	0.6000
Prestige Economics	0.6687	0.6136	0.7014	0.7021	0.7231	0.6250
Ritterbusch & Assoc	0.7339	0.6471	0.8442	0.6895	0.7608	0.6526
Societe Generale	0.5538	0.4722	0.1968	0.2125	0.5804	0.4571
Summit Energy	0.6022	0.6211	0.4296	0.6160	0.6985	0.6639

(continued on next page)

Table 2 (continued)
Forecasting firm herding behavior by subperiod

	BCK		CJ_Dev		CJ_Prob	
Forecasting Firm	3/4/2003-12/31/2012	1/1/2013-12/9/2021	3/4/2003-12/31/2012	1/1/2013-12/9/2021	3/4/2003-12/31/2012	1/1/2013-12/9/2021
SunTrust Robinson Humphrey	0.5979	0.7100	0.2399	0.5852	0.6118	0.7113
Tuohy Brothers	0.6769	0.6109	0.7732	0.4274	0.7500	0.6364
UBS	0.5874	0.5800	0.4107	0.1490	0.7123	0.5952
Wells Fargo Securities	0.7133	0.5085	0.8357	0.0791	0.7409	0.5068
Means	0.6223	0.5833	0.5581	0.4497	0.6806	0.6183
Medians	0.6000	0.5548	0.4922	0.3619	0.6906	0.6307
Test for Equality of Means (p-value)	(0.007)		(0.096)		(0.000)	
Correlation Coefficient	0.681^{***}		0.617^{***}		0.671^{***}	

BCK and CJ herding measures for various forecasting firms are reported for two subperiods: 3/4/2003-12/31/2012 and 1/1/2013-12/9/2021. BCK Herding Measure: BCK < .5 → Herding; > .5 → Anti-herding. CJ_Dev Herding Measure: CJ_Dev < 0 → Herding; > 0 → Anti-herding; = 0 → Neither. CJ_Prob Herding Measure: CJ_Prob < .5 → Herding; > .5 → Anti-herding. Values are computed across weeks for each firm displayed. We restrict the sample to forecasting firms with at least 20 forecasts in each subperiod. The last row of the table reports the correlation coefficient between the subperiods. Refer to Table 3 and Section 7 of the text for details on the computation of the herding measures. *** denotes significantly different from zero at the 1% level.

Table 3 Full Subsample (3/4/2003-12/9/2021), Period Interaction Specification
Forecast accuracy, herding strategy, and forecast timing; dependent variable firm-level average of standardized Relative Absolute Forecast Error

	Model 1			*Model 2*		
Variable	BCK	CJ_Dev	CJ_Prob	BCK	CJ_Dev	CJ_Prob
Constant	0.047	0.003	-0.002	0.047	0.003	-0.002
	(0.771)	(0.977)	(0.983)	(0.773)	(0.971)	(0.983)
Herding measure	0.600***	0.674***	0.633***	0.600***	0.696***	0.634***
	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Days before EIA release	0.589***	0.252***	0.200	0.591***	0.239***	0.199
	(0.001)	(0.002)	(0.020)	(0.001)	(0.002)	(0.027)
Herding measure x Days before EIA release	0.080	0.078	0.039	0.079	0.098	0.040
	(0.601)	(0.355)	(0.594)	(0.611)	(0.200)	(0.571)
D1	-0.079	-0.005	0.004	-0.076	-0.006	0.002
	(0.656)	(0.967)	(0.973)	(0.671)	(0.962)	(0.989)
(Herding measure) x D1	-0.022	-0.018	0.011	-0.012	-0.044	0.021
	(0.913)	(0.895)	(0.924)	(0.951)	(0.746)	(0.861)
(Days before EIA release) x D1	0.149	0.239	0.304	0.160	0.253	0.304
	(0.423)	(0.059)	(0.019)	(0.414)	(0.044)	(0.024)
(Herding measure x Days before EIA release) x D1	-0.196	-0.011	-0.058	-0.183	-0.029	-0.036
	(0.272)	(0.925)	(0.536)	(0.307)	(0.798)	(0.702)
Number of forecasts				-0.024	0.124	0.018
				(0.820)	(0.070)	(0.832)
(Number of forecasts) x D1				-0.096	-0.161	-0.157
				(0.467)	(0.124)	(0.169)
Adjusted r-square	0.435	0.595	0.566	0.431	0.595	0.567
Observations	107	107	107	107	107	107

Raw variables are standardized by subtracting the sample mean and dividing by the sample standard deviation. Results are 'between' regressions in which average values are regressed on average values, and averages are computed across non-zero observations. In Model 1, we regress the firm average standardized value of our measure of forecast accuracy RAFE on the firm average standardized values of the three herding measures, the standardized number of days the forecast is released prior to the EIA report release, and the interaction of the two explanatory variables, D1=1 for second subperiod (after 12/31/2012), the interaction of the dummy and herding measures, the interaction of the dummy and days prior to EIA release, and the interaction of the dummy, herding measures, and days before EIA release. The sample is the 81 firms with 20 or more forecasts in either subperiod. *Model 2* includes the firm average standardized value of the number of forecasts made by the forecasters, and an interaction variable with D1. The p-value for the test of the null that a coefficient equals zero is shown in parentheses and are calculated using Huber-White standard errors. *** indicates rejection of the null that the estimated coefficient equals zero at the 1% level.

Table 4 Full Sample ((3/4/2003-12/9/2021), Analyst-week panel specification
Analyst-week forecast accuracy, herding strategy, and forecast timing; dependent variable is analyst-week of standardized Relative Absolute Forecast Error

	Model 1			*Model 2*		
Variable	BCK	CJ_Dev	CJ_Prob	BCK	CJ_Dev	CJ_Prob
Intercept	0.001	0.012	0.005	0.487***	0.370***	0.370***
	(0.985)	(0.615)	(0.818)	(0.000)	(0.000)	(0.000)
Herding measure	0.104***	0.240***	0.249***	0.127***	0.193***	0.205***
	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Days before EIA release	0.068***	0.081***	0.072***	0.058***	0.066***	0.056***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Herding measure × Days before EIA Release	-0.018	0.016	0.002	-0.002	0.018	0.007
	(0.435)	(0.508)	(0.949)	(0.920)	(0.301)	(0.699)
Number of forecasts	-0.089***	-0.085***	-0.088***	0.044	-0.005	-0.009
	(0.000)	(0.000)	(0.000)	(0.109)	(0.830)	(0.615)
Year FE				YES	YES	YES
Adjusted r-square	0.036	0.069	0.071	0.074	0.092	0.097
Observations	17,465	17,465	17,465	17,465	17,465	17,465

Herding measures are standardized by subtracting the sample mean and dividing by the sample standard deviation. In *Model 1*, we regress the standardized value of forecast accuracy RAFE on the standardized values of the three herding measures, the number of days prior to the EIA release, the interaction of the two variables, and the experience of the forecast firms (the number of forecasts prior to week t). The sample consists of the 56 firms with 100 or more forecasts. Model 2 includes the year fixed-effect dummies. t-statistics are shown in parentheses and are calculated using Huber-White standard errors. ***indicates rejection of the null that the estimated coefficient equals zero at the 1% level.

Table 5 Full Sample ((3/4/2003-12/9/2021), Analyst-week panel herding specification
Analyst herding strategy; dependent variable analyst-level weekly forecast error, (F(i,t)-A(t))

Variable	*Model 1*	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
Intercept	-0.315***	-0.301***	-0.275***	-0.370***	-0.347***	-0.392***
	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)
Analyst Forecast minus Consensus	0.524***	0.529***	0.698***	0.547***	1.047***	0.984***
F(i,t)-C(t)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
D1			-0.054		0.120	0.108
			(0.763)		(0.424)	(0.524)
D1 × (F(i,t)-C(t))			-0.374***		-0.131	-0.025
			(0.000)		(0.213)	(0.780)
D1 × Top 20						0.176
						(0.484)
D1 × Bottom 20						-0.964**
						(0.020)
Top 20 × (F(i,t)-C(t))				-0.504***		-0.380***
				(0.000)		(0.000)
Bottom 20 × (F(i,t)-C(t))				0.453***		0.354***
				(0.000)		(0.000)
DA × (F(i,t)-C(t))					-0.901***	-0.745***
					(0.000)	(0.000)
DB × (F(i,t)-C(t))					-0.319***	-0.337***
					(0.000)	(0.000)
D1 × DA × (F(i,t)-C(t))					0.285	0.136
					(0.025)	(0.213)
D1 × DB × (F(i,t)-C(t))					-0.344***	-0.256***
					(0.001)	(0.002)
D1 × Top 20 × (F(i,t)-C(t))						-0.012
						(0.859)

Table 5 (continued)

Variable	*Model 1*	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
D1 × Bottom 20 × (F(i,t)-C(t))						0.081
						(0.357)
Firm FE	YES	NO	YES	NO	YES	YES
Analyst FE	NO	YES	NO	NO	NO	NO
Adj. R-squared	0.181	0.183	0.203	0.241	0.241	0.273
Observations	13,630	12,885	13,630	13,630	13,630	13,630

Individual analysts forecast errors are regressed on the deviations of their forecasts from the prior consensus forecast. Analyst forecast error is calculated as the difference between the forecast of analyst i at date t, F(i,t) and the actual change in natural gas storage in week t, A(t). The forecast deviation from the consensus is calculated as (F(i,t)-C(t)) where the consensus C(t) equals the median of existing forecasts for actual in week t. Model 1 includes firm fixed effect dummies Model 2, replaces the firm fixed effects with economist fixed effects. The dummy variable D1 equals 1 if the forecast was issued during the second subperiod and 0 otherwise. Model 3 includes D1, and the interaction of D1 and (F(i,t)-C(t)). 'Top 20' equals 1 if the forecaster is in the top 20 in terms of accuracy and 0 otherwise, while 'Bottom 20' equals 1 for the bottom 20 forecasters in terms of accuracy. Model 4 includes interaction variables that are based on interacting Top 20 and Bottom 20 with (F(i,t)-C(t)). The dummy variable DA equals 1 if the forecast was released 2 days prior to the EIA report release and 0 otherwise, and DB equals 1 if the forecast was release 1 day prior and 0 otherwise. Model 5 includes the interaction of DA and (F(i,t)-C(t)) and the interaction of DB and (F(i,t)-C(t)) and the three-way interactions of D1 and these two interaction variables. Model 6 includes all variables. Models 4 and 6 5 exclude firm fixed effects due to the multicollinearity of fixed effects with Top 20, Bottom 20. In the regressions including the firm fixed effect dummies, only forecasts from firms making more than 20 forecasts within 2 days of the EIA actual storage announcement days are used. In the regressions including the economist fixed effect dummies, only forecasts from economists who made more than 20 forecasts within 2 days of the EIA announcements are used. t-statistics are shown in parentheses and are calculated using Huber-White standard errors. *** indicates rejection of the null that the estimated coefficient equals zero, at the 1% level.

Table 6
Forecast dispersion by time until the EIA announcement by subperiod

	3/4/2003-12/31/2012			1/1/2013-12/9/2021		
Days Prior to EIA Release (1)	Weeks With Observations (2)	Mean Forecast Standard Deviation All Weeks (3)	Mean Forecast Standard Deviation Common Weeks (4)	Weeks With Observations (5)	Mean Forecast Standard Deviation All Weeks (6)	Mean Forecast Standard Deviation Common Weeks (7)
0	442	6.219	5.585	330	3.901	3.690
1	505	7.006	5.953	458	4.716	4.533
2	495	7.622	5.914	431	6.352	5.791
3	72	7.119		332	7.673	
Welch's F-value		6.13^{***}	0.41		19.57^{***}	33.62^{***}

Individual forecasts are grouped according to how far before the EIA's natural gas storage report the forecasts are released as reported on Bloomberg. Mean absolute forecast errors (AFE) are reported where forecast errors are measured as the difference (in bcf) between the forecast and the natural gas storage figure announced by Bloomberg. Means are also reported for relative absolute forecast errors (RAFE) where forecast errors are adjusted for seasonality as described in Section 5. F-values are reported for Welch tests of the null hypothesis that mean forecast errors are equal for all days prior to the EIA release. *** denotes rejection of the null hypothesis of no difference across days at the 1% level. Thirty-one observations where the forecasts were more than three days prior to the EIA release are excluded from the sample.

Appendix

Figure A1. Example Screenshot of the Ensemble of Forecasts of the Change in Natural Gas in Storage Surveyed and Assembled by Bloomberg L.P., Source: Bloomberg Finance L.P.

The release date of the actual change in natural gas storage is displayed at the top of the screen, in the example case, Thursday, August 25, 2022, the day and date of the EIA release. The Summary section reports statistics for the forecasts up to the date of the screen shot, units in billions cubic feet (Bcf). Reported at the bottom are the individual analysts' forecasts, including the economist's (analyst's) names, the firm the analyst is affiliate with, the value of the estimate (forecast), and the date the forecast was submitted. The data released by the EIA are measured as of the Friday prior to the Thursday on which the EIA report is released, in the example case the data are measured as of August 19, 2022. We collected the forecasts for each week from 3/4/2003 to 12/9/2021.

Table A1 Herding Measures

Herding Measure	Calculation Formulae	Herding Inference Indicator	Anti-Herding Inference Indicator	Neither Herding nor Anti-Herding
BCK	$= \left(\frac{\mathrm{N}(\mathrm{F_{jt}} > \mathrm{A_t} \text{and} \mathrm{F_{jt}} > \mathrm{C_t})}{\mathrm{N}(\mathrm{F_{jt}} > \mathrm{C_t})}\right)\left(\frac{\mathrm{N}(\mathrm{F_{jt}} > \mathrm{C_t})}{\{\mathrm{N}(\mathrm{F_{jt}} > \mathrm{C_t}) + \mathrm{N}(\mathrm{F_{jt}} < \mathrm{C_t})\}}\right) + \left(\frac{\mathrm{N}(\mathrm{F_{jt}} < \mathrm{A_t} \text{and} \mathrm{F_{jt}} < \mathrm{C_t})}{\mathrm{N}(\mathrm{F_{jt}} < \mathrm{C_t})}\right)\left(\frac{\mathrm{N}(\mathrm{F_{jt}} < \mathrm{C_t})}{\{\mathrm{N}(\mathrm{F_{jt}} > \mathrm{C_t}) + \mathrm{N}(\mathrm{F_{jt}} < \mathrm{C_t})\}}\right)$	< 0.5	> 0.5	$= 0.5$
CJ_Dev	$\left(\frac{k-h}{k}\right)_j$ in: $\mathrm{F_{jt}} - \mathrm{A_t} = \left(\frac{k-h}{k}\right)_j (\mathrm{F_{jt}} - \mathrm{C_t})$ The empirical estimate of CJ_Dev is the estimate β in the following regression model $\mathrm{F_{jt}} - \mathrm{A_t} = \alpha + \beta(\mathrm{F_{jt}} - \mathrm{C_t})$	$\left(\frac{k-h}{k}\right)_j < 0$	$\left(\frac{k-h}{k}\right)_j > 0$	$\left(\frac{k-h}{k}\right)_j = 0$
CJ_Prob	$= \frac{\#[(\mathrm{F_{jt}} - \mathrm{A_t}) \cdot (\mathrm{F_{jt}} - \mathrm{C_t}) > 0]}{\#[(\mathrm{F_{jt}} - \mathrm{A_t}) \cdot (\mathrm{F_{jt}} - \mathrm{C_t}) > 0] + \#[(\mathrm{F_{jt}} - \mathrm{A_t}) \cdot (\mathrm{F_{jt}} - \mathrm{C_t}) < 0]}$	< 0.5	> 0.5	$= 0.5$

BCK is due to Bernhardt, Campello, and Kutsoati (2006), CJ_Dev and CJ_Prob are due to Chen and Jiang (2006). Refer to Section 7 for background details. F_{jt} represents forecaster j's forecast of the change in the natural gas storage level at time t, A_t denotes the actual storage level change as announced by the EIA, and C_t denotes the median of all forecasts prior to j's forecast. # denotes 'the number of'. Herding in the case of the BCK measure can be identified by either of the first parts of the two multiplications, hence we employ a weighted average of the two in our analyses similar to what is employed by Bernhardt, Campello, and Kutsoati (2006).