

EXPECTATION FORMATION FOLLOWING LARGE, UNEXPECTED SHOCKS

Scott R. Baker, Tucker S. McElroy, and Xuguang S. Sheng*

Abstract—By matching a large database of individual macroforecaster data with the universe of sizable natural disasters across 54 countries, we identify a set of new stylized facts: forecasters are persistently heterogeneous in how often they issue or revise a forecast; information rigidity declines significantly following large, unexpected natural disaster shocks; and disagreement decreases among inattentive agents while it might increase for attentive ones. We develop a learning model that captures the two channels through which natural disaster shocks affect expectation formation: attention effect—the visibly large shocks induce immediate and synchronized updating of information for inattentive agents—and uncertainty effect—attentive agents might increase their acquisition of private information to compensate for the higher uncertainty after shocks.

I. Introduction

FORECASTS of macroeconomic variables drive government policy as well as corporate investment decisions. The micro-foundations of expectation formation are key to understanding the dynamics of these forecasts. Indeed, Manski (2018) concludes, “I urge measurement and analysis of the revisions to expectations that agents make following occurrence of unanticipated shocks.” This paper seeks to analyze forecaster responses to such shocks by matching a large database of individual macroforecaster data with the universe of natural disasters across 54 countries. We find that professional forecasters respond to the large, unexpected shocks in consistently different ways, depending on how often they issue or revise a forecast. As a result, the overall forecast accuracy and dispersion show interesting dynamics that has not been explored in the literature. We build a theory of information updating aimed at matching the stylized facts of expectations formation of professional forecasters.

Our theory has three key elements. First, agents are not interchangeable: attentive agents have a larger benefit from forecast accuracy and revise forecasts frequently; inattentive agents who do not benefit much from the accuracy of their forecasts make revisions infrequently and nonsystematically. Second, large and unexpected shocks induce inattentive agents to update to a greater degree, since the cost of not updating information is very high. Third, large shocks also induce attentive agents to purchase more private information.

Received for publication March 23, 2018. Revision accepted for publication December 12, 2018. Editor: Olivier Coibion.

*Baker: Northwestern University; McElroy: U.S. Census Bureau; Sheng: American University.

Versions of this paper were presented at AEA, the 5th Annual IAAE, SITE, the ECB, the Cleveland and Philadelphia Feds, the Bank of Finland, and Federal Reserve Board. We thank Philippe Andrade and Stefano Eusepi for discussions and anonymous referees and conference and seminar participants for many helpful suggestions. The views expressed here are our own and do not necessarily reflect the views of U.S. Census Bureau.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00826.

Our theory tells the following story about expectations formation following attention grabbing and unexpected shocks. First, visibly large shocks induce an immediate increase in updating of information for more inattentive agents (extensive margin). This attention effect is particularly pronounced for those with an outdated information set, resulting in a significant decline in their information rigidity. Second, for attentive agents, the occurrence of those shocks can generate increased uncertainty among them. To compensate for this higher uncertainty, some attentive agents increase their acquisition of private information (intensive margin). This acquisition is worthwhile if the benefit of doing so—the reduction of the mean squared error—is larger than the cost that resulted from buying additional pieces of private information.¹

The ingredients of our theory are motivated by our empirical findings. Our primary database of macroeconomic forecasts comes from Consensus Economics and covers a range of macro-variables across 54 countries. We focus on professional forecaster data primarily because such forecasters have a comparative advantage in allocating resources to acquire, absorb, and process information in forming expectations. As such, the degree of information rigidity found in professional forecasters likely forms a lower bound for other economic agents. Furthermore, the expectations of professional forecasters directly affect those of households (Carroll, 2003) and are used as inputs to the decisions of the representative agent (Ilut & Schneider, 2014).

Moreover, we focus on a survey that gives a significant amount of discretion to forecasters. That is, forecasters can choose to not report or not update their forecasts in any given month, and a significant number of forecasters exercise this capability. We find that forecasters are persistently heterogeneous in how often they issue or revise a forecast, with attentive agents submitting or revising forecasts every month, while inattentive agents provide and revise forecasts infrequently. The inattention channel that we study would be muted in surveys that require forecasters to report or update their forecasts for each period.

Another key element of our empirical analysis is the selection and identification of large, unexpected shocks. Our natural disaster data come from the Center for Research on the Epidemiology of Disasters and contain over 15,000 natural disasters. We limit our attention to unpredictable disasters like tornadoes, earthquakes, and storms rather than slower-moving disasters such as heat waves or epidemics. We

¹Our results are largely consistent with the work by Giacomini et al. (2017), who also explore the role of attention in forecast updating and forecast accuracy in both a theoretical and empirical setting in Brazil. They find that scarce attention is an important driver of forecaster behavior and that time-varying incentives on the part of the forecaster can greatly affect attention paid.

further limit our focus to significant disasters as measured by the number of people affected or killed and the monetary damages caused.²

To address the concerns that the disaster shocks are not fully unexpected and to better measure their impact, we construct a news-based measure of coverage of the disaster across several thousand English-language newspapers from the Access World News database. Using an index of the relative change in newspaper coverage regarding the disaster, we can flexibly identify disasters that have a relatively large impact and are unexpected. The large, unanticipated shocks identified in our paper are different from the so-called man-bites-dog signals in Nimark (2014), where observing those signals would change the probability distribution of the underlying variable. By contrast, the occurrence of those shocks increases all agents' uncertainty and induces synchronized information updating for most agents, but does not substantially change the underlying data-generating process in our framework.

By carefully matching the forecaster data with these natural disasters, we find that a large and unexpected shock induces a synchronized response and information updating among professional forecasters. Following those shocks, the overall dispersion among forecasters declines and forecast accuracy increases. These results may at first appear counterintuitive but are in fact a natural consequence of the unexpected shocks: disasters induce more inattentive agents to update, but also cause attentive agents to purchase more private information and be more attentive.

Our paper is closely related to the theoretical literature on expectations formation with information frictions. Notable examples include the sticky information model of Mankiw and Reis (2002) and Reis (2006); the noisy information model of Sims (2003), Woodford (2003), and Maćkowiak and Wiederholt (2009); the hybrid sticky-noisy information model of Andrade and Le Bihan (2013) and Andrade et al. (2016); and the Bayesian learning model of Lahiri and Sheng (2008) and Giacomini, Skreta, and Turen (2020).

In contrast to these papers, agents in our model are persistently heterogeneous in their type—attentive and inattentive—as seen in the data. In contrast to all previous work, we explicitly model agents' behavior following large, unexpected shocks. Our model also generalizes the noisy information model in two dimensions by allowing for (a) heterogeneous precision of private signals such that agents put different weights on private signals (relative to the same public information) and (b) time-varying precision of public signals in order to capture the increased uncertainty among economic agents following a shock. These features of the model enable us to measure state-dependent information rigidity in a multivariate context.

²Without this limitation, nearly every country is hit by at least one small natural disaster in nearly every period of our sample, muting any identifying variation.

Our empirical result that the degree of information rigidity significantly changes after the occurrence of large shocks adds to the literature relying on survey expectations to evaluate models with information frictions. Recent contributions, among many others, include Carroll (2003), Mankiw, Reis, and Wolfers (2004), Branch (2007), Coibion (2010), and Coibion and Gorodnichenko (2015). The findings from all of these papers firmly establish the presence of information rigidities in the expectations formation process. However, most papers treat the degree of inattention as a structural parameter, whereas we find that the visibly large shocks induce immediate and synchronized updating of information. This result supports state dependence in the information updating process as in Gorodnichenko (2008) and Maćkowiak and Wiederholt (2009).

Furthermore, all of these papers predict that following large shocks, disagreement among professional forecasters either increases or does not change significantly. However, using forecasts for a variety of macroeconomic variables across many countries, we document that disagreement can decrease following large shocks that affect forecaster attention. Our model is successful in explaining this apparent anomaly, while also matching other key features of the expectations formation process.

Finally, our paper makes contact with the literature on rare disasters.³ While the disasters used as shocks in our paper share some similarities to those in this literature, we turn our focus toward the effects on expectations formation rather than to any real economic effects.

The paper proceeds as follows. Section II describes the data used in this paper. We establish a set of new stylized facts about expectations formation following large, unexpected shocks in section III. We introduce the information structure that agents face in section IV. We propose a theory of expectation updating and illustrate the model implications through simulations in section V. Section VI concludes. Additional estimation and simulation results are relegated to the online appendix.

II. Data

A. Consensus Economics Forecast Data

Our database of macroeconomic forecasts comes from Consensus Economics. We use aggregated data from 1989 to 2014 covering 54 countries for which we can obtain both forecasts and true macroeconomic data.⁴ Consensus Economics

³For example, Cavallo, Cavallo, and Rigobon (2014) investigate the impact of two earthquakes in Chile and Japan on prices and supply disruptions. Orlik and Veldkamp (2014) posit that the realizations of disasters make economic agents more uncertain. Maćkowiak and Wiederholt (2015) develop a model in which agents make state-contingent plans in a rare event subject to information constraint. Baker, Bloom, and Terry (2017) study the effect of uncertainty on economic growth by using disasters of various types as natural experiments.

⁴The countries are Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, China, Colombia, Czech Republic, Denmark,

solicits the individual forecasts from professional forecasters: banks and financial firms, leading industrial companies, consulting firms, think tanks, and research groups. We observe the mean and standard deviation of individual forecasts for GDP and inflation in the current and the next calendar year. For these variables, panelists are asked about calendar-year predictions rather than a rolling period of twelve months. Consequently, forecasts mechanically become more accurate as forecasting horizon gets shorter. That is, forecasts for 2014 GDP will be significantly more accurate when solicited in December 2014 than in January 2014. Because of this feature, it is important to control for within-year variation in the timing of the surveys.

In addition to these aggregated data, we use individual forecast data for the G7 countries: Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States. The individual forecast data cover 1989 to 2014 and include forecasts for GDP, personal consumption, business investment, corporate profits, industrial production, producer prices, consumer prices, wages, car sales, housing starts, unemployment, current account balance, short- and long-term interest rates, and federal budget balances. Not all variables are covered for all countries, though forecasts for common variables such as GDP, inflation, and investment are well represented for the entire G7. There are 296 unique panelists across these countries, of which 225 have submitted at least 25 individual monthly forecasts to the survey.

Individual panelists are not required to submit a forecast each month and can choose the variables and countries to which they would like to respond. Additionally, panelists can submit an identical forecast from one month to the next. These features of this set of forecaster data are common across other widely used sources of forecaster data like Bloomberg forecasts and the Philadelphia Fed Survey of Professional Forecasters.

B. Disaster Data

Our natural disaster data have been obtained from the Center for Research on the Epidemiology of Disasters (CRED). These data contain over 15,000 extreme weather events such as droughts, earthquakes, epidemics, floods, extreme temperatures, insect infestations, avalanches, landslides, storms, volcanoes, fires, and hurricanes from 1960 to 2014. For each disaster, we can observe the event's category, its date and location, the number of deaths, the total number of people affected by the event, and the estimated monetary cost of the event. The CRED data include industrial and transportation accidents, which we exclude in our analysis.

For each country-month period, we give a value of 1 if a disaster has occurred and a 0 otherwise. This means that if a country has, for example, three earthquakes in one month, it receives a value of 1. The reason for this approach is to avoid double counting recurring but linked events within a month such as an earthquake with multiple aftershocks.

Because of the large number of disasters covered in the data, we need to apply a filter to focus only on major events. With this aim, we include a shock only if it fulfills at least one of the following conditions: (a) more than 0.00001% of a country's population dead (e.g., more than thirty dead in the United States); (b) more than \$10 million in damages, and (c) more than 50,000 people affected (e.g., made homeless, injured, substantial financial losses). Our results are robust to modification of filters for all three characteristics or by utilizing both relative and absolute filters. Below we discuss a weighting system to place higher weight on larger and more unexpected disasters.

Finally, we adjust the date of a disaster if it takes place after the Consensus Economics survey date in a given month. That is, if the June Consensus Economics forecast has already taken place, we attribute any further disasters in June to the month of July in terms of the first forecast that they could potentially affect.

C. Newspaper Data

Two potential concerns are that the disaster shocks that we use are not fully unexpected or unnoticed by a forecaster. In order to help alleviate these potential problems, we turn to a measure of unexpectedness and impact derived from news article mentions.

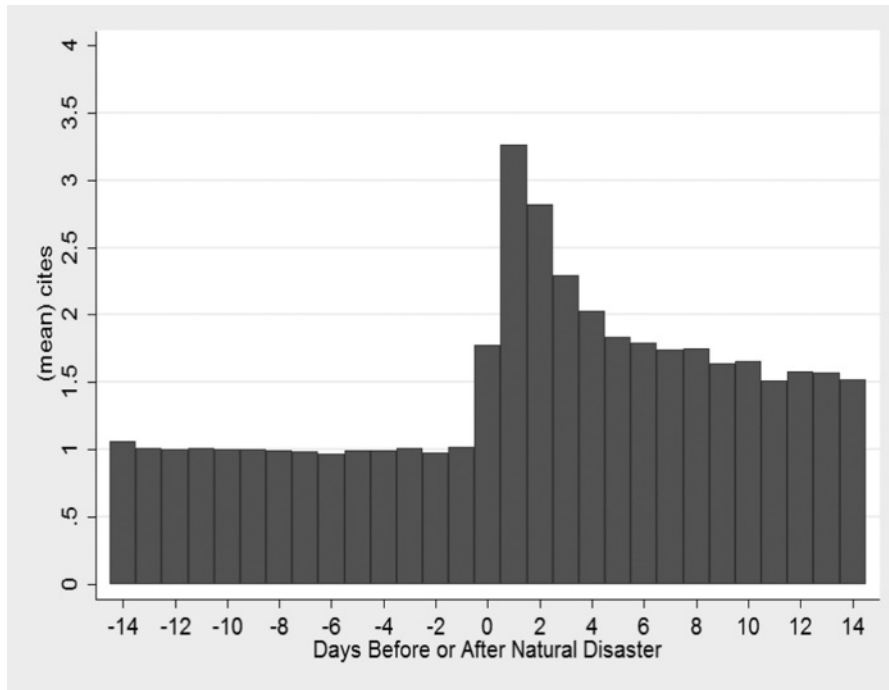
Using a database of newspapers from Access World News, we construct an index that measures the amount of news about a given country in the days surrounding each event.⁵ For each individual disaster, we search the newspaper archive for articles that mention the country where the disaster took place. For each of the fifteen days leading up to the disaster and fifteen days following it, we measure this count of articles and take the ratio of the postdisaster article count to the pre-disaster article count. Figure 1 shows an average of this series where each event's coverage has been normalized to 1 in the fifteen days prior to the event. A value of 2 at time 0 means that there are, on average, twice as many articles written that contain that country's name on that day relative to the pre-disaster average.

This process allows us to measure the change in attention, or at least newspaper attention, paid to a country following a disaster. This will enable us to flexibly distinguish between disasters that are relatively unimportant from those that are more newsworthy. Moreover, it will help us to filter out

Estonia, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Malaysia, Mexico, Netherlands, New Zealand, Norway, Pakistan, Peru, Philippines, Poland, Portugal, Romania, Russia, Singapore, Slovak Republic, Slovenia, South Africa, Spain, Sweden, Switzerland, Taiwan, Thailand, Turkey, United Kingdom, United States, Ukraine, and Venezuela.

⁵Access World News contains over 3,000 newspapers worldwide. We focus on newspapers from the United States, as they make up the majority of coverage in this database. For U.S. disasters, we look for changes in articles written that mention the state that the disaster occurs in rather than the country.

FIGURE 1.—CHANGES IN NEWSPAPER ARTICLES REGARDING AFFECTED COUNTRIES



Data obtained by searching approximately 2,500 English-language newspapers on Access World News. For each natural disaster, daily article counts of the number of articles written that contain the name of the affected country. This is averaged over all natural disasters studied in the regression analysis. For graphing purposes, the series for each event is normalized such that the preperiod has a mean of 1.

expected disasters that may have already been incorporated into previous forecasts. That is, if we observe a similar number of articles regarding the country before and after the event date, we can assume that the event was predicted ahead or it was not that important.

Our primary news-based scaling measure is the percentage increase in newspaper articles mentioning a given country in the five days after the event relative to the five days prior to the event. We use a relatively narrow window in order to minimize concerns about longer-term trends in coverage about various countries, but our results are robust to using up to fifteen-day windows. When using the news-weighted shocks, we turn to the shock with the highest jump in media citations for that category in that month.⁶

Table A1 displays some basic summary statistics regarding the indexes used to weight the disasters in our sample. We include two different scalings of the disaster index. The first is news scaling. The combined scaling refers to a combined z -score comprising of the news scaling, the monetary damages caused by the disaster, and the number of deaths caused by the disaster (mean of 1 and maximum of 4.5). The scaled disaster indexes are normalized to the same mean as the overall disaster index.

⁶For example, one such included disaster is Cyclone Xynthia, which hit France and other locations in Continental Europe on February 27, 2010. It caused over \$4 billion in damages, killed approximately fifty people, left over 1 million homes without power, and halted train and air traffic. Looking at newspaper coverage surrounding this event, we see a jump in articles after the storm hit of 53% relative to the period of five days before the storm.

Despite the filtering that we employ, most of the disasters in our data are not large enough to significantly affect the economies of the countries that we observe, especially when considering the offsetting stimulative aid or spending that follows a destructive event. That is, we do not see these natural disaster shocks as equivalent to large macroeconomic shocks (e.g., oil price shocks, changes in monetary policy, new trade deals or government spending initiatives) that would change the direction of a national economy. The primary impact on forecasters may be an increase in attention paid to a given country and its economy, with predicted growth remaining relatively stable. For larger macroeconomic shocks, uncertainty about future impacts on macrovariables may be significant enough to outweigh any decrease in forecast dispersion due to changes in attention among forecasters.

III. Empirical Results

A. State-Dependent Informational Rigidities

We first test for the presence of information rigidities in our sample of macroeconomic forecast data across G7 countries. Our specification takes the form noted in Coibion and Gorodnichenko (2015), utilizing data on the average forecasts across all individual forecasters in a given country-month, that is,

$$\begin{aligned} ForecastError_{i,t} = & \beta_1 ForecastRevision_{i,t} + Time_t \\ & + Country_i + \epsilon_{it}, \end{aligned} \quad (1)$$

TABLE 1.—STATE-DEPENDENT INFORMATIONAL RIGIDITIES (GDP)

Variables	(1) Forecast Error	(2) Forecast Error	(3) Forecast Error	(4) Forecast Error
Forecast Revision	0.514*** (0.0784)	0.555*** (0.0822)	0.530*** (0.0810)	0.541*** (0.0822)
Forecast Rev × Disaster		-0.286*** (0.104)		
Disaster		-0.0339 (0.0418)		
Forecast Rev × Disaster (News Scaling)			-0.106*** (0.0383)	
Disaster (News Scaling)			-0.0304 (0.0207)	
Forecast Rev × Disaster (Combined Scaling)				-0.183*** (0.0668)
Disaster (Combined Scaling)				-0.0617* (0.0319)
Observations	11,408	11,408	11,408	11,408
R ²	0.295	0.296	0.296	0.296
Number of groups	54	54	54	54
Time FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. Regressions performed for GDP forecasts across 54 countries. Forecast Error denotes the difference of the ex post true GDP growth value from the mean forecast. “Forecast Revision” denotes the difference of the mean GDP forecast from the previous month’s mean GDP forecast. “Disaster” is an indicator variable for a disaster occurring in a particular country-month for over 1,000 natural disasters in the sample. “News Scaling” refers to the ratio of news articles written about a country in the five days following a disaster to those written in the five days before a disaster (has a mean of 1 and a max value of 6). “Combined Scaling” for disasters refers to a combined z-score comprising the news scaling, the monetary damages caused by the disaster, and the number of deaths caused by the disaster (mean of 1 and maximum of 4.5). Standard errors are Driscoll-Kraay robust errors.

where $ForecastError_{i,t} = ActualValue_{i,t} - MeanForecast_{i,t}$ and $ForecastRevision_{i,t} = MeanForecast_{i,t} - MeanForecast_{i,t-1}$.

In the presence of information rigidities, β_1 would be predicted to be positive. That is, forecasters update periodically over time, and thus the mean forecast converges only slowly to the full-information forecast, driving a positive relationship between forecast revisions and the forecast error of any given period relative to the truth. To be as conservative as possible, we use Driscoll-Kraay standard errors throughout the empirical results.

In table 1, we restrict our analysis to forecasts of next-year GDP. In the first column, we find that the change in mean forecasts from month to month is strongly and positively related to the forecast error (table A1 displays summary statistics for forecast errors and forecast revisions). We interpret this as strong evidence for the presence of information rigidities in macroeconomic forecasting and that updating of forecasts is predictably less than complete in any given month. That is, because not all forecasters update in each period, the movement of the mean forecast goes only partway to the full-information forecast that is likely closer to the ex post true value for the period.

Columns 2 to 4 mirror this specification but add in interactions of $ForecastRevision_{i,t}$ with indicators for disasters and with disasters that are scaled by two different metrics. We find strong, negative coefficients on the interaction terms, demonstrating that in the month following a natural disaster, the correlation between forecast revisions and forecast errors weakens substantially. Column 2 uses a simple indicator for whether there was a natural disaster in country i in month t . In the month of a natural disaster, the strength of the relationship between forecast revisions and forecast errors falls by approximately 50% (-0.286/0.555).

Column 3 scales the disaster by the size of the increase in newspaper coverage surrounding the disaster. Here we see that not only do natural disasters tend to affect these informational rigidities, but they do so in a way related to the size or newsworthiness of the disaster. Given the maximum disaster “news scaling” is approximately 6, these coefficients indicate that a sufficiently large disaster reduces the relationship between forecast errors and forecast revisions to approximately 0. In contrast, a small disaster may not affect the relationship to any large degree, consistent with the idea that if a disaster does not merit mention in a newspaper, forecasters will likely be unaware as well.

Column 4 uses a scaling based on three factors: the number of deaths caused by a disaster, the monetary cost of the disaster, and the jump in news coverage. Each series is normalized to a standard deviation of 1, and then an average across all three metrics is taken. Both of the disaster scalings in columns 3 and 4 have an overall mean and standard deviation of 1 for nonzero values. Similar to our finding with only the news-based scaling, we again find that in general, more significant disasters drive down the relationship between forecast revisions and forecast errors. This result is consistent with the idea that professional forecasters may pay more attention to larger disasters and, as a result, update forecasts more frequently.⁷

Table 2 mirrors the earlier approach but demonstrates the correlation between forecast revisions and lagged forecast revisions, as in Nordhaus (1987). An advantage of Nordhaus’s test is that it is completely independent of the “true” values of the macroeconomic variables in question. Our findings follow

⁷These results are robust to the exclusion of any single type of natural disasters (e.g., excluding earthquakes or excluding hurricanes).

TABLE 2.—FORECAST REVISION LAGS (GDP)

Variables	(1) Forecast Rev	(2) Forecast Rev	(3) Forecast Rev	(4) Forecast Rev
Lagged Forecast Revision	0.139*** (0.0358)	0.149*** (0.0372)	0.145*** (0.0362)	0.148*** (0.0371)
Lag Forecast Rev × Disaster		−0.0736** (0.0296)		
Lagged Disaster		0.00223 (0.00976)		
Lag Forecast Rev × Disaster (News Scaling)			−0.0428*** (0.0143)	
Lagged Disaster (News Scaling)			−0.00943* (0.00557)	
Lag Forecast Rev × Disaster (Combined Scaling)				−0.0604** (0.0268)
Lagged Disaster (Combined Scaling)				−0.00893 (0.00744)
Observations	11,444	11,444	11,444	11,444
R ²	0.106	0.107	0.107	0.107
Number of groups	54	54	54	54
Time FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. Regressions performed for forecasts across 54 countries (including only forecasts of GDP). “Forecast Revision” denotes the difference of the mean forecast from the previous month’s mean forecast. “Disaster” is an indicator variable for a disaster occurring in a particular country-month for over 1,000 natural disasters in the sample. “News Scaling” refers to the ratio of news articles written about a country in the five days following a disaster to those written in the five days before a disaster (has a mean of 1 and a maximum value of 6). “Combined Scaling” for disasters refers to a combined z-score comprising the news scaling, the monetary damages caused by the disaster, and the number of deaths caused by the disaster (mean of 1 and maximum of 4.5). Standard errors are Driscoll-Kraay robust errors.

a similar pattern to our earlier results. We find that forecast revisions in the current month tend to consistently and positively predict those in the following month. This relationship diminishes substantially following a natural disaster, and the relationship between forecast revisions and lagged forecast revisions becomes increasingly weak as the disaster becomes larger. This again suggests that a large and unexpected shock precipitates a state-dependent response and an increase in information updating.

In both tables, we see that larger disasters can drive substantially larger responses by forecasters than smaller disasters. This is repeated throughout our results for both average and individual forecaster effects. Note that the largest of the disasters are several times larger than the median disaster, so even modest effects that can be observed by interacting simply with a disaster indicator are significantly magnified for many of the more notable and newsworthy disasters and events.

Tables A2 and A3 follow tables 1 and 2, but include forecast data across all variables in the sample (e.g., GDP, CPI, long- and short-run interest rates, unemployment, and consumption) and include forecast variable fixed effects. We find qualitatively similar results across all forecast variables as when restricting the analysis to GDP: all variables exhibit significant information rigidity that declines following large natural disasters.

B. Heterogeneous Individual Forecasters

Consensus Economics forecasts are also useful in that underlying forecast data from individual forecasters are available. Not only can we observe how the overall mean forecast

for a given country variable changes over time, but also how individual forecasters respond. Differences in the frequency and timing of forecast updates among individual forecasters can have significant impacts for the aggregate accuracy and dispersion of aggregate forecasts. With the individual forecaster data, we investigate the extent to which persistent heterogeneity among forecasters drives some of these differences.

We split forecasters into two groups. The “attentive” group is made up of forecasters who report a forecast for a given country variable in more than 95% of the months that they are present in the sample. The “inattentive” group is made up of forecasters who report forecasts less frequently (on average, reporting forecasts for 70% of months in the sample, with a minimum around 25%). This corresponds with approximately the top quintile and bottom four quintiles of forecaster reporting.⁸

Table 3 demonstrates some of the persistent differences across these two groups, controlling for time, country, and variable fixed effects. In columns 1 and 2, “Attentive Forecaster” is a binary indicator for being in the top quintile or bottom four quintiles of this measure of forecaster attentiveness. We measure how forecasters in these two groups are different from one another in two areas: absolute forecast error and absolute differences from the mean forecast. In both cases, we find that the more attentive forecasters have fewer persistent errors and deviations from the average forecast. Some of these errors are derived from more accurate updates

⁸We have performed similar exercises by measuring attentiveness as the fraction of forecasts that are different than the previous forecast for a given forecaster (e.g., Andrade & Le Bihan, 2013) and found qualitatively similar results.

TABLE 3.—INDIVIDUAL FORECAST DISPERSION

Variables	(1) Difference from Mean	(2) Forecast Error	(3) Difference from Mean	(4) Forecast Error
Attentive Forecaster	-0.0724*** (0.0135)	-0.363*** (0.107)		
Fraction Forecasts Reported			-0.177*** (0.054)	-0.779** (0.326)
Observations	292,391	156,718	292,391	156,718
R ²	0.054	0.120	0.054	0.120
Time FE	Yes	Yes	Yes	Yes
VAR FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Forecaster FE	Yes	Yes	Yes	Yes

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. “Forecast Error” denotes the absolute value of the difference of the individual forecast from the ex post true value. “Difference from Mean” means the absolute value of the difference between the individual forecast and the mean forecast for that country-variable-month. “Attentive Forecaster” is an indicator variable that notes that a forecaster is in the top quartile of fraction of forecasts reported. Data cover 280 forecasters and seven countries for GDP, CPI, consumption, short- and long-run interest rates, unemployment, wages, and producer prices. Standard errors are Driscoll-Kraay robust errors.

in forecasts, but they also stem from inattentive forecasters who fail to update for a given month and report outdated forecasts more often.

Columns 3 and 4 of table 3 dispense with the binary indicator of attentiveness and simply use the average fraction of total forecasts reported for a given forecaster while they were participating in the Consensus Economics panel. Again, we see that forecasters who submit more forecasts tend to have lower forecast errors relative to the truth and also smaller deviations from the mean forecast for any given country-variable-month.

A concern is that forecasters may shift between being attentive and inattentive over time. We find this type of switching to be uncommon. The top panel of figure 2 plots the fraction of forecasts reported against the fraction of forecasts reported in the previous year across all forecasters. We find a high degree of persistence in the reporting frequency across years, with inattentive forecasters likely to remain so over time, and the reverse is true for attentive forecasters. This may be driven by institutional features of the forecasters’ firm, where they may be assigned to update forecasts only infrequently and so do not respond to the Consensus Economics requests until a new forecast is made by the firm. The bottom panel of figure 2 plots the fraction of forecasts that are changed from month to month for a given forecaster rather than the fraction of forecasts reported.

C. Individual Forecaster Updating

In tables 1 and 2, we found that information rigidity, as measured through mean forecast revisions, changed significantly in response to natural disasters. Table 4 uses the individual forecaster data to examine the channels through which this reduction in information rigidity takes effect. In columns 1 and 2, we test the effect that natural disasters have on the individual likelihood of changing a forecast from the previous month. We include time, variable, country, and forecaster fixed effects to isolate the within-forecaster and within-variable impacts of these natural disasters. We find a positive impact (albeit insignificant, looking only at a natural disaster

indicator): natural disasters tend to increase the probability of changing an individual forecast by approximately 0.86 percentage points (on a mean likelihood of changing of approximately 50%). Column 2 demonstrates that this effect is larger for more newsworthy disasters, with disasters in the top decile driving an approximate 2.13 percentage point increase in the likelihood of a forecaster revising their previous month’s forecast (reflecting a scaled value for the top decile of disasters of approximately 3.5).⁹

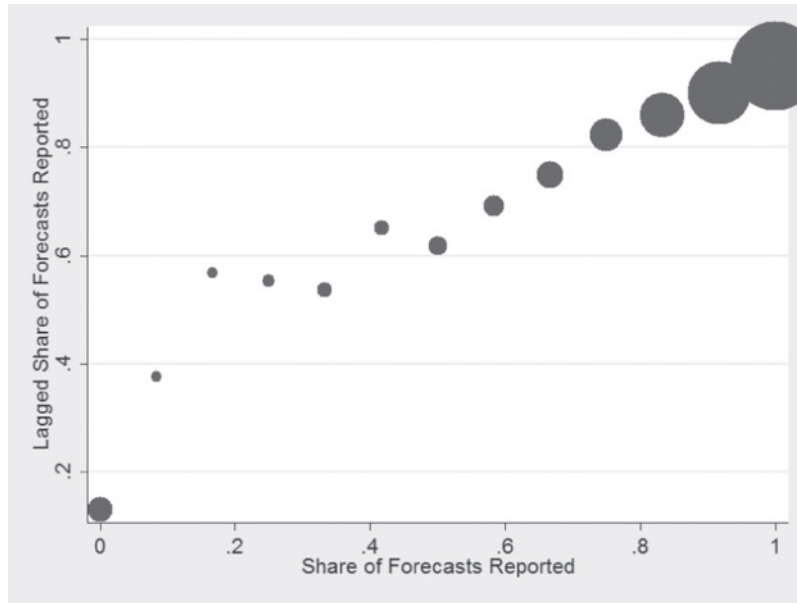
Columns 3 and 4 repeat this exercise but interacting with a measure of attention among forecasters. In both cases, we find that attentive forecasters are much less likely to update their forecasts in response to disasters. Inattentive forecasters tend to drive all of the combined effect, changing forecasts following disasters and especially following large disasters.

Table 5 demonstrates the counterintuitive result that disasters can actually decrease the dispersion, as measured by the squared distance from the mean forecast in a given month, among some groups of professional forecasts due to the effect that disasters have on inattentive forecasters. In columns 1 and 2, we see that following a disaster, attentive forecasters see moderate increases in forecaster dispersion (i.e., measuring the combined effect for quintile 5 of the disaster effect and the interaction effect). However, inattentive forecasters see a decline in this measure of dispersion, as they tend to update their forecasts after a natural disaster. In fact, the less attentive a forecaster is, the more his or her distance from the mean forecast decreases following a disaster.

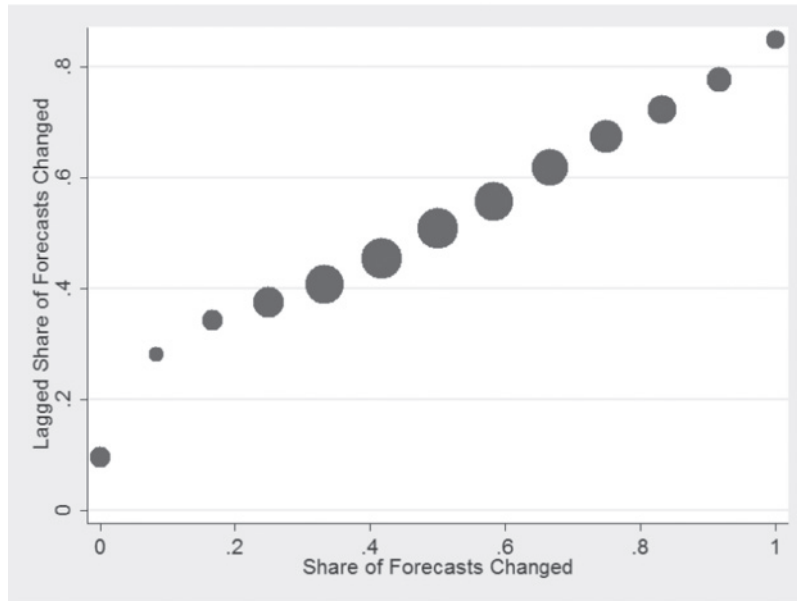
Columns 3 and 4 repeat this exercise but instead measure the impact of disasters across attentive quintiles on the squared distance from the true ex post value of a given forecast variable. Again, we find a slightly positive impact of natural disasters on forecast errors for attentive forecasters, while there are negative effects for the most inattentive forecasters. Even conditioning on a forecaster updating a forecast, inattentive forecasters tend to report more accurately following

⁹These results are robust to how we treat nonresponses—that is, if we fill in missing data with the previous month’s data for a forecaster who missed a month’s forecast or if we just exclude that month.

FIGURE 2.—PERSISTENCE IN FORECASTS



(a) Persistence in Forecast Reporting



(b) Persistence in Forecast Updating

In the top panel, the vertical axis represents the share of eligible months that a given forecaster reported a forecast for in year $t - 1$. The horizontal axis represents the share of eligible months that a given forecaster reported a forecast for in year t . Mean values of horizontal bins are plotted (each bin represents an increment of one month out of twelve months). Thus, a point on the 45-degree line means that, on average, forecasters in that group reported forecasts at the same frequency as the previous year. In the bottom panel, the vertical axis represents the share of eligible months where a given forecaster changed a forecast from the previous month in year $t - 1$. The horizontal axis represents the share of eligible months where a given forecaster changed the forecast from the previous month in year t . Mean values of horizontal bins are plotted (each bin represents an increment of one month out of twelve months). Plotted points are scaled by the number of forecasters in each bin.

a natural disaster than when changing their forecast without a disaster.

The magnitudes of these effects can be fairly large. For instance, for the least attentive quintile, the impact of the median disaster on accuracy is equivalent to the impact of being one month closer to the period the forecast is relevant for (e.g., making a forecast four months in advance rather than

five months in advance). For a disaster in the top decile of newsworthiness, the effect is equivalent to being almost four months nearer to the relevant period (e.g., making a forecast four months in advance rather than eight months in advance).

Overall, these results suggest that disasters induce inattentive agents to update their forecasts and, in doing so, actually move them closer to the mean forecast and also to the ex

TABLE 4.—INDIVIDUAL FORECAST CHANGES AND DISASTERS

Variables	(1) Change	(2) Change	(3) Change	(4) Change
Disaster	0.00857 (0.00872)	0.00192 (0.00631)	0.0311** (0.0143)	
Scaled Disaster		0.00576** (0.00283)		0.0157** (0.00617)
Disaster × Attentive Quintile			-0.00842** (0.00371)	
Scaled Disaster × Attentive Quintile				-0.00401*** (0.00150)
Observations	292,391	292,391	292,391	292,391
R ²	0.149	0.149	0.140	0.140
Time FE	Yes	Yes	Yes	Yes
VAR FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Forecaster FE	Yes	Yes	Yes	Yes

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. “Attentive Quintile” refers to the quintile of attentiveness a forecaster belongs to in terms of fraction of new forecasts that forecaster has reported during his or her tenure. “Disaster” is an indicator variable for a disaster occurring in a particular country-month for over 1,000 natural disasters in the sample. “Scaled Disaster” refers to the ratio of news articles written about a country in the five days following a disaster to those written in the five days before a disaster (has a mean of 1 and a maximum value of 6). “Change” is an indicator variable for whether that forecaster changed the forecast from the previous month for that country-variable. Data cover 280 forecasters and seven countries for GDP, CPI, consumption, short- and long-run interest rates, unemployment, wages, and producer prices. Standard errors are Driscoll-Kraay robust errors.

TABLE 5.—DISPERSION AND ACCURACY FOLLOWING DISASTERS

Variables	(1) Distance from Mean Forecast	(2) Distance from Mean Forecast	(3) Distance from Actual	(4) Distance from Actual
Scaled Disaster	-0.0109* (0.00586)	-0.0136* (0.00589)	-0.657** (0.275)	-0.675* (0.360)
Scaled Disaster × Attentive Quintile	0.00351 (0.00251)	0.00441* (0.00253)	0.139** (0.0491)	0.146** (0.0447)
Changed Forecast		-0.102*** (0.0130)		-1.341** (0.551)
Observations	292,335	292,335	156,686	156,686
R ²	0.056	0.057	0.101	0.102
Time FE	Yes	Yes	Yes	Yes
VAR FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Forecaster FE	Yes	Yes	Yes	Yes

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. “Attentive Quintile” refers to the quintile of attentiveness a forecaster belongs to in terms of fraction of new forecasts that the forecaster has reported during his or her tenure. “Scaled Disaster” refers to the ratio of news articles written about a country in the five days following a disaster to those written in the five days before a disaster (has a mean of 1 and a maximum value of 6). “Changed Forecast” is an indicator variable for whether that forecaster changed the forecast from the previous month for that country-variable. Columns 1 and 2 use the squared distance from the mean monthly forecast (excluding a forecaster’s own forecast) as the dependent variable. Columns 3 and 4 use the squared distance from the true (ex post) value of the forecast variable in question as the dependent variable. Data cover 280 forecasters and seven countries for GDP, CPI, consumption, short- and long-run interest rates, unemployment, wages, and producer prices in columns 1 and 2. Columns 3 and 4 include only data for GDP, CPI, consumption, and unemployment. Standard errors are Driscoll-Kraay robust errors.

post true value of the variable. Given the limited economic impact of most of the disasters in our sample, it is likely that for inattentive agents, natural disasters act mainly as an attention shock, prompting them to update outdated forecasts and converge toward a newer consensus value. To rationalize the empirical results, we develop a framework aimed at explaining the expectations formation process following natural disasters. In the next section, we introduce the information structure that agents face and derive the state-space representation. In section V, we explore some of the key properties of both types of agents, such as their information rigidity, forecast accuracy, and dispersion.

IV. Information Structure and State Space

A. Information Structure

We suppose that economic agents seek to forecast an m -dimensional signal process $\{\pi_t\}$ that is obfuscated by

noise. We envision the existence of both a public and a private channel for all agents.

Our model assumes that the i th agent ($1 \leq i \leq N$) observes the signal through the public channel contaminated by a common noise $\{\eta_t\}$, whereas the private channel provides the signal contaminated by private noise. The public noise $\{\eta_t\}$ is heteroskedastic and subject to potential first- and second-moment shocks. The public observation grid is described by a matrix B , which has m columns. Hence, the costless public data can be written as $B \pi_t + \eta_t$.

Private information is manifested through an observation grid A , a matrix of the same size as B ; here, it will be $m \times m$ -dimensional. One unit of private information consists of the signal contaminated by private noise, which is denoted v_t , and hence $A \pi_t + v_t$ would be observed. However, the agent has the option to purchase ℓ units of private information, which consists of ℓ i.i.d. replications of v_t , denoted by $v_t^{(\ell)}$. (Although ℓ depends on t , we suppress this in the notation.) Then letting $A^{(\ell)} = v_\ell \otimes I_m$, where v_ℓ is a vector of ℓ ones, the total private data received are $A^{(\ell)} \pi_t + v_t^{(\ell)}$. Hence the

overall data received, consisting of private and public portions, are

$$y_t^{(\ell)} = \begin{bmatrix} A^{(\ell)} \\ B \end{bmatrix} \pi_t + \begin{bmatrix} v_t^{(\ell)} \\ \eta_t \end{bmatrix}. \quad (2)$$

This notation makes the dependence on ℓ explicit, although it also depends on the profile of agent i . In fact, the next section describes how at each time t , an agent makes a selection of ℓ , based on preference for forecast accuracy and shocks in the economy.

Henceforth, we make the following assumptions about the processes. We suppose that $\{\pi_t\}$ follows a stationary VAR(p) stochastic process of dimension m ,

$$\pi_t = \Phi_1 \pi_{t-1} + \Phi_2 \pi_{t-2} + \dots + \Phi_p \pi_{t-p} + u_t, \quad (3)$$

where p is taken sufficiently large to approximate a generic signal. These can be generated at random using the bijective reparameterization of the stable VAR(p) class provided in Roy, McElroy, and Linton (2019).

The m -dimensional process $\{\eta_t\}$ is serially uncorrelated with a stochastic covariance matrix Σ_t . This assumption is designed to reflect changing uncertainty surrounding the signal, corresponding to epochs of heightened volatility following large shocks. The marginal distributions are Gaussian, except when there is a shock (see below). Each unit of private noise is an m -dimensional Gaussian process that is serially uncorrelated with covariance matrix Σ_v , a deterministic matrix particular to agent i . (We allow for each agent to have somewhat different sources of private information, and therefore Σ_v can depend on i .) Hence, the ℓ m -dimensional Gaussian process $\{v_t^{(\ell)}\}$ is serially uncorrelated with covariance matrix $I_\ell \otimes \Sigma_v$. Finally, we assume that private information is independent across agents.

Regarding the dynamics about the variance of public signal, Σ_t , there is literature on its specification using stochastic volatility processes (see, e.g., Chiu, Leonard, & Tsui, 1996, and Uhlig, 1997). We adopt the broad framework of Cogley and Sargent (2005) and Primiceri (2005), but with some modifications suggested by Neusser (2016). Specifically, consider the Cholesky decomposition $\Sigma_t = B_t \Omega_t B_t'$, where B_t is a unit lower triangular and Ω_t is diagonal. Each of the diagonal entries of Ω_t is assumed to follow an exponential random walk. The matrix B_t can be written as the matrix exponential of some C_t , where C_t is lower triangular with 0s on the diagonal. Each element of C_t is modeled as following an independent random walk, and $B_t = \exp\{C_t\}$. In this way, the process $\{\Sigma_t\}$ can be generated.

To generate a temporary shock, we consider both first- and second-moment effects. The first-moment facet corresponds to a jump (up or down) in the value of the public data, whereas the second-moment facet corresponds to increased variability in the public data; both effects occur in the public noise, which is heteroskedastic, while we model the signal

as being unchanged. (This reflects a view that the signal corresponds to the dynamics of a variable apart from temporary aberrations.)

The second-moment effect of a temporary shock at some time index τ can be generated by scaling the diagonal entries of a single Σ_t by some $a > 0$, but without altering B_t or Ω_t , so that the effect is transitory:

$$\Sigma_t = B_t \Omega_t B_t' \times (1 + a 1_{\{t=\tau\}}). \quad (4)$$

This ensures that Σ_τ has values multiplied by $1 + a$. However, the corresponding shock η_τ will not be large unless the random vector is generated from the right tail of the normal distribution. We proceed by generating m random variables independently from the marginal distribution $\mathbb{P}[Z > x | Z > b] = (1 - \Phi(x))/(1 - \Phi(b))$, and multiplying the corresponding vector by $\Sigma_\tau^{1/2}$ to obtain η_τ , thereby obtaining a first-moment shock. These modifications to η_t and Σ_t at time $t = \tau$ will be designated as a temporary shock, mimicking the first-moment and second-moment shock arising from natural disasters. In our simulations, we simulate a large shock with parameters $a = 19$ and $b = 2$.

B. State-Space Representation

We here give details about the Kalman filter for processing noisy information, assuming that Σ_t and the parameters governing $\{\pi_t\}$ are known. Likewise, the matrices B and $A^{(\ell)}$ are also assumed to be known. In practice, the dynamics of these processes would not be known to the forecasters. Instead, our viewpoint is that the state-space model reflects the essential facets of each agent's internal process for generating forecasts.

Suppose that the signal can be expressed as a component of a pm -dimensional Markovian state vector x_t , that is, there exists a matrix G such that $\pi_t = Gx_t$. (Here, p gives the number of states, each of which has dimension m and G is of dimension $m \times pm$.) The transition equation for this state vector is

$$x_t = \Phi x_{t-1} + \epsilon_t \quad (5)$$

for $t \geq 1$ and an initial value x_0 . Here, $x_t' = [\pi_t', \pi_{t-1}', \dots, \pi_{t-p+1}']$, $G = [I_m, 0, \dots]$, and

$$\Phi = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I_m & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I_m & 0 \end{bmatrix}.$$

The transition matrix, Φ , has eigenvalues less than 1 in absolute value by assumption. The signal innovations $\{\epsilon_t\}$ are assumed to be uncorrelated with x_0 , so that ϵ_t is uncorrelated with x_{t-1} for $t \geq 1$. The innovations' common covariance

matrix is denoted as Σ^ϵ . Let

$$\delta_t(\ell) = \begin{bmatrix} v_t^{(\ell)} \\ \eta_t \end{bmatrix}, \quad H(\ell) = \begin{bmatrix} A^{(\ell)} \\ B \end{bmatrix} G,$$

so that combining equation (2) with $\pi_t = G x_t$ yields the observation equation:

$$y_t(\ell) = H(\ell) x_t + \delta_t(\ell). \quad (6)$$

Evidently, $\{\delta_t(\ell)\}$ is heteroskedastic white noise, with covariance matrix S_t given by

$$S_t = \text{Var}[\delta_t(\ell)] = \begin{bmatrix} I_\ell \otimes \Sigma_v & 0 \\ 0 & \Sigma_t \end{bmatrix}. \quad (7)$$

Together, equations (6) and (5) describe the information structure in state-space form. For ease of notation, we suppress the dependence on ℓ .

We define the following quantities: the forecast of the state vector is $\widehat{x}_{t+1|t} = \mathbb{E}[x_{t+1}|y_1, \dots, y_t]$, and its mean squared error matrix is $P_{t+1|t} = \text{Var}[x_{t+1} - \widehat{x}_{t+1|t}]$. The residual is the data minus its forecast, namely, $e_t = y_t - \widehat{y}_{t|t-1}$, and its mean squared error matrix is denoted V_t . The Kalman gain is by definition $K_t = \text{Cov}[x_{t+1}, e_t] \text{Var}[e_t]^{-1}$ and plays a key role in updating a signal extraction estimate given new information. Initialization of the recursive Kalman filter algorithm is given by $\widehat{x}_{1|0} = 0$ and $P_{1|0} = \text{Var}[x_1]$, which are the correct quantities given a stationary state vector. In the case of a VAR(p) signal process, this initial variance can be computed directly from the companion form. Then for $1 \leq t \leq T$, we compute

$$e_t = y_t - H \widehat{x}_{t|t-1}, \quad (8)$$

$$V_t = H P_{t|t-1} H' + S_t, \quad (9)$$

$$K_t = \Phi P_{t|t-1} H' V_t^{-1}, \quad (10)$$

$$\widehat{x}_{t+1|t} = \Phi \widehat{x}_{t|t-1} + K_t e_t, \quad (11)$$

$$P_{t+1|t} = (\Phi - K_t H) P_{t|t-1} \Phi' + G' \Sigma^\epsilon G. \quad (12)$$

As an additional step, because the signal is a linear function of the state vector, we have

$$\widehat{\pi}_{t+1|t} = G \widehat{x}_{t+1|t}, \quad (13)$$

$$\text{Var}[\widehat{\pi}_{t+1|t} - \pi_{t+1}] = G P_{t+1|t} G'. \quad (14)$$

Equation (10) gives a recursive formula for the Kalman gain, and its dependence on the heteroskedastic noise is clearly given through V_t in equation (9). Moreover, equations (11) and (12) tell us how to update our one-step-ahead prediction and forecast error variance for the state vector. Again, because the Kalman gain depends on the heteroskedastic variance Σ_t , both the state vector forecast and its uncertainty will be affected. To understand the Kalman gain better, observe that

$$\widehat{x}_{t|t} = \widehat{x}_{t|t-1} + P_{t|t-1} H' V_t^{-1} e_t, \quad (15)$$

which follows by applying Φ^{-1} to equation (11); hence, $\Phi^{-1} K_t$ tells us the factor to multiply the new information e_t in order to update $\widehat{x}_{t|t-1}$ to the revised quantity $\widehat{x}_{t|t}$. Rearranging this relationship and using equation (13) yields

$$\begin{aligned} \widehat{\pi}_{t+1|t+1} &= \left(I_m - G \Phi^{-1} K_{t+1} \begin{bmatrix} A^{(\ell)} \\ B \end{bmatrix} \right) \widehat{\pi}_{t+1|t} \\ &\quad + G \Phi^{-1} K_{t+1} y_{t+1}. \end{aligned} \quad (16)$$

This can be compared with expressions in Coibion and Gorodnichenko (2012), which focused on the homoskedastic case. Formally, the signal seems to depend on past forecasts and new information in the same way; however, the Kalman gain is different from the homoskedastic case. We illustrate this difference below.

C. Information Rigidity

Extending the formulation of Coibion and Gorodnichenko (2012) for the homoskedastic public noise, equation (16) indicates that the old forecast $\widehat{\pi}_{t+1|t}$ is scaled by $I_m - R_t$, where

$$R_t = G \Phi^{-1} K_{t+1} \begin{bmatrix} A^{(\ell)} \\ B \end{bmatrix}. \quad (17)$$

Note that R_t is $m \times m$ -dimensional. When the Kalman gain is small, little modification to the old forecast is needed. From equations (9) and (10), clearly K_t is small when S_t is large; a sudden jump in Σ_t (irrespective of ℓ) will drive up V_t and thereby decrease R_t . In other words, shocks will have the effect that new information is received with high uncertainty, as the forecaster knows there is little signal content in the noisy data. As a result, the new forecast will closely resemble the previous period's forecast.

Formally, the information rigidity, defined as the sequence

$$r_t = \text{tr}[I_m - R_t]/m, \quad (18)$$

is high when the new data are deemed untrustworthy (i.e., when Σ_t is high). Note that $I_m - R_t$ scales the old forecast $\widehat{\pi}_{t+1|t}$ in equation (16), and this part of the equation can be expressed as $\lambda \widehat{\pi}_{t+1|t}$ when the old forecast is an eigenvector of $I_m - R_t$ with eigenvalue λ . Hence, we might interpret the eigenvalues as the degree of information acquisition, and therefore the trace (which is the sum of the eigenvalues) divided by m yields the average degree of information acquisition. Other measures are possible, such as the product of the eigenvalues, or $\det[I_m - R_t]$, which in analogy with the forecasting literature may be described as a total degree of information acquisition. We have explored in simulation the properties of the total information, and its behavior is similar

to that of the average information r_i ; we focus on the latter for the remainder of the paper.

Our measure in equation (18) represents the average degree of information rigidity when predicting many variables. This definition is in line with the empirical evidence in Coibion and Gorodnichenko (2015) that forecast revisions of other variables have little predictive power for the forecast errors of each variable, that is, the absence of statistical evidence for the importance of off-diagonal elements of the matrix $I_m - R_t$ in our context.

We emphasize three features of this definition. First, information rigidity is defined in a multivariate context. This is important because imperfect information theories of the business cycle typically require the existence of inattention for consumers, firms, and workers, not their inattention to a single variable, such as inflation. Second, information rigidity is allowed to differ across agents to reflect differences in the weight attached on prior beliefs (e.g., Lahiri & Sheng, 2008). Third, information rigidity is allowed to change over time in response to the increased uncertainty due to large shocks (e.g., Bloom, 2009).

V. A Framework for Expectation Formation

A. Agent Preference and Loss Function

Our framework depends on the cost of private information, the preference of each agent for forecasting accuracy, and the impact of a shock. As indicated by our empirical findings, there are two types of forecasters, attentive and inattentive, who are not interchangeable and correspond to different forecasting profiles. We suppose that each agent has an associated β coefficient, a positive number that governs the importance of forecast error to the agent. This is essentially the same as the marginal benefit of reducing forecast MSE, as described in Giacomini et al. (2017).

A unit of private information costs α , so that ℓ units incur a total cost of $\alpha\ell$. The agent's forecast MSE at time t depends on the number of units of private information and is denoted $M_t(\ell)$; it is computed via $GP_{t+1|t}G'$ at time t , using equation (12). Note that equation (13) implicitly relies on ℓ , the number of units purchased. In the special case that $m = 1$ and the signal is an AR(1) with $G = 1$ (and with A and B equal to 1), we can show that $M_t(\ell)$ is decreasing with ℓ and increasing in σ_ϵ^2 .¹⁰ More generally, for $m > 1$, we must study a scalar measure of $M_t(\ell)$, such as the trace. Such a measure is weakly decreasing in ℓ as the conditioning set is increased, and the lower bound of 0 indicates that $M_t(\ell)$ is convex. Thus, as an agent purchases more private information, his MSE will decrease. But as he purchases each additional piece, the decline in his MSE becomes smaller.

The total cost to an agent at time t who has purchased ℓ units of private information is

$$C_t(\ell) = \beta \text{tr}[M_t(\ell)] + \alpha\ell. \quad (19)$$

Therefore, a decrease to $\text{tr}[M_t(\ell)]$ that resulted from increases to ℓ will be offset with the increased cost $\alpha\ell$. Each agent endeavors to minimize the cost function, but that person's behavior depends on β , which is different for each agent. Low values of β correspond to indifference to forecast performance, and such agents are said to be inattentive. At the extreme end, $\beta = 0$ indicates that forecast performance is irrelevant to the agent, and private information will never be purchased. The other extreme corresponds to $\beta = \infty$, where forecast performance is of crucial importance; then the cost is minimized by taking ℓ as large as possible. Hence, higher values of β correspond to attentive agents.¹¹ While we assume heterogeneity in gains and keep the cost parameter the same across agents, this simplified assumption is not central to the results. Indeed, all we can identify in the data is the heterogeneity in the ratio of benefit to cost. Thus, β can be interpreted as "normalized gain parameter."

The forecasting process is described as follows. At time t , all agents observe the costless public data. They also observe whether there has been a shock, which has first- and second-moment aspects. The agent's internal decision-making and forecasting process is modeled through the state-space form, where forecasts and MSE are computed according to assumptions about the public noise. The agent computes $C_t(\ell)$ for all $0 \leq \ell \leq L$ and chooses ℓ^* (integer valued), which minimizes cost. Clearly this minimizing choice can be different at each time t . The agent then purchases ℓ^* units of private information and generates the corresponding forecasts.

This entire process might be expected to evolve smoothly over times t , with $1 \leq t \leq T$, when no shocks are present. Attentive agents are buying a few units of private information, and this quantity would plausibly be roughly constant over time. In contrast, inattentive agents buy little, or even no, private information. This scenario is interrupted by the occurrence of a shock (at time t), because $\text{MSE}(\text{tr}[M_t(\ell)])$ increases, thereby increasing $C_t(\ell)$. By increasing ℓ , the agent can reduce $\text{tr}[M_t(\ell)]$, and whether this is worthwhile is determined by the offset of the cost $\alpha\ell$.

If the shock is sufficiently dramatic, inattentive agents will be moved to purchase some private information, and their resulting $\text{tr}[M_t(\ell^*)]$ will be lower than it would be otherwise. Whether it is lower than its value at time $t - 1$ depends on whether the shock's effect on forecast error is offset by the purchase of private information. The attentive agent might also purchase information; however, because $\text{tr}[M_t(\ell)]$ is convex, there is a diminishing return obtained by getting more private information. As a result, it may not be worthwhile for the attentive agent to purchase even more information when the shock occurs.

¹¹We suppose there is some upper-bound L on the quantity of private information that can be purchased; in our simulations, we have set $L = 20$, although this has no impact because the maximum threshold is not attained by any agent.

¹⁰To save space, the illustration of this special case is omitted here.

B. Assessing Forecasting Performance

From the preceding mechanism, each agent i generates a set of forecasts $\widehat{\pi}_{t+1|t}(i)$. This depends, for each t , on ℓ^* , which in turn depends on the cost profile of agent i through β_i and the variability of private information Σ_v . The forecast and its error covariance $G P_{t+1|t}(i) G'$ can be computed from equations (13) and (14). If we have interest in some linear composite of agents' results, say $\sum_{i=1}^N w_i \widehat{\pi}_{t+1|t}(i)$ for given weights w_i , then the corresponding target is $\sum_{i=1}^N w_i \pi_{t+1}$, which equals π_{t+1} when the weights sum to 1. This composite is called the consensus forecast when the weights all equal $1/N$:

$$\bar{\pi}_{t|t-1} = N^{-1} \sum_{i=1}^N \widehat{\pi}_{t|t-1}(i). \quad (20)$$

Other quantities of interest are:

$\widehat{\pi}_{t t-1}(i) - \pi_t$:	Forecast error of agent i
$\bar{\pi}_{t t-1} - \pi_t$:	Error of consensus forecast
$\widehat{\pi}_{t t-1}(i) - \bar{\pi}_{t t-1}$:	Discrepancy of agent i from consensus forecast

We note that the forecast error of agent i is orthogonal to the public data and their own private data. The mean square of the forecast error of agent i is identified with the quantity $M_t(\ell)$ above, but here it is convenient to denote it by M_i instead, as we are fixing the time index for this discussion and wish to stress the dependence on agent i . As each agent has forecast MSE of M_i , we can get an average measure via

$$\widetilde{M} = N^{-1} \sum_{i=1}^N M_i. \quad (21)$$

In order to compute the mean square error of equation (20), the consensus forecast MSE $\widetilde{M}_{t+1|t}$, it is necessary to calculate forecast error covariances $\text{Cov}[\widehat{x}_{t+1|t}(j) - x_{t+1}, \widehat{x}_{t+1|t}(k) - x_{t+1}]$, denoted $Q_{t+1|t}^{(jk)}$. For $j = k$, this covariance is just $P_{t+1|t}$. Otherwise, the following recursion can be used for computation. Note that the Kalman gains $K_t(j)$ and observation matrices $H(j)$ depend on the j th Kalman filter calculation and that such an index j in turn determines the ℓ^* needed in the formulas.

Proposition 1. *The covariance of prediction errors across agents, $Q_{t+1|t}^{(jk)}$, can be computed recursively by*

$$\begin{aligned} Q_{t+1|t}^{(jk)} &= [\Phi - K_t(j)H(j)] Q_{t|t-1}^{(jk)} [\Phi - K_t(k)H(k)]' \\ &\quad + \Sigma^\epsilon + K_t(j) \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_t \end{bmatrix} K_t(k)', \end{aligned} \quad (22)$$

with the initialization $Q_{1|0}^{(jk)} = \text{Var}[x_1]$ for all j and k .

Proof. See the online appendix.

We can now determine the consensus forecast MSE. Letting

$$C_{jk} = \text{Cov}[\widehat{\pi}_{t+1|t}(j) - \pi_{t+1}, \widehat{\pi}_{t+1|t}(k) - \pi_{t+1}],$$

it follows that $C_{jk} = G Q_{t+1|t}^{(jk)} G'$. Hence, because $\bar{\pi}_{t+1|t} - \pi_{t+1} = N^{-1} \sum_{i=1}^N (\widehat{\pi}_{t+1|t}(i) - \pi_{t+1})$, we obtain

$$\widetilde{M}_{t+1|t} = N^{-2} \sum_{j,k=1}^N C_{jk}. \quad (23)$$

Next, the mean square of the discrepancy of agent i is denoted D_i . It is a measure of disagreement for agent i versus the consensus and will be small if that agent behaves like her cohort. To compute the disagreement, observe that

$$\begin{aligned} \widehat{\pi}_{t+1|t}(i) - \bar{\pi}_{t+1|t} &= (\widehat{\pi}_{t+1|t}(i) - \pi_{t+1}) \\ &\quad - N^{-1} \sum_{j=1}^N (\widehat{\pi}_{t+1|t}(j) - \pi_{t+1}) \\ &= (1 - N^{-1}) (\widehat{\pi}_{t+1|t}(i) - \pi_{t+1}) \\ &\quad - N^{-1} \sum_{j \neq i} (\widehat{\pi}_{t+1|t}(j) - \pi_{t+1}), \end{aligned}$$

and hence with $\widetilde{w}_j = -1/N$ for $j \neq i$ and $\widetilde{w}_i = 1 - 1/N$, we have $D_i = \sum_{j,k=1}^N \widetilde{w}_j \widetilde{w}_k C_{jk}$. The average disagreement, or dispersion, is defined via

$$\widetilde{D} = N^{-1} \sum_{i=1}^N D_i. \quad (24)$$

C. Summary of Model Implications

To summarize, we propose a model of information acquisition in which all agents behave the same way but are different in the benefit that they gain from the accuracy of their beliefs. The marginal benefit from a revision (compared to the marginal cost) is much larger for attentive agents with higher gain parameters. Within this model, disaster shocks are specified in the same way for all agents as an increase in the variance of the public signal. To compensate for this higher uncertainty, both inattentive and (some) attentive agents will increase their acquisition of private information, leading to declined information rigidity after large shocks. Inattentive agents buy a lot of private information after a disaster shock because it is cheaper for a more uninformed agent to know more about the fundamental relative to an agent who already knows a lot. For attentive agents, however, this acquisition is worthwhile only if the benefit of doing so—the reduction of

the mean squared error—is larger than the cost from buying additional pieces of private information.

In particular, our model generates the following implications for the MSE and information rigidity (IR):

- For attentive agents (high β): Normally their forecast MSE M_i and IR is low. When a shock occurs, there may be a slight increase to both (if they forgo buying additional private information) or a slight decrease to both (this happens if they purchase additional private information, which can offset the shock's impact on MSE). If the shock is more moderate, they are unlikely to alter their behavior.
- For inattentive agents (low β): Normally their forecast MSE and IR is high. When a shock occurs, they usually purchase some private information, and it is typically more than sufficient to offset the impact of the shock, hence, forecast MSE and IR will decrease. They tend to react even to more moderate shocks.

The consensus forecast MSE is a mixture of the forecast MSEs of individual agents, and, hence, its overall behavior depends on the correlation of forecast errors for various agents, some of whom may be purchasing additional private information. It is therefore related to disagreement. In particular, for disagreement our model predicts:

- For attentive agents (high β): If they do not purchase private information, disagreement is moderate and increases slightly during a shock, while the correlation of their forecast errors decreases. If they do purchase private information during the shock, the disagreement will decrease while forecast errors become more correlated.
- For inattentive agents (low β): The disagreement is high initially, and drops during the shock (because they purchase information). Also, forecast errors become more correlated. For an individual agent i who does not purchase information and defies the consensus, D_i can increase, but the overall measure \tilde{D} will tend to decrease.

D. Simulation Results

We illustrate the above model implications through the results of simulations. Our data-generating process (DGP) is described as follows. We set $m = 3$, and for the signal considered, a VAR(2) based on empirical fits of industrial production, inflation, and federal funds rate data. This is merely intended to furnish a reasonable signal DGP. The coefficients are

$$\Phi_1 = \begin{bmatrix} 0.7745 & -0.1472 & -0.4292 \\ 0.0556 & 0.3139 & 0.1116 \\ 0.1066 & 0.0582 & 1.3309 \end{bmatrix}$$

$$\Phi_2 = \begin{bmatrix} -0.0517 & 0.0979 & 0.4708 \\ 0.0170 & 0.1813 & -0.0151 \\ -0.0054 & 0.0384 & -0.3872 \end{bmatrix}.$$

The innovation covariance was set to I_3 . Samples of size $T = 100$ were generated, with a burn-in period of 500 observations. The observation matrices A and B were set equal to identity matrices I_3 . To generate the public noise Σ_t , the innovations of the exponential random walk Ω_t , and the matrix exponential C_t were Gaussian of standard deviation .01 for all three dimensions. A temporary shock at time $\tau = 50$ was generated in the manner described in equation (4).

To generate the private noise, an overall dispersion coefficient with value .001 controls the spread of entries in the covariance matrix, whereas the overall scale is determined via multiplication by .1. This low value makes private information more valuable. The settings were determined by empirically examining various cases and studying the resulting behavior of simulations. The number of agents was set to $N = 10$, and two different profiles were considered. For inattentive agents, we drew β uniform on $[9, 10]$, whereas for attentive agents, β was uniform on $[45, 55]$. The cost of private information is $\alpha = 1$.

We generate a large shock with parameters $a = 19$ and $b = 2$ as described in section IVA. The simulation results are shown in figure 3 for information rigidity, figure 4 for mean squared error, and figure 5 for forecast disagreement.¹²

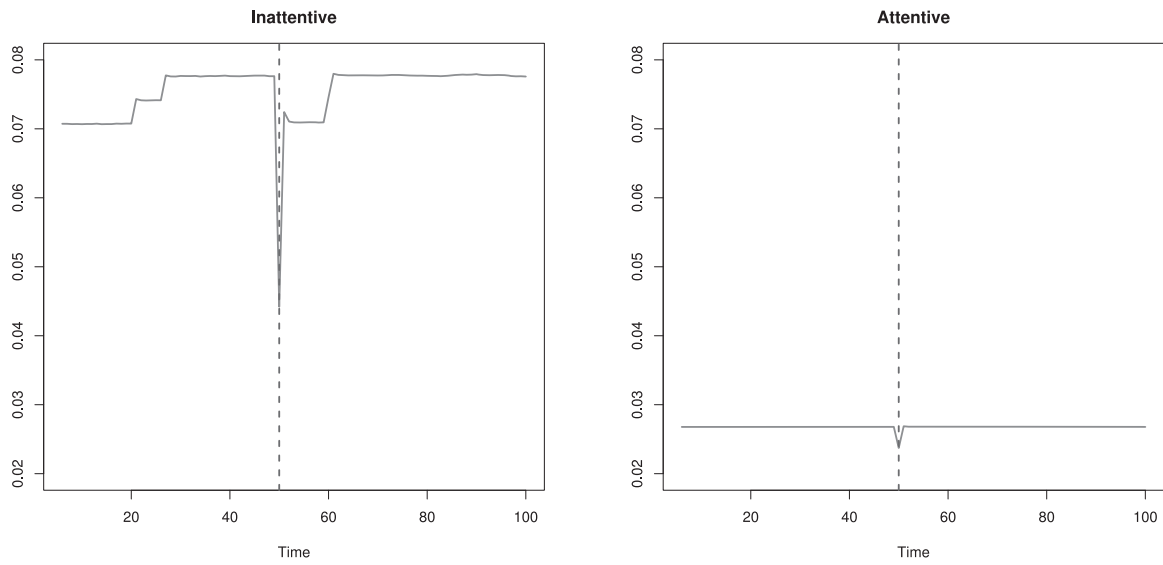
We first examine the behavior of inattentive agents in the scenario of this large shock. The shock is so large that all the agents buy information, moving from one unit to two units uniformly. As a result, individual and aggregate measure of MSE and disagreement are in concordance (figures 4 and 5), uniformly dipping down at the shock and then reverting to former levels. Agents 3 and 10 continue to buy two units of information for some time after the shock, reverting to stingier behavior only at time points 60 and 59, respectively.

The impact of this behavior can be seen in the aggregate plots of MSE and disagreement (figures 4 and 5). Also, the behavior of M_3 and M_{10} (as well as D_3 and D_{10}) shows a sustained plunge at the time of the shock, and for several times thereafter, followed by a return to former levels of MSE and disagreement. Information rigidity (figure 3) begins at a moderate level (high relative to attentive forecasters) and dips considerably during the shock, also displaying the same effects due to agents 3 and 10 immediately afterward.

Next, we consider the impact of this large shock on the attentive agents. These forecasters are more conservative with regard to shocks, and half of them (agents 1 through 5) purchase additional information, again moving from three to four units; the other five agents make no adjustment to their strategy during this very large shock. As a result, the behavior

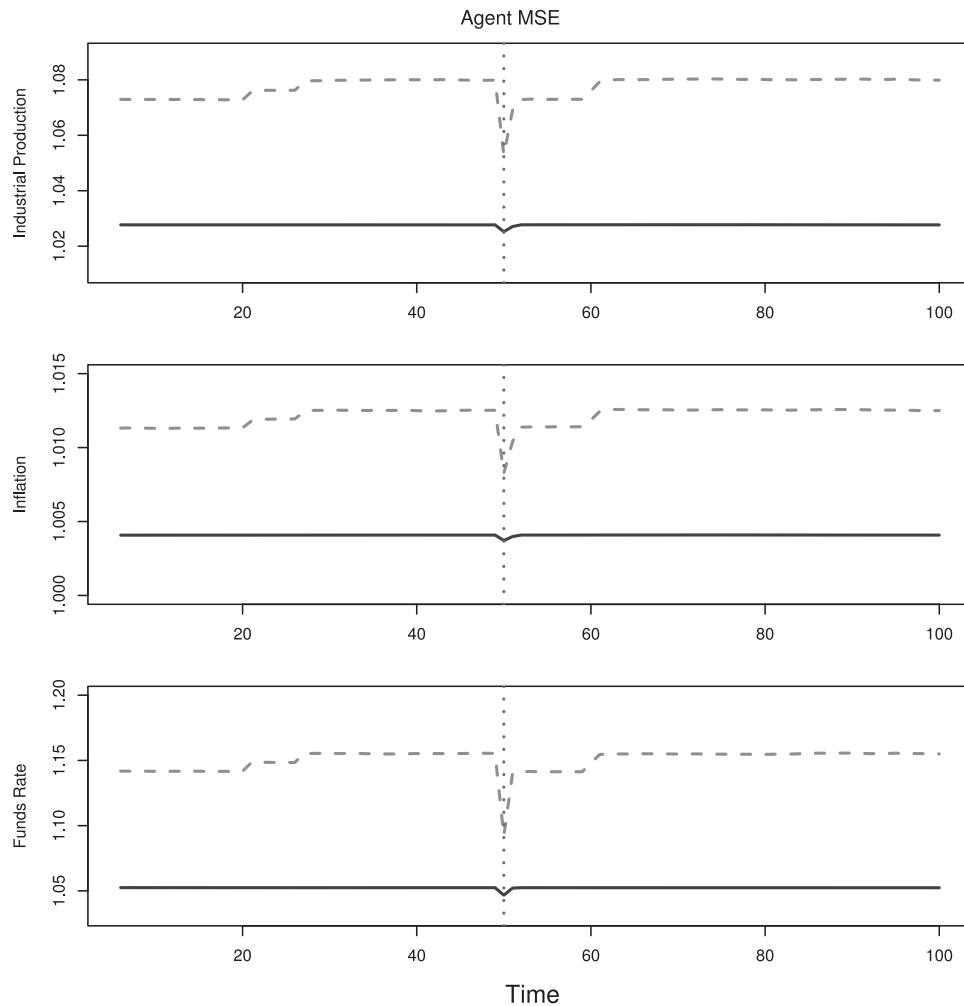
¹²We describe and simulate the impacts of moderate shocks in the online appendix. We find qualitatively similar results, though of somewhat reduced magnitude.

FIGURE 3.—INFORMATION RIGIDITY FOLLOWING A LARGE SHOCK



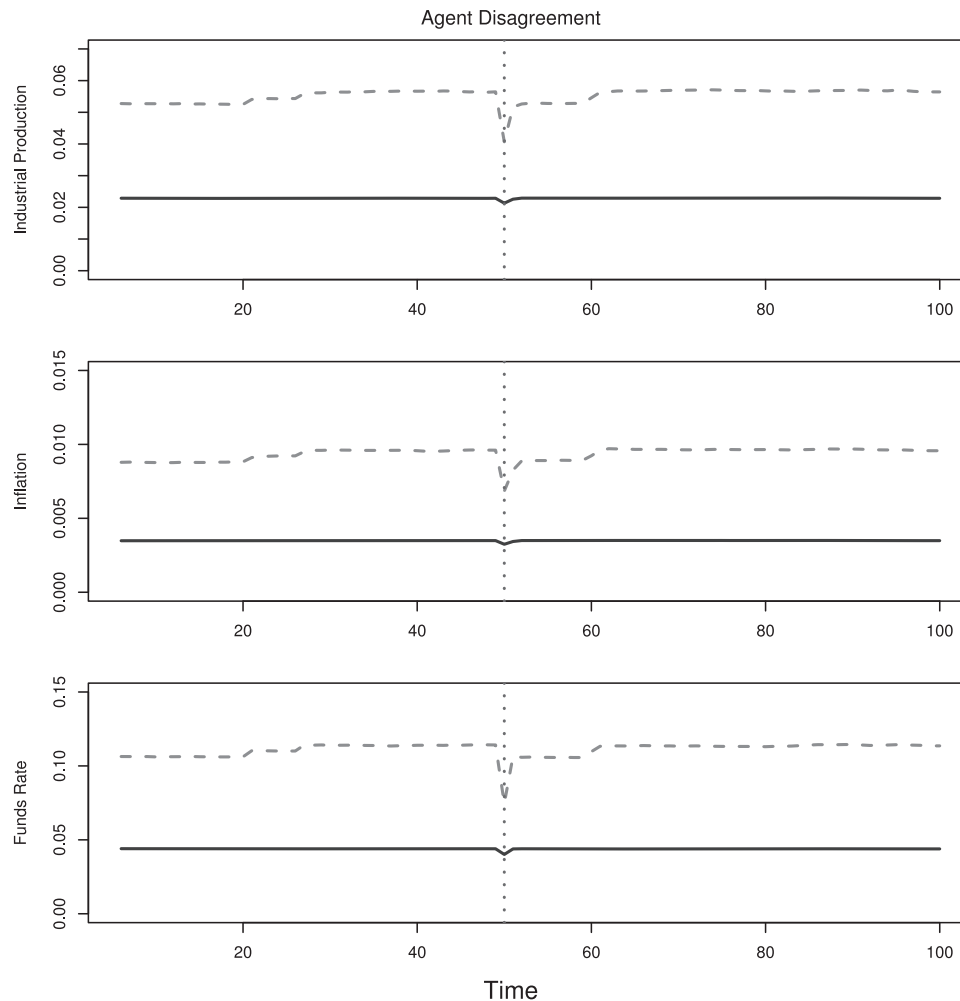
The average degree of information rigidity in predicting three variables in the context of a large shock is displayed, calculated as the scaled trace of the matrix of the weights attached to the agent's previous forecast relative to new information as in equation (18). The left panel shows results for inattentive agents, while the right panel shows results for attentive agents. The shock time $\tau = 50$ is indicated by a vertical dashed line.

FIGURE 4.—MEAN SQUARED ERROR FOLLOWING A LARGE SHOCK



These plots show the average mean squared error (MSE) for inattentive (dotted line) and attentive (solid line) agents in the presence of a large shock. The first, second, and third rows plot the MSE in predicting industrial production, inflation, and federal funds rate, respectively. The MSE \bar{M} is calculated according to equation (21). The shock time $\tau = 50$ is indicated by a vertical dashed line.

FIGURE 5.—FORECAST DISAGREEMENT FOLLOWING A LARGE SHOCK



These plots show forecast disagreement among inattentive (dotted line) and attentive (solid line) agents in the presence of a large shock. The first, second, and third rows plot the disagreement in predicting industrial production, inflation, and federal funds rate, respectively. Disagreement \bar{D} is calculated according to equation (24). The shock time $\tau = 50$ is indicated by a vertical dashed line.

of M_1 through M_5 , as well as D_1 through D_5 , exhibits a downward movement during the shock; the opposite behavior occurs for the nonpurchasing agents. Because they swiftly revert to their former behavior after the shock, the effect on information rigidity (figure 3) is to just generate a downward drop at the time of the shock. The aggregate MSE \tilde{M} (figure 4) and aggregate disagreement \tilde{D} (figure 5) display the same features.

We need to point out that the effect of a large, unexpected shock on overall (including both attentive and inattentive agents) mean squared error and disagreement depends on the size of the shock, agents' preference for forecast accuracy, and the proportion of inattentive agents. Our empirical results show that inattentive agents tend to dominate, and thus, overall mean squared error and disagreement decline following large, unexpected shocks (at least given that the uncertainty component of the shock is not sufficiently large to outweigh the effects on the inattentive agents). Our model, however, has rich implications for different paths of overall mean squared error and disagreement following shocks.

VI. Conclusion

This paper provides a new view on what drives the behavior of macroeconomic forecasters. We find that individual forecasters are persistently heterogeneous in how often they revise or even issue a forecast. Given that many commonly used macroeconomic forecasts are derived from the average forecast from a selected set of forecasters, these differences in the frequency of revision have the potential to bias average forecasts and change the dynamics of forecasts. Accounting for this heterogeneity helps explain the information rigidity observed in many commonly utilized macroeconomic forecasts and gives a reason why forecast dispersion may actually decline following a large shock that increases uncertainty about future economic performance.

We demonstrate a significant degree of information rigidity in forecasts, driven by the fact that many forecasters choose not to update their forecast in successive time periods. Matching forecasts from a panel of 54 countries to a detailed set of natural disasters, we show that this information

rigidity declines significantly following natural disasters. At an individual level, this effect seems consistent with an attention shock affecting the forecasters, where newsworthy disasters induce formerly inattentive forecasters to update their forecasts. This may result in a counterintuitive result, where shocks to countries can increase uncertainty but decrease forecaster dispersion.

We model this phenomenon with a learning model that incorporates both an attention effect and an uncertainty effect. Our theory has three key elements. First, agents are not interchangeable. Attentive agents have a larger benefit from forecast accuracy and revise forecasts frequently. Inattentive agents who do not benefit much from the accuracy of their forecasts make revisions infrequently and nonsystematically. Second, large and unexpected shocks induce inattentive agents to update to a greater degree. Third, large shocks might induce attentive agents to purchase more private information and be more attentive.

Our model explains a world in which large shocks like natural disasters induce an immediate increase in updating of information for inattentive agents (extensive margin). This attention effect is particularly pronounced for those with an outdated information set, resulting in a significant decline in information rigidity. To compensate for the higher uncertainty following shocks, some attentive agents increase their acquisition of private information (intensive margin). These findings warn against treating the degree of information rigidity as a structural parameter and suggest that future research should explore state dependence in the information updating process. To this end, our paper moves one step forward by introducing time-varying uncertainty in expectations formation framework and proposing a measure of state-dependent information rigidity in response to uncertainty.

REFERENCES

- Andrade, Philippe, and Hervé Le Bihan, "Inattentive Professional Forecasters," *Journal of Monetary Economics* 60 (2013), 976–982.
- Andrade, Philippe, Richard Crump, Stefano Eusepi, and Emanuel Moench, "Fundamental Disagreement," *Journal of Monetary Economics* 83 (2016), 106–128.
- Baker, Scott R., Nick Bloom, and Stephen Terry, "Does Uncertainty Reduce Growth? Using Disasters as Natural Experiments," NBER working paper 19475 (2017).
- Bloom, Nick, "The Impact of Uncertainty Shocks," *Econometrica* 77 (2009), 623–685.
- Branch, William A., "Sticky Information and Model Uncertainty in Survey Data on Inflation Expectations," *Journal of Economic Dynamics and Control* 31 (2007), 245–276.
- Carroll, Christopher D., "Macroeconomic Expectations of Households and Professional Forecasters," *Quarterly Journal of Economics* 118 (2003), 269–298.
- Cavallo, Alberto, Eduardo Cavallo, and Roberto Rigobon, "Prices and Supply Disruptions during Natural Disasters," *Review of Income and Wealth* 60 (2014), S449–S471.
- Chiu, Tom Y. M., Tom Leonard, and Kam-Wah Tsui, "The Matrix-Logarithmic Covariance Model," *Journal of the American Statistical Association* 91:433 (1996), 198–210.
- Cogley, Timothy, and Thomas J. Sargent, "Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Dynamics* 8 (2005), 262–302.
- Coibion, Olivier, "Testing the Sticky Information Phillips Curve," this REVIEW 92 (2010), 87–101.
- Coibion, Olivier, and Yuriy Gorodnichenko, "What Can Survey Forecasts Tell Us about Informational Rigidities?" *Journal of Political Economy* 120 (2012), 116–159.
- , "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review* 105 (2015), 2644–2678.
- Giacomini, Raffaella, Wagner Piazza Gaglianone, Joao Victor Issler, and Vasiliki Skreta, "Incentive-Driven Inattention," University College London, Department of Economics mimeograph (2017).
- Giacomini, Raffaella, Vasiliki Skreta, and Javier Turén, "Heterogeneity, Inattention and Bayesian Updates," *American Economic Journal: Macroeconomics* 12:1 (2016), 282–309.
- Gorodnichenko, Yuriy, "Endogenous Information, Menu Costs, and Inflation Persistence," NBER working paper 14184 (2008).
- Ilut, Cosmin L., and Martin Schneider, "Ambiguous Business Cycles," *American Economic Review* 104 (2014), 2368–2399.
- Lahiri, Kajal, and Xuguang Sheng, "Evolution of Forecast Disagreement in a Bayesian Learning Model," *Journal of Econometrics* 144 (2008), 325–340.
- Maćkowiak, Bartosz, and Mirko Wiederholt, "Optimal Sticky Prices under Rational Inattention," *American Economic Review* 99 (2009), 769–803.
- , "Inattention to Rare Events," European Central Bank working paper series 1841 (2015).
- Mankiw, N. Gregory, and Ricardo Reis, "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics* 117 (2002), 1295–1328.
- Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers, "Disagreement about Inflation Expectations" (pp. 209–248), in M. Gertler and K. Rogoff, eds., *NBER Macroeconomics Annual 2003* (Cambridge, MA: MIT Press, 2004).
- Manski, Charles F., "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise" (pp. 411–471), in M. S. Eichenbaum and J. Parker, eds., *NBER Macroeconomics Annual 2017* (Chicago: University of Chicago Press, 2018).
- Neusser, Klaus, "A Topological View on the Identification of Structural Autoregressions," *Economics Letters* 144 (2016), 107–111.
- Nimark, Kristoffer P., "Man-Bites-Dog Business Cycles," *American Economic Review* 104 (2014), 2320–2367.
- Nordhaus, William, "Forecasting Efficiency: Concepts and Applications," this REVIEW (1987), 667–674.
- Orlik, Anna, and Laura Veldkamp, "Understanding Uncertainty Shocks and the Role of Black Swans," NBER working paper 20445 (2014).
- Primiceri, Giorgio E., "Time Varying Structural Autoregressions and Monetary Policy," *Review of Economic Studies* 72 (2005), 821–852.
- Reis, Ricardo, "Inattentive Producers," *Review of Economic Studies* 73 (2006), 793–821.
- Roy, Anindya, Tucker S. McElroy, and Peter Linton, "Constrained Estimation of Causal Invertible VARMA," *Statistica Sinica* 29 (2019), 455–478.
- Sims, Christopher A., "Implications of Rational Inattention," *Journal of Monetary Economics* 50 (2003), 665–690.
- Uhlig, Harald, "Bayesian Vector Autoregressions with Stochastic Volatility," *Econometrica* 65 (1997), 59–73.
- Woodford, Michael, "Imperfect Common Knowledge and the Effects of Monetary Policy," in P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds., *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund Phelps* (Princeton, NJ: Princeton University Press, 2003).