Expectation Formation Following Large Unexpected Shocks

Scott R. Baker* Tucker S. McElroy† Xuguang S. Sheng‡

November 5, 2017

Abstract

By matching a large database of individual forecaster data with the universe of sizable natural disasters across 54 countries, we identify a set of new stylized facts: (i) forecasters are persistently heterogeneous in how often they issue or revise a forecast; (ii) information rigidity declines significantly following large, unexpected natural disaster shocks; (iii) the response of forecast disagreement displays interesting patterns: attentive forecasters tend to move away from the previous consensus following a disaster while the opposite is true for inattentive forecasters. 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 – the occurrence of those shocks generates increased uncertainty among attentive agents.

JEL Classification: C32, C53, D83, D84, E17, E37

Keywords: Expectation Formation, Heterogeneous Agents, Information Rigidity, Learning, Natural Disasters, Uncertainty Shock

*Department of Finance, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. E-mail: s-baker@kellogg.northwestern.edu.

†Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233 E-mail: tucker.s.mcelroy@census.gov.

‡Department of Economics, American University 4400 Massachusetts Avenue, NW, Washington, DC 20016, E-mail: sheng@american.edu.

1 Introduction

Expectations matter. Yet, how economic agents form expectations remains an open question, as evidenced by the comprehensive survey in Manski (2017). Indeed, Manski concludes, "To make progress, I urge measurement and analysis of the revisions to expectations that agents make following occurrence of unanticipated shocks." We answer this call by matching a large database of individual forecaster 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, which 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. (i) Agents are not interchangeable: attentive agents revise forecasts frequently and inattentive agents make revisions infrequently and non-systematically. Assumed here is that the fraction of inattentive agents is exogenous and persistent, consistent with what we see in the data. (ii) Inattentive agents update sporadically during the normal times à la Mankiw and Reis (2002). However, following large and unexpected shocks, the cost of not updating information is high, and accordingly the majority of inattentive agents will update their information sets and make new forecasts. (iii) Attentive agents always combine public and private signals in making their forecasts, as described in the noisy information model à la Sims (2003), Woodford (2003) and Maćkowiak and Wiederholt (2009).

Our theory tells the following story about expectations formation following large unexpected shocks. Visibly large shocks induce an immediate increase in updating of information for most inattentive agents. This attention effect is particularly pronounced for those with an outdated information set, resulting in a significant decline in information rigidity. This is the first channel through which natural disaster shocks affect expectation formation. For attentive agents, the occurrence of those shocks generates increased uncertainty among them and as a result, their new forecasts closely resemble the previous period's forecasts. This is the second channel through which natural disaster shocks affect expectation formation.

The ingredients of our theory are motivated by our empirical findings. Our primary database of macroeconomic forecasts comes from Consensus Economics and covers a dozen 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 and Schneider, 2014).

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 or revise forecasts only infrequently. This latter finding is consistent with the so-called predisposition effects in Branch (2004) that as long as there is no substantial evidence that would dramatically surprise those inattentive agents, they will not revise their previous forecasts.

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 disease outbreaks. We further limit our focus only on significant disasters as measured by the number of people affected or killed and the monetary damages caused.[1]

To further address the concerns that the disaster shocks are not fully unexpected or relatively small in magnitude, we construct a news-based measure of coverage of the disaster across several thousand English-language newspapers from the Access World News database. This index allows us to measure the change in attention, or at least newspaper attention, paid to a country following a disaster. This enables us to flexibly distinguish between disasters that are relatively unimportant from those that may be much more so. Since we expect that unexpected disasters should have the highest increases in news coverage, this approach also helps to filter out more predicted disasters that may have been "baked in" to previous forecasts.

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 counter-intuitive but are in fact a natural consequence of the dominance of inattentive agents – disasters induce inattentive forecasters to update their forecasts and, in doing so, actually move the majority of them closer to the mean forecast and also to the ex-post true value of the variable.

Our paper makes several contributions to the literature. First, we build a theory of expectation updating with heterogeneous agents that extends the existing literature in a number of ways. Our model is closely related to the theoretical literature on expectations formation with information frictions. For instance, Mankiw and Reis (2002) propose the sticky information model that explains agents' rational inattention in terms of limited resources and the cost of updating information sets. Reis (2006) generalizes this basic model to allow for state-dependent setup in which the length between information update is optimally chosen as

[1] Without this limitation, nearly every country is hit by a natural disaster in nearly every period of our sample.

a function of shocks. Sims (2003), Woodford (2003) and Maćkowiak and Wiederholt (2009) advocate the noisy information model that emphasizes the limited ability of economic agents to process new information from noisy signals. Andrade and Le Bihan (2013) develop the hybrid sticky-noisy information model and Andrade, et al. (2016) extend the model in a multivariate context. Giacomini, Skreta and Turen (2016) formulate a theory of expectation updating in which agents can be inattentive but, when updating, they follow Bayes' rule and assign homogeneous weights to public information. In contrast to these papers, agents in our model are persistently heterogeneous in their type – attentive and inattentive. In contrast to all previous work, we also explicitly model agents' behavior following large, unexpected shocks; see, e.g. Ortoleva (2012) and Nimark (2014).[2]

Our model also generalizes the noisy information model in two dimensions by allowing for (i) heterogeneous precision of private signals such that interpretation of the same public signal differs across agents,[3] and (ii) time-varying precision of public signals in order to capture the increased uncertainty among attentive agents due to large shocks. The latter feature of our model is motivated by the recent literature on the impact of uncertainty shocks on real economic activities.[4]

Another novel feature of our paper is to propose a new concept of information rigidity within the noisy information model and document high-frequency time variations in the

[2] Ortoleva (2012) models the non-Bayesian reactions to unexpected news. Our model differs from Ortoleva in that we specify the change of paradigm for inattentive agents whereas for attentive agents, they still face the same signal extraction problem, albeit with much noisier public signals following large shocks. The large, unexpected shocks identified in our paper are also 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 agents' uncertainty and induces synchronized information updating, but does not change the underlying data generating process in our framework.

[3] Lahiri and Sheng (2008) estimate a Bayesian learning model with heterogeneity and find that differential interpretation of public information explains from almost nothing to 70% of forecast disagreement as forecast horizon gets shorter from 24- to 1-month ahead. Acemoglu, Chernozhukov and Yildiz (2016) formulate a theory in which starting with heterogeneous prior beliefs, agents will not converge to a consensus even after observing the same infinite sequence of signals when there is uncertainty in the public information.

[4] Important contributions in this literature include Abel (1983), Bernanke (1983), Bloom (2009), Pastor and Veronesi (2012), Jurado, Ludvigson and Ng (2015), Baker, Bloom and Davis (2016), and Baker, Bloom and Terry (2017), among others.

estimated degree of information rigidity after the occurrence of large shocks. Coibion and Gorodnichenko (2012) interpret the weight on agent's previous forecast (relative to new information) as the degree of information rigidity, which is univariate and constant over time. We generalize their measure in a multivariate context and define information rigidity as the scaled trace of the matrix of the weights attached to agent's previous forecast. Our measure represents the average degree of information rigidity when predicting many variables and 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.

Using this new measure, we find that the estimated degree of information rigidity significantly changes after the occurrence of large shocks. This finding adds to the literature relying on survey expectations to evaluate models with information frictions.[5] The findings from all of these papers firmly establish the presence of information rigidities in 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). Finally, 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 indeed decreases following the occurrence of large shocks. Our model is successful in explaining this

[5]Within the sticky information framework, Mankiw, Reis and Wolfers (2004) and Branch (2007) use the aggregate forecasts to calibrate information rigidity, while Carroll (2003) estimates the degree of inattention by matching consensus forecasts of households and professionals, and Coibion (2010) estimates the inattention degree in a sticky-information Phillips curve. In contrast, Andrade and Le Bihan (2013) and Andrade, et al. (2016) provide a direct, micro-data estimate of this inattention parameter. Patton and Timmermann (2010) estimate a learning model and find that heterogeneity in priors plays the most important role in explaining disagreement up to two year ahead. Coibion and Gorodnichenko (2012) compare the sticky and noisy information models by estimating the response of forecast errors and disagreement to structural shocks. Coibion and Gorodnichenko (2015) propose a new approach to test the full-information rational expectations hypothesis and quantify the underlying degree of information rigidity.

seemingly anomaly and matching other key features of the expectations formation process.

Our paper proceeds as follows. Section 2 describes the dataset used in this paper. We establish a set of new stylized facts about expectations formation following large unexpected shocks in section 3. We introduce the information structure faced by agents in section 4. We propose a theory of expectation updating in section 5. Section 6 presents the simulation results. Finally, section 7 concludes. The appendix includes proofs and additional tables.

2 Data

2.1 Consensus Forecast Data

Our primary database of macroeconomic forecasts comes from Consensus Economics. We utilize aggregated data regarding 54 countries back to 1989 for which we can obtain sufficient forecast data alongside information regarding the actual GDP and inflation for a given country-year.[6] Consensus Economics obtains the individual forecasts from professional forecasters including banks and financial firms, leading industrial companies, consulting firms, and think tanks and research groups. These forecasts cover the means and standard deviations of individual forecasts for GDP and inflation in the current calendar year as well as the next calendar year. Because panelists are asked about calendar years rather than a rolling period of 12 months, as forecasts are solicited later in the year, they mechanically become more accurate. That is, forecasts for 2014 GDP will be significantly more accurate when solicited in December of 2014 than in January of 2014.

In addition to this aggregated forecast data, we utilize individual forecasts from Con-

[6]Countries include Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, China, Colombia, Czech Republic, Denmark, 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, UK, US, Ukraine, and Venezuela.

sensus Economics. This subset of data solely covers the G7 countries: Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States. The individual forecast data cover 1989 - 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. The forecasts for the majority of these variables are expressed in terms of growth rates rather than in levels. There are 296 unique panelists, 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 many commonly used sources of forecaster data like Bloomberg forecasts and the Philadelphia Fed Survey of Professional Forecasters.

2.2 Disaster Data

Our natural disaster data has been obtained from the Center for Research on the Epidemiology of Disasters (CRED). This data contains 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 economic cost of the event. The CRED data includes industrial and transportation accidents which we exclude in our analysis.

For each country-month period, we give a value of one if a disaster has occurred and a zero otherwise. This means that if a country has, for example, three earthquakes in one month, it only receives a value of one. 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 diasters 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: 1. More than 0.00001% of a country's population dead (eg. more than 30 dead in the United States); 2. More than \$10M in damages; 3. More than 50,000 people 'affected' (eg. 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. Furthermore, 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 survey date in a given month. That is, if the June Consensus Forecast has already taken place, we attribute any further disasters in June to the month of July.

2.3 Newspaper Data

Two natural concerns are that the disaster shocks that we utilize are not fully unexpected or are relatively small in magnitude. In order to help alleviate both of these potential problems, we turn to a measure of unexpectedness and impact derived from news article mentions of the countries in question.

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.[7]

[7] Access World News contains over 2,500 newspapers. We focus on newspapers from the US, as they make up the majority of coverage in this database. For US disasters, we look for changes in articles written that

For each individual disaster, we search the newspaper archive for news articles that mention the country in question. For each of the 15 days leading up to the disaster and 15 days following it, we measure this count of articles. We then measure the relative change in the number of articles written about the country. Figure 1 shows an average of this series where each event's coverage has been normalized to 1 in the 15 days prior to the event. A value of 2 following time zero 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 may be much more so. Moreover, it will help us to filter out expected disasters that may have been 'baked in' to previous forecasts, as only unexpected disasters should have the highest increases in news coverage. 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 and/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 15 days after the event relative to the 15 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 5-day or 30-day windows. When using the media-weighted shocks, we use the shock with the highest jump in media citations for that category in that month.

mention the state that the disaster occurs in rather than the country.

3 Empirical Results

3.1 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:

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

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 relationship between forecast revisions and the forecast error of any given period relative to the truth.

In Table 1, we restrict our analysis to forecasts of next-year GDP. In the first column of Table 1, we find that the change in mean forecasts from month to month is strongly and positively related to the forecast error. 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 part-way to the full-information forecast that is likely closer to the ex-post true value for the period.

Columns 2-4 mirror this specification but add in interactions of $ForecastRevision_{i,t}$ with disasters and disasters that are scaled utilizing 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 aforementioned 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 impact the relationship to any large degree.

Finally, 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 one 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. Similarly to our finding with only the news-based scaling, we again find that, in general, larger 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, their information rigidities significantly decrease following such disasters.

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' test is that it is completely independent of the 'true' values of the macroeconomic variables in question. Our findings follow 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. However, this relationship breaks down following a natural disaster in the country in question. Moreover, the relationship between forecast revisions and lagged forecast revisions becomes increasingly weak as the disaster in question becomes larger. This again suggests that a large and unexpected shock precipitates a state-dependent response and an increase in information updating.

Tables A1 and A2 follow Tables 1 and 2, but include forecast data across all variables in the sample (GDP, CPI, long- and short-run interest rates, unemployment, and consumption) and include forecast variable fixed effects. We find qualitatively similar results across all forecasted variables as when restricting the analysis to GDP: all variables exhibit significant information rigidity that declines following large natural disasters.

3.2 Heterogeneous Individual Forecasters

Consensus Economics forecasts are also useful in that underlying forecast data from individual forecasters are available. So, not only can we observe how the overall mean forecast for a given country-variable changes over time, but we can also observe 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). This corresponds with approximately the top quintile and bottom four quintiles of forecaster reporting.[8]

[8] We have performed similar exercises measuring attentiveness as the fraction of forecasts which are different than the previous forecast for a given forecaster and found qualitatively similar results.

Table 3 demonstrates some of the persistent differences across these two groups. In columns 1 and 2, 'Attentive Forecaster' is simply 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 persistent errors and deviations for the more 'attentive' forecasters.

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 errors relative to the truth and also lower deviations from the mean forecast for any given country-variable-month.

One possible concern with our analysis is that forecasters may frequently shift between being attentive and inattentive. We find this type of switching to be uncommon for individual forecasters. 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, 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. Figure 3 does the same but for the fraction of forecasts that are changed from month to month for a given forecaster.

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 utilizes the individual forecaster data to identify the channels through which this reduction in information rigidity takes effect. In columns 1 and 2, we regress the likelihood of a forecaster changing their forecast in a given month relative to their forecast in the previous month on disasters and scaled disasters. 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 significant and positive impact, where a disaster increases the probability of changing an individual forecast by approximately 0.86 percentage points (on average likelihood of changing of approximately 52%). Column 2 demonstrates that this effect is larger for larger and 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.[9]

Columns 3-6 repeat this exercise for our previously defined groups of attentive and inattentive forecasters. In columns 3 and 5, we find that the likelihood of attentive forecasters updating their forecast is unaffected by a natural disaster hitting a given country. In columns 4 and 6, we find the opposite is true for inattentive forecasters. This group drives the entirety of the combined effect we found in columns 1 and 2, with inattentive forecasters being much more likely to update their forecast following a natural disaster. Column 7 demonstrates this same phenomenon across quintiles of attentiveness, where we see that the most attentive forecasters are the least likely to change their forecast following a natural disaster.

Table 5 demonstrates the counter-intuitive result that disasters can actually decrease the dispersion of professional forecasts due to the effect that disasters have on inattentive forecasters. Here we see that, following a disaster, attentive forecasters see little change in the relationship between their own forecast and the mean forecast. However, inattentive forecasters see a decline in this measure of dispersion, as they tend to update their forecasters after a natural disaster. This holds true for both the difference between an individual's forecast and the overall mean forecast as well as the difference between the individual's forecast and their own-attentiveness-group's forecast. These results suggest that disasters

[9]These results are all robust to how we treat non-responses. 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 for them. This only affects about 8% of forecaster-months in the data.

induce inattentive forecasters to update their forecasts and, in doing so, actually moves them closer to the mean forecast and also to the ex-post true value of the variable.

Finally, Table 6 estimates the impacts of disasters on forecast accuracy relative to the ex-post true values of the macroeconomic variables. We perform these tests for all forecasters, attentive forecasters, and inattentive forecasters. Again, we find that there is little impact of natural disasters on forecast errors for attentive forecasters, while there are strong negative effects for inattentive forecasters. Even conditioning on a forecaster updating their forecast, inattentive forecasters tend to report more accurately following a natural disaster than when changing their forecaster without a disaster.

Overall, these results seem to demonstrate that, for inattentive forecasters, natural disasters act mainly as an attention shock, prompting them to update stale forecasts and converge towards a newer consensus value. To match 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 faced by agents and derive the state space representation. In section 5, we explore some of the key properties of both types of agents, such as their information rigidity, forecast accuracy and dispersion.

4 Information Structure and State Space Representation

We suppose that economic agents seek to forecast a signal process $\{\pi_t\}$ that is obfuscated by noise. We envision the existence of both a public and a private channel for attentive agents. Our model assumes that the ith attentive 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 a private noise $\{\nu_t(i)\}$. Hence the observation process for the ith agent is

$$y_t(i) = \begin{bmatrix} A^{(i)} \\ B \end{bmatrix} \pi_t + \begin{bmatrix} \nu_t(i) \\ \eta_t \end{bmatrix}. \tag{2}$$

for $1 \leq t \leq T$. The signal process is common to all agents, but the noise processes depend on i. The matrix $A^{(i)}$ corresponds to the private manifestation of the signal; for many applications these matrices are given by an identity matrix. Inattentive agents ($N+1 \leq i \leq N+M$) can be described through a similar framework, albeit there is no private channel and $A^{(i)}$ is null. Hence, the mathematical results derived below can be applied to the inattentive case as well, with some adjustments.

We suppose that $\{\pi_t\}$ is a stationary Markov process; principal interest focuses on the case that the signal follows a VAR(p) stochastic process of dimension m. Also we suppose that $\{\eta_t\}$ 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. In contrast, the private noises $\{\nu_t(i)\}$ are each serially uncorrelated with covariance matrix $\Sigma^{(i)}$; this covariance is deterministic, yielding a homoscedastic error.

4.1 State Space Representation with Known Parameters

We here give details about the Kalman filter for processing noisy information, in the case that Σ_t and the parameters governing $\{\pi_t\}$ are known. The matrices B and $A^{(i)}$ are also assumed to be known. Of course 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. Therefore, it is sufficient for our purposes to treat all the parameters as known, although it will be convenient to generate viable examples of heteroscedastic noise via a stochastic covariance process, described in section 4.2.

Suppose that the signal can be expressed as a component of a Markovian state vector x_t, i.e., there exists a matrix G such that $\pi_t = G\, x_t$. The transition equation for this state vector is

$$x_t = \Phi\, x_{t-1} + \epsilon_t \tag{3}$$

for $t \geq 1$ and an initial value x_0. Here Φ is the transition matrix, which by assumption has eigenvalues less than one in absolute value. 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 Σ^ϵ. Let

$$\delta_t(i) = \begin{bmatrix} \nu_t^{(i)} \\ \eta_t \end{bmatrix} \qquad H(i) = \begin{bmatrix} A^{(i)} \\ B \end{bmatrix} G,$$

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

$$y_t(i) = H(i)\, x_t + \delta_t(i). \tag{4}$$

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

$$S_t = \mathrm{Var}[\delta_t(i)] = \begin{bmatrix} \Sigma^{(i)} & 0 \\ 0 & \Sigma_t \end{bmatrix}. \tag{5}$$

Note that the only data available to the ith agent is $\{y_t(i)\}$, and so estimates of the signal are to be constructed on this basis, without reference to the data available to some other agent j. For ease of notation below, we suppress the dependence on i. Together, equations (4) and (3) describe the information structure in state space form (ssf). Then $x_t' = [\pi_t', \pi_{t-1}', \ldots, \pi_{t-p+1}']$ with $G = [I_m, 0, \ldots]$ (and I_m is the m-dimensional identity matrix) corresponds to the companion form, yielding a VAR(1) for the state vector, writing

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

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 square error matrix is $P_{t+1|t} = \mathrm{Cov}[x_{t+1} - \widehat{x}_{t+1|t}]$. The residual is the data minus its forecast, namely $e_t = y_t - \widehat{y}_{t|t-1}$, and its mean square error matrix is denoted V_t. The Kalman gain is by definition $K_t = \mathrm{Cov}[x_{t+1}, e_t]\,\mathrm{Var}[e_t]^{-1}$, 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 $P_{1|0} = \mathrm{Var}[x_1]$, 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

$$e_t = y_t - H\,\widehat{x}_{t|t-1} \tag{6}$$

$$V_t = H\,P_{t|t-1}\,H' + S_t \tag{7}$$

$$K_t = \Phi\,P_{t|t-1}\,H'\,V_t^{-1} \tag{8}$$

$$\widehat{x}_{t+1|t} = \Phi\,\widehat{x}_{t|t-1} + K_t\,e_t \tag{9}$$

$$P_{t+1|t} = (\Phi - K_t\,H)\;P_{t|t-1}\,\Phi' + G'\,\Sigma^{\epsilon}\,G. \tag{10}$$

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

$$\widehat{\pi}_{t+1|t} = G\,\widehat{x}_{t+1|t} \tag{11}$$

$$\mathrm{Cov}[\widehat{\pi}_{t+1|t} - \pi_{t+1}] = G\,P_{t+1|t}\,G'. \tag{12}$$

Equation (8) gives a recursive formula for the Kalman gain, and its dependence on the heteroscedastic noise is clearly given through V_t in (7). Moreover, equations (9) and (10) tell us how to update our one-step ahead prediction and forecast error variance for the state vector. Again, because the Kalman gain depends upon the heteroscedastic variance Σ_t, both the state vector forecast and its uncertainty will be impacted. To understand the Kalman gain better, observe that

$$\widehat{x}_{t|t} = \widehat{x}_{t|t-1} + P_{t|t-1} \, H' \, V_t^{-1} \, e_t, \tag{13}$$

which follows by applying Φ^{-1} to (9); hence $\Phi^{-1} K_t$ tells us the factor to multiply the new information e_t by in order to update $\widehat{x}_{t|t-1}$ to the revised quantity $\widehat{x}_{t|t}$. Rearranging this relationship, and utilizing (11) yields

$$\widehat{\pi}_{t+1|t+1} = \left(I_m - G \, \Phi^{-1} \, K_{t+1} \begin{bmatrix} A^{(i)} \\ B \end{bmatrix} \right) \widehat{\pi}_{t+1|t} + G \, \Phi^{-1} \, K_{t+1} \, y_{t+1}. \tag{14}$$

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

4.2 Estimation with Unknown Parameters

We describe below how the signal can be optimally estimated from the observed data, given a noise process that allows for time-varying volatility – this process is designed to mimic the ways in which agents might incorporate new information into their current forecasts.

In order to effectively describe the forecasting process, it is vital to give a flexible and broad specification for the signal and noise processes. If the signal is stationary, a flexible model is given by the VAR(p) class, where p is taken sufficiently large to approximate a generic signal. However, Θ must be drawn so as to ensure the stability of the resulting VAR(p) polynomial.[10] A bijective reparametrization of the stable VAR(p) class is provided in Roy, McElroy and Linton (2017), where a prior distribution can be simply placed upon unrestricted Euclidean space. The bijection maps arbitrary values in Euclidean space (of appropriate dimension) to a member of the stable VAR(p) class. A simple prior is obtained by adopting a diffuse Gaussian distribution. The variance matrix Σ^{ϵ} is also described in this mapping. The private noise has variance matrix $\Sigma^{(i)}$, which can be parameterized in the same manner as Σ^{ϵ}.

Regarding the dynamics about the variance of public signal, Σ_t, there is an extant literature on its specification using stochastic volatility processes.[11] 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 unit lower triangular and Ω_t is diagonal. The diagonal entries of Ω_t are assumed to each 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 zeroes 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.

Our own implementation of this framework utilizes the VAR(p) class with the exponential random walk model for volatility. Different user parameters dictate the priors, essentially determining the dispersion for the Gaussian prior for the VAR(p) coefficients, and the dis-

[10] Note that the Minnesota prior of Doan, Litterman and Sims (1984) does not ensure this, and hence cannot be utilized.

[11] Chiu, Leonard and Tsui (1996) modeled Σ_t as the matrix exponential of a symmetric matrix A_t (which can take negative values), whose vech was modeled as a VAR process. Uhlig (1994, 1997) modeled Σ_t via a generalized Cholesky decomposition, wherein the diagonal factor followed a positive random walk model. A Wishart autoregressive process was studied in Gouriéroux, Jasiak and Sufana (2009).

persion for the random walk increments in the volatility process.

A temporary shock can be generated by adding a diagonal matrix to a single Σ_t, but without altering B_t or Ω_t, so that the effect is transitory. This can be done by taking the exponential of a large variance random vector, and inserting the results into a diagonal matrix. This will be designated as a temporary shock, mimicking the shock arising from natural disasters.[12]

5 A Framework for Expectation Formation

Our framework has three key ingredients: (i) heterogeneous forecasters, i.e., attentive versus inattentive, (ii) stochastic volatility, and (iii) Kalman filter updating. As indicated by our empirical findings, there are two types of forecasters, attentive and inattentive, who are not interchangeable. Without loss of generality, let $i = 1, \ldots N$ denote the indices of attentive forecasters and $i = N + 1, \ldots, N + M$ correspond to inattentive forecasters. The key assumption here is that the fraction $q = \frac{N}{N+M}$ is exogenous.

5.1 Modeling Expectation Formation for Attentive Agents

We first develop the case of attentive agents. As noted earlier, the ith attentive agent $(1 \leq i \leq N)$ observes both public and private signals and makes the forecast through a signal extraction process. Given the data $\{y_t(i)\}$ for $1 \leq t \leq T$, for each agent i, we obtain $\widehat{\pi}^A_{t+1|t}(i)$ and its error covariance $G\, P_{t+1|t}(i)\, G'$ from (11) and (12), with the superscript A denoting attentive agents. If we have interest in some linear composite of attentive agents' results, say $\sum_{i=1}^N w_i \widehat{\pi}^A_{t+1|t}(i)$ for given weights w_i, then the corresponding target is $\sum_{i=1}^N w_i \pi_{t+1}$, which

[12]In contrast, a permanent shock can be generated by inserting a large variance innovation to the exponential random walk in Ω_t, at some particular time t. This generates a large upward jump in the diagonal entries of Σ_t, or equivalently in the trace of Σ_t. Such a device can replicate the impact of sustained shocks upon the variability of public information noise. The discussion regarding how a permanent shock can affect expectations formation is beyond the scope of the current paper, and we leave it for future research.

equals π_{t+1} when the weights sum to one. The variance of the discrepancy between such a weighted average and π_{t+1} is the mean squared error (MSE), given by

$$MSE^A_{t+1|t} = G \sum_{i,j=1}^{N} w_i \, w_j Q^{(ij)}_{t+1|t} \, G', \tag{15}$$

where by definition $Q^{(ij)}_{t+1|t} = \text{Cov}[\widehat{x}^A_{t+1|t}(i) - x_{t+1}, \widehat{x}^A_{t+1|t}(j) - x_{t+1}]$. For $i = j$, this covariance is just $P_{t+1|t}$. Otherwise, the following recursion can be used for computation; note that the Kalman gains $K_t(i)$ and observation matrices $H(i)$ depend upon the ith Kalman filter calculation.

Proposition 1 *The covariance of prediction errors across attentive agents, $Q^{(ij)}_{t+1|t}$, can be computed recursively by*

$$Q^{(ij)}_{t+1|t} = [\Phi - K_t(i)\, H(i)]\, Q^{(ij)}_{t|t-1} [\Phi - K_t(j)\, H(j)]' + G'\, \Sigma^{\epsilon}\, G - K_t(i) \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_t \end{bmatrix} K_t(j)', \tag{16}$$

with the initialization $Q^{(ij)}_{1|0} = Var[x_1]$ for all i and j.

Typically we are interested in the case that attentive agents contribute equally to the composite forecast, in which the MSE is given by (15), where $Q^{(ij)}_{t+1|t}$ is computed via equation (16) and all the weights are $w_i = 1/N$. In this case, the average forecast is defined to be

$$\overline{\pi}^A_{t+1|t} = N^{-1} \sum_{i=1}^{N} \widehat{\pi}^A_{t+1|t}(i). \tag{17}$$

The disagreement across attentive agents is defined as the sample variability of the forecasts across forecasters, i.e.,

$$D^A_{t+1|t} = N^{-1} \sum_{i=1}^{N} \left(\widehat{\pi}^A_{t+1|t}(i) - \overline{\pi}^A_{t+1|t}\right) \left(\widehat{\pi}^A_{t+1|t}(i) - \overline{\pi}^A_{t+1|t}\right)'. \tag{18}$$

We now discuss the model implications for understanding information rigidity for attentive agents. Extending the formulation of Coibion and Gorodnichenko (2012) for the homoscedastic public noise, (14) indicates that the old forecast $\widehat{\pi}_{t+1|t}$ is scaled by $I_m - R_t$, where

$$R_t = G\,\Phi^{-1}\,K_{t+1}\begin{bmatrix} A^{(i)} \\ B \end{bmatrix}. \tag{19}$$

Note that R_t is $m \times m$ dimensional. When the Kalman gain is small, little modification to the old forecast is needed. From (7) and (8), clearly K_t is small when S_t is large – a sudden jump in Σ_t (irrespective of $\Sigma^{(i)}$) 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. This is the uncertainty channel through which natural disaster shocks affect expectation formation: the occurrence of those shocks generates increased uncertainty among economic agents.

Formally, the information rigidity, defined as the sequence

$$r_t = \operatorname{tr}\left[I_m - R_t\right]/m, \tag{20}$$

is high when the new datum is deemed untrustworthy, i.e., when Σ_t is high. The factor of m in the definition of information rigidity normalizes for the dimension m of the signal $\{\pi_t\}$. Clearly, r_t offers a scalar normalization of the quantity $I_m - R_t$ that multiplies the old forecast. Note that information rigidity is also a function of the private information variability; the relative strength of $\Sigma^{(i)}$ and Σ_t must play a role, as these quantities are featured in S_t and hence V_t. For example, a lower value of $\Sigma^{(j)}$ for agent j means this forecaster pays close attention to finer fluctuations in the public data source volatility, and their information rigidity adapts more readily to nuanced changes in the market. Their rigidity is lower overall, and is less susceptible to sudden spikes to public information variability than inattentive forecasters. Our measure in (20) 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 differential interpretation of public information, as documented in Lahiri and Sheng (2008) and Acemoglu, Chernozhukov and Yildiz (2016). Third, information rigidity is allowed to change over time in order to capture the increased uncertainty among attentive agents due to large shocks. This feature is consistent with the recent studies on uncertainty shocks, e.g. Bloom (2009), Jurado, Ludvigson and Ng (2015), among others.

5.2 Modeling Expectation Formation for Inattentive Agents

We now develop the case of inattentive agents. The ith inattentive agent, $N+1 \leq i \leq N+M$, observes the signal $\{\pi_t\}$ through an observation matrix B, and obfuscated by a common noise $\{\eta_t\}$, yielding the observation equation

$$y_t = B\,\pi_t + \eta_t.$$

The matrix B represents any linear combinations of the signal that dictate how the signal is manifested publicly to the agent. During normal times, inattentive agents only update sporadically. Following Reis (2006), we let $\lambda \in [0,1]$ denote the probability that any given inattentive agent fails to update their information set, and operates under an older vintage of information. We refer to λ as the degree of information rigidity for inattentive agents. When they update, inattentive agents face the signal extraction problem. For those who do not update, they simply set their forecasts to the ones in the previous period.

Let $\mathcal{G}_t$ denote the sigma-field generated by $\{y_s, s \leq t\}$, i.e., the information about the process available up to time t. Then if an agent at time $t+1$ has just updated his information, he would utilize the signal extraction forecast $\mathbb{E}[\pi_{t+1}|\mathcal{G}_t]$. However, if that agent had last updated two periods ago then his forecast would be $\mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-1}]$. In general, we denote such a forecast by inattentive agent i by $\widehat{\pi}^{IA}_{t+1|t}(i)$, with the notation indicating that a forecast of π_{t+1} is furnished at time t, even though with probability λ the forecast is generated with older vintages. The average of such forecasts is $M^{-1}\sum_{i=N+1}^{N+M}\widehat{\pi}^{IA}_{t+1|t}(i)$; breaking the sum into subsets of agents that last updated their forecast at time $t-k$ (with $k \geq 0$) – and with A_k denoting the number of such – we obtain

$$M^{-1}\sum_{i=N+1}^{N+M}\widehat{\pi}^{IA}_{t+1|t}(i) = M^{-1}\sum_{k=0}^{\infty} A_k\,\mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-k}].$$

When M is large, the proportion A_k/M approximates the probability of not updating for k time periods, followed by finally updating at the next time – this probability is given by the geometric distribution, and equals $(1-\lambda)\,\lambda^k$. Hence an asymptotic approximation to the average forecast is

$$\overline{\pi}^{IA}_{t+1|t} = (1-\lambda)\sum_{k=0}^{\infty}\lambda^k\,\mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-k}]. \tag{21}$$

The degree of variability across inattentive agents with respect to the group mean is called the disagreement, and is computed by averaging squared discrepancies of each $\widehat{\pi}^{IA}_{t+1|t}(i)$ with

$\overline{\pi}^{IA}_{t+1|t}$. Using the same approximation for large M that was used to derive (21) yields

$$D^{IA}_{t+1|t} = (1-\lambda)\sum_{k=0}^{\infty}\lambda^k \left(\mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-k}] - \overline{\pi}^{IA}_{t+1|t}\right)\left(\mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-k}] - \overline{\pi}^{IA}_{t+1|t}\right)'. \tag{22}$$

In order to assess the accuracy of the group inattentive mean in an ensemble sense, we can calculate the mean squared error as follows:

$$\begin{aligned} \mathrm{MSE}^{IA}_{t+1|t} &= \mathbb{E}\left[\left(\overline{\pi}^{IA}_{t+1|t} - \pi_{t+1}\right)\left(\overline{\pi}^{IA}_{t+1|t} - \pi_{t+1}\right)'\right] \\ &= (1-\lambda)^2 \sum_{k,\ell\geq 0}\lambda^{k+\ell}\,\mathrm{Cov}\left(\mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-k}] - \pi_{t+1}, \mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-\ell}] - \pi_{t+1}\right). \end{aligned} \tag{23}$$

Apparently, lower values of λ indicate that more timely forecasts receive higher weight in (21), and the MSE will decrease. However, other costs (such as maintaining a database, and the resources required to do analysis) may render attractive the behavior of using a higher value of λ during normal times. In contrast, following large and unexpected shocks, the cost of not updating the information vintage is very high, and accordingly the majority of inattentive agents will update their information and make new forecasts. This is the attention channel through which natural disaster shocks affect expectation formation: visibly large shocks induce immediate updating of information for most agents. This attention effect is particularly pronounced for those with an outdated information set, resulting in a significant decline in λ.

Finally, it is possible using the state space form to provide a specific formula for the signal extraction error covariance featured in equation (23), the MSE for inattentive agents. Here the formula is for a single agent, so we do not insert the i index in the notation. Replacing the conditioning on the semi-infinite sigma-field $\mathcal{G}_t$ by projection on a finite past, the MSE is modified to the formula

$$\text{MSE}^{IA}_{t+1|t} = (1-\lambda)^2 \sum_{k,\ell=0}^{t-1} \lambda^{k+\ell}\, G\, R_{k,\ell}\, G' \tag{24}$$

with $R_{k,\ell} = \text{Cov}\left[\widehat{x}^{IA}_{t+1|t-k} - x_{t+1}, \widehat{x}^{IA}_{t+1|t-\ell} - x_{t+1}\right]$, and the dependence of $R_{k,\ell}$ on t has been suppressed in the notation. The formula, given below, is only needed for the case of inattentive agents where no private noise is present, so the ssf quantities (such as Kalman gain) are to be interpreted accordingly by omitting $A^{(i)}$ and $\nu_t(i)$.

Proposition 2 *The covariance of prediction errors across forecast horizons, $R_{k,\ell}$, can be computed if $k \leq \ell$ by*

$$R_{\ell,\ell} = \Phi^{\ell}\, P_{t+1|t}\, \Phi'^{\ell}$$

$$R_{\ell-1,\ell} = \Phi^{\ell-1}\,(\Phi - K_{t+1-\ell}\, B)\, P_{t+1-\ell|t-\ell}\, \Phi'^{\ell} + \sum_{j=0}^{\ell-1} \Phi^{j}\, \Sigma^{\epsilon}\, \Phi'^{j}$$

$$R_{k,\ell} = \Phi^{k} \prod_{j=k}^{\ell-1} (\Phi - K_{t-j}\, B)\, P_{t+1-\ell|t-\ell}\, \Phi'^{\ell} + \sum_{i=2}^{\ell-k} \Phi^{k} \prod_{j=k}^{\ell-i} (\Phi - K_{t-j}\, B)\, \Sigma^{\epsilon}\, \Phi'^{\ell-i+1} + \sum_{j=0}^{k} \Phi^{j}\, \Sigma^{\epsilon}\, \Phi'^{j},$$

where $k \leq \ell - 2$ in the last case, and where the matrix products are computed with the lowest index matrix first, and multiplying on the right by matrices of higher index.

6 Monte Carlo Simulation

6.1 Summary of Model Implications

We now summarize the model implications for understanding information rigidity, mean forecast, mean squared error and disagreement among agents. As above we set $i = 1, \ldots N$ to be the indices of attentive agents and $i = N+1, \ldots, N+M$ correspond to inattentive agents, with $q = \frac{N}{N+M}$. Let τ denote the time index of a shock. Besides the individual treatment of attentive and inattentive agents, it is of interest to study the composite of these two groups.

i. Attentive agents all form their expectations through a signal extraction process, as described in section 4. Their information rigidity is described in equation (20), while the mean forecast is (17). The MSE is given by (15), where $Q^{(ij)}_{t+1|t}$ is defined in equation (16) and all the weights are $w_i = 1/N$. The disagreement among attentive agents is provided by (18).

ii. Inattentive agents face a constant probability λ of not updating their information (e.g., $\lambda = 0.7$), as described in section 5.2. We refer to λ as the degree of information rigidity for inattentive agents. When updating, they face a signal extraction problem. In periods of not updating, they simply set their forecasts to the previous ones. Their mean forecast is defined in equation (21) by assuming equal weights. Assuming that these inattentive agents update their information independently of each other, the MSE is given by (24), where the forecast error covariance is given in Proposition 2. Moreover, the disagreement among inattentive agents is provided by (22).

iii. Combining attentive and inattentive agents, the average forecast, $\overline{\pi}^{ALL}_{t+1|t}$, is defined as the average of forecasts from all $N + M$ agents, and this can be approximated (for large M) by the weighted sum

$$\overline{\pi}^{ALL}_{t+1|t} = q\, N^{-1} \sum_{i=1}^{N} \widehat{\pi}^{A}_{t+1|t}(i) + (1-q)\,(1-\lambda) \sum_{k=0}^{\infty} \lambda^k\, \mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-k}]. \tag{25}$$

The exact MSE is difficult to compute, because of correlations between the attentive and inattentive forecast errors, but we can get a heuristic measure by weighting each contribution towards MSE:

$$\mathrm{MSE}^{ALL}_{t+1|t} = q\, N^{-2} \sum_{i,j=1}^{N} G\, Q^{(ij)}_{t+1|t}\, G' + (1-q)\, \mathrm{MSE}^{IA}_{t+1|t}. \tag{26}$$

Finally, the overall disagreement, $D^{ALL}_{t+1|t}$ can be expressed as:

$$\begin{aligned} D^{ALL}_{t+1|t} &= q\, N^{-1} \sum_{i=1}^{N} \left(\widehat{\pi}^{A}_{t+1|t}(i) - \overline{\pi}^{A}_{t+1|t}\right) \left(\widehat{\pi}^{A}_{t+1|t}(i) - \overline{\pi}^{A}_{t+1|t}\right)' \\ &+ (1-q)\,(1-\lambda) \sum_{k=0}^{\infty} \lambda^k \left(\mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-k}] - \overline{\pi}^{IA}_{t+1|t}\right) \left(\mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-k}] - \overline{\pi}^{IA}_{t+1|t}\right)' \\ &+ q\,(1-q) \left\| N^{-1} \sum_{i=1}^{N} \widehat{\pi}^{A}_{t+1|t}(i) - (1-\lambda) \sum_{k=0}^{\infty} \lambda^k\, \mathbb{E}[\pi_{t+1}|\mathcal{G}_{t-k}] \right\|^2, \end{aligned} \tag{27}$$

where $\|\cdot\|$ denotes the ℓ_2 norm of a vector. Clearly, overall disagreement in (27) comes from heterogeneity within attentive agents (first line), within inattentive agents (second line) and between two group mean forecasts (third line).

6.2 Data Generating Process

We illustrate these assertions 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 upon empirical fits of industrial production, inflation, and federal funds rate data. This is merely intended to furnish a reasonable signal DGP. The coefficients are

$$\Phi_1 = \begin{bmatrix} 0.7745 & -0.1472 & -0.4292 \\ 0.0556 & 0.3139 & 0.1116 \\ 0.1066 & 0.0582 & 1.3309 \end{bmatrix} \qquad \Phi_2 = \begin{bmatrix} -0.0517 & 0.0979 & 0.4708 \\ 0.0170 & 0.1813 & -0.0151 \\ -0.0054 & 0.0384 & -0.3872 \end{bmatrix},$$

and 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 $A^{(i)}$ 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 were Gaussian of standard deviation .01 for all three dimensions. In the case of a shock, its impact on the public noise is modeled by increasing the standard deviation to .2 (twenty times as large).

To generate the private noise, an overall dispersion coefficient with value .01 controls the spread of entries in the covariance matrix, whereas the overall scale is determined via multiplication by $20\,Y_i$, where Y_i has a χ^2 distribution of one degree of freedom, generated independently for all attentive agents. This gives some spread and variability to the private noise; the settings were determined by empirically examining various cases and studying the resulting behavior of simulations. The number of attentive agents was set to $N = 100$. For purposes of understanding the case of combining attentive and inattentive agents, we set $M = 200$ so that $q = 1/3$.

6.3 Simulation Results

Figure 4 presents the degrees of information rigidity for both types of agents. For inattentive agents, 70% do not update their information sets during the normal times. After observing a shock occurred at time $\tau = 50$, the majority of inattentive agents update information and revise their forecasts, which accordingly leads to a decline of information rigidity from 0.7 to 0.2.[13] The opposite is true for attentive agents: during the normal times, they place about 0.33 weight to previous forecasts relative to new information, but observing the shock significantly increases the average degree of their information rigidity from 0.33 to 0.45 in

[13]For simplicity, we assume that the probability of not updating information, λ, is the same across three variables for inattentive agents. Allowing for different levels of λ across variables (e.g. Andrade, et al. 2016) does not change the results qualitatively for the effect of a large unexpected shock on information processing of inattentive agents.

predicting three variables. Because the shock is temporary, information rigidity at time $\tau = 51$ for both types of agents returns to the pre-shock level.

The dynamics of mean squared error is shown in Figure 5. Generally speaking, inattentive agents make much larger forecast errors than attentive ones and as a result, the overall MSE is dominated by the former. However, agents respond differently in terms of their MSE following the shock. In particular, the MSE declines for inattentive agents, since the visibly large shock induces immediate and synchronized updating of information and moves their forecasts much closer to the truth. In contrast, the large unexpected shock generates increased uncertainty among attentive agents and increases their MSE. Following the shock, the overall MSE declines due to the dominance of inattentive agents, confirming the empirical results presented earlier.

Forecast disagreement across agents shows a similar pattern, as illustrated in Figure 6. There exists substantial disagreement among both types of agents in predicting any of three variables. For inattentive agents, they disagree mainly because they use different vintages of dataset; for attentive agents, they disagree since they receive different private signals and interpret the same public signal differently. Inattentive agents have much higher levels of disagreement and accordingly play a dominant role in the overall disagreement. Following the shock, most of inattentive agents have updated their outdated information and moved their forecasts closer to the average. Attentive agents, on the other hand, disagree more after observing the shock due to differential interpretation of public information. The overall disagreement declines following the shock, which appears counter-intuitive but in fact is a natural consequence of the dominance of inattentive agents.

We need to point out that the effect of a large, unexpected shock on overall mean squared error and disagreement depends on the magnitude of the shock, the proportion of inattentive agents and the average deviation of inattentive agents' mean forecast from the truth. Both our empirical and simulation results show that inattentive agents tend to dominate and thus, overall mean squared error and disagreement decline following large, unexpected shocks. Our model, however, has implications for different paths of overall mean squared error and disagreement following large, unexpected events.

7 Conclusion

This paper provides a new view on what drives macroeconomic forecasters to update and provide forecasts. We find that individual forecasters are persistently heterogeneous in how often they revise or even issue a forecast over time. Given that many commonly utilized 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.

We demonstrate that there is a significant degree of information rigidity in forecasts over time, driven by the fact that many forecasters choose to not 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. Moreover, following such disasters, previously inattentive forecasters move their stale forecasts closer to the mean forecast, decreasing the dispersion in forecasts. This may result in a counter-intuitive result, where shocks to countries can increase uncertainty but decrease forecaster dispersion due to the fact that inattentive forecasters may be induced to update their stale information sets.

We model this phenomenon with a learning model that incorporates these two channels: attention effect – the visibly large shocks induce immediate and synchronized updating of information for inattentive agents, and uncertainty effect – the occurrence of those shocks generates increased uncertainty among attentive agents. Our theory has three key elements. (i) There are two persistent types of agents: attentive and inattentive, with attentive agents revising forecasts frequently and inattentive agents generally revising forecasts infrequently and non-systematically. (ii) Inattentive agents update sporadically during the normal times. However, following large and unexpected shocks, the cost of not updating information is high, and accordingly the majority of inattentive agents will update their information sets and make new forecasts. (iii) Attentive agents optimally combine public and private signals in making their forecasts, as described in the noisy information model.

Our model yields a world in which large shocks like natural disasters induce an immediate increase in updating of information for most inattentive agents. This attention effect is particularly pronounced for those with an outdated information set, resulting in a significant decline in the information rigidity. For attentive agents, the occurrence of those shocks generates increased uncertainty and as a result, their new forecasts closely resemble the previous period's forecasts. 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 accordingly proposing a measure of time-varying information rigidity in a multivariate context.

Finally, our paper suggests that there is room to improve so-called consensus or mean forecasts. We find that many individual forecasters are persistently more inattentive or inactive than others. Accounting for this heterogeneity across forecasters can help explain both the information rigidity observed in many commonly-utilized macroeconomic forecasts and a reason why forecast dispersion may actually decline, depending on the portion of inattentive forecasters, following a national 'shock' that increases uncertainty about future economic performance.

References

[1] Abel, Andrew (1983). Optimal investment under uncertainty. *American Economic Review*, 73, 228–233.

[2] Acemoglu, Daron, Chernozhukov, Victor, and Yildiz, Muhamet (2016). Fragility of asymptotic agreement under Bayesian learning. *Theoretical Economics* 11, 187–225.

[3] Andrade, Philippe and Le Bihan, Hervé (2013). Inattentive professional forecasters. *Journal of Monetary Economics* 60, 976–982.

[4] Andrade, Philippe, Crump, Richard, Eusepi, Stefano, and Moench, Emanuel (2016). Fundamental disagreement. *Journal of Monetary Economics* 83, 106–128.

[5] Baker, Scott R., Bloom, Nick, and Davis, Steve J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics* 131, 1593–1636.

[6] Baker, Scott R., Bloom, Nick, and Terry, Stephen (2017). Does uncertainty reduce growth? Using disasters as natural experiments. *NBER Working Paper*, No. 19475.

[7] Bernanke, Ben S. (1983). Irreversibility, uncertainty and cyclical investment. *Quarterly Journal of Economics*, 97, 85–106.

[8] Bloom, Nick (2009). The impact of uncertainty shocks. *Econometrica* 77, 623–685.

[9] Branch, William A. (2004). The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations. *Economic Journal* 114, 592–621.

[10] Branch, William A. (2007). Sticky information and model uncertainty in survey data on inflation expectations. *Journal of Economic Dynamics and Control* 31, 245-276.

[11] Carroll, Christopher D. (2003). Macroeconomic expectations of households and professional forecasters. *Quarterly Journal of Economics* 118, 269–298.

[12] Chiu, Tom Y.M., Leonard, Tom, and Tsui, Kam-Wah (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association* 91(433), 198–210.

[13] Cogley, Timothy and Sargent, Thomas J. (2005). Drift and volatilities: monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics* 8, 262–302.

[14] Coibion, Olivier (2010). Testing the sticky information Phillips curve. *Review of Economics and Statistics* 92(1), 87–101.

[15] Coibion, Olivier and Gorodnichenko, Yuriy (2012). What can survey forecasts tell us about informational rigidities? *Journal of Political Economy* 120, 116–159.

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

[17] Doan, Thomas, Litterman, Robert B., and Sims, Christopher A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3(1), 1–100.

[18] Giacomini, Raffaella, Skreta, Vasiliki, and Turén, Javier (2016). Models, inattention and Bayesian updates. *mimeo*

[19] Gorodnichenko, Yuriy (2008). Endogenous information, menu costs, and inflation persistence. *NBER Working Paper* No. 14184.

[20] Gouriéroux, Christian, Jasiak, Joann, and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics* 150, 167–181.

[21] Ilut, Cosmin L., and Schneider, Martin (2014). Ambiguous business cycles. *American Economic Review* 104(8), 2368-2399.

[22] Jurado, Kyle, Ludvigson, Sydney C., and Ng, Serena (2015). Measuring uncertainty. *American Economic Review* 105(3), 1177–1216.

[23] Lahiri, Kajal, and Sheng, Xuguang (2008). Evolution of forecast disagreement in a Bayesian learning model. *Journal of Econometrics* 144, 325–340.

[24] Maćkowiak, Bartosz, and Wiederholt, Mirko (2009). Optimal sticky prices under rational inattention. *American Economic Review* 99, 769–803.

[25] Mankiw, N. Gregory, and Reis, Ricardo (2002). Sticky information versus sticky prices: a proposal to replace the new Keynesian Phillips curve. *Quarterly Journal of Economics* 117, 1295–1328.

[26] Mankiw, N. Gregory, Reis, Ricardo, and Wolfers, Justin (2004). Disagreement about inflation expectations. In *NBER Macroeconomics Annual 2003*, edited by Gertler, M., and Rogoff, K., 209–248. Cambridge, MA: MIT Press.

[27] Manski, Charles F. (2017). Survey measurement of probabilistic macroeconomic expectations: progress and promise. *NBER Working Paper* No. 23418.

[28] Neusser, Klaus (2016). A topological view on the identification of structural autoregressions. *Economics Letters* 144, 107–111.

[29] Nimark, Kristoffer P. (2014). Man-bites-dog business cycles. *American Economic Review* 104, 2320–2367.

[30] Nordhaus, William (1987). Forecasting efficiency: concepts and applications. *Review of Economics and Statistics* 69(4), 667–674.

[31] Ortoleva, Pietro (2012). Modeling the change of paradigm: non-Bayesian reactions to unexpected news. *American Economic Review* 102, 2410–2436.

[32] Pastor, Lubos, and Veronesi, Pietro (2012). Uncertainty about government policy and stock prices. *Journal of Finance*, 67, 1219–1264.

[33] Patton, Andrew J., and Timmermann, Allan (2010). Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics* 57, 803–820.

[34] Primiceri, Giorgio E. (2005). Time varying structural autoregressions and monetary policy. *Review of Economic Studies* 72: 821–852.

[35] Reis, Ricardo (2006). Inattentive producers. *Review of Economic Studies* 73, 793–821.

[36] Roy, Anindya, McElroy, Tucker S., and Linton, Peter (2017). Constrained estimation of causal invertible VARMA. Forthcoming in *Statistica Sinica*.

[37] Sims, Christopher A. (2003). Implications of rational inattention. *Journal of Monetary Economics* 50, 665–690.

[38] Uhlig, Harald (1994). On singular Wishart and singular multivariate beta distributions. *The Annals of Statistics* 22, 395–405.

[39] Uhlig, Harald (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica* 65, 59–73.

[40] Woodford, Michael (2003). Imperfect common knowledge and the effects of monetary policy. In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund Phelps*, edited by Aghion, P., Frydman, R., Stiglitz, J., and Woodford, M. Princeton, NJ: Princeton Univ. Press.

Table 1. Information Rigidities Based on GDP Forecast Error

VARIABLES	(1) Forecast Error	(2) Forecast Error	(3) Forecast Error	(4) Forecast Error
Forecast Revision	0.514***	0.555***	0.530***	0.541***
	(0.0286)	(0.0307)	(0.0294)	(0.0301)
Forecast Rev * Disaster		-0.286***		
		(0.0787)		
Disaster		-0.0339		
		(0.0341)		
Forecast Rev * Disaster (News Scaling)			-0.106**	
			(0.0448)	
Disaster (News Scaling)			-0.0304	
			(0.0196)	
Forecast Rev * Disaster (Combined Scaling)				-0.183***
				(0.0649)
Disaster (Combined Scaling)				-0.0617**
				(0.0282)
Observations	11,408	11,408	11,408	11,408
R^2	0.295	0.296	0.296	0.296
Time FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: 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 5 days following a disaster to those written in the 5 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 comprised 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).

Table 2. Information Rigidities Based on GDP Forecast Revision

VARIABLES	(1) Forecast Rev	(2) Forecast Rev	(3) Forecast Rev	(4) Forecast Rev
Lagged Forecast Revision	0.139***	0.149***	0.145***	0.148***
	(0.00924)	(0.00988)	(0.00949)	(0.00972)
Lag Forecast Rev * Disaster		-0.0736***		
		(0.0259)		
Lagged Disaster		0.00223		
		(0.0115)		
Lag Forecast Rev * Disaster (News Scaling)			-0.0428***	
			(0.0150)	
Lagged Disaster (News Scaling)			-0.00943	
			(0.00661)	
Lag Forecast Rev * Disaster (Combined Scaling)				-0.0604***
				(0.0214)
Lagged Disaster (Combined Scaling)				-0.00893
				(0.00946)
Observations	11,444	11,444	11,444	11,444
R^2	0.106	0.107	0.107	0.107
Time FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: 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 5 days following a disaster to those written in the 5 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 comprised 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).

Table 3. Individual Forecast Dispersion

VARIABLES	(1) Diff. From Mean	(2) Fore. Error	(3) Diff. From Mean	(4) Fore. Error
Attentive Forecaster	-0.0724***	-0.395***		
	(0.00717)	(0.0416)		
Fraction Forecasts Reported			-0.177***	-1.147***
			(0.0173)	(0.148)
Observations	292,390	176,985	292,390	176,985
R^2	0.054	0.128	0.054	0.128
Time FE	YES	YES	YES	YES
VAR FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: 'Forecast Error' denotes the absolute value of the difference of the individual forecast from the ex-post true GDP growth 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 covers 280 forecasters and 7 countries for GDP, CPI, consumption, short- and long-run interest rates, unemployment, wages, and producer prices.

Table 4. Individual Forecast Changes and Disasters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Change	Change	Change - A	Change - I	Change - A	Change - I	Change
Scaled Disaster * Attentive Quintile							-0.00285***
							(0.000701)
Scaled Disaster							0.0120***
							(0.00252)
Attentive Quintile							0.0877***
							(0.00134)
Disaster	0.00857***	0.00192	0.00336	0.0114***	0.000857	0.000183	
	(0.00223)	(0.00310)	(0.00413)	(0.00259)	(0.00541)	(0.00373)	
Scaled Disaster		0.00576***			0.00193	0.0101***	
		(0.00187)			(0.00270)	(0.00241)	
Observations	292,390	292,390	64,255	228,135	64,255	228,135	292,335
R^2	0.149	0.149	0.069	0.113	0.069	0.113	0.162
Time FE	YES	YES	YES	YES	YES	YES	YES
VAR FE	YES	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES	YES
Forecaster FE	YES	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: 'Attentive Quintile' refers to the quintile of attentiveness a forecaster belongs to in terms of fraction of new forecasts that forecaster has reported during their 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 5 days following a disaster to those written in the 5 days before a disaster (has a mean of 1 and a max value of 6). 'Change' is an indicator variable for whether that forecaster changed their forecast from the previous month for that country-variable. Columns 3 and 5 restrict the sample to forecasters in the top quintile of attentiveness while columns 4 and 6 restrict to the bottom 4 quintiles of attentiveness. Data covers 280 forecasters and 7 countries for GDP, CPI, consumption, short- and long-run interest rates, unemployment, wages, and producer prices.

Table 5. Dispersion Following Disasters

	(1)	(2)	(3)	(4)	(5)	(6)
	Diff from Overall Mean			Diff from Attentive Group Mean		
VARIABLES	All Forecasters	Attentive Fore.	Inattentive Fore.	All Forecasters	Attentive Fore.	Inattentive Fore.
Scaled Disaster	-0.00669	-0.00180	-0.00876*	-0.00426	0.00111	-0.0172**
	(0.00408)	(0.00679)	(0.00503)	(0.00504)	(0.00635)	(0.00744)
Observations	288,997	63,455	225,542	284,749	224,064	60,684
R^2	0.022	0.044	0.019	0.019	0.019	0.035
Time FE	YES	YES	YES	YES	YES	YES
VAR FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
Forecaster FE	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: 'Attentive Forecaster' refers to a forecaster being in the top quintile of attentiveness in terms of fraction of new forecasts that forecaster has reported during their tenure. 'Scaled Disaster' refers to the ratio of news articles written about a country in the 5 days following a disaster to those written in the 5 days before a disaster (has a mean of 1 and a max value of 6). 'Difference from the Overall Mean' refers to the difference between the individual forecast and the mean forecast across all forecasters. 'Difference from Attentive Group Mean' refers to the difference between the individual forecast and the mean forecast across all forecasters in their attentiveness group (eg. top quintile of attentiveness or bottom 4 quintiles of attentiveness). Data covers 280 forecasters and 7 countries for GDP, CPI, consumption, short- and long-run interest rates, unemployment, wages, and producer prices.

Table 6. Forecast Accuracy and Disasters

VARIABLES	(1) Fore. Error	(2) Fore. Error - A	(3) Fore. Error - I	(4) Fore. Error	(5) Fore. Error - A	(6) Fore. Error - I
Changed Forecast				-0.0497***	-0.0343**	-0.0575***
				(0.00565)	(0.0167)	(0.00601)
Scaled Disaster	-0.0252***	-0.00440	-0.0260***	-0.0258***	-0.00458	-0.0265***
	(0.00682)	(0.0175)	(0.00741)	(0.00682)	(0.0175)	(0.00740)
Observations	154,274	25,350	128,923	154,274	25,350	128,923
R^2	0.085	0.110	0.088	0.085	0.110	0.089
Time FE	YES	YES	YES	YES	YES	YES
VAR FE	YES	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	YES	YES	YES
Forecaster FE	YES	YES	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: ‘Forecast Error’ denotes the absolute value of the difference of the individual forecast from the ex-post true GDP growth value. ‘Disaster’ is an indicator variable for a disaster occurring in a particular country-month for over 1,000 natural disasters in the sample. ‘Changed Forecast’ is an indicator variable for whether that forecaster changed their forecast from the previous month for that country-variable. Columns 2 and 5 restrict the sample to forecasters in the top quintile of attentiveness while columns 3 and 6 restrict to the bottom 4 quintiles of attentiveness. Data covers 280 forecasters and 7 countries for GDP, CPI, consumption, short- and long-run interest rates, unemployment, wages, and producer prices.

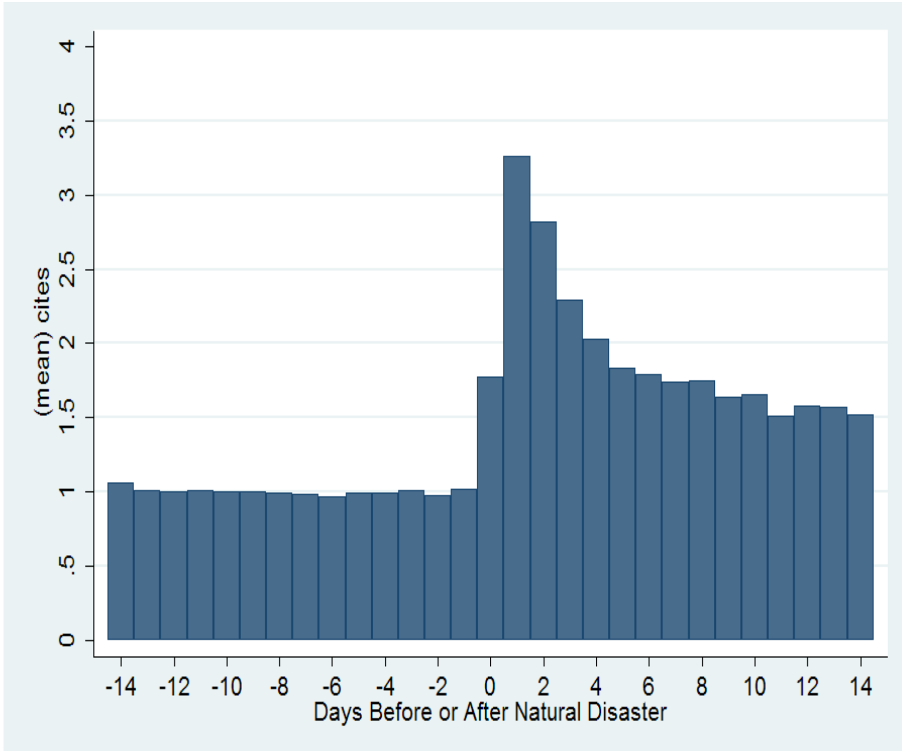


Figure 1: Changes in Newspaper Articles Regarding Affected Countries

Notes: 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 pre-period has a mean of one.

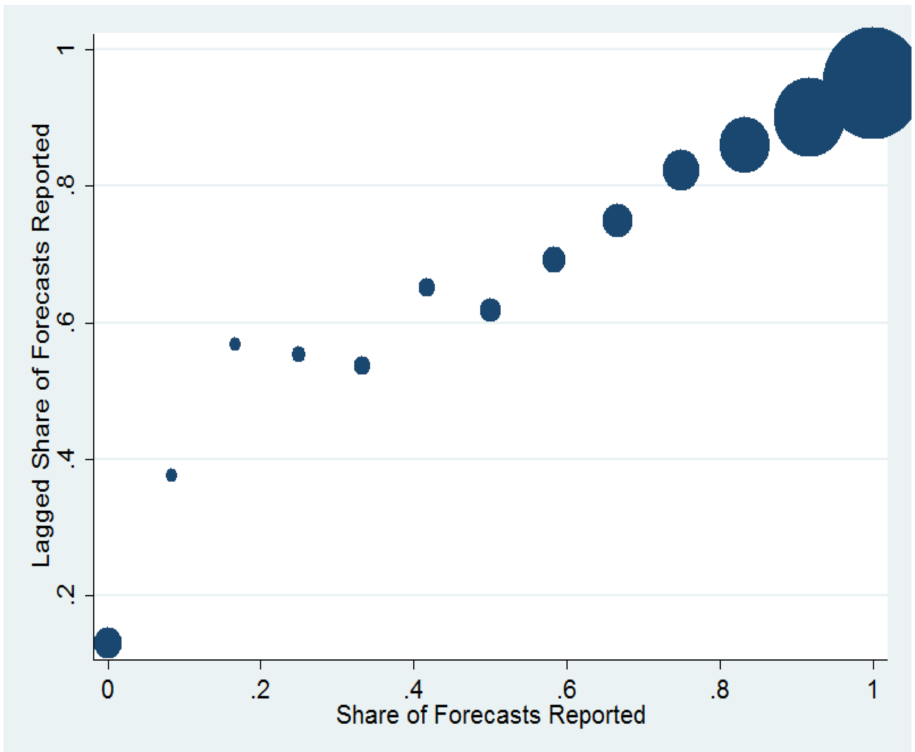


Figure 2: Persistence in Forecast Reporting

Notes: 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 a one month out of twelve months increment). Thus, a point on the 45-degree line means that, on average, forecasters in that group reported forecasts at the same frequency as last year. Plotted points are scaled by the number of forecasters in each bin.

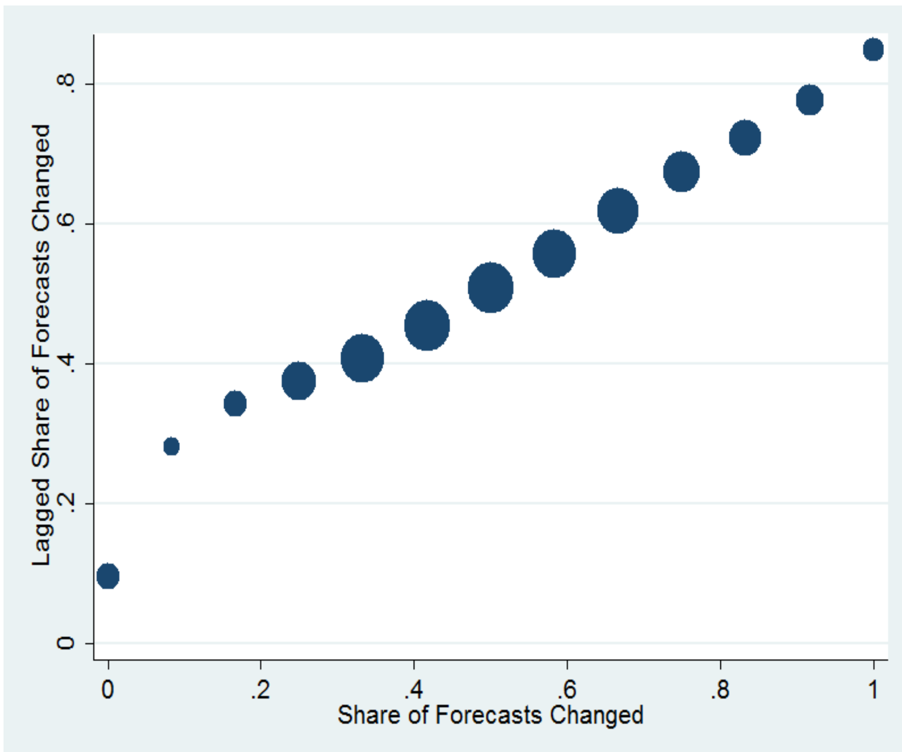
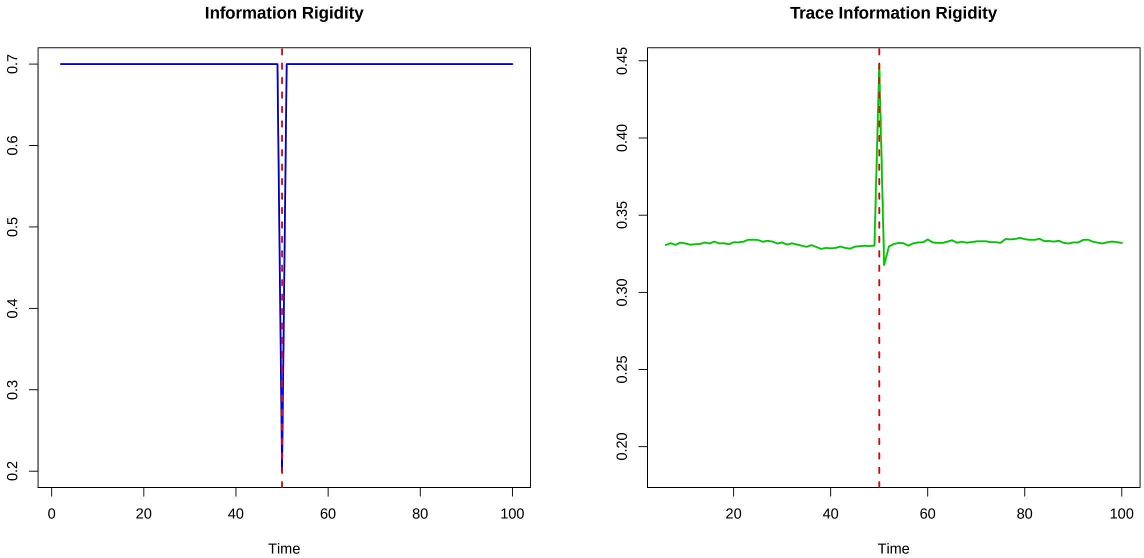


Figure 3: Persistence in Forecast Updating

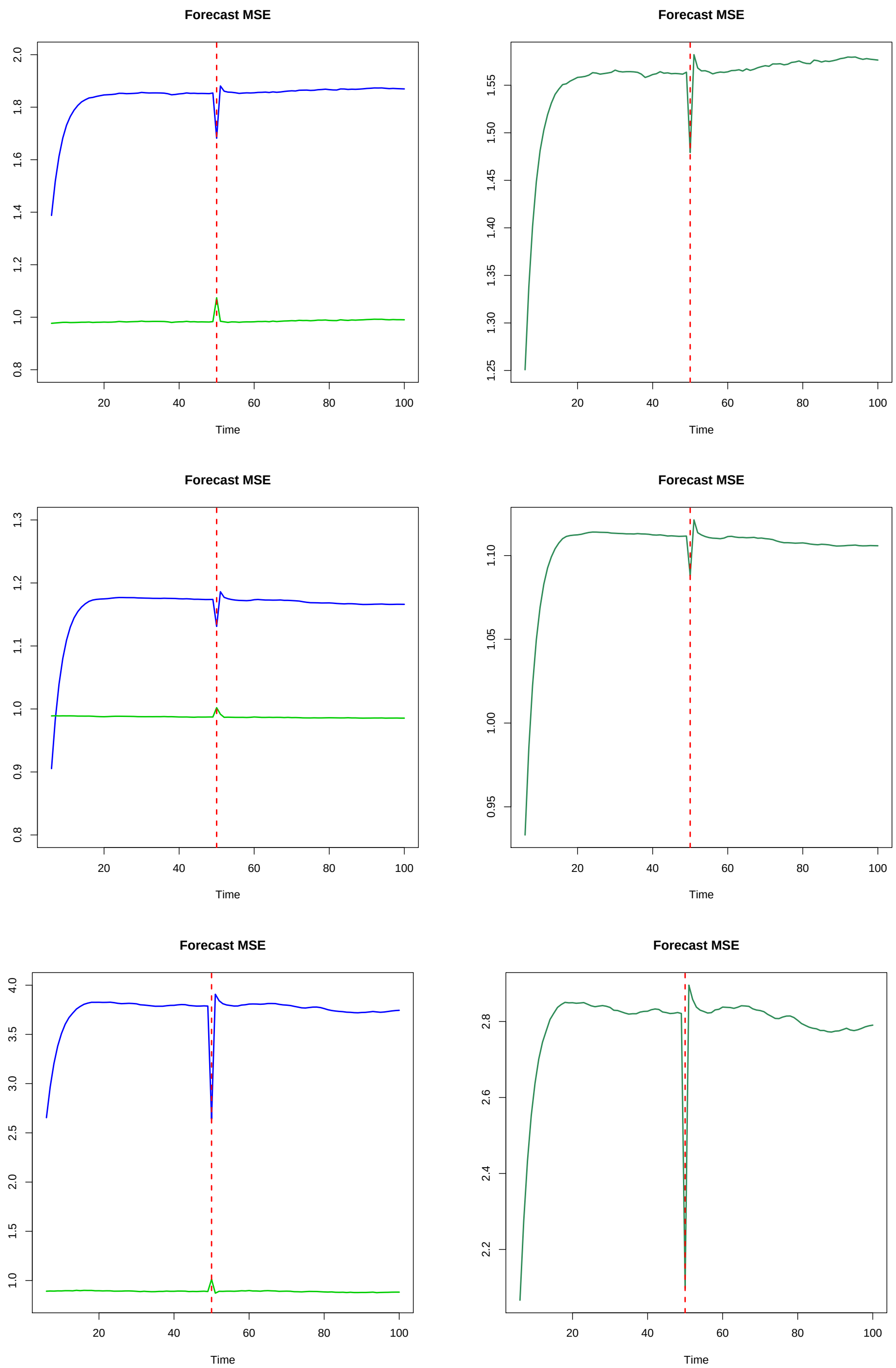
Notes: The vertical axis represents the share of eligible months where a given forecaster changed their forecast from the previous month in year $t-1$. The horizontal axis represents the share of eligible months where a given forecaster changed their forecast from the previous month in year t. Mean values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

Figure 4: Information Rigidity: Inattentive (left) vs. Attentive Agents (right)



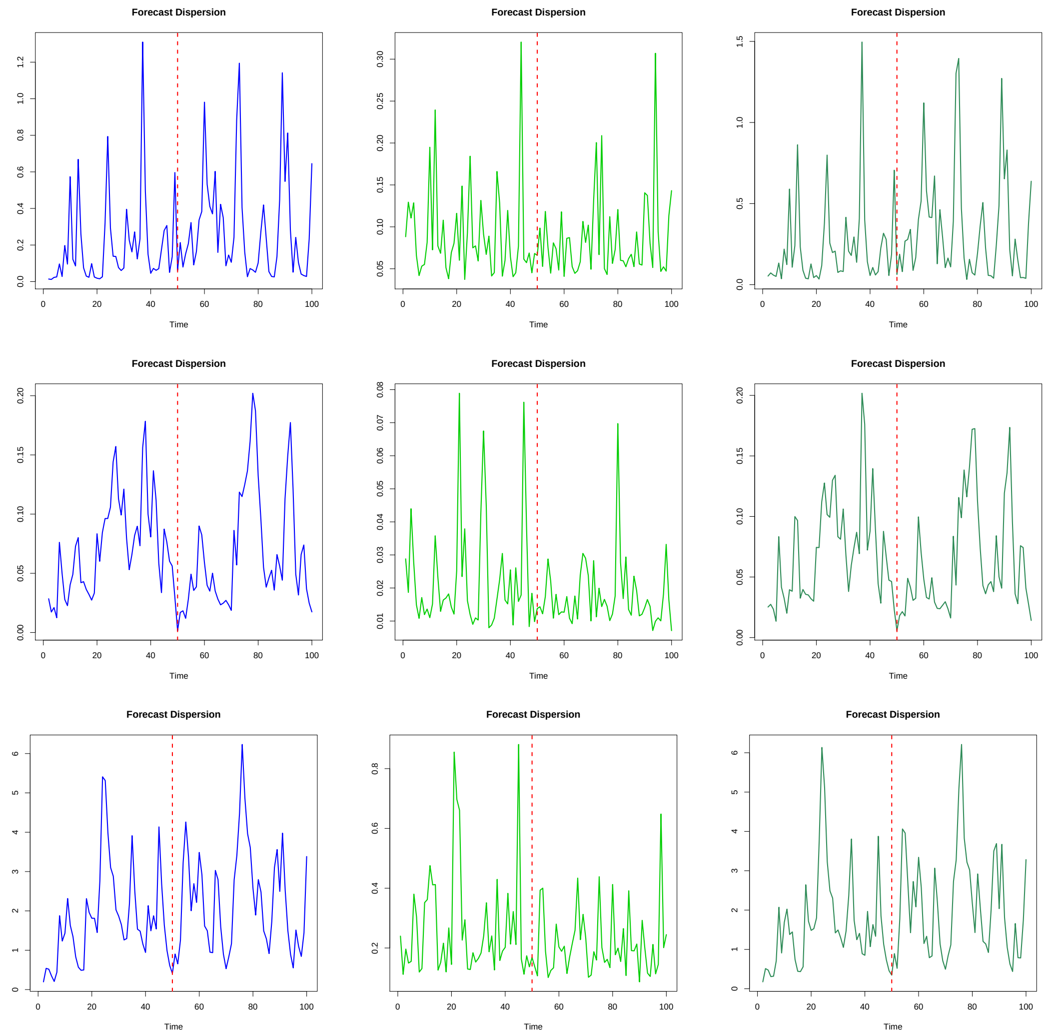
Notes: The left panel shows the degree of information rigidity for inattentive agents, measured as the percentage of agents who do not update their information sets. The right panel plots the average degree of information rigidity for attentive agents in predicting three variables, calculated as the scaled trace of the matrix of the weights attached to agent's previous forecast relative to new information.

Figure 5: Mean Squared Error: Inattentive (left, blue line), Attentive (left, green line) and Overall (right)



Notes: These plots show the mean squared error (MSE) in predicting three variables. The first (second or third) row plots the MSE in predicting variable 1 (2 or 3). The MSE is calculated according to (15) for attentive agents, (24) for inattentive agents, and (26) for all agents.

Figure 6: Forecast Disagreement: Inattentive (left), Attentive (middle) and Overall (right)



Notes: These plots show forecast disagreement in predicting three variables. The first (second or third) row plots the disagreement in predicting variable 1 (2 or 3). Disagreement is calculated according to (18) for attentive agents, (22) for inattentive agents, and (27) for all agents.

8 Appendix

Proof of Proposition 1. Equation (9) can be rendered with i superscripts for the ith agent, yielding

$$\widehat{x}_{t+1|t}(i) - x_{t+1} = \Phi\,(\widehat{x}_{t|t-1}(i) - x_t) - \epsilon_{t+1} + K_t(i)\, e_t(i)$$
$$e_t(i) = y_t(i) - H(i)\,\widehat{x}_{t|t-1}(i) = -H(i)\,(\widehat{x}_{t|t-1}(i) - x_t) + \delta_t(i).$$

Note that $\delta_t(i)$ is orthogonal to $y_s(j)$ for all $s \leq t-1$ and all j; moreover $\{\delta_t(i)\}$ is uncorrelated with $\{x_t\}$. Hence

$$Q^{(ij)}_{t+1|t} = \mathrm{Cov}(\widehat{x}_{t+1|t}(i) - x_{t+1}, \widehat{x}_{t|t-1}(j) - x_t)\,\Phi' - \mathrm{Cov}(\widehat{x}_{t+1|t}(i) - x_{t+1}, e_t(j))\,K_t(j)' + \Sigma^{\epsilon}, \quad (28)$$

because ϵ_{t+1} is orthogonal to $\widehat{x}_{t+1|t}(i)$ and $\mathrm{Cov}(x_{t+1}, \epsilon_{t+1}) = \Sigma^{\epsilon}$. Next,

$$\begin{aligned}
&\mathrm{Cov}(\widehat{x}_{t+1|t}(i) - x_{t+1}, \widehat{x}_{t|t-1}(j) - x_t) \\
&= \Phi\, Q^{(ij)}_{t|t-1} + K_t(i)\,\mathrm{Cov}(e_t(i), \widehat{x}_{t|t-1}(j) - x_t) \\
&= \Phi\, Q^{(ij)}_{t|t-1} - K_t(i)\, H(i)\, Q^{(ij)}_{t|t-1}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\mathrm{Cov}(\widehat{x}_{t+1|t}(i) - x_{t+1}, e_t(j)) \\
&= \Phi\,\mathrm{Cov}(\widehat{x}_{t|t-1}(i) - x_t, e_t(j)) + K_t(i)\,\mathrm{Cov}(e_t(i), e_t(j)) \\
&= -\Phi\, Q^{(ij)}_{t|t-1}\, H(j)' + K_t(i)\left(H(i)\, Q^{(ij)}_{t|t-1}\, H(j)' + \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_t \end{bmatrix}\right).
\end{aligned}$$

Substituting into (28) yields

$$Q^{(ij)}_{t+1|t} = [\Phi - K_t(i)\,H(i)]\,Q^{(ij)}_{t|t-1}\,\Phi' + \Sigma^{\epsilon}$$
$$+ [\Phi - K_t(i)\,H(i)]\,Q^{(ij)}_{t|t-1}\,H(j)'\,K_t(j)' - K_t(i)\begin{bmatrix} 0 & 0 \\ 0 & \Sigma_t \end{bmatrix} K_t(j)',$$

which simplifies to the stated formula. To initialize, we compute the $t = 0$ case, and observe that

$$Q^{(ij)}_{1|0} = \mathrm{Cov}(\widehat{x}_{1|0}(i) - x_1, \widehat{x}_{1|0}(j) - x_1) = \mathrm{Var}[x_1]. \quad \square$$

Proof of Proposition 2. The multi-step ahead forecasting of the state vector is s-traightforward: $\widehat{x}_{t+1|t-k} = \Phi^k\,\widehat{x}_{t+1-k|t-k}$. Also, iteration of (3) yields $x_{t+1} = \Phi^k\,x_{t+1-k} + \sum_{j=0}^{k-1}\Phi^j\,\epsilon_{t+1-j}$. Therefore the multi-step ahead forecasting error can be expressed in terms of one-step ahead forecasting error, via

$$\widehat{x}_{t+1|t-k} - x_{t+1} = \Phi^k\,\left(\widehat{x}_{t+1-k|t-k} - x_{t+1-k}\right) - \sum_{j=0}^{k-1}\Phi^j\,\epsilon_{t+1-j}.$$

We can also express the one-step ahead forecasting error in terms of prior such errors. Suppose $k \le \ell$:

$$\widehat{x}_{t+1-k|t-k} - x_{t+1-k} = \Phi\,\left(\widehat{x}_{t-k|t-k-1} - x_{t-k}\right) + K_{t-k}\,e_{t-k} - \epsilon_{t-k+1}$$
$$\cdots = \Phi^{\ell-k}\,\left(\widehat{x}_{t+1-\ell|t-\ell} - x_{t+1-\ell}\right) + \sum_{j=0}^{\ell-k-1}\Phi^j\,\left(K_{t-k-j}\,e_{t-k-j} - \epsilon_{t+1-k-j}\right).$$

This indicates that the multi-step ahead forecasting error is related to past one-step ahead forecasting errors as follows, when $k < \ell$:

$$\widehat{x}_{t+1|t-k} - x_{t+1} = \Phi^{\ell} \left(\widehat{x}_{t+1-\ell|t-\ell} - x_{t+1-\ell}\right) - \sum_{j=0}^{\ell-1} \Phi^{j} \, \epsilon_{t+1-j} + \sum_{j=0}^{\ell-k-1} \Phi^{j+k} \, K_{t-k-j} \, e_{t-k-j}.$$

The errors e_t have the following form (using the assumption that no private information is present):

$$e_{t-k-j} = B \left(x_{t-k-j} - \widehat{x}_{t-k-j|t-k-j-1}\right) + \eta_{t-k-j}. \tag{29}$$

For $0 \leq j \leq \ell - k - 1$, η_{t-k-j} is uncorrelated with $\widehat{x}_{t+1-\ell|t-\ell} - x_{t+1-\ell}$, and moreover for $0 \leq j \leq \ell - 1$, ϵ_{t+1-j} is uncorrelated with $\widehat{x}_{t+1-\ell|t-\ell} - x_{t+1-\ell}$. Now using (29), we can re-express the one-step ahead forecasting errors as

$$\begin{aligned}
\widehat{x}_{t+1-k|t-k} - x_{t+1-k} &= (\Phi - K_{t-k} \, B) \left(\widehat{x}_{t-k|t-k-1} - x_{t-k}\right) + K_{t-k} \, \eta_{t-k} - \epsilon_{t-k+1} \\
\cdots &= \prod_{j=k}^{\ell-1} (\Phi - K_{t-j} \, B) \left(\widehat{x}_{t-\ell+1|t-\ell} - x_{t-\ell+1}\right) \\
&+ \prod_{j=k}^{\ell-2} (\Phi - K_{t-j} \, B) \left(K_{t-\ell+1} \, \eta_{t-\ell+1} - \epsilon_{t-\ell+2}\right) \\
&\cdots + (\Phi - K_{t-k} \, B) \left(K_{t-k-1} \, \eta_{t-k-1} - \epsilon_{t-k}\right) + K_{t-k} \, \eta_{t-k} - \epsilon_{t-k+1}.
\end{aligned}$$

The convention regarding the product symbols is as discussed in Proposition 2. Hence the multi-step ahead forecasting errors can be re-expressed as

$$
\begin{aligned}
\widehat{x}_{t+1|t-k} - x_{t+1} &= \Phi^k \prod_{j=k}^{\ell-1} (\Phi - K_{t-j}\,B)\,\left(\widehat{x}_{t-\ell+1|t-\ell} - x_{t-\ell+1}\right) \\
&\quad + \Phi^k \prod_{j=k}^{\ell-2} (\Phi - K_{t-j}\,B)\,(K_{t-\ell+1}\,\eta_{t-\ell+1} - \epsilon_{t-\ell+2}) \\
&\quad \cdots + \Phi^k\,(\Phi - K_{t-k}\,B)\,(K_{t-k-1}\,\eta_{t-k-1} - \epsilon_{t-k}) + \Phi^k\,K_{t-k}\,\eta_{t-k} - \sum_{j=0}^{k} \Phi^j\,\epsilon_{t+1-j}
\end{aligned}
\tag{30}
$$

when $k \leq \ell - 2$; when $k = \ell - 1$ the simpler formula is

$$
\widehat{x}_{t+1|t-k} - x_{t+1} = \Phi^k\,(\Phi - K_{t-k}\,B)\,\left(\widehat{x}_{t-k|t-k-1} - x_{t-k}\right) + \Phi^k\,K_{t-k}\,\eta_{t-k} - \sum_{j=0}^{k} \Phi^j\,\epsilon_{t+1-j}.
$$

From these expressions, the formulas for $R_{k,\ell}$ can now be deduced. The case $R_{\ell,\ell}$ is standard, whereas for $R_{\ell-1,\ell}$ indicates we should set $k = \ell - 1$, and together with

$$
\widehat{x}_{t+1|t-\ell} - x_{t+1} = \Phi^\ell\,\left(\widehat{x}_{t+1-\ell|t-\ell} - x_{t+1-\ell}\right) - \sum_{j=0}^{\ell-1} \Phi^j\,\epsilon_{t+1-j}
\tag{31}
$$

we find $R_{\ell-1,\ell}$ has the stated expression. This uses the fact that $\widehat{x}_{t+1-\ell|t-\ell} - x_{t+1-\ell}$ is uncorrelated with $\eta_{t-\ell+1}$, as well as ϵ_{t+1-j} for $0 \leq j \leq \ell - 1$. Next, for $k \leq \ell - 2$ we compute $R_{k,\ell}$ using (31) together with (30). $\square$

Table A1. Information Rigidities Based on Forecast Error

VARIABLES	(1) Forecast Error	(2) Forecast Error	(3) Forecast Error	(4) Forecast Error
Forecast Revision	0.486***	0.509***	0.500***	0.500***
	(0.0162)	(0.0177)	(0.0168)	(0.0173)
Forecast Rev * Disaster		-0.139***		
		(0.0425)		
Disaster		-0.00725		
		(0.0169)		
Forecast Rev * Disaster (News Scaling)			-0.0870***	
			(0.0272)	
Disaster (News Scaling)			-0.00853	
			(0.0106)	
Forecast Rev * Disaster (Combined Scaling)				-0.0838**
				(0.0353)
Disaster (Combined Scaling)				-0.0227
				(0.0140)
Observations	31,981	31,981	31,981	31,981
R^2	0.141	0.141	0.141	0.141
Time FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES
VAR FE	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: Regressions performed for forecasts across 54 countries (including forecasts of GDP, CPI, long- and short-run interest rates, unemployment, and consumption). Forecast Error denotes the difference of the ex-post true value from the mean forecast. 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 5 days following a disaster to those written in the 5 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 comprised 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).

Table A2. Information Rigidities Based on Forecast Revision

VARIABLES	(1) Forecast Rev	(2) Forecast Rev	(3) Forecast Rev	(4) Forecast Rev
Lagged Forecast Revision	0.154***	0.159***	0.158***	0.158***
	(0.00549)	(0.00599)	(0.00571)	(0.00586)
Lag Forecast Rev * Disaster		-0.0332**		
		(0.0147)		
Lagged Disaster		0.00687		
		(0.00593)		
Lag Forecast Rev * Disaster (News Scaling)			-0.0248***	
			(0.00949)	
Lagged Disaster (News Scaling)			0.000560	
			(0.00373)	
Lag Forecast Rev * Disaster (Combined Scaling)				-0.0293**
				(0.0122)
Lagged Disaster (Combined Scaling)				0.00133
				(0.00490)
Observations	31,955	31,955	31,955	31,955
R^2	0.058	0.058	0.058	0.058
Time FE	YES	YES	YES	YES
Country FE	YES	YES	YES	YES
VAR FE	YES	YES	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: Regressions performed for forecasts across 54 countries (including forecasts of GDP, CPI, long- and short-run interest rates, unemployment, and consumption). 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 5 days following a disaster to those written in the 5 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 comprised 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).