# WORKING PAPER

## News-driven inflation expectations and information rigidities

VEGARD H. LARSEN,
LEIF ANDERS THORSRUD
AND JULIA ZHULANOVA

NORGES BANK

**Working papers fra Norges Bank, fra 1992/1 til 2009/2 kan bestilles over e-post:**
FacilityServices@norges-bank.no

Fra 1999 og senere er publikasjonene tilgjengelige på www.norges-bank.no

Working papers inneholder forskningsarbeider og utredninger som vanligvis ikke har fått sin endelige form. Hensikten er blant annet at forfatteren kan motta kommentarer fra kolleger og andre interesserte. Synspunkter og konklusjoner i arbeidene står for forfatternes regning.

**Working papers from Norges Bank, from 1992/1 to 2009/2 can be ordered by e-mail**
FacilityServices@norges-bank.no

Working papers from 1999 onwards are available on www.norges-bank.no

Norges Bank's working papers present research projects and reports (not usually in their final form) and are intended inter alia to enable the author to benefit from the comments of colleagues and other interested parties. Views and conclusions expressed in working papers are the responsibility of the authors alone.

# News-driven inflation expectations and information rigidities*

Vegard H. Larsen†     Leif Anders Thorsrud‡     Julia Zhulanova§

This version: 20 February 2019

## Abstract

We investigate the role played by the media in the expectations formation process of households. Using a novel news-topic-based approach we show that news types the media choose to report on, e.g., fiscal policy, health, and politics, are good predictors of households' stated inflation expectations. In turn, in a noisy information model setting, augmented with a simple media channel, we document that the underlying time series properties of relevant news topics explain the time-varying information rigidity among households. As such, we not only provide a novel estimate showing the degree to which information rigidities among households varies across time, but also provide, using a large news corpus and machine learning algorithms, robust and new evidence highlighting the role of the media for understanding inflation expectations and information rigidities.

**JEL-codes: C11, C53, D83, D84, E13, E31, E37**

**Keywords:** Expectations, Media, Machine Learning, Inflation

# 1 Introduction

In most democracies the fourth estate, i.e., the news media, plays an important role in society. The media not only has the capacity of advocacy and implicit ability to frame political and economic issues, but it is also the primary source from which most people get information.[1] In macroeconomics, expectations are center stage. But, expectations are shaped by information, and information does not travel unaffected through the ether. Rather, it is digested, filtered, and colored by the media. Surprisingly, however, the potential independent role of the media in the expectation formation process has received relatively little attention in macroeconomics, both in theory and in applied work.[2]

In this paper we build on a growing literature providing evidence for a departure from the full information rational expectation (FIRE) assumption towards a theory of information rigidities (Coibion and Gorodnichenko (2012), Dovern et al. (2015), Coibion and Gorodnichenko (2015a), Armantier et al. (2016)), and investigate the potential role played by the media for households' inflation expectations in this setting.

Our analysis is motivated by two particular views giving rise to information rigidities. First, we take the view that acquiring information can be looked upon as a choice variable (Sims (2003), Gorodnichenko (2008), Woodford (2009), Mackowiak and Wiederholt (2009)), implying that the degree of information rigidity is time-varying. For example, in times of trouble it might be worth devoting more resources towards staying informed than in more regular time periods. Second, as in Nimark and Pitschner (2018), we assume that no agent has the resources to monitor all events that are potentially relevant for her decision, and thereby delegate their information choice to specialized news providers who report only a curated selection of events. That is, the media works as "information intermediaries" between agents and the state of the world. In a general, but abstract, theoretical model, Nimark and Pitschner (2018) show that this delegation is optimal when the information flow is overwhelming, and that media's news selection functions and distributions of events jointly determine the degree to which knowledge about an event is common among agents.[3]

We merge and test these two views within the setup proposed by Coibion and Gorod-

---

[1] See, for example, Blinder and Krueger (2004), Curtin (2007), and Fullone et al. (2007).

[2] As an (anecdotal) example, in the first "Handbook of Media Economics" (Simon P. Anderson and Strömberg (2015)) there is a separate chapter about "The Role of Media in Finance" (Tetlock (2015)), but no equivalent chapter about "The Role of Media in Macroeconomics". However, notable exceptions, like Carroll (2003) and Nimark and Pitschner (2018), do exist. We discuss these in greater detail below.

[3] We note that this view differs from the costly information literature (Grossman and Stiglitz (1980), Veldkamp (2006), Nieuwerburgh and Veldkamp (2010)). Rather than agents ex-ante deciding on the expected usefulness of a particular signal, knowledge of events is jointly determined ex-post through the delegated information choice mechanism.
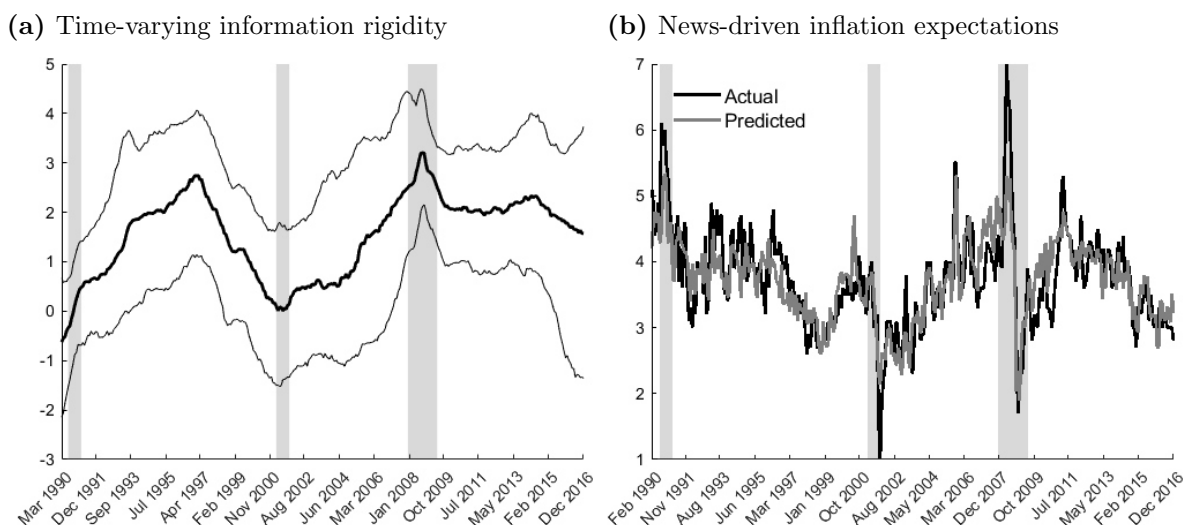
**(a)** Time-varying information rigidity     **(b)** News-driven inflation expectations

**Figure 1.** Figure 1a illustrates the estimated time-varying information rigidity among households (Michigan Survey of Consumers). Median estimates together with 68 percent probability bands are reported. Figure 1b reports the households' reported inflation expectations together with predicted expectations using news. Gray shaded areas illustrate recession periods as defined by NBER (U.S.).

nichenko (2015a). Starting from a noisy information framework, where agents form and update beliefs about the underlying fundamentals via a signal extraction problem, they show that agents' forecast revisions are a sufficient statistic when testing for information rigidities in forecasting efficiency regressions. In Section 2 we augment this framework by introducing time variation in the underlying parameters, and an explicit, but simple, role for the media. It then follows that information rigidity is a function of the time-varying persistence in media coverage and the noise-to-signal ratio in the signal extraction problem. The mechanics of the model are straight forward. When an important event happens, media coverage potentially becomes more concentrated and persistent around this event, and perhaps easier to filter (less noisy) for the agents. Accordingly, information rigidity is reduced as agents put more weight on new information relative to their previous forecasts. This contrasts with (a time-varying version of) the conventional model, where the degree of information rigidity would be determined by the time series properties of inflation itself, but mirrors our assumption that the media works as "information intermediaries" between agents and the state of the world.

Focusing on households' one-year-ahead expectations of inflation, measured by the University of Michigan Surveys of Consumers, we proceed empirically in four successive steps. First, in Section 3, we use a time-varying parameter model to fit households' forecast errors with their forecast revisions, and assess whether information rigidities among households actually show high-frequency variation across time. As seen from the solid black line in Figure 1a, which illustrates our estimate of time-varying information rigidity among households, we provide a confirmatory answer to this question. On average,

our estimate of the degree of information rigidity is just above one. Interpreted through the lens of the signal extraction model outlined in Section 2, this implies that households put a weight of roughly 0.4 on new information. In turn, this estimate is in line with the existing literature, but as seen from the figure, it is far from constant across the sample.

Second, in Sections 4.1 and 4.2, we use techniques from the Natural Language Processing (NLP) literature to construct 80 measures of the news topics the media writes about, i.e., the different types of news reporting, and map these high dimensional data to actual inflation expectations using penalized linear regressions of the LASSO type (Tibshirani (1996)). The intuition for why we focus on news topics, and how, is discussed further below, but as illustrated in Figure 1b, which graphs inflation expectations together with the fitted values from the LASSO, news topics written about in the media have high predictive power for consumers' inflation expectations. While the degree of sparsity is large, from 80 potential news topic candidates only roughly 10 are selected and significant, we later show that the narrative realism of the news-topic-based approach is good. Additional results strongly indicate that this type of textual data contain information not captured by a large set of conventional economic indicators, including inflation itself, confirming that the media is an important source for information among households.

Third, in Section 4.3, we combine the results obtained above, and investigate whether the evolution of households' information rigidity can be explained by the time-varying time series properties, i.e., the persistence and noise-to-signal ratio, in the statistically selected set of news topics. Running simple linear regressions, but taking aboard all posterior uncertainty, we confirm that it can. The regression fit is dominated by the persistence measure, and when media persistence (noise-to-signal) is high (low), information rigidities tend to be low (high), and vice versa, as theory predicts.

We later show, in a falsification experiment, that this result is highly unlikely to be obtained by chance. In particular, looking at the persistence in news topics that are not selected by the LASSO, i.e., not relevant for households' inflation expectations, we obtain mostly insignificant results. Further analysis, in Section 4.4, also confirms, in line with our hypothesis that households' inflation expectations are news-driven, that a negative correlation is not found between the persistence of inflation itself and information rigidity.[4]

As alluded to above, important business cycle events and information rigidities are closely related. In state-dependent models of information rigidity, the degree of information rigidity should be inversely related to business cycle developments, and much lower after a large and visible shock than during normal periods (Mackowiak et al. (2018)). At

---

[4]In contrast, results presented in Appendix E show that if we instead focus on expectations measured by the Survey of Professional Forecasters, the media does not matter, but the persistence in inflation itself does. This is in line with the intuition that the media matters foremost for households, and less so for professionals.

the same time, in the related delegated information choice view of the world (Nimark and Pitschner (2018)), the distribution of events and media's reporting on those events jointly determine the degree of information rigidity. In Section 4.5 we report estimates from Vector Autoregressive (VAR) models, confirming these predictions. We document a strong dynamic interaction between the business cycle and the media-based estimates of information rigidity proposed here. For policy institutions aiming at managing consumers' expectations to stabilize economic fluctuations, e.g., monetary policy (Galí (2008)), this also highlights the role of the media for their communication strategies.[5]

Finally, as seen from Figure 1a, the Great Recession period was associated with a high level of information rigidity, and not low, as would be the prediction from our model. The importance of news media is further documented in Section 5 when we try to understand why? In the process we document that the significant relationship between households' expectations and news topics withstand out-of-sample evaluation. Still, during the Great Recession period the predictive relationship between news topics and households' expectations breaks down, but strengthens when using news topics to predict actual inflation. This suggest that the quality of information was good, but that something happened with how households consumed it. Using newspaper circulation statistics, which showed the largest cyclical fall since the 1940s during the Great Recession period, we find evidence consistent with what we label an information diffusion story: If the quality of information is good, but information diffusion is poor, the quality of information matters less. An alternative interpretation of this result is that the households consumed a different type of news during the $GR$ period, e.g., (free) social media, and not that information diffusion was lower per se. Irrespective of interpretation, this is in accordance with our main conclusions about news media's important role in the expectations formation process.

The contribution, and novelty, of our analysis is threefold. First, we provide direct evidence of time-variation in the degree of information rigidity among households in the U.S. Although results reported in Loungani et al. (2013), Coibion and Gorodnichenko (2015a), and Dovern et al. (2015) point towards low frequency state-dependence in the degree of information rigidity, we are, to the best of our knowledge, the first to provide a quantitative measure of high-frequency changes in the degree of information rigidity among households in the U.S.

Second, we are the first to relate measures of households information rigidity to the time series properties of the news, i.e., the persistence and noise-to-signal ratio. This is important because it puts our analysis well within an established theoretical framework

---

[5]As Blinder et al. (2008) write: "...if researchers are interested in testing market responses to communication, it may make sense to focus on statements that actually reach market participants, and on the content as conveyed by the media."

used to test and explain information rigidities (Coibion and Gorodnichenko (2015a)). However, by analyzing the relationship between information frictions and the media our analysis also speaks directly to work by Carroll (2003), Doms and Norman (2004), Pfajfar and Santoro (2008), Pfajfar and Santoro (2013), Lamla and Lein (2014), Dräger and Lamla (2017). The seminal contribution by Carroll (2003) is particularly well known. He shows in an epidemiological model of inflation expectations that households update their beliefs towards professionals, assumed to express their views through the media and to be fully informed, more frequently in periods of intense media reporting on inflation. The epidemiological view, however, has later been questioned by Pfajfar and Santoro (2013), who show that available and perceived news stories do not help at restricting the forecast gap between professionals and households, but rather widen it.

Finally, we make an important contribution in how we use text as data to better understand the expectations formation process among households. In contrast to the earlier studies in the literature cited above, which almost exclusively have derived quantitative media measures by counting terms related to *inflation* in the news, we work with the assumption that many news items might be of relevance for inflation expectations, even without explicitly mentioning terms related to inflation. As such, we hypothesize that when the media writes extensively about topics related to, e.g., politics, even without explicitly mentioning terms related to inflation, this reflects that something is happening in this area that potentially has economy-wide effects, and might therefore also affect inflation expectations. In line with this, we find that news stories about, e.g., fiscal policy, health, and politics, significantly affects households' inflation expectations. Importantly, we also show that the news-topic-based approach adopted here delivers results in accordance with theory, while an approach relying on simple word counts does not.

Technically, the news-topic-based approach is operationalized by estimating a topic model, of the Latent Dirichlet Allocation (Blei et al. (2003)) class, on a large news corpus extracted from the *Dow Jones Newswires Archive* (DJ). Following Larsen and Thorsrud (2018b), the topic decomposition is then transformed into tone-adjusted time series, measuring how much each topic is written about in the media at any given point in time.[6]

In sum, the analysis conducted here provide positive evidence in favor of the state-dependent information rigidity view, as advocated in, e.g., Sims (2003), Mackowiak and Wiederholt (2009), and Coibion and Gorodnichenko (2015a). However, as in Nimark and Pitschner (2018), our analysis emphasizes the role of information providers. As such, this study also speaks to the literature trying to identify the causal effect of the media. While

---

[6]Thorsrud (2018) and Larsen and Thorsrud (2018a) show that a similar topic decomposition of economic news can be used to construct daily business cycle indicators with very good classification and nowcasting properties for GDP growth. See also Hansen and McMahon (2015), Larsen (2017), Hansen et al. (2018) and Dybowski and Adämmer (2018) for related economic applications of the LDA technology.

this has been relatively unexplored in macroeconomics, it has received some attention in finance (Dougal et al. (2012), Peress (2014), Larsen and Thorsrud (2017)), and much more attention in other branches of the literature and in other sciences (Gentzkow and Shapiro (2010), Gentzkow et al. (2011), Shiller (2017), King et al. (2017), Prat (2018)).

# 2 Information rigidities in theory

There are many theoretical models that predict a departure from the full information part of the FIRE assumption used in text book economics. For example, Mankiw and Reis (2002) propose a sticky-information model where agents update their information sets infrequently as a result of fixed cost to the acquisition of information, while Sims (2003) and Woodford (2003) have proposed mechanisms categorized as noisy information models. In Sims (2003), for example, the underlying assumption is that people's ability to process information is constrained and thereby rationally choose what information to pay attention to (rational inattention).

Starting from a noisy information framework, where agents form and update beliefs about the underlying fundamentals via a signal extraction problem, Coibion and Gorodnichenko (2015a) show that economic agents' forecast revisions are a sufficient statistic to test whether expectations are rational, in the FIRE sense, or inhabit information rigidities consistent with (all) the theories mentioned above. We take the view that acquiring information can be looked upon as a choice variable (Sims (2003), Gorodnichenko (2008), Woodford (2009), Mackowiak and Wiederholt (2009)), implying that the degree of information rigidity is time-varying. We also take the view that media works as "information intermediaries" between agents and the state of the world (Nimark and Pitschner (2018)). Below we merge and incorporate these two assumptions in the Coibion and Gorodnichenko (2015a) framework.

## 2.1 The role of media

We start by making the assumption that most people do not follow inflation as measured by the statistical agency per se, but get information about inflation through the media. While this information-object is high dimensional, letting $\pi_t^N$ denote an aggregated measure of relevant media coverage, the signal agent $i$ receives about inflation at time period $t$ can be written as:

$$y_{it} = \pi_t^N + \omega_{it}, \tag{1}$$

where $\omega_{it} \sim N(0, \sigma_{\omega t}^2)$ is idiosyncratic noise capturing heterogeneity in forecasting "models" across agents, while potential heterogeneity in the signal noise across time is captured

by time dependence in $\sigma_{\omega t}^2$. In our framework, the noise term could be thought of as capturing heterogeneity in how different agents weigh and interpret different news sources and items. For example, not all agents (if any) read, and interpret correctly, all articles relevant for forecasting inflation.

We further assume that media actually fulfills its purpose in informing the public about important developments in society, including inflation. However, exactly how the media does this, e.g., the systematic editorial decisions, resources used, and discussions within the media houses, are not observable to the agents. Thus, the relationship between actual inflation, $\pi_t$, and media's coverage of inflation is specified as:

$$\pi_t^N = \pi_t + \alpha_t, \tag{2}$$

where $\alpha_t$ is a time-fixed effect, capturing for example potential media biases (Pfajfar and Santoro (2008), Lamla and Lein (2014)). Importantly, as the agents only observe the left-hand side of (2), they do not know that the news they receive, with noise, does not map one-for-one to actual inflation.[7]

To introduce dynamics into the model we specify the time series properties of media coverage as a simple autoregressive process of order one:

$$\pi_t^N = \rho_t^N \pi_{t-1}^N + \nu_t^N, \tag{3}$$

where $\nu_t^N \sim N(0, \sigma_{\nu t}^2)$. Again, we allow for potential time dependencies in the process by letting both $\rho_t^N$ and $\sigma_{\nu t}^2$ depend on the time index $t$, where variation in, e.g., $\rho_t^N$, can be due to major economic or political events that become extensively covered by the media.

Together, equations (1) and (3) constitute a conditional linear state space system, and, removing the time-varying parameter specification and replacing the $\pi_t^N$ terms with $\pi_t$, the system becomes identical to the one in Coibion and Gorodnichenko (2015a). Using the standard Kalman filter recursions, the variance of the prediction error, $\Psi_t \equiv P_{t|t-1} \equiv E(\pi_t^N - \pi_{t|t-1}^N(i))(\pi_t^N - \pi_{t|t-1}^N(i))'$, can be written as:

$$\Psi_t = (\rho_t^N)^2(\Psi_t - \Psi_t(\Psi_t + \sigma_{\omega t}^2)^{-1}\Psi_t) + \sigma_{\nu t}^2, \tag{4}$$

which is known as the Riccatti equation. From this it follows that the Kalman Gain, capturing the weight assigned to new information about $\pi_t^N$ contained in the prediction

---

[7] We show in Appendix C that if agents also form an expectation about the time-varying constant in (2), the time-varying information rigidity will be a function of the time series properties of inflation itself, and not the news, as in our setup (shown below). However, as documented in Section 4.4, using actual inflation gives results at odds with theory, suggesting that such an assumption is questionable. In general, the assumptions behind equations (1) and (2) are also consistent with a substantial literature showing that people are not fully informed about their, e.g., tax credit (Chetty and Saez (2013)), returns to schooling (Jensen (2010), Wiswall and Zafar (2014)), and their marginal price for basic consumption goods such as electricity and water (Carter and Milon (2005).

error, is given by:

$$K_t = \rho_t^N \Psi_t (\Psi_t + \sigma_{\omega t}^2)^{-1}. \tag{5}$$

As seen from (5), this weight depends on the persistence of media coverage, $\rho_t^N$, and on the amount of noise in the signal, $\sigma_{\omega t}^2$. The forecast for the unobservable state is then given by:

$$\pi_{t|t}^N(i) = \pi_{t|t-1}^N(i) + K_t(y_t - y_{t|t-1}(i)). \tag{6}$$

Averaging equation (1) and (6) across agents, and iterating $h$ periods forward, equation (6) becomes:

$$\pi_{t+h}^N - F_t \pi_{t+h}^N = \frac{1 - K_t}{K_t}(F_t \pi_{t+h}^N - F_{t-1} \pi_{t+h}^N) + \nu_{t+h,t}^N, \tag{7}$$

where $\nu_{t+h,t}^N = \sum_{j=1}^h (\rho_t^N)^{h-j} \nu_{t+j}^N$, and $F_t \pi_{t+h}^N$ is the agents' expected future media coverage. We observe neither $\pi_{t+h}^N$ nor $F_t \pi_{t+h}^N$. However, using (2), we can write equation (7) as:

$$\pi_{t+h} - F_t \pi_{t+h} = c_t + \beta_t(F_t \pi_{t+h} - F_{t-1} \pi_{t+h}) + e_t \tag{8}$$

where $F_t \pi_{t+h}$ is the households' expressed expectation of future inflation, through, e.g., surveys, and $\beta_t = \frac{1-K_t}{K_t}$, $e_t = \nu_{t+h,t}^N$, and $c_t = F_t \alpha_{t+h} - \alpha_{t+h} + \frac{1-K_t}{K_t}(F_t \alpha_{t+h} - F_{t-1} \alpha_{t+h})$.[8]

As in Coibion and Gorodnichenko (2015a), equation (8) describes the relationship between ex-post forecast errors and ex-ante mean forecast revisions. Although individuals form their forecasts rationally conditional on their information set, the ex-post mean forecast error across agents is systematically predictable using ex-ante mean forecast revisions due to gradual adjustment of beliefs to new information. A higher value of $\beta_t$ implies a higher degree of information rigidity. Conversely, if $\beta_t = c_t = 0$, we are back in the world of FIRE.

We depart from this earlier literature by introducing an explicit media channel and time variation in (1) and (3). This implies that the degree of information rigidity, $\beta_t$, is time-varying and depends on the time series properties of media coverage. In particular, because $\beta_t$ is a function of the Kalman Gain (5), information rigidity is decreasing if the persistence of media coverage ($\rho_t^N$) is high, and increasing if the amount of noise in the signal received by households ($\sigma_{\omega t}^2$) is high (relative to $\sigma_{\nu t}^2$). In contrast, in the conventional model, without a media channel, agents are assumed to follow inflation directly. Thus, the degree of information rigidity would be determined by properties of inflation itself.

## 2.2 Two testable hypothesis

Together, equations (5) and (8) deliver two testable hypothesis. First, one can estimate (8) to gauge the degree of time variation in the parameters. Cross-sectional results in

---

[8]When households are not aware of the gap between $\pi_t$ and $\pi_t^N$, their expectations of $\alpha_t$ is implicitly equal to zero. Hence, $F_t \alpha_{t+h} = F_{t-1} \alpha_{t+h} = 0$, and $c_t = -\alpha_{t+h}$.

Coibion and Gorodnichenko (2015a) point towards state dependent low-frequency variation in information rigidity among professional forecasters. However, to the best of our knowledge, nobody have tested the degree to which information rigidity among households (in the U.S.) varies across time. We do so in Section 3.

Second, and conditional on time-variation in $\beta_t$, one can use (5) and test if the underlying time-varying persistence and noise-to-signal ratio in media coverage explain the evolution of $\beta_t = \frac{1-K_t}{K_t}$. However, to operationalize such a test, one needs a measure of media coverage that is relevant for households' inflation expectations. In Section 4 we first propose a measure of media coverage and evaluate its relevance for inflation expectations. Then, we test if the underlying time-varying persistence and noise-to-signal ratio in media coverage relevant for inflation expectations can explain the evolution of $\beta_t$.

# 3 Information rigidities in the data

To bridge our analysis with the earlier literature, we start by estimating a static version of equation (8), and then turn to the more advanced time-varying parameter specification towards the end of the section. In both cases, we use the Michigan Survey of Consumers (MSC) monthly measure of one year ahead CPI inflation as $F_t \pi_{t+h}$, and U.S. headline CPI inflation as a measure of realized inflation $\pi_{t+h}$. Because the MSC only contain households' forecast of inflation over the course of the next year, revisions to these forecasts will not have perfectly overlapping time periods. Accordingly, the static model we estimate is:

$$\pi_{t+12,t+1} - F_t \pi_{t+12,t+1} = c + \beta(F_t \pi_{t+12,t+1} - F_{t-1} \pi_{t+11,t}) + e_t, \qquad (9)$$

where $\pi_{t+12,t+1}$ is actual inflation over the next year, and $F_t \pi_{t+12,t+1}$ is the households expectations, at time $t$, of inflation over the next year. Thus, the left-hand and right-hand side variables in (9) are the forecast errors and revisions, respectively. Both variables are graphed in Figure 8 in Appendix B.

Table 1 reports the results of estimating (9) using the IV estimator.[9] The parameter of foremost interest is $\beta$. A rejection of the null hypothesis of $\beta = 0$ indicates potential information rigidities. As seen from column $I$, the $\beta$ estimate is large and significant. Moreover, the results reported in columns $II$ to $VI$ show that this finding remains robust

---

[9]Due to the non-overlapping time periods in forecast revisions, OLS estimates of (9) will be biased since the error term consist of the rational expectations forecast error. To avoid this issue, we follow Coibion and Gorodnichenko (2015a), and instead apply an IV estimator using the (log) change in the monthly price of oil as an instrument. Table 6, in Appendix B, reports the first stage regression results when the (log) change in the monthly oil price is used as an instrument for the households' forecast revisions. As seen from the table, the instrument is strong and relevant. In the remaining part of this analysis, we will therefore be using the households' forecast revisions instrumented by the price of oil.

**Table 1.** Michigan Survey of Consumers inflation forecast errors and revisions. Each column reports the results of the following regression: $\pi_{t+12,t+1} - F_t\pi_{t+12,t+1} = c + \beta(F_t\pi_{t+12,t+1} - F_{t-1}\pi_{t+11,t}) + \delta z_{t-1} + u_t$. Since the forecast revisions contain non overlapping forecast horizons, an instrument variable approach is used to avoid having rational expectation forecast errors in the error term. Newey-West corrected standard errors are reported in parenthesis. *, **, and ***, indicate that coefficients are statistically significant at the 10, 5, 1 percent level, respectively. See the text for details about the additional controls $z_{t-1}$ and the first-stage IV estimates. The sample period is 1990:01-2016:12.

| | I | Inflation II | Unem-ployment III | Oil price growth IV | T-bill rate V | All (to the left) VI | Factors ×10 VII | Double selection VIII |
|---|---|---|---|---|---|---|---|---|
| | | | | | | **Additional controls: $z_{t-1}$** | | |
| $\beta$ | 1.87*** | 1.63*** | 1.83*** | 1.83*** | 1.73*** | 1.51*** | 1.44*** | 1.10** |
| | (0.62) | (0.59) | (0.61) | (0.62) | (0.61) | (0.56) | (0.53) | (0.49) |
| $\delta$ | | -0.17** | -0.15*** | 0.00 | 0.63 | - | - | - |
| | | (0.07) | (0.04) | (0.01) | (0.66) | - | - | - |
| $R^2$ | 0.03 | 0.05 | 0.05 | 0.02 | 0.03 | 0.08 | 0.19 | 0.32 |
| N | 325 | 324 | 324 | 324 | 324 | 324 | 324 | 324 |

when including the same type of control variables as used in Coibion and Gorodnichenko (2015a) analyzing information rigidity among professional forecasters, namely lagged values of inflation, unemployment, oil price growth, and the T-bill rate.

To further test the robustness of this results, we first control for a total of 10 factors obtained from the monthly *FRED-MD* database developed by McCracken and Ng (2016).[10] As it is well known that factors like these capture a large bulk of the co-movement among macroeconomic indicators, potential omitted variable biases should be less severe than when only controlling for subjectively chosen single indicators. Next, for the same reason, but to avoid the reliance on factor estimates, we also run a double selection procedure (Belloni et al. (2014)). In short, the double selection algorithm is implemented as follows: First, we regress the treatment (forecast revisions) and the dependent (forecast errors) variables on all the variables in the *FRED-MD* data set using the LASSO estimator (described in greater detail in Section 4.2). Next, after these two separate penalized regressions, we run an OLS regression on the dependent variable, including the treatment variable and the union of the control variables selected in step one. The final parameter estimates of $\beta$ from these two additional tests are reported in columns *VII* and *VIII* of Table 1. While the point estimates become somewhat smaller when controlling for a larger

---

[10]The *FRED-MD* is a much used data set in macroeconomics, and contains roughly 130 monthly economic indicators. The data is briefly described in Appendix A. The 10 factors extracted from the *FRED-MD* data set are obtained using conventional Principal Components Analysis (PCA), see, e.g., Stock and Watson (1989).

set of variables, the results are still significant at either the 1 or 5 percent level.

The estimates in Table 1 strengthen the conclusions drawn in earlier research about the presence of information rigidities. However, the theory models we implicitly build on (Sims (2003), Gorodnichenko (2008), Woodford (2009), Mackowiak and Wiederholt (2009)), and the model in Section 2, imply that the degree of information rigidity varies across time.[11] We turn to this next.

## 3.1 Time-varying information rigidities?

By construction, allowing the parameters in 9 to change through time will deliver a better model fit. On the other hand, controlling for a large set of other relevant variables, as in columns *VI* to *VIII* in Table 1, becomes substantially more difficult. Time-varying parameter models are already highly parameterized (one parameter for each time period), and increasing the model size with more explanatory variables makes this challenge even more severe. For this reason, we estimate the time-varying version of 9:

$$\pi_{t+12,t+1} - F_t\pi_{t+12,t+1} = c_t + \beta_t(F_t\pi_{t+12,t+1} - F_{t-1}\pi_{t+11,t}) + e_t, \tag{10}$$

using the Latent Threshold Model (LTM) idea by Nakajima and West (2013). Here, dynamic sparsity is enforced on the system through a latent threshold mechanism, which shrinks the parameters towards zero whenever they are not contributing significantly to the model fit. Accordingly, the time-varying parameter model we specify is parsimonious in its size, we only include forecast revisions as explanatory variables, but also, due to the threshold mechanism, faithful to the null hypothesis of full information, i.e., $\beta_t = 0$.

Formally, dynamic sparsity is enforced on the system through the time-varying parameters $c_t$ and $\beta_t$. For, e.g., $\beta_t$, the LTM structure can be written as:

$$\beta_t = \beta_t^*\varsigma_{\beta,t} \quad \varsigma_{\beta,t} = I(|\beta_t^*| \geq d_\beta), \tag{11}$$

where we let $\beta_t^*$ follow a random walk process:

$$\beta_t^* = \beta_t^* + \upsilon_{\beta,t}, \tag{12}$$

with $\upsilon_{\beta,t} \sim i.i.d.N(0, \sigma_{\beta,\upsilon}^2)$. In (11) $\varsigma_{\beta,t}$ is a zero one variable, whose value depends on the indicator function $I(|\beta_t^*| \geq d_\beta)$. If $|\beta_t^*|$ is above the the threshold value $d_\beta$, then $\varsigma_{\beta,t} = 1$, otherwise $\varsigma_{\beta,t} = 0$, and $\beta_t$ shrinks to zero. For completeness, we assume a similar, but independent, structure for the $c_t$ parameter, which can be obtained by replacing all $\beta$ terms in (11) and (12) with $c$.

---

[11]Although their power to detect multiple structural changes is low, simple Cusum tests (Brown et al. (1975)) also indicate that there are at least two breaks in the sample (see Figure 9 in Appendix B).

As is common for many time-varying parameters models, estimation of (10) is done by drawing from the conditional posterior distribution using MCMC simulations. In the interest of preserving space, details about priors, initialization, and the estimation algorithm are relegated to Appendix G.1.

The time-varying posterior estimates of $\beta_t$ were plotted already in Figure 1a in Section 1. We clearly see that the degree of information rigidity varies substantially across the sample. During the U.S. recessions in the early 1990s it started out low, but then increased sharply both during and in the years after the recession end. The mid 1990s were associated with a high degree of information rigidity. Then, well before the 2001 recession episode, information rigidity started falling, and remained low until the mid 2000s, before it increased substantially again in the years prior to the Great Recession. Since then it has remained at a relatively high level, albeit with small drops after the Great Recession period and towards the end of the sample. Although not our primary focus, we also observe that the $c_t$ parameter is downward trending, see Figure 10a in Appendix B. As (10) is basically a (time-varying) forecast efficiency regression, this suggests a departure not only from full information, but also from rational expectations where forecast errors should be white noise. In light of the theory model in Section 2.1, one interpretation of this parameter is that it captures media biases and that such biases are not constant across time, as also suggested by findings in Souleles (2004).

How large is the degree of information rigidity? The static results reported in Table 1 indicate that the degree of information rigidity is substantial. Interpreted through the lens of the model in Section 2.1, our estimates suggest that agents put a weight of less than 0.35-0.47 on new information, and more than 0.53-0.65 on their previous forecasts, i.e., $\hat{K} = 1/(1 + \hat{\beta})$. These numbers are well in line with Coibion and Gorodnichenko (2015a), who find similar magnitudes for the Survey of Professional Forecasters forecast of the GDP deflator. Alternatively, in the context of sticky-information models, Coibion and Gorodnichenko (2015a) show that these estimates equivalently imply an updating frequency every six to eight months. This is close to twice as frequent as in the epidemiological model estimated by Carroll (2003), and more than twice as frequent compared to results presented by Dräger and Lamla (2017) and Doepke et al. (2008) for U.S. and European households, respectively. But, as shown above, the estimates of information rigidity is time-varying, and therefore also sample dependent.

Looking at the time-varying parameter estimates in Figure 1a, we obtain results similar to the static ones on average. However, the weight put on new information, relative to old, varies from basically 1, during the recessionary periods in the early 1990s and 2000s, to less than 0.25 during the Great Recession period. Again, interpreted in context of sticky-information models, this implies updating frequencies ranging from every month

to roughly every 10th month. Compared to Lamla and Sarferaz (2012), who provide evidence of time variation in information rigidity among German households ranging from 2 to 33 months, these estimates are still modest. But, as discussed in Coibion and Gorodnichenko (2015a), the magnitudes of information rigidity we document here have profound macroeconomic effects in theoretical models incorporating information frictions.

# 4 Expectations and news

Having established that the degree of information rigidity varies significantly across time, we now turn to our second question: Can the underlying time-varying persistence and noise-to-signal ratio in media coverage explain the evolution of $\beta_t$? To address this question we proceed in three steps. First, we introduce quantitative measures of news coverage using a statistical topic model. Next, we construct a mapping between the derived news topics and households' inflation expectations by running penalized linear regressions. Under the assumption that only news topics with predictive power for actual expectations should be relevant for describing the information households care about, the idea is to construct an approximation to the high dimensional object $\pi_t^N$ in equation (1).[12] Finally, we test if the underlying time-varying persistence and noise-to-signal ratio in media coverage, relevant for inflation expectations, can explain the evolution of $\beta_t$.

## 4.1 The news

The main raw media data used in this analysis consist of roughly 5 million news articles from the *Dow Jones Newswires Archive* (DJ), covering the period 1990 to 2016. The database covers a large range of Dow Jones' news services, including content from *The Wall Street Journal*. All text is business-focused and written in English.

Arguable, most households do not likely read, e.g., *The Wall Street Journal*. However, it is very likely that news stories relevant for inflation expectations are covered by this type of source, and that such coverage spills over to news sources that households follow more directly. King et al. (2017), for example, provide a convincing randomized experiment showing that even articles reported in small media outlets affect the nationwide discussion of the articles' specific subjects. Moreover, *The Wall Street Journal* is one of the largest newspapers in the United States in terms of circulation, and therefore leaves a large footprint in the U.S. media landscape. Ideally, of course, one would want to work with the exact media content people consume (if that was measurable), together with their

---

[12]Conversely, if one had looked at news with predictive power for actual inflation in this setting, one might have ended up using news items that households never read or cared about. Such a null set can not plausibly explain households updating behavior.

individually stated inflation expectations. As a second best, we use aggregated inflation expectations, and one important news source.

To make the high-dimensional, and unstructured, textual data applicable for time series analysis, i.e., to explain time-varying information rigidities, we follow Larsen and Thorsrud (2018b) and Thorsrud (2018), and work with the simple assumption that the more intensive a given topic is represented in the media at a given point in time, the more likely it is that this topic represents something of importance for the economy's current and future needs and developments, including inflation expectations. This assumption is operationalized by doing a topic decomposition of the news corpus, i.e., all the text and articles, using a Latent Dirichlet Allocation (LDA) model (Blei et al. (2003)). This model can be looked upon as a factor model applied to text, where each article is treated as a mixture of topics, while each topic is treated as a mixture of words (terms). The LDA model is also one of the most popular clustering algorithms in the NPL literature because of its simplicity, and because it has proven to classify text in much the same manner as humans would do (Chang et al. (2009)). Thus, the topic decomposition transforms something that is large and complex, i.e., the corpus, into something that is relatively small, dense, and interpretable.

As common in this literature, and prior to estimation, the news corpus is cleaned (Gentzkow et al. (2017)). We remove stop-words, do stemming, and apply term frequency - inverse document frequency calculations. A more detailed description of these steps is given in Appendix F.1. Likewise, in the interest of preserving space, we describe the technical details related to the LDA and estimation in Appendix F.2. Here we note that, based on Larsen and Thorsrud (2018b) and Thorsrud (2018), we extract 80 different topics in total, and use the average of the 10 last iterations of the Gibbs simulations, used to estimate the LDA, as our measure of article weights and topics. Using the output from the LDA, the topic decomposition is transformed into time series, measuring how much each topic is written about at any given point in time. We note that, by definition, on a given day, more coverage of one particular news topic leads to less coverage of other topics, i.e., the topic probabilities sum to 1 on each day in the sample. Across time, however, there can be large variation in the topic contributions. Finally, we compute the tone of the news using a simple dictionary-based approach, counting positive relative to negative words in articles relevant for each news topic, and sign adjust the topic frequencies accordingly. A more detailed description of this latter step is relegated to Appendix F.3.[13]

To build intuition, Figure 2 illustrates the output from the above steps for six of the 80

---

[13]The results presented in Thorsrud (2018) highlight that tone-adjusted topic frequencies perform much better for nowcasting GDP growth than un-adjusted topic frequencies do. I.e., whether or not the news is positive or not matters. In Appendix D we show that the same applies in the current setting.

**Figure 2.** Topic categorization and time series. For each word cloud the size of a word reflects the probability of this word occurring in the topic. Each word cloud only contains a subset of all the words in the topic distribution. Topic labels are subjectively given, and the topic time series are normalized.

topics. A full list of the estimated topics is given in Table 5, in Appendix A. First, the LDA produces two outputs; one distribution of topics for each article in the corpus, and one distribution of words for each of the topics. The latter distributions are illustrated using word clouds in Figure 2. A bigger font illustrates a higher probability for the terms. As the LDA estimation procedure does not give the topics any name, labels are subjectively given to each topic based on the most important terms associated with each topic. How much each topic is written about at any given point in time, and its tone, is illustrated in the graphs below each word cloud. The graphs should be read as follows: Progressively more positive values means the media writes more about this topic, and that the tone of reporting on this topic is positive. Conversely, progressively more negative values means the media writes more about this topic, but that the tone of reporting is negative. Across topics, our simple hypothesis is that these fluctuations can tell us something important about which narratives dominate in the public discourse at different points in time.

Starting with Carroll (2003), the conventional method for quantifying media coverage of inflation has been to apply Boolean techniques. That is, simply counting (subjectively defined) terms related to inflation in every news article (or headline), and then constructing time series based on aggregated daily or monthly (normalized) counts. Although we also apply this method, in Section 4.4, our preferred method for quantifying media coverage relevant for inflation expectations relies on the topic-based decomposition. The advantage of this procedure is that articles that are relevant for inflation, but do not use the term *inflation* explicitly, might be captured by the more general topics. Still, a large amount of topics are needed to describe the news corpus, making the mapping between inflation expectations and news a high-dimensional variable selection problem. We turn to this next.

## 4.2   News-driven inflation expectations?

To find the set $S$ of news topics relevant for households' inflation expectations, we run linear predictive regressions like:

$$F_t \pi_{t+12,t+1} = \alpha + \sum_{n=1}^{M} b_n X_{n,t-1} + \epsilon_t, \tag{13}$$

where $F_t \pi_{t+12,t+1}$ is the households' expectations, at time $t$, of inflation over the next year, and $M$ are the number of news topics $X_{n,t-1}$. Each news topic is lagged one period relative to $F_t \pi_{t+12,t+1}$ to avoid simultaneity issues and look-ahead-biases. Our results are also robust to inclusion of more than one lag, and we will later augment the right hand side of (13) with a large set of hard economic indicators.

Among the large set of $M$ predictors, we are interested in those that contribute significantly in predicting $F_t \pi_{t+12,t+1}$. However, as the number of explanatory variables $M$ is high relative to the number of periods $T$ in our data sample, the standard ordinary least squares (OLS) estimator is inappropriate. Instead, we use the Least Absolute Shrinkage and Selection Operator (LASSO) method, first proposed by Tibshirani (1996). In contrast to OLS, this method is built for high-dimensional variable selection problems, and shrinks parameter estimates for unimportant variables towards zero. The LASSO thereby encourages simple and sparse models.

Formally, letting $\boldsymbol{y} = (F_1 \pi_{1+12,2}, \ldots, F_T \pi_{T+12,T+1})'$ be a $T \times 1$ response variable, and $\boldsymbol{X} = [X_0, \ldots, X_{T-1}]$ be the $T \times M$ matrix of predictors, the LASSO algorithm solves the constrained least squares problem:

$$\hat{\boldsymbol{B}} = \underset{\boldsymbol{B}}{\operatorname{argmin}} \frac{1}{T} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{B}\|_2^2 + \lambda \|\boldsymbol{B}\|_1, \tag{14}$$

where $\lambda \geq 0$ is a tuning parameter, controlling the amount of regularization. If $\lambda = 0$, (14) yields the OLS solution. If $\lambda > 0$, coefficients will be shrunk towards 0. As is

common in the literature, we choose the optimal value of the tuning parameter using 5-fold cross-validation (CV), and minimum mean squared error (MSE) loss. To further avoid over-fitting, we choose the sparsest model within one standard error of the minimum loss, but note that our results are robust to choosing instead the more highly parameterized MSE solution. Prior to estimation, all variables are standardized to make the penalized regressions invariant to scale. Finally, as LASSO parameter estimates will be pulled towards zero, and thereby be biased, we follow Belloni and Chernozhukov (2013) and run the post-LASSO routine, i.e., OLS on the selected variable set, when reporting the results and making inference.

The column labeled *News* in Table 2 summarizes our main results. Among 80 potential news topics, 11 are selected by LASSO. Of these, 6 are significant at the 10 percent level or lower. As the LASSO regressions are just predictive relationships used for variable selection, we do not spend time on interpreting the sign of the coefficients. The adjusted $R^2$ statistic, however, is informative and as high as 41 percent. Thus, as also seen in Figure 1b, the selected news topics explain a relatively large fraction of the total variation in households' inflation expectations. The relevant set includes topics like *Health, Internet, Clients, Aviation, Labor market, The White House, and M&A*, while the partial $R^2$ statistics suggest that the topics *Health* and *Internet* contribute the most to the regression fit.

Of course, many news articles are just reporting on hard economic indicators the households might actively follow. However, the independent relevance of news topics for describing households' inflation expectations is robust to augmenting the news topic regressors in (13) with the 130 variables in the *FRED-MD* database (described in Section 3), and re-estimating the LASSO. As seen from the column labeled *News and Hard* in Table 2, while some of the topics selected in the news-only regression become insignificant when controlling for the hard economic indicators, most of them survive (and only one new topic gets selected). And, the adjusted $R^2$ only increases from 41 to 61 percent. In other words, the news topics capture aspects of households' inflation expectations that are not captured by hard economic indicators. Interestingly, however, among the hard economic indicators that gets selected, we find many variables already focused on in the earlier literature, like, production and consumption indicators (Pfajfar and Santoro (2008) and Ehrmann et al. (2017)), volatility measures and spreads (Dräger and Lamla (2017)), and consumer sentiment (Doms and Norman (2004) and Ehrmann et al. (2017)).

The (significant) topics in Table 2 might not have been given names that intuitively link them to households' inflation expectations. However, the narrative realism of the approach becomes evident when we query the news corpus for articles where each of the significant topics have a particularly high topic weight. This is illustrated in Table 3.

**Table 2.** Post-LASSO regression results. The dependent variable is the Michigan Survey of Consumers stated inflation expectations over the next year. In the column labeled *News*, news topics are used as the only predictors. In the column labeled *News and Hard* the set of potential predictors is augmented to also include roughly 130 hard economic variables from the McCracken and Ng (2016) database. The table only reports the subset of news variables selected in either of the two regressions. For the hard economic indicators we only report those variables that are selected at the 10 percent significance level. The last raw reports the number of significant variables relative to total number of chosen variables for each model. *, **, and ***, indicate that coefficients are statistically significant at the 10, 5, 1 percent level, respectively. The sample period is 1990:01-2016:12.

| | News | | News and Hard | |
| --- | --- | --- | --- | --- |
| | $b$ | $Partial R^2$ | $b$ | $Partial R^2$ |
| Fiscal policy | - | - | -0.52*** (0.17) | 0.03 |
| M&A | 0.34*** (0.13 ) | 0.02 | 0.18 (0.11) | 0.01 |
| Clients | 0.43*** (0.14 ) | 0.03 | 0.05 (0.12) | 0.00 |
| Health | -1.15*** (0.20 ) | 0.09 | -0.46*** (0.18) | 0.02 |
| Internet | 0.71*** (0.14 ) | 0.08 | 0.42*** (0.11) | 0.05 |
| The White House | 0.30 (0.18 ) | 0.01 | 0.36** (0.15) | 0.02 |
| Labor market | 0.27* (0.16 ) | 0.01 | - | - |
| Aviation | -0.44** (0.20 ) | 0.02 | -0.33** (0.16) | 0.01 |
| Strategy | 0.18 (0.14 ) | 0.01 | 0.08 (0.13) | 0.00 |
| News service | 0.14 (0.22 ) | 0.00 | 0.09 (0.18) | 0.00 |
| Stock indices | 0.08 (0.21 ) | 0.00 | - | - |
| Events | 0.29 (0.19 ) | 0.01 | - | - |
| Real personal consumption expenditure | | | -16.88** (7.91) | 0.01 |
| IP:Durable consumption goods | | | -7.15** (2.79) | 0.02 |
| Employment Mining | | | 7.88* (4.22) | 0.01 |
| Building Permits in the South Census Region | | | -0.70*** (0.12) | 0.11 |
| Unfilled Orders for Durable Goods | | | 5.99 (3.66) | 0.01 |
| Real M2 Money Stock | | | -26.95*** (6.79) | 0.05 |
| 1-Year Treasury Rate | | | 0.90*** (0.27) | 0.04 |
| 5-Year Treasury Rate | | | 0.06 (0.19) | 0.00 |
| 3-Month Treasury Bill Minus Federal Funds Rate | | | -0.86*** (0.14) | 0.11 |
| Consumer Sentiment Index | | | -0.03*** (0.01) | 0.07 |
| Volatility Index | | | -0.02*** (0.01) | 0.04 |
| Adjusted $R^2$ | 0.41 | | 0.61 | |
| N | 6.00 of 11.00 | | 14.00 of 22.00 | |

**Table 3.** Narrative realism and story examples. The examples are found by querying the corpus for news articles where the news topics listed in column one receive a high weight. The date of publication is printed in parenthesis.

| | |
|---|---|
| **Aviation** | *(2013-01-24) Want a quick 30% discount on your family's trip to Europe or Hawaii? In the crazy airfare world, sometimes buying two tickets is cheaper than one. Pairing two discounted tickets together to create your own connecting itinerary can often be less expensive than flying on one ticket, if you take advantage of airlines' city-specific specials, or create your own route using discount airlines.* |
| **White House** | *(2003-09-04)U.S. Trade Representative Robert Zoellick is pressing Japan to open up its rice market under a new round of global trade talks that have stalled primarily over the politically sensitive farm trade among rich countries, Kyodo News reported Friday. To bring the global trade talks to a successful conclusion, Japan needs to expand the minimum volume of rice it is obliged to import under WTO rules, Zoellick said at a...* |
| **Internet** | *(2011-12-02) Google's plan to partner with major retailers and shippers to help online shoppers get products delivered within a day signals a ratcheting-up its rivalry with e-commerce king Amazon. But the move likely won't come as a surprise to Amazon CEO... Bezos's initial fears about Google were realized when the fast-rising search engine launched its first price-comparison service, Froogle, in 2002...* |
| **Health** | *(2006-08-16) An experimental blood test has shown a glimmer of promise of one day addressing a major health-care challenge: detecting lung cancer at an early stage. The test, developed by researchers at the University of Kentucky, is designed to identify ... Mr. Cohen of 20/20 GeneSystems estimates the cost of the blood test would be less than $200. CT scans can cost between $300 and $1,000 and usually aren't covered by...* |
| **Fiscal policy** | *(2008-06-12) On the surface, it sounds appealing. The Internal Revenue Service recently announced that economic stimulus payments directly deposited to eligible tax-favored accounts, including traditional and Roth IRAs, may be withdrawn tax- and penalty-free. In other words, if your $1,000 stimulus payment was directly deposited into your IRA, you can take out up to $1,000 from that account tax- and penalty-free...* |

From the roughly five million articles in the corpus, we extract the 10 most important, according to the article weights, and subjectively chose one of these for the table. Although none of the stories contain explicit *inflation* terms, it is clear that media coverage of topics related to, and labeled, *Aviation*, *White house*, *Internet*, *Health*, and *Fiscal policy* might all plausibly affect how economic agents consider inflation developments. The *Aviation* story, for example, talks about airline prices and discounts, while the *Health* and *Fiscal policy* stories talk about costs and taxes.

In sum, we find that a relatively large fraction of the variation in household's inflation expectations can be explained by between 5 and 11 news topic time series. Most of the news variables survive when controlling for a large set of hard economic indicators, suggesting that economic news play an independent role in shaping inflation expectations. In turn, these news topics provide a plausible narrative for what type of news articles households pay attention to when forming their inflation expectations.

## 4.3 News-driven information rigidities?

It follows from the theory model in Section 2 that the degree of information rigidity among households should be a function of the persistence and noise-to-signal ratio in the signal extraction problem. Using a set $S$ of relevant news topics, derived from the LASSO procedure in the previous section, we now test this relationship by running the following regression:

$$\beta_t = c + \gamma_1 \rho_t + \gamma_2 \kappa_t + u_t. \tag{15}$$

Here, $\beta_t$ is the median time-varying information rigidity, reported in Figure 1a, while $\rho_t$ and $\kappa_t$ are the persistence and noise-to-signal ratio in the underlying information set. Based on the regression results in the previous section, we construct quantitative measures of these as follows.

First, for each of the news topics in the selected variable set $S$, we run simple Autoregressive $(AR(p))$ models. To introduce time dependencies, we allow both the volatility of the $AR(p)$ innovations and the autoregressive parameters to be time dependent. The parameters follow random walk processes, and we set $p = 1$ to avoid over-fitting. This model structure, together with the Gibbs simulations used for estimation, is standard in the time series literature (see, e.g. Primiceri (2005)), and described in greater detail in Appendix G.3. For future reference, we let $\hat{\rho}_{i,t}$ and $\hat{\sigma}_{i,t}$ denote the estimated posterior draws of the time-varying persistence and volatility for news topic $i$. As higher and more persistent coverage of one type of news leads to less coverage of other news items by definition in the LDA model, time-variation in $\hat{\rho}_{i,t}$ also captures the predictions from Nimark and Pitschner (2018) that news-coverage will be homogeneous, for example around major events.

Second, for each of the news topics in the selected variable set $S$, we construct a measure of the noise in the signal by calculating the standard deviation of the posterior topic estimates. That is, when calculating the importance of each news topic for each day in the sample, posterior uncertainty will introduce variability in the article weight distributions and in the articles selected to tone-adjust the news topic time series. Using the sum of the standard deviation of these distributions, we obtain a measure of the noise, denoted $\hat{\omega}_{i,t}$. Intuitively, this noise measure can be interpreted as follows: If it is difficult for the topic model algorithm to classify with high precision the topic proportions of a given article, it would likely be difficult for a human as well. Thus, uncertainty regarding what the news is about increases. Likewise, uncertainty increases if articles differ in terms of their tone.

Finally, aggregating across all the news topics $i$ in the set $S$, and combining the output

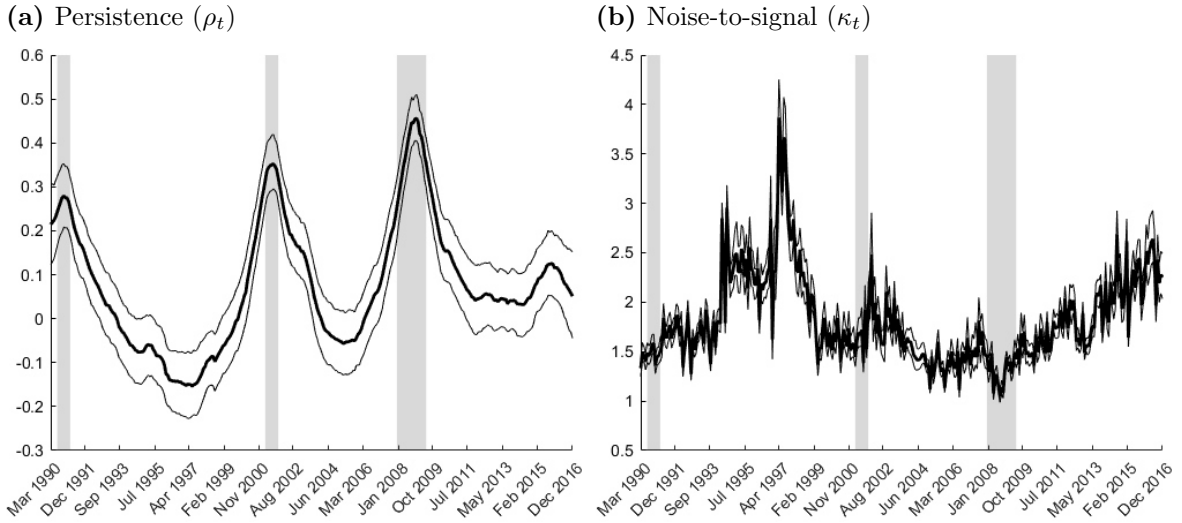**(a)** Persistence ($\rho_t$)    **(b)** Noise-to-signal ($\kappa_t$)

**Figure 3.** Time-varying media persistence and noise-to-signal ratio. The black solid lines are the median estimates, while the gray solid lines are 68 percent probability bands. The gray shaded areas illustrate recession periods as defined by NBER (U.S.).

from these two steps, we obtain:

$$\hat{\rho}_t = \sum_{i=1}^{n} w_i \hat{\rho}_{i,t} \quad \hat{\kappa}_t = \sum_{i=1}^{n} w_i \hat{\kappa}_{i,t} \text{ with } \hat{\kappa}_{i,t} = \frac{\hat{\omega}_{i,t}}{\hat{\sigma}_{i,t}} \text{ and } w_i = \frac{R_i^2}{\sum_{i=1}^{n} R_i^2}, \quad (16)$$

where $w_i$ is the normalized partial $R^2$ statistic for variable $i$ in Table 2. Thus, variables that are more important in terms of explaining the variation in households' inflation expectations are given a larger weight when constructing $\hat{\rho}_t$ and $\hat{\kappa}_t$.

The results presented in Table 2 give some degrees-of-freedom in choosing the relevant set S. To avoid using news topics potentially just capturing information highly correlated with hard economic indicators, and to avoid relying on test statistics computed in stage two of a variable selection problem, we define the set S to include all the selected topics from the LASSO regression controlling for hard economic indicators. Figure 3 graphs the posterior draws of the estimates in (16) for this set. The average degree of persistence in the news varies significantly, and tends to be especially high around recession periods, as one would expect. In contrast, and perhaps surprising, the noise-to-signal ratio is not very high around recessions. Instead, it seems to associate (in particular) the mid 1990s as a "noisy period". We do not have a good explanation for this pattern, but note that while the method used to construct this variable is intuitive, it is also somewhat sensitive to the representativeness of the raw corpus data. If some time periods contain news extracts from fewer, or different types, of articles, this might contaminate our noise-to-signal measure.[14]

---

[14]The topic extraction itself is less prone to this issue because the topic distributions are based on information from the whole sample. We also note that the estimates in Figure 3, and the results presented below, are very similar if we instead define the set S as; Only the selected and significant topics from

Column $I$ in Figure 4a reports the results of estimating (15) for each draw of $\hat{\rho}_t$ and $\hat{\kappa}_t$, using non-informative natural conjugate priors. Accordingly, parameter estimates correspond to the OLS solution, but taking into account the generated regressors issue by sampling from the full distribution of $\hat{\rho}_t$ and $\hat{\kappa}_t$. Going forward, we label this our *Benchmark* model. As seen from the figure, the coefficient estimates have the correct sign. A higher persistence and lower noise-to-signal ratio lead to a reduction in information rigidity. However, given that we work in a high-dimensional time-varying parameter setting, the posterior uncertainty is naturally large, especially for the persistence parameter.

The results presented in Table 4 highlight that the patterns documented above are very unlikely to be obtained by chance, at least for the persistence parameter. We show this by running a falsification experiment. First, 100 different sets of news topics, not including those used in constructing the set $S$ above, are constructed. Then, for each of these alternative sets, we calculate $\hat{\rho}_t$ and $\hat{\kappa}_t$, and redo the estimation of (15). Since the news topics used in this experiment are "irrelevant" for households' inflation expectations (according to the LASSO), we also expect the posterior distributions to be less informative. And, as illustrated in Table 4, they are. The posterior probability of $\hat{\gamma}_1 < 0$ is roughly 85 percent for the *Benchmark* model, while only 31 percent of the alternative regressions have a posterior probability greater than 50 percent for this outcome. In fact, only roughly 6 percent of the alternative regressions have a posterior probability of $\hat{\gamma}_1 < 0$ that is equal to, or greater, than 80 percent. For the noise-to-signal ratio, this result is less strong, and almost 83 percent of the alternative regressions have roughly the same high posterior probability of $\hat{\gamma}_2 > 0$ as the *Benchmark* model. This suggests that there is a large common component in the noise-to-signal variable, which might be attributed to the measurement issue mentioned earlier.

In columns $II$ and $III$ in Figure 4a we augment the *Benchmark* model specification in (15) with either; 10 factors from the macro economic *FRED-MD* database (column $II$); run a double selection algorithm controlling for all the roughly 130 variables in this database (column $III$). How we constructed the factors and implement the double selection algorithm was discussed in Section 3.[15] These augmented regressions naturally strengthens the results presented in column $I$ of Figure 4a by increasing the adjusted $R^2$ statistic considerably, but also makes the news-based persistence parameter more negative and highly significant. As such, taking aboard potentially omitted variables strengthens our results further.

the news-only regression; All selected topics from the news-only regression; Only the selected and significant topics from the regression controlling for hard economic indicators. These additional results can be obtained on request.

[15]Here, the factor augmented regressions and the double selection algorithm are done for each draw of $\hat{\rho}_t$ and $\hat{\kappa}_t$, while revision uncertainties in the hard economic indicators are not accounted for.
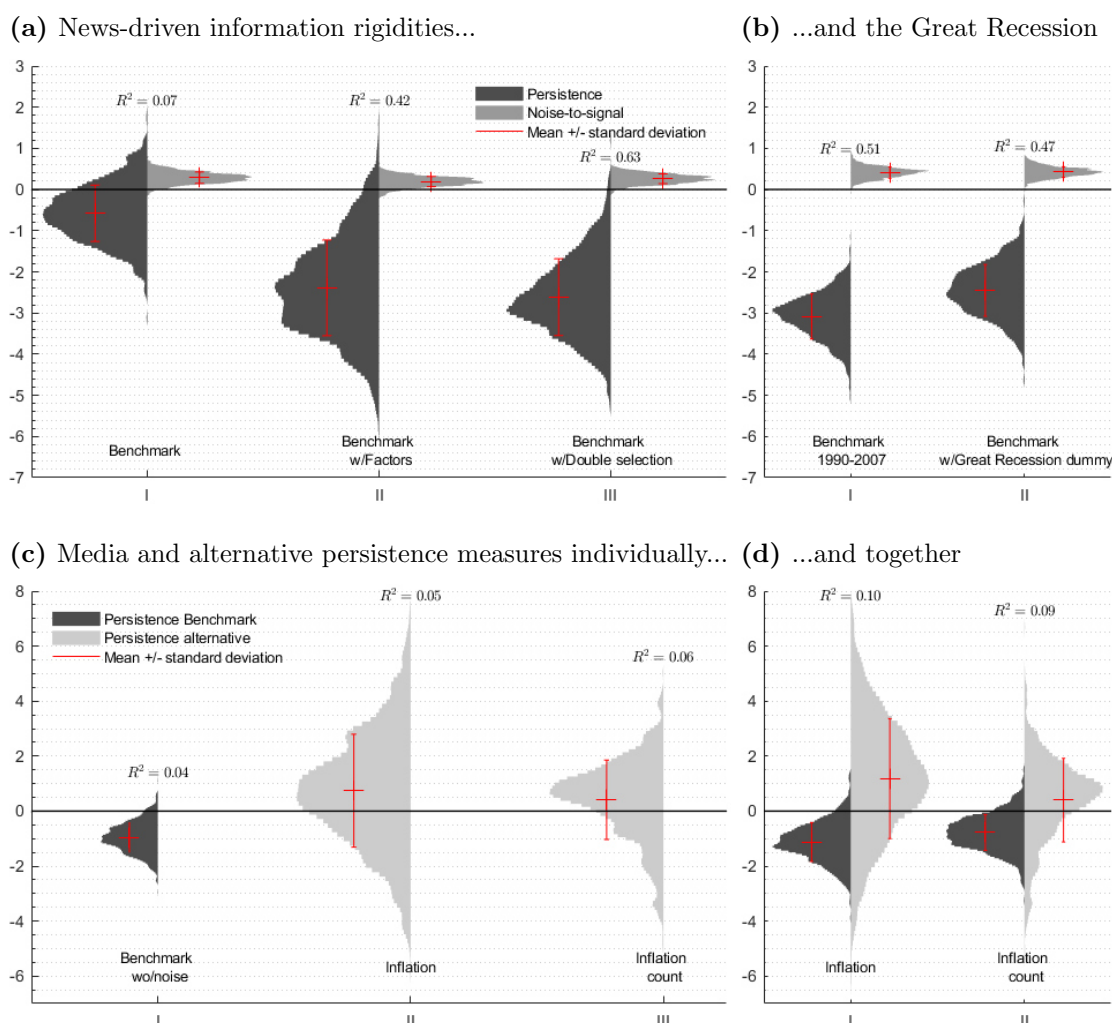
**(a)** News-driven information rigidities...

**(b)** ...and the Great Recession

**(c)** Media and alternative persistence measures individually...

**(d)** ...and together

**Figure 4.** Violin plots of posterior distributions. Column *I* in Figure 4a reports the estimates from the *Benchmark* model, while columns *II* and *III* report estimates from an augmented version of the *Benchmark* model using either: 10 factors from the *FRED-MD* database, or a double selection algorithm controlling for all the variables in this database. Dark and light gray areas correspond to the distributions for $\gamma_1$ and $\gamma_2$, from equation (15), respectively. The red crosses mark the mean estimate +/- one standard deviation. The mean adjusted $R^2$ statistic is reported above the distributions. Figure 4b reports *Benchmark* estimates from a truncated sample, and when including a Great Recession dummy variable to the regression. In Figure 4c equation (15) is estimated; Using the *Benchmark* news-based model without including the noise-to-signal term (*I*); Using the persistence in inflation (*II*) and using the persistence in the alternative inflation count series (*III*). Figure 4d reports the same estimates as in Figure 4c after also controlling for the *Benchmark* news-based persistence measure.

The results presented in Figure 4a summarizes well our main result: Media coverage plays an important role for describing the degree of information rigidity among households. However, from visual inspection of the co-movement between the estimated information rigidity (Figure 1a) and the media persistence (Figure 3a) we observe that something is different during the Great Recession period. For this time period, both information rigidity and media persistence is high, suggesting a positive correlation, and not negative,

**Table 4.** Cumulative histogram. Equation (15) is estimated for 100 randomly selected sets of news topics (not in the set of topics used to generate the *Benchmark* parameter distributions). The table reports the fraction of draws that have posterior probabilities $Pr(\gamma_1 < 0) \geq x$ and $Pr(\gamma_2 > 0) \geq x$, where $x$ refers to a bin in the histogram. The bin associated with the posterior probability for the *Benchmark* model is marked in gray.

| $x$ | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
|---|---|---|---|---|---|---|---|---|---|
| $Pr(\gamma_1 < 0) \geq x$ | 0.81 | 0.59 | 0.46 | 0.40 | 0.31 | 0.20 | 0.17 | 0.06 | 0.03 |
| $Pr(\gamma_2 > 0) \geq x$ | 0.96 | 0.95 | 0.94 | 0.91 | 0.91 | 0.91 | 0.91 | 0.87 | 0.83 |

as implied by theory. This is also one reason for why the results in columns *II* and *III* in Figure 4a are so strong relative to those in column *I*. For example, the double selection procedure consistently chooses control variables that experienced particularly large swings during the Great Recession period, i.e., housing starts, spreads, weekly hours, and employment figures (see Figure 11a in Appendix B). The results presented in Figure 4b illustrate more directly the peculiarity of the Great Recession period. That is, estimating the *Benchmark* model up until 2007, or including a dummy variable for the Great Recession period, yields very similar results to those presented in columns *II* and *III* in Figure 4a. We return to the Great Recession discussion in Section 5. Before that we explore if alternative inflation-based variables can explain the time-varying degree of information rigidity among households, and then examine the dynamic interaction between information rigidity, the media, and macro economic developments in greater detail.

## 4.4 Inflation and an alternative news measure

First, in the standard theoretical framework developed by Coibion and Gorodnichenko (2015a), there is no role for the media, and it is the underlying time series properties of inflation itself that should determine the degree of information rigidity. To test this no-media alternative we construct a quantitative measure of the persistence in inflation as we did for the news topics, i.e., using a time-varying $AR(1)$ model, and then re-estimate (15) using this persistence measure instead of the news-based one. As we do not have a good measure for noise in the inflation series, the noise-to-signal ratio is not included in the regression.[16]

As seen from column *II* in Figure 4c, the persistence parameter for the inflation-based regression is not significant, and if anything, has the wrong sign. For comparison, using the news-based persistence measure as the only explanatory variable in (15) we obtain an estimate with the correct sign and where the 68 percent posterior intervals exclude

---

[16]In Coibion and Gorodnichenko (2015a) the real-time revisions in the outcome series of interest is used as a proxy for noise. As (CPI) inflation is not revised, this is not possible in our setting.

zero, see column *I* in Figure 4c. This finding also holds after controlling for inflation and the news-based persistence measure in the same regression, see column *I* in Figure 4d. Accordingly, both results confirm that there is an important independent role for the media in explaining households' expectations formation process.[17]

In the literature we speak to, the conventional method used to measure the intensity of media reporting relevant for households' inflation expectations has been to count the number of terms related to inflation in the corpus' articles (headlines) (Carroll (2003), Pfajfar and Santoro (2013), Lamla and Lein (2014)). In our view, and as alluded to in Section 4.1, this method builds on a rather stringent assumption. Many items in the news might be of relevance for households' inflation expectations, even without explicitly mentioning the term *inflation*. This motivates our news-topic-based approach.

The results presented in the last columns in Figures 4c and 4d illustrate that the news-topic-based approach also provides a better description of the time-varying information rigidity observed among households than the count-based method does. The results are produced as follows. First, we construct an alternative media measure based on counting terms related to inflation in articles using the wild-card search *inflation\**. This count metric is then summed for each day in the sample, and normalized by the article count for that day. Next, we follow the same procedure as used above to measure media persistence, i.e., estimating a time-varying $AR(1)$ model for the alternative media measure, and then redo the estimation of (15). Figure 12, in Appendix B, reports the alternative inflation count measure together with the time-varying persistence parameter. Re-estimating the LASSO regression in Section 4.2, including the inflation count measure as an additional variable, we observe that it is not selected (results not shown). Further, as seen from column *III* in Figure 4c and column and *II* in Figure 4d, the persistence for this alternative measure of inflation news does not explain the evolution of households' information rigidity, and if anything, has the wrong sign.

As an additional test of our research design, we have also applied our methodology to the Survey of Professional Forecasters (SPF) CPI inflation forecasts. A-priori we conjecture that the media plays a much smaller role for professional forecasters than for households,[18] and our results confirms this conjecture. The media-based news topics do not predict SPF expectations. However, in line with the model outlined in Section 2, but without a media channel, the time series properties of inflation itself can (partly)

---

[17]In unreported results we have also tested if persistence in macroeconomic variables relevant for households' inflation expectations (confer Table 2), can explain households' time-varying information rigidity, finding that they can not.

[18]That is, professional forecasters have much less need to delegate their information choice to the media. They surely know and follow actual CPI inflation, and they have much more resources than households to help them stay informed about the different (economic) states of the world.

explain the degree of time-varying information rigidity among profession forecasters. In the interest of preserving space, the details of this experiment are relegated to Appendix E.

## 4.5 Information rigidity, business cycles, and the media

The mechanics of the model outlined in Section 2 are straight forward. When an important (business cycle) event happens, media coverage potentially becomes more concentrated and persistent around this event and, if the noise-to-signal ratio is also reduced, information rigidity falls. Accordingly, the degree of information rigidity and business cycle developments are closely related, and documenting this relationship is important because theoretical models with information frictions give different policy implications than models assuming, e.g., FIRE. For example, in state-dependent models of information rigidity (Gorodnichenko (2008), Woodford (2009), Mackowiak and Wiederholt (2009)), the degree of information rigidity should be much lower after a large and visible shock than during normal periods. And, when acquiring information is looked upon as a choice variable, the degree of information rigidity should be inversely related to business cycle developments (Mackowiak et al. (2018)). At the same time, in the related delegated information choice view of the world (Nimark and Pitschner (2018)), the distribution of events and media's reporting on those events jointly determine the degree of information rigidity.

When using aggregate (monthly) business cycle data, these views give at least three testable predictions. First, because predictors of information rigidity and the business cycle are jointly determined, they imply that these variables should Granger cause each other. Second, and for the same reason, they imply that an exogenous shock to information rigidity cannot be separately identified from a large economic event, e.g., a recessionary shock. Or, in other words, these two shocks are the same. Finally, they imply that following a more regular business cycle disturbance, the business cycle and information rigidity should be negatively correlated.

Below we test each of these predictions in a simple Vector Autoregressive (VAR) framework using our news-based information rigidity predictors. This framework allows us to conduct standard Granger causality tests (Granger (1969)) and use impulse response analysis to investigate the dynamic interaction between information rigidity, business cycles, and the media.

As a measure of the business cycle we follow a large literature, see, e.g. Stock and Watson (1989), Bjørnland et al. (2017), and McCracken and Ng (2016), and use the first principal component estimated from the monthly *FRED-MD* database. This variable is graphed in Figure 11b in Appendix B. The VAR is specified with six lags, and includes the business cycle indicator, denoted *BC*, the news-based persistence and noise-to-signal
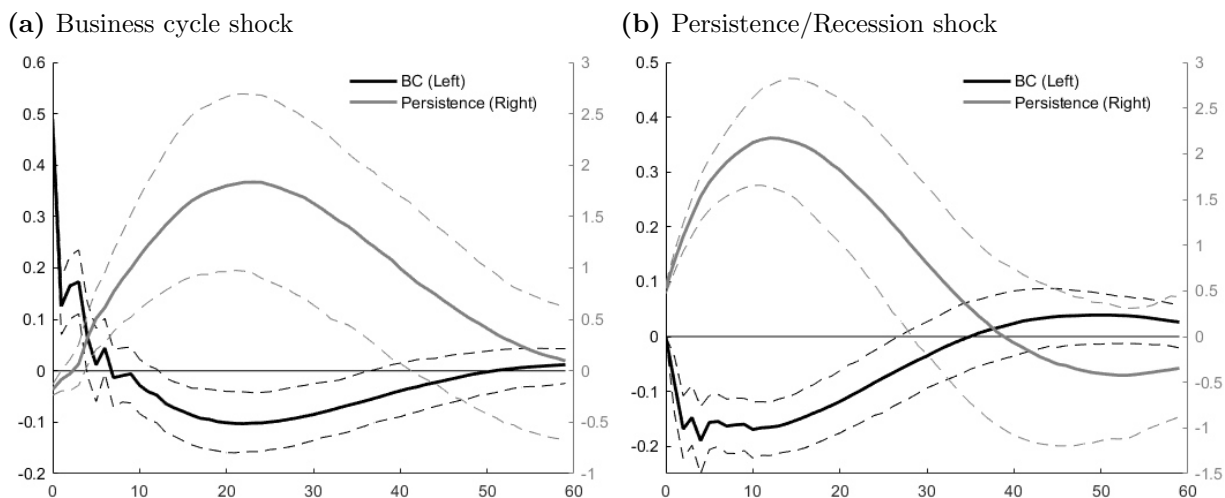
**(a)** Business cycle shock

**(b)** Persistence/Recession shock

**Figure 5.** Impulse responses following a one standard deviation exogenous shock in a three-variable VAR identified using a recursive ordering ($[BC_t, \rho_t, \kappa_t]$). The solid lines are mean estimates, while the dashed lines report 95 percent confidence bands, estimated using a residual bootstrap.

ratio time series, in that order.[19]

According to the Granger causality test we find strong support, significant at the one percent level, for the first prediction, namely that Granger causality runs in both directions between the business cycle and the news-based persistence variable. In contrast, the news-based noise-to-signal ratio is not Granger caused by any of the other two variables, and does not itself Granger cause neither the business cycle indicator nor the persistence measure.

Figure 5a reports the response in the business cycle indicator and the news-based persistence variable following an exogenous shock to the $BC$ indicator, identified using a recursive ordering (Cholesky).[20] In line with the third prediction from above, there is an inverse relationship between the two response paths. On impact, the persistence variable falls slightly, indicating that information rigidity increases. Then, as the business cycle boom cools off, persistence gradually increases, and reaches its peak when the business cycle is at its trough, after roughly 25 months. Thus, according to these estimates, and given the negative relationship between information rigidities and the news-based persistence measure, information rigidities are at their lowest during economic downturns.

The implications from Figure 5a, namely that information rigidities are at their lowest during recessionary periods, is also documented in Loungani et al. (2013), who look

---

[19] For simplicity, the mean estimates of the news-based persistence and noise-to-signal ratio time series are used in the regression. The VAR lag length is determined from likelihood ratio tests, while parameter uncertainty in the VAR is estimated using a residual bootstrap.

[20] As implied by the Granger causality tests, the noise-to-signal ratio is more or less exogenous, and its impulse responses are more or less insignificant. In the interest of brevity, and for visual clarity, we do not report them.

at GDP growth forecasts of professional forecasters in 46 countries, and by Coibion and Gorodnichenko (2015a), who analyze inflation expectations among professional forecasters. However, neither of these studies considers the role of the media in this setting.

A consequence of the second prediction from above, i.e., that an exogenous shock to information rigidity cannot be separately identified from a large economic event, is that an immediate increase in the news-based persistence measure should be associated with an immediate drop in the business cycle indicator. Figure 5b confirms this prediction. An unexpected increase in the news-based persistence measure, implying a reduction in information rigidity, is associated with a sharp fall in the business cycle, i.e., a recession. In Figure 5b, and by construction, the drop first occurs with a one-period lag. However, when reversing the order of the variables in the VAR, we show in Figure 13, in Appendix B, that the fall is immediate. Importantly, these responses differ from those obtained after a more regular business cycle disturbance, graphed in Figure 5a, where the relationship between economic activity and the news-based persistence measure is more gradual. We also note that the results reported in Figure 5 are robust. Augmenting the VAR with typical recession indicators like the spread between long and short horizon interest rate maturities, the VIX index, oil prices, or differencing the highly persistent news-based variables, do not qualitatively change our results.

In sum, while it is well known that models with information frictions affect the dynamics of the business cycle and have important policy implications (Mackowiak et al. (2018)), the results documented here are new because they highlight the role played by the media as "information intermediaries" in this setting. For institutions conducting countercyclical policies (partly) by managing expectations (Galí (2008)), this suggest that media should play an important role in their communication strategies, as also emphasized in, e.g., Berger et al. (2011) and Haldane (2017).

# 5 The Great Recession period

The results presented in Section 4.3 documented that there is a significant and theory consistent relationship between households' information rigidity and media persistence and noise. However, as was illustrated in Figure 4b, the Great Recession period, hereafter referred to as the $GR$ period, perturbed this relationship. The question then becomes, why is there a $GR$ "puzzle"? In this final section, we propose two potential explanations, which not only shed further light on the news-driven inflation expectations and information rigidities relationship, but also allow us to conduct a (economic) meaningful out-of-sample evaluation of the predictive relationships documented earlier.

First, one could argue, as for example in Imbs (2010) and Bjørnland et al. (2017), that

the *GR* period was the first truly global recession period in decades, and that inflation behaved differently during this period relative to before, as discussed in, e.g., Ball and Mazumder (2011) and Coibion and Gorodnichenko (2015b). Accordingly, media coverage might simply have been less informative for describing inflation, and inflation expectations, during the *GR* period relative to before (and after). In this environment, households might rationally have chosen not to follow the media to update their expectations. We call this the bad media quality explanation.

Second, one (of potentially many) alternative explanation is that media coverage was good and relevant, also during the *GR* period, but that households paid less attention to news altogether. That is, if the quality of information is good, but information diffusion is poor, the quality of information matters less. While information diffusion is unobserved, a reasonable first order approximation can be obtained by using newspaper circulation statistics, where simple statistics seem to give some support to this view.[21] In particular, it is well known that newspaper circulation numbers have been falling dramatically in the U.S. (and in other countries) since the early 1990s, but the cyclical patterns of newspaper circulation are less known. However, irrespective of whether one detrends the non-stationary circulation statistics with a second order polynomial, a Hodrick-Prescott (HP) filter (Hodrick and Prescott (1997)), or simply looks at the yearly difference of the series, one obtains the same qualitative answer: During recessionary periods, circulation numbers often, but not always, drop below trend or fall. But, during the *GR* period, the negative gap was over three times larger than during any other recession in the U.S. since the 1940s, see Figure 14, in Appendix B. Accordingly, the *GR* period might have been different because information diffusion from news media was much worse during this time period than in any other recession the last decades.

## 5.1 Out-of-sample performance

We start by addressing the bad quality story. To this end we rely on a quasi out-of-sample (OOS) predictive experiment for both households' inflation expectations and actual CPI inflation. We first estimate a LASSO regression, as described in Section 4.2, for each of these outcome variables for the sample period 1990-2000, predict one period forward, and redo this analysis for the remaining part of the sample, using an expanding estimation window. As explanatory variables, we either use all the news topics, or the hard economic indicators from the *FRED-MD* database. As such, the OOS experiment allows us to track how the prediction error evolves, and thereby pinpoint when potential breaks occur,

---

[21]The newspaper circulation statistics are collected from Pew Research Center and their *Newspaper Fact Sheet* at http://www.journalism.org/fact-sheet/newspapers/. We refer to their web-pages for more documentation on how the circulation statistics are collected and compiled.

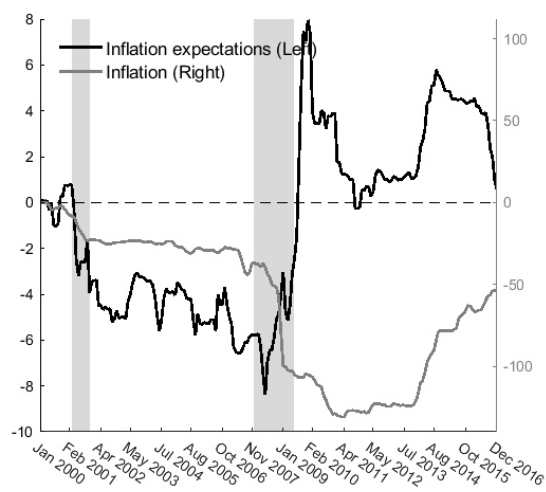**(a)** In-sample stability and news topic duration
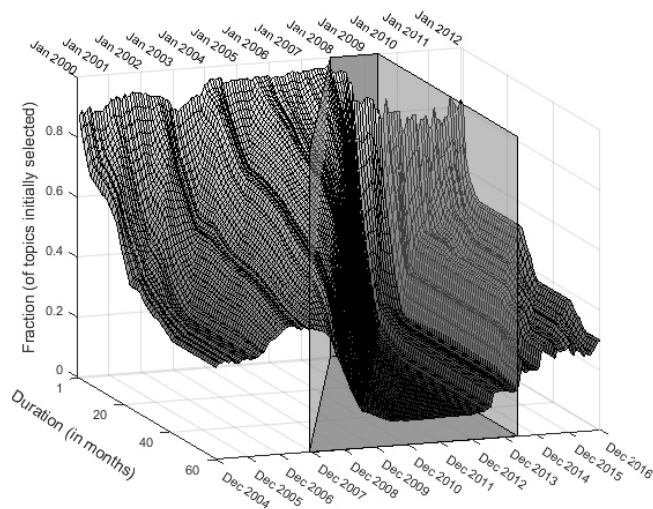
**(b)** CSPED

**Figure 6.** Figure 6a report the (recursively estimated) fraction of topics that are still being selected by the LASSO 1 to 60 months after they were initially selected as predictors for households' inflation expectations. Figure 6b reports the out-of-sample cumulative squared prediction error differences (CSPED) between the news-based-only and hard-economic-only based LASSO regressions. Two outcome variables are considered, namely households' inflation expectations and actual CPI inflation. See the text for further details.

compute in-sample (recursive) statistics related to the stability of the relevant variable sets, and compare the news-based forecasting performance for both inflation expectations and actual inflation.

Figure 6 summarizes the main results from this experiment. Figure 6a shows that the $GR$ period is associated with a large change in the relationship between news topics and inflation expectations. The figure should be read as follows: For each estimation vintage, ending as illustrated on the top x-axis, we calculate the fraction of news topic predictors that are still being selected by the LASSO algorithm 1-60 months later, where the end date for the 60 period duration is illustrated on the lower x-axis. For example, during the mid 2000s, roughly 40 percent of the initially selected news topics were still useful predictors for households' inflation expectations even 5 years (60 months) after they where initially selected. Thus, the predictive relationship was fairly stable and long lasting. Going into the $GR$ period, this pattern changes markedly. From Figure 6a we observe that less than 10 percent of the selected news topic predictors are still being used after just 20-40 months. After the $GR$ period, however, there seems to be a return towards the same patterns we observe prior to the $GR$ period. Thus, there is clear evidence of in-sample instability during the $GR$ period.

The black line in 6b puts further evidence behind this reasoning. It reports the cumulative squared prediction error difference (CSPED) between the news-based out-of-sample predictions and those based on the model including only the hard economic indicators.

A value above zero indicates that the latter specification is better than the news-based one, and vice versa. As clearly seen in the figure, the news-based out-of-sample predictions were better until 2008, but then deteriorated substantially. However, starting from roughly 2010, we again observe that the news-based model improves relative to the one based on hard economic indicators, confirming that the $GR$ period was different. Indeed, if the evaluation sample had started after the $GR$ period, the news-based model would have been better also in absolute terms.

At the same time, it is hard to argue that media coverage became of worse quality during the $GR$ period. In particular, when we do the OOS experiment for actual CPI inflation, and compare the news-based performance to one based on hard economic indicators, we actually observe that the news-based approach is superior, see the gray line in Figure 6b. And, in contrast to the results for household expectations, the news-based approach improves further during the $GR$ period.[22]

In sum, and consistent with the bad media quality explanation, Figure 6 documents clear in-sample instability and an out-of-sample deterioration of the news and inflation expectation relationship during the $GR$ period. At the same time, however, our results suggest that the news was highly informative of actual inflation developments during this period. In other words, households could have followed the news to form good updates of their inflation expectations, but do not seem to have done so. This points us towards the second potential explanation mentioned above, where media coverage was good and relevant, also during the $GR$ period, but households paid less attention to news altogether.

## 5.2 An information diffusion story?

We formally address what we label as the information diffusion explanation by estimating an augmented version of equation (15):

$$\beta_t = c + \gamma_1 \rho_t + \gamma_2 \kappa_t + \gamma_3 circ_t + \gamma_4(circ_t \times \rho_t) + \gamma_5(circ_t \times \kappa_t) + u_t \qquad (17)$$

where $circ_t$ is the HP-filtered newspaper circulation variable, used in (17) as a crude approximation for information diffusion.[23] All variables, $\rho_t$, $\kappa_t$ and $circ_t$, are measured as

---

[22]Overall, the root mean squared forecasting error (RMSFE) for the news-based model is 1.2, while the LASSO regression entertaining the *FRED-MD* data obtains a RMSFE score of 1.3. Interestingly, both statistics are much better than 1.9, which is the score obtained when we estimate the much used unobserved-components stochastic volatility inflation forecasting model suggested by Stock and Watson (2007). The finding that the news-topic-based approach outperforms the usage of hard economic indicators when predicting actual CPI inflation is a novel finding in itself. We leave it for future research to explore in greater detail how news data can be used to predict inflation developments.

[23]The circulation statistic is recorded on a yearly frequency. We obtain monthly numbers by using shape-preserving piecewise cubic interpolation on the yearly trend-adjusted estimates.
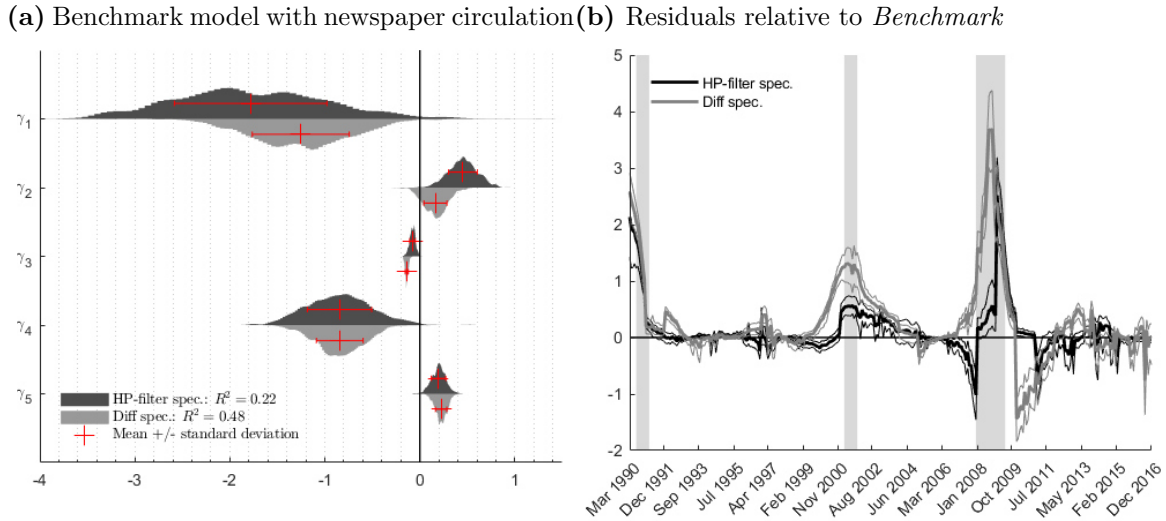
**(a)** Benchmark model with newspaper circulation **(b)** Residuals relative to *Benchmark*

**Figure 7.** Figure 7a reports the posterior estimates for two versions of equation (17). See the text for details. The red crosses mark the mean estimate +/- one standard deviation. Figure 7b reports the squared residual difference between the *Benchmark* model, estimated from (15), and the one including interaction terms, estimated from (17). Median estimates together with 68 percent probability bands are reported. The gray shaded areas illustrate recession periods as defined by NBER (U.S.).

deviations from their means. Accordingly, $\gamma_3$ measures the direct effect of a higher than usual information diffusion on information rigidity, while the interaction terms capture the idea that the effects of the media persistence and noise-to-signal ratio might be conditional on the degree of information diffusion.

Figure 7 summarizes the results. Four findings stand out. First, as seen from the dark-colored density estimates in Figure 7a, and consistent with a large body of past work outside economics consistently finding that those who consume print media are better informed about current events than those who do not (see, e.g., Finnegan and Viswanath (1996)), we find that the independent effect of the circulation variable is negative and significant. More importantly here, the persistence and noise parameters have a significant negative and positive sign, respectively, and the interaction terms suggest that these patterns are stronger in times of relatively higher newspaper circulation, i.e., better information diffusion, as hypothesized at the beginning of this section. Second, the mean adjusted $R^2$ statistic for (17) is 0.24, which is substantially higher than 0.04 obtained for the *Benchmark* model in column *I* in Figure 4a. Thus, controlling for the circulation statistic improves the model's explanatory power considerably. Third, from Figure 7b, which reports the squared error difference between the *Benchmark* model and the one in (17), we learn that most of the improvement in fit is due to smaller errors during recessionary periods, and in particular during the *GR* period. Finally, although there are many ways to calculate the cyclical patterns of a non-stationary variable, the qualitative results reported above are not very sensitive to our choice of using the com-

33

mon HP-filter for detrending the newspaper circulation statistics. In fact, when instead using the simple difference in the newspaper circulation statistics, we obtain the posterior estimates reported in light gray in Figure 7, where the model fit improves further.

Of course, as news consumers today have the option to substitute from print to online news, one could argue that newspaper circulation statistics no longer serves as a good approximation for information diffusion, and that the large cyclical drop during the *GR* period is due to a substitution towards cheaper (free) online news or social media. As such, an alternative interpretation of the results reported above is that the households consumed a different type of news during the *GR* period, and not that information diffusion was lower per se. Although we can not rule out this explanation, which is highly interesting in itself, we note that during the early 1990s, and partly during the early 2000s, the substitution effect must have been minor because online news did not exist. Still, the increase in explanatory power for the augmented model, relative to the *Benchmark* model, is almost just as large during the 1990s recession as in the *GR* period, see Figure 7b, and it is exactly during these recession periods the circulation statistic falls substantially (confer Figure 14, in Appendix B). Accordingly, what we label as the information diffusion explanation is consistent across the sample, while the substitution explanation is not.

We conclude from the above analysis that the predictive relationship between news topics and households' inflation expectations withstand out-of-sample testing. Moreover, in terms of information rigidities, we find that one potential explanation for the *GR* "puzzle" is related to the amount of information households' consumed during this period, or, alternatively, the type of media content. Irrespective of interpretation, this is in accordance with the results presented in earlier sections about news media's important role in the expectations formation process. Going forward, the preceding analysis motivates more work investigating how the relationship between media coverage and macroeconomic expectations might be affected by a changing media landscape.

# 6  Conclusion

We investigate the role of media for understanding inflation expectations and information rigidities among U.S. households. Taking the view that acquiring information can be looked upon as a choice variable (Sims (2003), Gorodnichenko (2008), Woodford (2009), Mackowiak and Wiederholt (2009)), and that the media work as "information interme-diaries" (Nimark and Pitschner (2018)), we augment the testing framework introduced by Coibion and Gorodnichenko (2015a) with a simple media channel, and find empirical support for the following: First, the degree of information rigidity among households is far from constant, but varies significantly across time (over the business cycle). Second,

using a novel news-topic-based approach, we show that the news types the media choose to report on are good predictors of households' stated inflation expectations. Finally, we show that the underlying time series properties of news topics relevant for households' expectations explain the time-varying information rigidity among households. When media persistence (noise-to-signal) is high (low), information rigidities tend to be low (high), and vice versa, as the theory model predicts.

A number of robustness tests document that these results are very unlikely to have been obtained by chance, and that similar findings are not found when using the time series properties of inflation itself, or a measure of media coverage based on counting inflation terms in the news. In line with theory, we also document significant interactions between the business cycle and the part of households' information rigidity that can be explained by the media. For policy institutions aiming at managing consumers' expectations, e.g., monetary policy, this emphasizes the media as an important channel for their communication strategies.

Our study speaks to a growing literature documenting significant information rigidities and a departure from the full information rational expectations hypothesis. Thus far, however, the role of the media has only gotten limited (formal) attention in the this literature. Using a large news corpus and machine learning algorithms we contribute by focusing on media's role within a well-established (theoretical) testing framework. Our results highlight media's importance in such a setting.

Arguably, our analysis is only partial. We do not specify a full model for how potentially profit maximizing "information intermediaries" interact with utility maximizing consumers etc. Given information and expectations prevalent role in economics, an interesting avenue for further research would be to incorporate the delegated information choice mechanism, and the role of the media as information providers, in a general equilibrium framework.

# References

Armantier, O., S. Nelson, G. Topa, W. van der Klaauw, and B. Zafar (2016). The Price Is Right: Updating Inflation Expectations in a Randomized Price Information Experiment. *The Review of Economics and Statistics 98*(3), 503–523.

Ball, L. M. and S. Mazumder (2011). Inflation dynamics and the great recession. Working Paper 17044, National Bureau of Economic Research.

Belloni, A. and V. Chernozhukov (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19*(2), 521 – 547.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives 28*(2), 29–50.

Berger, H., M. Ehrmann, and M. Fratzscher (2011). Monetary policy in the media. *Journal of Money, Credit and Banking 43*(4), 689–709.

Bjørnland, H. C., F. Ravazzolo, and L. A. Thorsrud (2017). Forecasting GDP with global components: This time is different. *International Journal of Forecasting 33*(1), 153 – 173.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM 55*, 77–84.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent Dirichlet Allocation. *J. Mach. Learn. Res. 3*, 993–1022.

Blinder, A. S., M. Ehrmann, M. Fratzscher, J. De Haan, and D.-J. Jansen (2008). Central bank communication and monetary policy: A survey of theory and evidence. *Journal of Economic Literature 46*(4), 910–45.

Blinder, A. S. and A. B. Krueger (2004). What does the public know about economic policy, and how does it know it? Working Paper 10787, National Bureau of Economic Research.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological) 37*(2), 149–192.

Carroll, C. D. (2003). Macroeconomic Expectations of Households and Professional Forecasters. *The Quarterly Journal of Economics 118*(1), 269–298.

Carter, C. K. and R. Kohn (1994). On Gibbs Sampling for State Space Models. *Biometrika 81*(3), 541–553.

Carter, D. W. and J. W. Milon (2005). Price knowledge in household demand for utility services. *Land Economics 81*(2), 265–283.

Chang, J., S. Gerrish, C. Wang, J. L. Boyd-graber, and D. M. Blei (2009). Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22*, pp. 288–296. Cambridge, MA: The MIT Press.

Chetty, R. and E. Saez (2013). Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients. *American Economic Journal: Applied Economics 5*(1), 1–31.

Coibion, O. and Y. Gorodnichenko (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy 120*(1), 116–159.

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

Coibion, O. and Y. Gorodnichenko (2015b). Is the phillips curve alive and well after all? inflation expectations and the missing disinflation. *American Economic Journal: Macroeconomics 7*(1), 197–232.

Curtin, R. (2007). *What US Consumers Know About Economic Conditions*. OECD.

Doepke, J., J. Dovern, U. Fritsche, and J. Slacalek (2008). The dynamics of european inflation expectations. *The B.E. Journal of Macroeconomics 8*(1).

Doms, M. and M. Norman (2004). Consumer sentiment, the economy, and the news media. Working Papers in Applied Economic Theory.

Dougal, C., J. Engelberg, D. Garcia, and C. A. Parsons (2012). Journalists and the stock market. *Review of Financial Studies 25*(3), 639–679.

Dovern, J., U. Fritsche, P. Loungani, and N. Tamirisa (2015). Information rigidities: Comparing average and individual forecasts for a large international panel. *International Journal of Forecasting 31*(1), 144–154.

Dräger, L. and M. J. Lamla (2017). Imperfect information and consumer inflation expectations: Evidence from microdata. *Oxford Bulletin of Economics and Statistics 79*(6), 933–968.

Dybowski, T. and P. Adämmer (2018). The economic effects of u.s. presidential tax communication: Evidence from a correlated topic model. *European Journal of Political Economy 55*, 511 – 525.

Ehrmann, M., D. Pfajfar, and E. Santoro (2017). Consumers' Attitudes and Their Inflation Expectations. *International Journal of Central Banking 13*(1), 225–259.

Finnegan, J. R. and K. Viswanath (1996). The knowledge gap hypothesis: Twenty-five years later. *Annals of the International Communication Association 19*(1), 187–228.

Fullone, F., M. Gamba, E. Giovannini, and M. Malgarini (2007). *What Do Citizens Know about Statistics*. OECD.

Galí, J. (2008). Introduction to Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework. In *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Introductory Chapters. Princeton University Press.

Gentzkow, M., B. T. Kelly, and M. Taddy (2017). Text as data. Working Paper 23276, National Bureau of Economic Research.

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? evidence from u.s. daily newspapers. *Econometrica 78*(1), 35–71.

Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review 101*(7), 2980–3018.

Gorodnichenko, Y. (2008). Endogenous information, menu costs and inflation persistence. NBER Working Papers 14184, National Bureau of Economic Research, Inc.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*(3), 424–38.

Griffiths, T. L. and M. Steyvers (2004). Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America 101*(Suppl 1), 5228–5235.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American Economic Review 70*(3), 393–408.

Haldane, A. G. (2017). A little more conversation, a little less action. *Bank of England-Speech*.

Hansen, S. and M. McMahon (2015). Shocking Language: Understanding the macroeconomic effects of central bank communication. Discussion Papers 1537, Centre for Macroeconomics (CFM).

Hansen, S., M. McMahon, and A. Prat (2018). Transparency and deliberation within the fomc: A computational linguistics approach. *The Quarterly Journal of Economics 133*(2), 801–870.

Heinrich, G. (2009). Parameter estimation for text analysis. Technical report, Fraunhofer IGD.

Hodrick, R. J. and E. C. Prescott (1997). Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking 29*(1), 1–16.

Imbs, J. (2010). The First Global Recession in Decades. *IMF ECONOMIC REVIEW 58*(2), 327–354. Conference on Economic Linkages, Spillovers, and the Financial Crisis, Paris, FRANCE, JAN 28-29, 2010.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics 125*(2), 515–548.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies 65*(3), 361–93.

King, G., B. Schneer, and A. White (2017). How the news media activate public expression and influence national agendas. *Science 358*(6364), 776–780.

Lamla, M. J. and S. M. Lein (2014). The role of media for consumers' inflation expectation formation. *Journal of Economic Behavior & Organization 106*, 62 – 77.

Lamla, M. J. and S. Sarferaz (2012). Updating inflation expectations. KOF working papers No. 301.

Larsen, V. H. (2017, April). Components of uncertainty. Working Paper 2017/5, Norges Bank.

Larsen, V. H. and L. A. Thorsrud (2017). Asset returns, news topics, and media effects. Working Paper 2017/17, Norges Bank.

Larsen, V. H. and L. A. Thorsrud (2018a). Business Cycle Narratives. Working Paper 2018/03, Norges Bank.

Larsen, V. H. and L. A. Thorsrud (2018b). The Value of News for Economic Developments. *Journal of Econometrics (Forthcoming)*.

Loungani, P., H. Stekler, and N. Tamirisa (2013). Information rigidity in growth forecasts: Some cross-country evidence. *International Journal of Forecasting 29*(4), 605 – 621.

Mackowiak, B. and M. Wiederholt (2009). Optimal sticky prices under rational inattention. *The American Economic Review 99*(3), 769–803.

Mackowiak, B. A., F. Matejka, and M. Wiederholt (2018). Survey: Rational Inattention, a Disciplined Behavioral Model. CEPR Discussion Papers 13243, C.E.P.R. Discussion Papers.

Mankiw, N. G. and R. Reis (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics 117*(4), 1295–1328.

McCracken, M. W. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Nakajima, J. and M. West (2013). Bayesian Analysis of Latent Threshold Dynamic Models. *Journal of Business & Economic Statistics 31*(2), 151–164.

Nieuwerburgh, S. V. and L. Veldkamp (2010). Information Acquisition and Under-Diversification. *Review of Economic Studies 77*(2), 779–805.

Nimark, K. P. and S. Pitschner (2018). News Media and Delegated Information Choice. CEPR Discussion Papers 11323, C.E.P.R. Discussion Papers.

Pang, B., L. Lee, and S. Vaithyanathan (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, Stroudsburg, PA, USA, pp. 79–86. Association for Computational Linguistics.

Peress, J. (2014). The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *The Journal of Finance 69*(5), 2007–2043.

Pfajfar, D. and E. Santoro (2008). Asymmetries in inflation expectation formation across demographic groups. Cambridge working papers in economics, Faculty of Economics, University of Cambridge.

Pfajfar, D. and E. Santoro (2013). News on Inflation and the Epidemiology of Inflation Expectations. *Journal of Money, Credit and Banking 45*(6), 1045–1067.

Prat, A. (2018). Media power. *Journal of Political Economy 126*(4), 1747–1783.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies 72*(3), 821–852.

Shiller, R. J. (2017). Narrative economics. *American Economic Review 107*(4), 967–1004.

Simon P. Anderson, J. W. and D. Strömberg (Eds.) (2015). Volume 1 of *Handbook of Media Economics*. North-Holland.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665 – 690. Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information.

Souleles, N. S. (2004). Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the michigan consumer sentiment surveys. *Journal of Money, Credit and Banking 36*(1), 39–72.

Stock, J. H. and M. W. Watson (1989). New indexes of coincident and leading economic indicators. In O. J. Blanchard and F. Stanley (Eds.), *NBER Macroeconomics Annual*, NBER Chapters, pp. 351–394. Cambridge, MA: The MIT Press.

Stock, J. H. and M. W. Watson (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking 39*(s1), 3–33.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance 62*(3), 1139–1168.

Tetlock, P. C. (2015). Chapter 18 - the role of media in finance. In J. W. Simon P. Anderson and D. Strömberg (Eds.), *Handbook of Media Economics*, Volume 1 of *Handbook of Media Economics*, pp. 701 – 721. North-Holland.

Thorsrud, L. A. (2018). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics (Forthcoming)*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological) 58*(1), 267–288.

Veldkamp, L. L. (2006). Media Frenzies in Markets for Financial Information. *American Economic Review 96*(3), 577–601.

Wiswall, M. and B. Zafar (2014). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies 82*(2), 791–824.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. In J. S. Philippe Aghion, Roman Frydman and M. Woodford (Eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, pp. 25–58. Princeton University Press.

Woodford, M. (2009). Information-constrained state-dependent pricing. *Journal of Monetary Economics 56*(S), 100–124.

# Appendices

## Appendix A  *FRED-MD* data

The *FRED-MD* database was compiled by McCracken and Ng (2016), and contains a wide range of economic variables used for macroeconomic monitoring. The extraction used here covers the sample 1990:01-2016:12, and contains 134 monthly variables. The database classifies the data into 8 groups: (1) Output and Income, (2) Labor Market, (3) Consumption and Orders, (4) Orders and Inventories, (5) Money and Credit, (6) Interest rate and Exchange Rates, (7) Prices, and (8) Stock Market. To obtain stationary data, and to prepare the data for PCA analysis, we use the same data transformations as suggested in McCracken and Ng (2016).

## Appendix B  Additional results
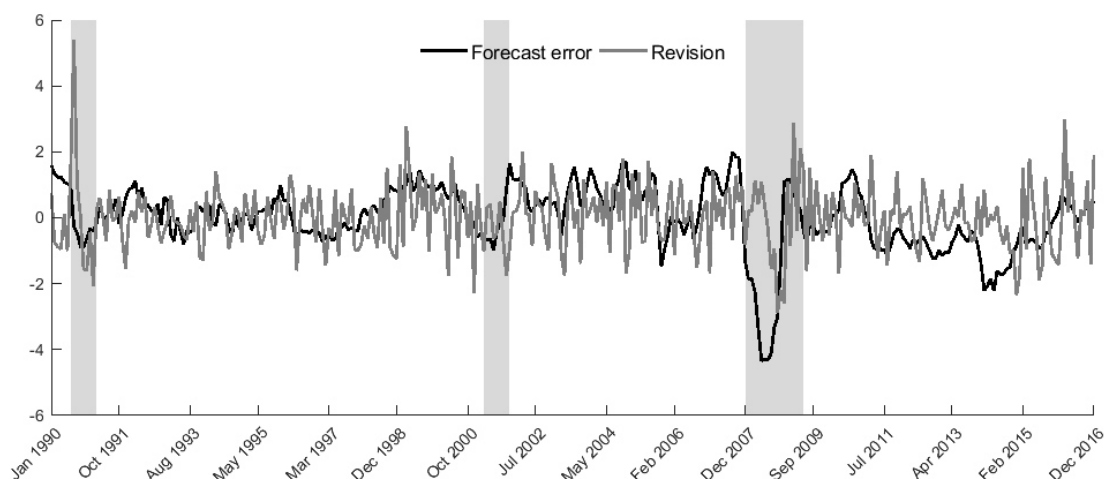


**Figure 8.** The figure reports households' forecast errors and forecast revisions of CPI inflation. Households' expectations are taken from the Michigan Survey of Consumers monthly survey of households' CPI inflation expectations over the course of the next year. The data are normalized. The gray shaded areas illustrate recession periods as defined by NBER (U.S.).

**(a)** Forward test

**(b)** Backward test
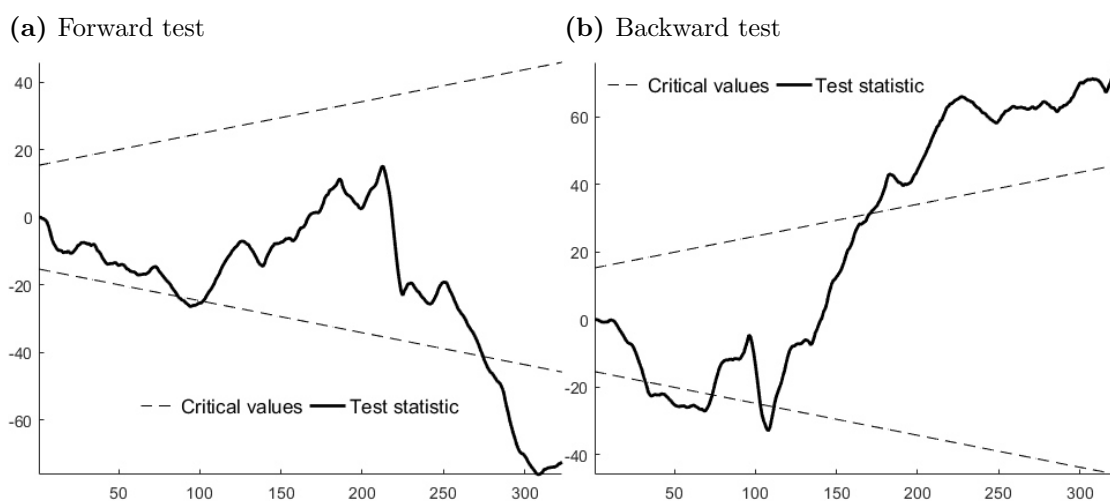
**Figure 9.** Cusum test (Brown et al. (1975)) of parameter stability in equation (9).



**(a)** Time-fixed effect $(c_t)$

**(b)** Information rigidity $(\beta_t)$
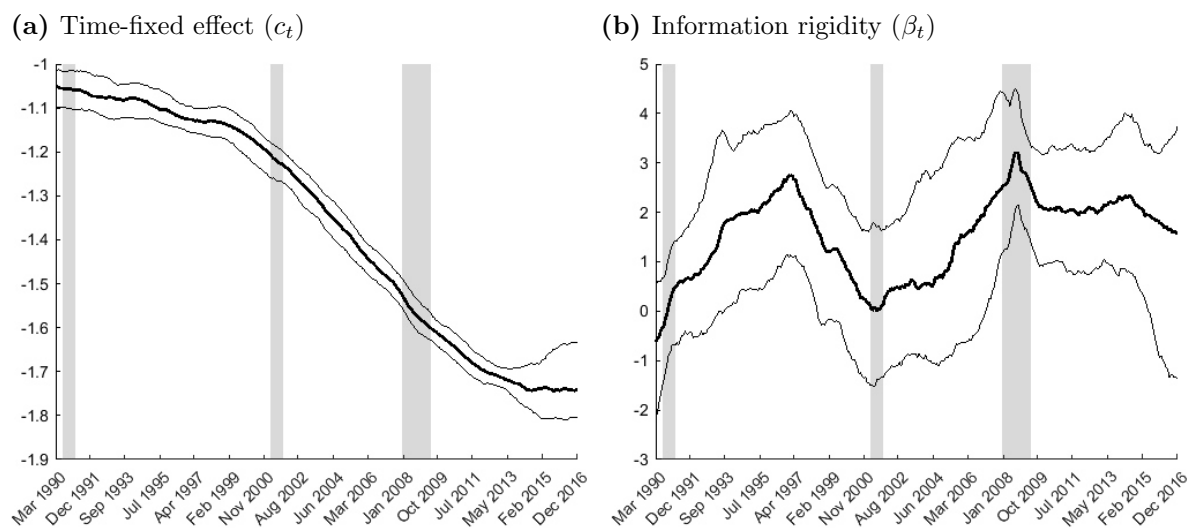
**Figure 10.** Michigan Survey of Consumers inflation expectations; Time-varying information rigidity and time-fixed effects (media bias). The black solid lines are the median, while the gray solid lines are 68 percent probability bands. The gray shaded areas illustrate recession periods as defined by NBER (U.S.).

43

**(a)** Selected control variables
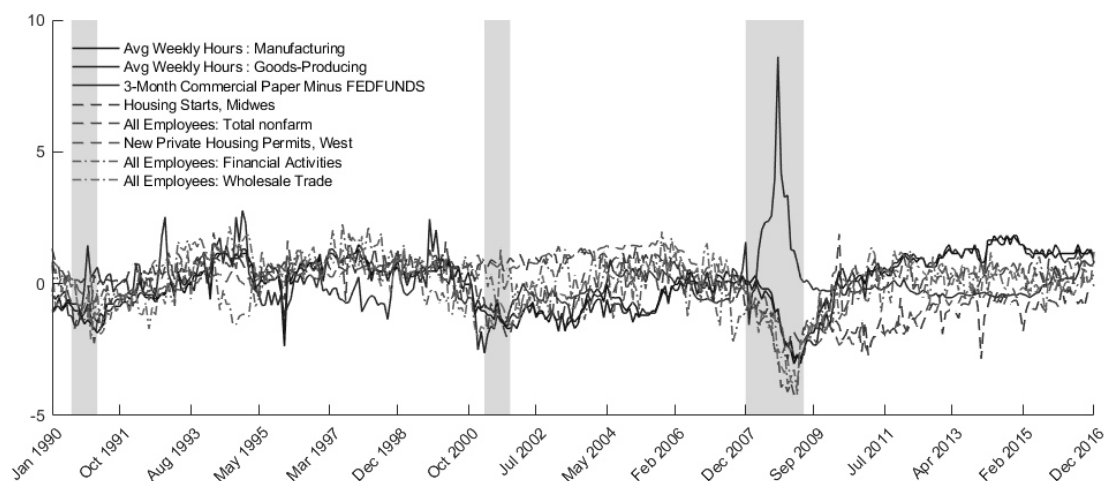


**(b)** Business cycle factor
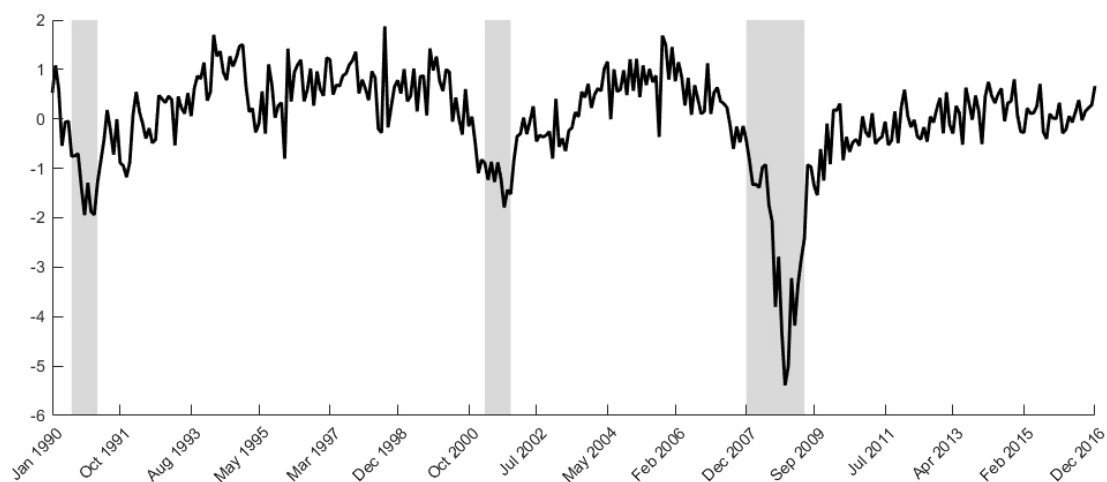


**Figure 11.** Figure 11a reports the hard economic indicators, from the *FRED-MD* database, that were selected more than 90 percent of the time in the double selection algorithm implemented in Section 4.3. Figure 11b reports the first principal component estimated from the monthly *FRED-MD* database. All data are standardized. The gray shaded areas illustrate recession periods as defined by NBER (U.S.)

**(a)** Inflation count measure
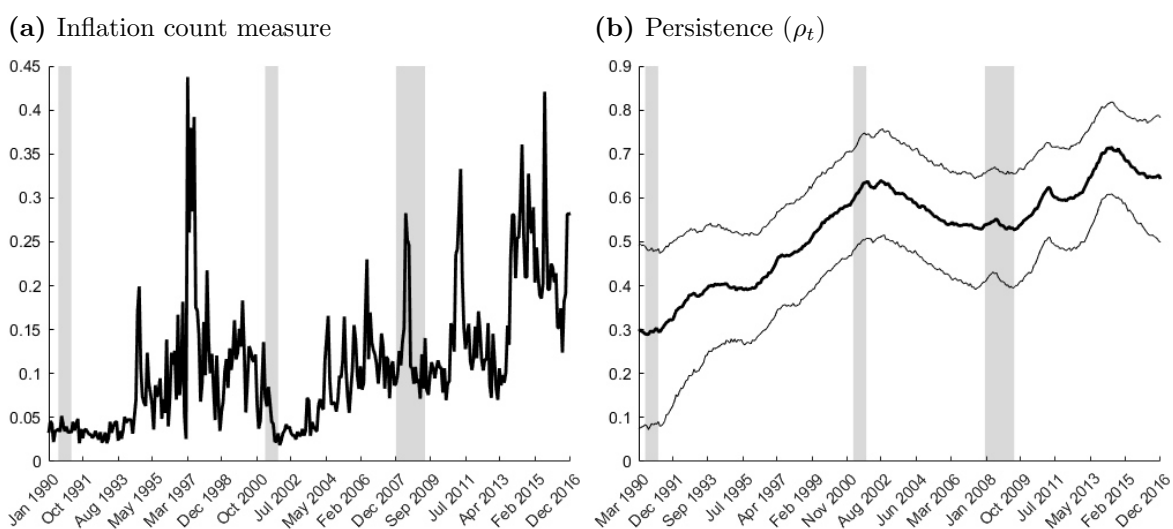
**(b)** Persistence ($\rho_t$)

**Figure 12.** Figure 12a reports the alternative inflation intensity count series, while Figure 12b reports the time-varying persistence in this series. The solid black line is the median estimate, while the solid gray lines are 68 percent probability bands. The gray shaded areas illustrate recession periods as defined by NBER (U.S.).

**(a)** Business cycle shock

**(b)** Persistence/Recession shock
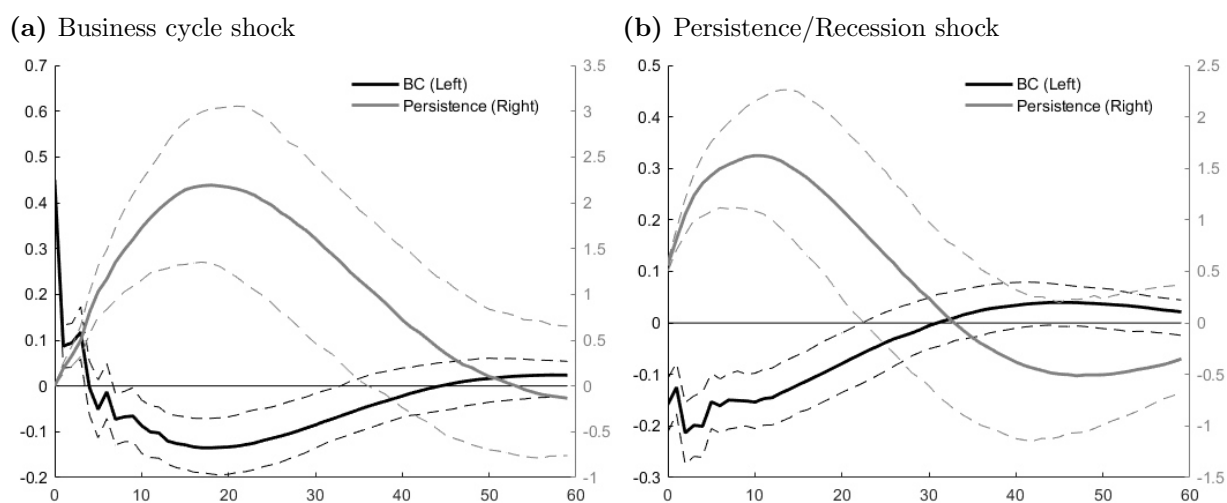
**Figure 13.** Impulse responses following a one standard deviation exogenous shock in a three variable VAR identified using a recursive ordering ($[\rho_t, \kappa_t, BC_t]$). The solid lines are mean estimates, while the dashed lines report 95 percent confidence bands, estimated using a residual bootstrap.

45
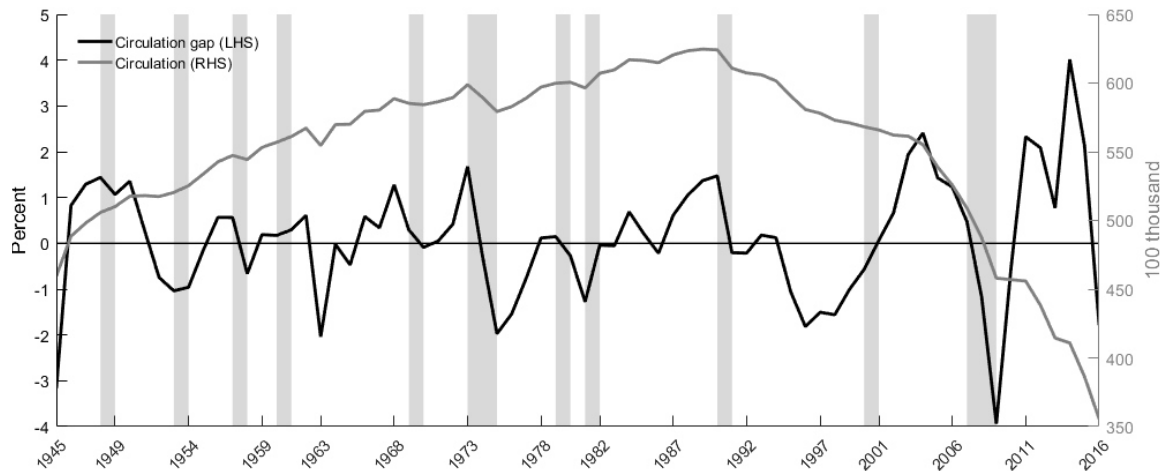
**Figure 14.** Yearly U.S. newspaper circulation statistics, in level and detrended with the Hodrick-Prescott (HP) filter (Hodrick and Prescott (1997)) with $\lambda = 100$. The circulation statistic is collected from Pew Research Center and their *Newspaper Fact Sheet* at http://www.journalism.org/fact-sheet/newspapers/. We refer to their web-pages for more documentation on how the circulation statistics are collected and compiled.

**Table 5.** US news topics. Subjective labeling and the most important words. Weights in parenthesis.

| Id | Label | Top words (word probability) |
|----|-------|------------------------------|
| 0 | Monetary policy | inflat, 0.056, monetari, 0.018, bernank, 0.018, technic, 0.015, greenspan, 0.014, resist, 0.012, minut, 0.011 |
| 1 | Fiscal policy | budget, 0.059, save, 0.033, deficit, 0.028, balanc, 0.014, social, 0.014, ir, 0.013, trillion, 0.012, reduct |
| 2 | Education | school, 0.044, ms, 0.035, univers, 0.031, famili, 0.028, student, 0.024, educ, 0.022, children, 0.016, colleg |
| 3 | Funding | loan, 0.106, mortgag, 0.082, borrow, 0.029, lend, 0.025, articl, 0.024, analysi, 0.022, lender, 0.022, link |
| 4 | Entertainment | book, 0.014, film, 0.013, art, 0.013, music, 0.011, movi, 0.011, star, 0.01, play, 0.01, artist, 0.006, theater |
| 5 | Telecommunication | network, 0.048, wireless, 0.043, phone, 0.033, mobil, 0.029, telecom, 0.025, telecommun, 0.022, verizon, 0.022 |
| 6 | Agriculture | edt, 0.062, corn, 0.014, est, 0.011, crop, 0.011, farmer, 0.01, ceo, 0.01, agricultur, 0.009, farm, 0.009 |
| 7 | Environment | water, 0.031, fuel, 0.029, environment, 0.024, emiss, 0.016, clean, 0.016, solar, 0.014, renew, 0.014, wast |
| 8 | Strategy | strategi, 0.026, opportun, 0.019, expand, 0.018, success, 0.016, focu, 0.016, challeng, 0.013, focus, 0.012 |
| 9 | Trading | vol, 0.361, avg, 0.167, ttl, 0.158, blk, 0.081, prev, 0.041, nm, 0.02, zero, 0.017, uptick, 0.017, nyse, 0.016 |
| 10 | Pharmaceutical | drug, 0.097, pharmaceut, 0.024, treatment, 0.02, fda, 0.019, patient, 0.019, trial, 0.015, cancer, 0.014, studi |
| 11 | Media | media, 0.039, tv, 0.032, cabl, 0.029, advertis, 0.025, network, 0.025, warner, 0.023, televis, 0.023, broadcast |
| 12 | Petroleum | crude, 0.083, barrel, 0.073, gasolin, 0.041, inventori, 0.02, nymex, 0.019, gallon, 0.018, heat, 0.018, opec |
| 13 | Public safety | polic, 0.034, fire, 0.025, kill, 0.015, death, 0.013, protest, 0.013, man, 0.011, gun, 0.011, safeti, 0.01 |
| 14 | Employment | employe, 0.08, worker, 0.076, union, 0.064, employ, 0.036, pension, 0.033, labor, 0.032, strike, 0.021, wage |
| 15 | Iraq | militari, 0.037, iraq, 0.036, war, 0.023, iraqi, 0.022, troop, 0.02, armi, 0.014, attack, 0.013, afghanistan |
| 16 | Market perform | merril, 0.012, usd, 0.011, neutral, 0.01, nasdaq, 0.01, tg, 0.01, ep, 0.01, valuat, 0.007, djia, 0.007 |
| 17 | Health care | health, 0.116, care, 0.086, hospit, 0.028, medic, 0.026, insur, 0.021, medicar, 0.019, coverag, 0.016 |
| 18 | News service | thomson, 0.039, guidanc, 0.036, exclud, 0.032, segment, 0.023, adjust, 0.023, item, 0.022, gross, 0.021, prior |
| 19 | Energy | electr, 0.069, util, 0.065, plant, 0.057, ga, 0.02, capac, 0.018, california, 0.013, facil, 0.012, transmiss |
| 20 | Natural gas | ga, 0.071, natur, 0.033, pipelin, 0.025, bp, 0.023, drill, 0.021, field, 0.019, explor, 0.016, refineri, 0.014 |
| 21 | China | china, 0.152, chines, 0.058, asia, 0.037, hong, 0.025, kong, 0.023, asian, 0.022, beij, 0.019, export, 0.016 |
| 22 | M&A | bid, 0.063, merger, 0.055, stake, 0.05, acquir, 0.032, transact, 0.025, combin, 0.021, takeov, 0.019, familiar |
| 23 | Advisory | trust, 0.053, brown, 0.031, bancorp, 0.018, advisor, 0.016, dj, 0.013, branch, 0.011, ohio, 0.011, mgmt, 0.01 |
| 24 | Smartphones | appl, 0.053, devic, 0.025, iphon, 0.019, game, 0.018, phone, 0.016, mobil, 0.014, app, 0.013, smartphon, 0.012 |
| 25 | Clients | client, 0.047, email, 0.043, assum, 0.036, dilut, 0.035, advis, 0.032, reader, 0.031, either, 0.029, along |
| 26 | Persuasion | know, 0.022, realli, 0.012, someth, 0.012, got, 0.011, happen, 0.011, cannot, 0.01, tell, 0.009, sure, 0.008 |
| 27 | Elections | elect, 0.041, campaign, 0.033, obama, 0.031, democrat, 0.025, republican, 0.023, parti, 0.023, vote, 0.021 |
| 28 | Software | softwar, 0.074, microsoft, 0.05, comput, 0.041, ibm, 0.02, network, 0.015, window, 0.015, oracl, 0.014, applic |
| 29 | Electronics | chip, 0.044, comput, 0.029, intel, 0.028, electron, 0.027, dell, 0.023, semiconductor, 0.021, equip, 0.018, pc |
| 30 | Regulations | regul, 0.061, sec, 0.044, practic, 0.018, regulatori, 0.016, law, 0.016, investig, 0.015, audit, 0.013, act |
| 31 | Food | food, 0.059, restaur, 0.028, brand, 0.021, chain, 0.014, mcdonald, 0.011, drink, 0.011, coffe, 0.01, cola, 0.01 |
| 32 | Justice | court, 0.078, law, 0.036, judg, 0.029, lawsuit, 0.028, legal, 0.026, claim, 0.024, settlement, 0.023, appeal |
| 33 | Economic crisis | crisi, 0.028, reform, 0.019, imf, 0.018, institut, 0.017, emerg, 0.015, stabil, 0.014, commit, 0.011, rubin |
| 34 | Retail | brand, 0.03, mart, 0.025, wal, 0.024, chain, 0.022, holiday, 0.017, discount, 0.015, shop, 0.015, apparel, 0.014 |
| 35 | Europe | london, 0.046, plc, 0.029, india, 0.024, franc, 0.022, french, 0.021, ag, 0.019, deutsch, 0.018, german, 0.017 |
| 36 | Leadership | ceo, 0.034, vice, 0.033, serv, 0.025, join, 0.022, resign, 0.021, replac, 0.019, role, 0.019, appoint, 0.017 |
| 37 | Terrorism | attack, 0.033, al, 0.023, terrorist, 0.021, terror, 0.019, israel, 0.017, palestinian, 0.014, pakistan, 0.013 |
| 38 | Stocks | common, 0.086, symbol, 0.057, issuer, 0.054, regist, 0.041, titl, 0.035, filer, 0.031, ownership, 0.03, outstand |
| 39 | Health | test, 0.039, studi, 0.023, dr, 0.019, diseas, 0.016, human, 0.013, health, 0.012, patient, 0.011, cancer, 0.009 |
| 40 | Insurance | insur, 0.121, life, 0.032, aig, 0.023, premium, 0.021, deposit, 0.018, re, 0.014, claim, 0.014, cover, 0.011 |
| 41 | Russia | russia, 0.039, russian, 0.03, minist, 0.022, nato, 0.018, kosovo, 0.014, prime, 0.013, eu, 0.012, moscow, 0.01 |
| 42 | Brokerage firms | morgan, 0.082, goldman, 0.056, stanley, 0.046, merril, 0.037, sach, 0.035, citigroup, 0.031, lynch, 0.03, ge |
| 43 | Stock indices | nasdaq, 0.063, composit, 0.03, nyse, 0.024, advanc, 0.023, lost, 0.023, poor, 0.016, climb, 0.016, ralli, 0.015 |
| 44 | Documentation | letter, 0.048, review, 0.038, request, 0.025, document, 0.021, correct, 0.019, wrote, 0.017, sent, 0.017, respond |
| 45 | Internet | onlin, 0.042, googl, 0.04, internet, 0.03, search, 0.025, user, 0.025, yahoo, 0.023, facebook, 0.022, ventur |
| 46 | Commentary | blog, 0.044, david, 0.033, steven, 0.025, pm, 0.023, paul, 0.021, onlin, 0.021, miller, 0.015, morn, 0.012 |
| 47 | The White House | bush, 0.069, white, 0.037, clinton, 0.034, georg, 0.02, secretari, 0.019, negoti, 0.016, leader, 0.014, free |
| 48 | East Asia | japan, 0.092, north, 0.077, south, 0.059, korea, 0.054, japanes, 0.046, dn, 0.033, tokyo, 0.029, korean, 0.024 |
| 49 | Natural gas | ga, 0.04, natur, 0.035, weather, 0.03, la, 0.022, casino, 0.022, winter, 0.016, normal, 0.015, vega, 0.014 |
| 50 | Currencies | euro, 0.085, currenc, 0.075, yen, 0.048, zone, 0.02, ecb, 0.013, japan, 0.011, london, 0.011, greec, 0.01, franc |
| 51 | Weapons | nuclear, 0.037, iraq, 0.035, iran, 0.033, weapon, 0.028, council, 0.028, sanction, 0.022, resolut, 0.016 |
| 52 | Results | dec, 0.05, dividend, 0.048, aug, 0.044, sept, 0.043, asknewswir, 0.03, item, 0.029, exclud, 0.027, oct, 0.027 |
| 53 | Volatility | auction, 0.036, hedg, 0.027, volatil, 0.022, spread, 0.017, fix, 0.017, dealer, 0.016, bid, 0.016, particip |
| 54 | Argumentation | often, 0.013, exampl, 0.012, rather, 0.01, approach, 0.01, actual, 0.009, fact, 0.008, argu, 0.008, studi, 0.007 |
| 55 | Labor market | economist, 0.05, labor, 0.026, revis, 0.024, claim, 0.023, unemploy, 0.022, employ, 0.02, read, 0.019, payrol |
| 56 | Real estate | properti, 0.054, estat, 0.045, hotel, 0.031, squar, 0.02, center, 0.019, leas, 0.017, park, 0.015, owner, 0.013 |
| 57 | Australia | mine, 0.033, australia, 0.025, chemic, 0.025, ltd, 0.024, materi, 0.021, coal, 0.02, australian, 0.019, ton |
| 58 | Fear | recess, 0.017, slow, 0.015, recoveri, 0.014, worri, 0.012, warn, 0.01, fear, 0.01, confid, 0.01, bad, 0.008 |
| 59 | Events | yesterday, 0.019, bp, 0.013, yr, 0.013, ep, 0.012, jh, 0.011, pjv, 0.01, chanc, 0.01, dec, 0.008 |
| 60 | California | san, 0.047, california, 0.046, counti, 0.031, calif, 0.027, lo, 0.027, angel, 0.027, francisco, 0.023, jersey |
| 61 | Bonds | moodi, 0.039, matur, 0.03, poor, 0.024, spread, 0.023, downgrad, 0.02, plu, 0.02, swap, 0.017, grade, 0.017 |
| 62 | Market talk | edt, 0.038, kevin, 0.029, kingsburi, 0.025, est, 0.024, premarket, 0.021, ep, 0.012, ceo, 0.012, kevinkingsburi |
| 63 | Latin America | mexico, 0.043, brazil, 0.036, de, 0.021, mexican, 0.02, latin, 0.019, brazilian, 0.017, local, 0.016, peso, 0.016 |
| 64 | Automobiles | car, 0.061, auto, 0.056, gm, 0.048, vehicl, 0.047, ford, 0.038, motor, 0.036, truck, 0.019, chrysler, 0.017 |
| 65 | Bankruptcies | bankruptci, 0.065, facil, 0.028, protect, 0.028, restructur, 0.027, creditor, 0.024, chapter, 0.019, court |
| 66 | Weather | storm, 0.023, hurrican, 0.022, florida, 0.019, west, 0.017, island, 0.016, damag, 0.015, coast, 0.015, texa |
| 67 | Sports | game, 0.035, team, 0.032, play, 0.02, season, 0.019, player, 0.018, sport, 0.014, win, 0.012, leagu, 0.01, coach |
| 68 | Investigations | investig, 0.037, alleg, 0.022, attorney, 0.02, prosecutor, 0.017, fraud, 0.014, crimin, 0.013, lawyer, 0.013 |
| 69 | Aircrafts | defens, 0.035, boe, 0.034, aircraft, 0.027, engin, 0.024, air, 0.02, commerci, 0.017, jet, 0.016, space, 0.016 |
| 70 | Options | option, 0.136, grant, 0.028, william, 0.028, lee, 0.021, tobacco, 0.02, exercis, 0.019, expir, 0.019, philip |
| 71 | Design | design, 0.012, ms, 0.009, room, 0.007, wear, 0.005, color, 0.005, style, 0.004, glass, 0.004, light, 0.004 |
| 72 | Investing | portfolio, 0.038, mutual, 0.031, cap, 0.022, percent, 0.02, etf, 0.019, institut, 0.011, fidel, 0.011, track |
| 73 | Transportation | steel, 0.04, ship, 0.036, transport, 0.028, port, 0.018, train, 0.015, shipment, 0.014, rail, 0.012, pacif |
| 74 | Commodities | gold, 0.096, metal, 0.034, commod, 0.031, ounc, 0.026, silver, 0.025, copper, 0.022, chicago, 0.021, cme, 0.016 |
| 75 | Aviation | airlin, 0.078, air, 0.037, flight, 0.035, travel, 0.034, carrier, 0.025, airport, 0.024, passeng, 0.02, pilot |
| 76 | Canada | canada, 0.059, canadian, 0.051, toronto, 0.016, td, 0.011, surpris, 0.009, jame, 0.008, ben, 0.008, whose, 0.008 |
| 77 | Transactions | fee, 0.06, card, 0.055, access, 0.046, payment, 0.043, compens, 0.033, visit, 0.03, paid, 0.024, kit, 0.018 |
| 78 | Congress | senat, 0.054, committe, 0.041, congress, 0.033, republican, 0.03, legisl, 0.03, vote, 0.03, democrat, 0.03 |
| 79 | Medical equip. | johnson, 0.042, devic, 0.027, boston, 0.025, medic, 0.025, st, 0.017, heart, 0.016, scientif, 0.014, stent, 0.011 |

**Table 6.** First stage instrumental variable regressions. The dependent variable is the forecast revisions of household. The instrument used is the monthly (log) change in the price of oil. The regression includes 334 time series observations. The $R^2 = 0.12$, and the F-stat = 45.80.

|  | Estimate | SE | t-stat | p-value |
|---|---|---|---|---|
| Intercept | -0.02 | 0.02 | -0.85 | 0.40 |
| Change in oil price | 0.02 | 0.00 | 6.77 | 0.00 |

# Appendix C   Endogenous bias

In the model developed in Section 2 we assumed that the relationship between actual inflation, $\pi_t$, and media's coverage of inflation was specified as:

$$\pi_t^N = \pi_t + \alpha_t, \tag{18}$$

where $\alpha_t$ is a time-fixed effect, capturing for example potential media biases (Pfajfar and Santoro (2008), Lamla and Lein (2014)). Importantly, as agents were assumed to only observe the left hand side of (18), they do not know that the news they receive, with noise, do not map one-for-one to actual inflation.

Here we consider a model where we assume that agents are aware of the gap between $\pi_t$ and $\pi_t^N$. In this case the signal extraction problem can be formulated as the following:

$$y_{it} = \boldsymbol{a}\boldsymbol{x_t} + \omega_{it} \tag{19}$$

$$\boldsymbol{x_t} = \boldsymbol{B_t}\boldsymbol{x_{t-1}} + \boldsymbol{\nu_t}, \tag{20}$$

where $\boldsymbol{x_t'} = \begin{pmatrix} \pi_t & \alpha_t \end{pmatrix}$, $\boldsymbol{a'} = \begin{pmatrix} 1 & 1 \end{pmatrix}$, $\boldsymbol{B_t} = \begin{pmatrix} \rho_t & 0 \\ 0 & 1 \end{pmatrix}$, and $\boldsymbol{\nu_t'} = \begin{pmatrix} \nu_t^\pi & \nu_t^\alpha \end{pmatrix}$, and $\boldsymbol{x_t}$ is the unobservable state vector predicted by agents, implying that agents predict both $\pi_t$ and $\alpha_t$.

As before, the variance of the prediction error can be found from the Ricatti equation, which now is $\boldsymbol{\Psi_t} = \boldsymbol{B_t}(\boldsymbol{\Psi_t} - \boldsymbol{\Psi_t}\boldsymbol{a}(\boldsymbol{a'}\boldsymbol{\Psi_t}\boldsymbol{a} + \sigma_{\omega t})^{-1}\boldsymbol{a'}\boldsymbol{\Psi_t})\boldsymbol{B_t'} + \boldsymbol{\Sigma_{\nu_t}}$. The Kalman Gain follows as:

$$\boldsymbol{K_t} = \begin{pmatrix} k_t^\pi & k_t^\alpha \end{pmatrix} = \boldsymbol{B_t}\boldsymbol{\Psi_t}\boldsymbol{a}(\boldsymbol{a'}\boldsymbol{\Psi_t}\boldsymbol{a} + \sigma_{\omega t})^{-1} = \begin{pmatrix} \frac{\rho_t\psi_{11}}{\psi_{11}+\psi_{22}+\sigma_{\omega t}} & \frac{\psi_{22}}{\psi_{11}+\psi_{22}+\sigma_{\omega t}} \end{pmatrix} \tag{21}$$

In this case, and in contrast to under the assumptions in Section 2, the Kalman Gain weight, used to update the inflation forecast ($k_t^\pi$), depends on the persistence of inflation itself, $\rho_t$, and on the amount of noise in the signal, $\sigma_{\omega t}$.

Following the same procedure as in Section 2 we can show that:

$$\pi_{t+h} - F_t\pi_{t+h} = c_t + \beta_t(F_t\pi_{t+h} - F_{t-1}\pi_{t+h}) + e_t \tag{22}$$

where $c_t = \alpha_{t|t-1} - \alpha_t$, $\beta_t = \frac{1-k_t^\pi}{k_t^\pi}$, and $e_t = \sum_{j=1}^{h}(\rho_t)^{h-j}\nu_{t+j}^\pi$. Accordingly, the $\beta_t$ coefficient, measuring the degree of information rigidities, depends on the properties of inflation itself, and not the news as in equation 8. Thus, while the expressions in equations 8 and 22 are the same, the assumptions underlying the reduced form coefficients $c_t$ and $\beta_t$ differ, and we document in Section 4.4 that using the time series properties of actual inflation to describe the evolution of $\beta_t$ gives results at odds with theory.

# Appendix D   Alternative news-topic time series

Results presented in Thorsrud (2018) highlight that tone-adjusted topic frequencies perform much better for nowcasting GDP growth than un-adjusted topic frequencies do. I.e., whether or not the news is positive or not matter. In this paper we have followed suit, and work with time series of news topics that are tone-adjusted, i.e., topic frequencies are adjusted depending on whether the news is positive or negative (see Appendix F.3 for the technical details). Below we show that tone adjustment of the topic frequencies also adds value in the current setting.

In particular, Table 7 presents results for the same post-LASSO regressions as in Table 2 in Section 4.2. However, in contrast to the results presented there, we have used news topic time series that are not tone-adjusted. In the news-only regressions we observe that many news topics are selected, and that the model fit is actually better than the comparable column in Table 2. Still, when we control for the hard economic variables in the *FRED-MD* database, almost all of the news-based variables drop out, and the adjusted $R^2$ remains very similar. This contrasts with the results presented for tone-adjusted news topics in Section 4.2, where controlling for hard economic variables does not significantly change which news topics gets selected by the LASSO algorithm. Accordingly, without tone adjustment, the news topic time series do not seem to capture the same independent media component as the tone-adjusted topic time series do. Moreover, if we run the LASSO regression including both tone-adjusted topic frequencies, un-adjusted topic frequencies, and hard economic indicators, the obtained result turns out to be very similar to the one presented in the *News and Hard* column in Table 2. That is, very few un-adjusted news topic times series are selected.

In sum, these results again highlight the independent role of the media for understanding households' inflation expectation process. The results also speak directly to the claim by Sims (2003) that the tone of economic reporting affect sentiment beyond the economic information contained in the reporting itself. Related empirical work by Doms and Norman (2004), Lamla and Lein (2014), and Pfajfar and Santoro (2008) also show how the tone of the news matter. Note here however, and as described in Appendix F.3, that our

**Table 7.** Post-LASSO regression results and unadjusted news topic times series. The dependent variable is the Michigan Survey of Consumers stated inflation expectations over the next year. In the column labeled *News*, unadjusted news topics are used as the only predictors. In the column labeled *News and Hard* the set of potential predictors is augmented to also include roughly 130 hard economic variables from the McCracken and Ng (2016) database. The table only reports the subset of news variables selected in either of the two regressions. To preserve space, the selected hard economic indicators are not reported. The last raw reports the number of significant variables relative to total number of chosen variables for each model. *, **, and ***, indicate that coefficients are statistically significant at the 10, 5, 1 percent level, respectively. The sample period is 1990:01-2016:12.

| | News | | News and Hard | |
|---|---|---|---|---|
| | $b$ | $PartialR^2$ | $b$ | $PartialR^2$ |
| Monetary policy | 0.41*** (0.12 ) | 0.04 | - | - |
| Agriculture | 0.37*** (0.11 ) | 0.04 | - | - |
| China | -0.28** (0.12 ) | 0.02 | - | - |
| M&A | 0.20** (0.09 ) | 0.02 | 0.36*** (0.06) | 0.11 |
| Clients | 0.37*** (0.08 ) | 0.07 | 0.02 (0.07) | 0.00 |
| Justice | 0.31** (0.12 ) | 0.02 | - | - |
| Health | -0.15 (0.11 ) | 0.01 | -0.22*** (0.06) | 0.04 |
| Russia | -0.21** (0.08 ) | 0.02 | - | - |
| The White House | 0.48*** (0.10 ) | 0.07 | - | - |
| Weapons | 0.28*** (0.07 ) | 0.05 | - | - |
| Volatility | -0.21** (0.10 ) | 0.01 | - | - |
| Automobiles | -0.59*** (0.10 ) | 0.10 | - | - |
| Aircrafts | 0.33** (0.13 ) | 0.02 | - | - |
| Investing | 0.24** (0.11 ) | 0.02 | - | - |
| Commodities | 0.26*** (0.09 ) | 0.03 | - | - |
| Adjusted $R^2$ | 0.63 | | 0.59 | |
| N | 14.00 of 25.00 | | 10.00 of 14.00 | |

tone adjustment procedure explicitly uses the output from the topic model. Thus, in our analysis, it is not the case that we look at only the overall tone, or sentiment, of the news. The topic decomposition matter.

# Appendix E   Survey of Professional Forecasters

Applying our methodology to the Survey of Professional Forecasters (SPF) CPI inflation forecasts serves two purposes. First, although we a-priori conjecture that the media should

**Table 8.** Survey of Professional Forecasters inflation forecast errors and revisions. Each column reports the results of the following regression: $\pi_{t+3,t} - F_t\pi_{t+3,t} = c + \beta(F_t\pi_{t+3,t} - F_{t-1}\pi_{t+3,t}) + \delta z_{t-1} + u_t$. Newey-West corrected standard errors are reported in parenthesis. *, **, and ***, indicate that coefficients are statistically significant at the 10, 5, 1 percent level, respectively. See the text for details about the additional controls $z_{t-1}$.

| | | Additional controls: $z_{t-1}$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | I | Inflation II | Unem-ployment III | Oil price growth IV | T-bill rate V | All (to the left) VI | Factors ×10 VII | Double selection VIII |
| $\beta$ | 1.55*** | 1.54*** | 1.53*** | 1.29*** | 1.48*** | 1.26*** | 0.95*** | 1.29*** |
| | (0.18) | (0.20) | (0.26) | (0.28) | (0.26) | (0.28) | (0.32) | (0.28) |
| $z_{t-1}$ | | -0.05 | -0.01 | 0.02 | 0.65 | - | - | - |
| | | (0.06) | (0.05) | (0.02) | (0.78) | - | - | - |
| $R^2$ | 0.36 | 0.36 | 0.28 | 0.29 | 0.29 | 0.28 | 0.40 | 0.29 |
| N | 142 | 141 | 108 | 108 | 108 | 108 | 108 | 108 |

play a much smaller role for professional forecasters than for households, this experiment is as a good robustness check of our research design. Second, if it is true that the media play less of a role for describing information rigidity among professionals, it should also be true that the difference between information rigidity among households and professionals should be partly explained by the media (because the media has been shown to matter for households). As such, looking at information rigidity among professional forecasters potentially allows us to perform one additional test regarding media's independent role in the expectations formation process among households.

Re-doing our analysis using SPF data, we reach four main conclusions.[24] First, when running the static regression in (9) using the SPF data, we obtain (significant) coefficient estimates that are similar to those reported in the earlier literature, see in particular Coibion and Gorodnichenko (2015a). Moreover, this finding complements existing evidence in the literature because we show that it is robust to doing the more restricted factor augmented and double selection procedures described earlier, see Table 8.

Second, when allowing for time-varying information rigidities, as in (10), we confirm that the degree of information rigidity changes through time, also for the SPF data, see Figure 15.

Third, when re-doing the LASSO regression in Section 4.2 using the SPF data, we

---

[24]The SPF is a quarterly survey, so the number of observations available for estimation becomes lower than for the household regressions. However, in contrast to the Michigan Survey of Consumers, forecasts for multiple horizons, up to one year ahead, are available for SPF. For this reason, we do not need to apply an instrument variable approach, but restrict ourself to $\pi_{t+3,t}$ inflation rates. The news topics data are converted from monthly to quarterly frequency by taking the quarterly mean of the monthly variables.
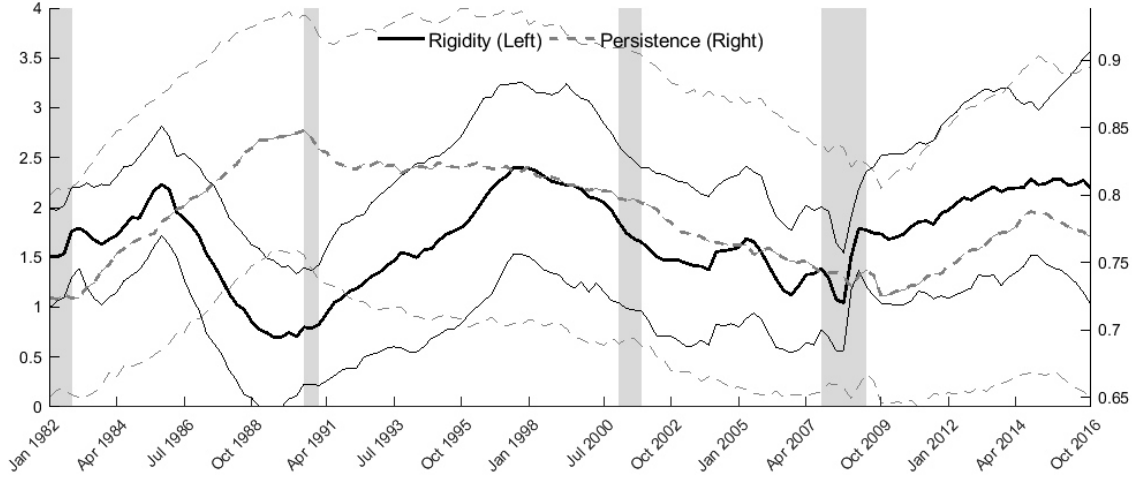
**Figure 15.** Survey of Professional Forecasters and information rigidity. The figure reports the estimated time-varying information rigidity among professionals together with the time-varying persistence in actual CPI inflation. Median estimates together with 68 percent probability bands are reported. The gray shaded areas illustrate recession periods as defined by NBER (U.S.)

actually find that none of the news topics are selected as predictors. Thus, our hypothesis regarding media's role in shaping households' inflation expectations, but not the expectations of professional forecasters, is supported by the data.

Finally, when testing whether the degree of time-varying persistence in actual inflation can explain the evolution of information rigidity among professional forecasters, we confirm the predictions from the theory model in Section 2 (without a media channel). As seen from column $I$ in Figure 16a, there is a negative relationship between the persistence parameter for inflation and the degree of information rigidity among professional forecasters. However, compared to the number of monthly observations available for estimation for the households' data, the number of observation available for the quarterly SPF data is much more restrictive. Therefore, the results also become more uncertain, and the posterior distribution have a substantial mass above zero. Interestingly, however, looking at the sub-samples 1982-2007 and 2008-2016 separately, we observe clear signs of multi-modality in the posterior distribution, with one clear peak both in the positive and negative area. In the context of households' expectations, we discuss this issue in Section 5.

We conclude from this experiment that our research design, estimating time-varying information rigidities and then linking them to the underlying time series properties of the relevant information sets (news or inflation), seems valid.

Turning to the second motivation for looking at the SPF data, we run the following regression:

$$|\beta_t^H - \beta_t^P| = c + \delta_1 \rho_t + \delta_2 \kappa_t + e_t, \tag{23}$$

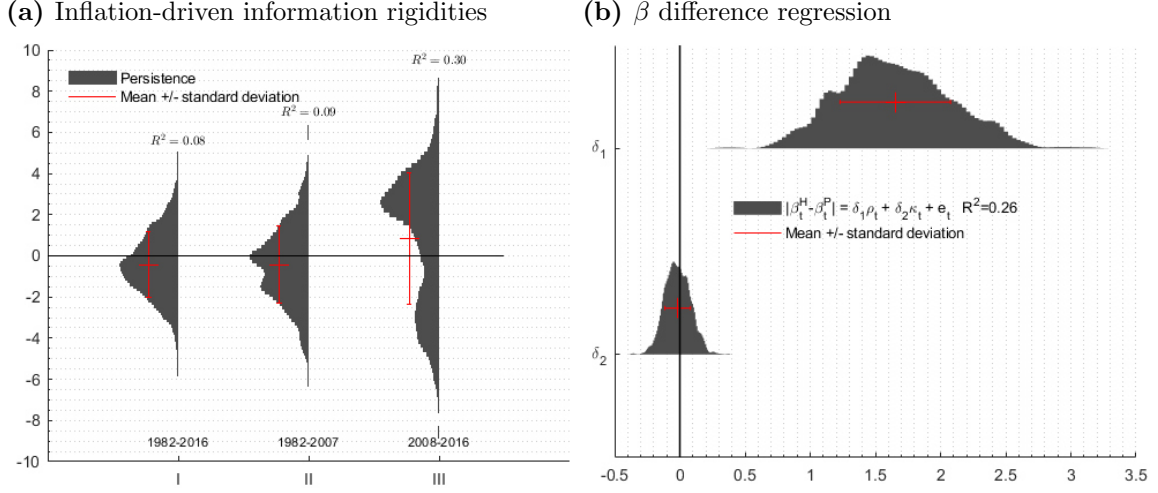where $|\beta_t^H - \beta_t^P|$ is the absolute difference between the time-varying information rigidity

**(a)** Inflation-driven information rigidities

**(b)** $\beta$ difference regression

**Figure 16.** Survey of Professional Forecasters and inflation persistence. Figure 16a reports the posterior distributions for the $\gamma_1$ parameter from equation (15) when $\beta_t$ is the time-varying information rigidity among the Survey of Professional Forecasters, and $\rho_t$ refers to the persistence in actual CPI inflation. Three different estimation samples are considered. Figure 16b reports the posterior estimates from equation 23. See the text and the notes to Figure 4 for additional details.

among households and professional forecasters, and $\rho_t$ and $\kappa_t$ are the news-based persistence and noise-to-signal variables defined earlier.[25] As we have shown that the media matters for households, but not for professional forecasters, the news-based persistence and noise measures should be able to explain the difference between the information rigidity among households and professional. Figure 16b graphs the results, and shows that we reach a positive answer. Although we find that the noise-to-signal parameter is centered around zero, the posterior estimate of the persistence parameter is highly significant. Moreover, the regression fit is relatively good, with an adjusted $R^2$ statistic of 0.26.

# Appendix F  Feature selection and topic time series

## F.1  Cleaning

The share size of the three datasets makes statistical computations challenging. However, as is customary in the Natural Language Processing (NPL) literature, some steps are

---

[25]By regressing the news-based indicators on the difference between households and professionals, equation (23) might look like the testing equation used in the epidemiological model for inflation expectations introduced by Carroll (2003). However, while similar in spirit, it differs fundamentally. Carroll (2003) assumes that households adjust their expectations toward rational professionals when media intensity is high. Thus, he uses the absolute difference in inflation expectations as dependent variable. In contrast, our results suggest that information rigidities are present also among the professionals, but that they cannot be explained by the media. Thus, we use absolute information rigidity differences as dependent variable to test for the role of the media among households.

taken to clean and reduce the raw dataset before estimation (Gentzkow et al. (2017)). First, a stop-word list is employed. This is a list of common words not expected to have any information relating to the subject of an article. Examples of such words are *the*, *is*, *are*, and *this*. In total, the stop-word list together with the list of common surnames and given names removed roughly 1800 unique tokens from the corpus. Next, an algorithm known as stemming is run. The objective of this algorithm is to reduce all words to their respective word stems. A word stem is the part of a word that is common to all of its inflections. An example is the word *effective* whose stem is *effect*. Finally, a measure called *tf-idf*, which stands for term frequency - inverse document frequency, is calculated. This measures how important all the words in the complete corpus are in explaining single articles. The more often a word occurs in an article, the higher the *tf-idf* score of that word. On the other hand, if the word is common to all articles, meaning the word has a high frequency in the whole corpus, the lower that word's *tf-idf* score will be. Around 150 000 of the stems with the highest *tf-idf* score are kept, and used as the final corpus.

## F.2 Topic extraction

The "cleaned", but still unstructured, datasets are decomposed into news topics using a Latent Dirichlet Allocation (LDA) model (Blei et al. (2003)). The LDA model is one of the most popular clustering algorithms in the NPL literature because of its simplicity, and because it has proven to classify text in much the same manner as humans would do (Chang et al. (2009)).

The LDA is an unsupervised topic model that clusters words into topics, which are distributions over words, while at the same time classifying articles as mixtures of topics. A unsupervised learning algorithm is an algorithm that can discover an underlying structure in the data without being given any labeled samples to learn from. The term "latent" is used because the words, which are the observed data, are intended to communicate a latent structure, namely the subject matter (topics) of the article. The term "Dirichlet" is used because the topic mixture is drawn from a conjugate Dirichlet prior.[26]

Different algorithms exist for solving the LDA model. We follow Griffiths and Steyvers (2004), and estimate the model using Gibbs simulations. Technical details and a short description of estimation and prior specifications are described in Appendix G.2. Here we note that we extract $K = 80$ topics from each of the three cleaned datasets. We

---

[26] As such, the LDA shares many features with latent (Gaussian) factor models used in conventional econometrics, but with factors (representing topics) constrained to live in the simplex and fed through a multinomial likelihood at the observation equation. Blei (2012) provides a nice layman introduction to topic modeling. More technical expositions of the LDA approach can be found in Blei et al. (2003) and Griffiths and Steyvers (2004).

subjectively chose $K = 80$ for two reasons. First, this was the choice showing the best statistical results in Larsen and Thorsrud (2018b) and Thorsrud (2018). Second, we have experimented with estimating both fewer and more topics. It is our experience that with $K$ substantially higher than 80, each topic starts to become highly event specific, i.e., there are signs of over-fitting. Conversely, extracting substantially fewer than 80 topics results in too general topics. Thus, in sum, our choice of $K = 80$ is based on a compromise between fitting the corpus well, getting interpretable topics, as well as earlier experience.

## F.3 Topic time series

Given knowledge of the topics (and their distributions), the topic decompositions are translated into tone-adjusted time series. To do this, we proceed in three steps. In short, we first collapse all the articles for a particular day into one document, and then compute, using the estimated word distribution for each topic, the topic frequencies for this newly formed document. See the end of Appendix G.2 for the technical details. This yields a set of $K$ daily time series. Then, for each day and topic, we find the article that is best explained by each topic, and from that identify the tone of the topic, i.e., whether or not the news is positive or negative. This is done using an external word list and simple word counts, similar to in Tetlock (2007). The word list used here classifies positive/negative words as defined by the *Harvard IV-4 Psychological Dictionary*. For each day, the count procedure delivers a statistic containing the normalized difference between positive and negative words associated with a particular article. These statistics are then used to sign-adjust the topic frequencies computed in step one.

Notice from the description above that also the tone adjustment procedure explicitly uses the output from the topic model. Still, the method used for identifying the tone of the news using dictionary based techniques is simple, and could potentially be improved upon with more sophisticated algorithms (Pang et al. (2002)). While leaving such endeavors for future research, Thorsrud (2018) shows that working with topic frequencies without tone adjustment results in a loss of important information.

# Appendix G   Models

## G.1   The Latent Threshold Model

In Section 3.1 of the main paper we describe the Latent Threshold Model (LTM). Here we provide the estimation details. For convenience we first repeat the model, which can

be written in general notation as follows:

$$y_t = \boldsymbol{x}_t' \boldsymbol{b}_t + u_t \qquad\qquad u_t \sim N(0, \sigma_u^2) \qquad\qquad (24a)$$

$$\boldsymbol{b}_t = \boldsymbol{\beta}_t \boldsymbol{\varsigma}_t \qquad\qquad \boldsymbol{\varsigma}_t = I(|\boldsymbol{\beta}_t| \geq \boldsymbol{d}) \qquad\qquad (24b)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\Xi} \boldsymbol{\beta}_{t-1} + \boldsymbol{e}_t \qquad\qquad \boldsymbol{e}_t \sim N(0, \boldsymbol{\Sigma}_e) \qquad\qquad (24c)$$

where $t$ is the time index, $\boldsymbol{x}_t$ is a $(n \times 1)$ vector of variables used for prediction, $\boldsymbol{b}_t$ a $(n \times 1)$ vector of time-varying parameters, and $\boldsymbol{\varsigma}_t$ is a zero one variable, who's value depends on the indicator function $I(|\boldsymbol{\beta}_t| \geq \boldsymbol{d})$. If the $i$th element in $|\boldsymbol{\beta}_t|$ is above the $i$th element in the $(n \times 1)$ threshold vector $\boldsymbol{d}$, then the $i$th element in $\boldsymbol{\varsigma}_t = 1$, otherwise $\boldsymbol{\varsigma}_t = 0$. $\boldsymbol{\Xi}$ equals the identity matrix in our application, but does not need to do so, see the discussion below. Finally, $\boldsymbol{\Sigma}_e$ is a diagonal matrix, and the error terms, $\boldsymbol{e}_t$ and $u_t$, are assumed to be independent. Apart from equation (24b), the system in (24) has a standard state space form.

To simulate from the conditional posterior of $\boldsymbol{\beta}_t$ and $\boldsymbol{d}$ in (24b), we follow the procedure outlined in Nakajima and West (2013). That is, conditional on all the data and hyper-parameters in the model, $\boldsymbol{x}^T = [\boldsymbol{x}_1, ..., \boldsymbol{x}_T]'$, $\boldsymbol{d}$, $\boldsymbol{\Sigma}_e$ and $\sigma_u^2$, we draw the conditional posterior of $\boldsymbol{\beta}_t$ sequentially for $t = 1 : T$ using a Metropolis-Hastings (MH) sampler. As described in Nakajima and West (2013), the MH proposals come from a non-thresholded version of the model specific to each time $t$, as follows. Fixing $\boldsymbol{\varsigma}_t = 1$, takes proposal distribution $N(\boldsymbol{\beta}_t | \boldsymbol{m}_t, \boldsymbol{M}_t)$ where:

$$\boldsymbol{M}_t^{-1} = \sigma_u^{-2} \boldsymbol{x}_t \boldsymbol{x}_t' + \boldsymbol{\Sigma}_e^{-1}(I + \boldsymbol{\Xi}'\boldsymbol{\Xi}) \qquad\qquad (25a)$$

$$\boldsymbol{m}_t = \boldsymbol{M}_t[\sigma_u^{-2} \boldsymbol{x}_t y_t + \boldsymbol{\Sigma}_e^{-1}\{\boldsymbol{\Xi}(\boldsymbol{\beta}_{t-1} + \boldsymbol{\beta}_{t+1}) + (I - 2\boldsymbol{\Xi} + \boldsymbol{\Xi}'\boldsymbol{\Xi})\boldsymbol{\beta}_0\}] \qquad (25b)$$

for $t = 2 : T - 1$. For $t = 1$ and $t = T$, a slight modification is needed. Details can be found in Nakajima and West (2013). The candidate is accepted with probability:

$$\alpha(\boldsymbol{\beta}_t, \boldsymbol{\beta}_t^p) = min\left\{1, \frac{N(y_t|\boldsymbol{x}_t'\boldsymbol{b}_t^p, \sigma_u^2)N(\boldsymbol{\beta}_t|\boldsymbol{m}_t, \boldsymbol{M}_t)}{N(y_t|\boldsymbol{x}_t'\boldsymbol{b}_t, \sigma_u^2)N(\boldsymbol{\beta}_t^p|\boldsymbol{m}_t, \boldsymbol{M}_t)}\right\} \qquad (26)$$

where $\boldsymbol{b}_t = \boldsymbol{\beta}_t \boldsymbol{\varsigma}_t$ is the current state, and $\boldsymbol{b}_t^p = \boldsymbol{\beta}_t^p \boldsymbol{\varsigma}_t^p$ is the candidate.

The independent latent thresholds in $d$ can then be sampled conditional on the data and the hyper-parameters. For this, a direct MH algorithm is employed. Let $\boldsymbol{d}_{-j} = \boldsymbol{d}_{0:s}\backslash \boldsymbol{d}_j$. A candidate is drawn from the current conditional prior, $\boldsymbol{d}_j^p \sim U(0, |\boldsymbol{\beta}_0| + K)$, where $K$ is described below, and accepted with probability:

$$\alpha(\boldsymbol{d}_j, \boldsymbol{d}_j^p) = min\left\{1, \Pi_{t=1}^T \frac{N(y_t|\boldsymbol{x}_t'\boldsymbol{b}_t^p, \sigma_u^2)}{N(y_t|\boldsymbol{x}_t'\boldsymbol{b}_t, \sigma_u^2)}\right\} \qquad (27)$$

where $\boldsymbol{b}_t$ is the state based on the current thresholds $(\boldsymbol{d}_j, \boldsymbol{d}_{-j})$, and $\boldsymbol{b}_t^p$ the candidate based on $(\boldsymbol{d}_j^p, \boldsymbol{d}_{-j})$.

Lastly, conditional on the data, the hyper-parameters and the time-varying parameters, we can sample $\sigma_u^2$ and the diagonal elements in $\boldsymbol{\Sigma}_e$ using the standard inverse Gamma distribution. For each of these elements we use a degrees of freedom prior of 10, and set the prior variance to 0.01.

In essence, the MH steps described above are identical to those described by Nakajima and West (2013). We only differ in the assumptions we make about the processes for the elements in $\boldsymbol{\beta}_t$, which follow stationary processes in their application, but independent Random Walks in our application. In turn, this simplifies inference, since $\boldsymbol{\Xi}$ is an identity matrix. However, this model formulation has consequences for the choice of $K$. The $K$ parameter, used to draw $\boldsymbol{d}_j^p$, controls our prior belief concerning the marginal sparsity probability. A neutral prior will support a range of sparsity values in order to allow the data to inform on relevant values. For stationary processes for $\boldsymbol{\beta}_t$, results in Nakajima and West (2013) suggest that setting $K = 3$ is a reasonable choice. Because of the non-stationarity assumption adopted here, this recommendation can not be followed. We set $K = 1$, which, based on a reference model without latent threshold dynamics, seems to be a reasonable prior. Finally, the prior mean and covariance for the initial states are set to zero and the identity matrix, respectively.

## G.2 The Latent Dirichlet Allocation Model

The Latent Dirichlet Allocation (LDA) is estimated using the algorithm described in Griffiths and Steyvers (2004). Before giving an overview of the procedure, we need to introduce some notation. First, let the corpus consists of $M$ distinct documents, where $N = \sum_{m=1}^{M} N_m$ is the total number of words in all documents, $K$ is the total number of latent topics, and $V$ is the size of the vocabulary. Each document consists of a repeated choice of topics $Z_{m,n}$ and words $W_{m,n}$. Let $t$ be a term in $V$, and denote $P(t|z = k)$, the mixture component, one for each topic, by $\boldsymbol{\Phi} = \{\boldsymbol{\varphi}_k\}_{k=1}^{K}$. Finally, let $P(z|d = m)$ define the topic mixture proportion for document $m$, with one proportion for each document $\Theta = \{\boldsymbol{\theta}_m\}_{m=1}^{M}$. The goal of the algorithm is then to approximate the distribution:

$$P(\boldsymbol{Z}|\boldsymbol{W}; \alpha, \beta) = \frac{P(\boldsymbol{W}, \boldsymbol{Z}; \alpha, \beta)}{P(\boldsymbol{W}; \alpha, \beta)} \tag{28}$$

using Gibbs simulations, where $\alpha$ and $\beta$ are the (hyper) parameters controlling the prior conjugate Dirichlet distributions for $\boldsymbol{\theta}_m$ and $\boldsymbol{\varphi}_k$, respectively. A very good explanation for how this method works is found in Heinrich (2009). The description below provides a brief summary only.

With the above definitions, the total probability of the model can be written as:

$$P(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\Theta}, \boldsymbol{\Phi}; \alpha, \beta) = \prod_{k=1}^{K} P(\boldsymbol{\varphi}_i; \beta) \prod_{m=1}^{M} P(\boldsymbol{\theta}_m; \alpha) \prod_{t=1}^{N} P(z_{m,t}|\boldsymbol{\theta}_m) P(w_{m,t}|\boldsymbol{\varphi}_{z_{m,t}}) \tag{29}$$

Integrating out the parameters $\varphi$ and $\theta$:

$$P(\boldsymbol{Z}, \boldsymbol{W}; \alpha, \beta) = \int_{\boldsymbol{\Theta}} \int_{\boldsymbol{\Phi}} P(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\Theta}, \boldsymbol{\Phi}; \alpha, \beta) \, d\boldsymbol{\Phi} \, d\boldsymbol{\Theta}$$

$$= \int_{\boldsymbol{\Phi}} \prod_{k=1}^{K} P(\boldsymbol{\varphi}_k; \beta) \prod_{m=1}^{M} \prod_{t=1}^{N} P(w_{m,t} | \boldsymbol{\varphi}_{z_{m,t}}) \, d\boldsymbol{\Phi} \int_{\boldsymbol{\Theta}} \prod_{m=1}^{M} P(\boldsymbol{\theta}_m; \alpha) \prod_{t=1}^{N} P(z_{m,t} | \boldsymbol{\theta}_m) \, d\boldsymbol{\Theta} \quad (30)$$

In (30), the terms inside the first integral do not include a $\theta$ term, and the terms inside the second integral do not include a $\varphi$ term. Accordingly, the two terms can be solved separately. Exploiting the properties of the conjugate Dirichlet distribution it can be shown that:

$$\int_{\boldsymbol{\Theta}} \prod_{m=1}^{M} P(\boldsymbol{\theta}_m; \alpha) \prod_{t=1}^{N} P(z_{m,t} | \boldsymbol{\theta}_m) \, d\boldsymbol{\Theta} = \frac{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \frac{\prod_{k=1}^{K} \Gamma(n_m^{(k)} + \alpha_k)}{\Gamma\left(\sum_{k=1}^{K} n_m^{(k)} + \alpha_k\right)} \quad (31)$$

and

$$\int_{\boldsymbol{\Phi}} \prod_{k=1}^{K} P(\boldsymbol{\varphi}_k; \beta) \prod_{m=1}^{M} \prod_{t=1}^{N} P(w_{m,t} | \boldsymbol{\varphi}_{z_{m,t}}) \, d\boldsymbol{\Phi} = \prod_{k=1}^{K} \frac{\Gamma\left(\sum_{t=1}^{V} \beta_t\right)}{\prod_{t=1}^{V} \Gamma(\beta_t)} \frac{\prod_{t=1}^{V} \Gamma(n_k^{(t)} + \beta_t)}{\Gamma\left(\sum_{t=1}^{V} n_k^{(t)} + \beta_t\right)} \quad (32)$$

where $n_m^{(k)}$ denotes the number of word tokens in the $m^{th}$ document assigned to the $k^{th}$ topic, and $n_k^{(t)}$ is the number of times the $t^{th}$ term in the vocabulary has been assigned to the $k^{th}$ topic.

Since $P(\boldsymbol{W}; \alpha, \beta)$, in (28), is invariable for any of Z, the conditional distribution $P(\boldsymbol{Z} | \boldsymbol{W}; \alpha, \beta)$ can be derived from $P(\boldsymbol{W}, \boldsymbol{Z}; \alpha, \beta)$ directly using Gibbs simulation and the conditional probability:

$$P(Z_{(m,n)} \mid \boldsymbol{Z}_{-(m,n)}, \boldsymbol{W}; \alpha, \beta) = \frac{P(Z_{(m,n)}, \boldsymbol{Z}_{-(m,n)}, \boldsymbol{W}; \alpha, \beta)}{P(\boldsymbol{Z}_{-(m,n)}, \boldsymbol{W}; \alpha, \beta)} \quad (33)$$

where $Z_{(m,n)}$ denotes the hidden variable of the $n^{th}$ word token in the $m^{th}$ document, and $Z_{-(m,n)}$ denotes all Zs but $Z_{(m,n)}$. Denoting the index of a word token by $i = (m, n)$, and using the expressions in (31) and (32), cancellation of terms (and some extra manipulations exploiting the properties of the gamma function) yields:

$$P(Z_i = k \mid \boldsymbol{Z}_{-(i)}, \boldsymbol{W}; \alpha, \beta) \propto (n_{m,-i}^{(k)} + \alpha_k) \frac{n_{k,-i}^{(t)} + \beta_t}{\sum_{t=1}^{V} n_{k,-i}^{(t)} + \beta_t} \quad (34)$$

where the counts $n_{\cdot,-i}^{(\cdot)}$ indicate that token $i$ is excluded from the corresponding document or topic. Thus, sampling topic indexes using equation (34) for each word in a document and across documents until convergence allows us to approximate the posterior distribution given by (28). As noted in Heinrich (2009), the procedure itself uses only five larger data structures; the count variables $n_m^{(k)}$ and $n_k^{(t)}$, which have dimension $M \times K$

and $K \times V$, respectively, their row sums $n_m$ and $n_k$, as well as the state variable $z_{m,n}$ with dimension $W$.

With one simulated sample of the posterior distribution for $P(\boldsymbol{Z}|\boldsymbol{W}; \alpha, \beta)$, $\varphi$ and $\theta$ can be estimated from:

$$\hat{\varphi}_{k,t} = \frac{n_k^{(t)} + \beta_t}{\sum_{t=1}^{V} n_k^{(t)} + \beta_r} \tag{35}$$

and

$$\hat{\theta}_{m,k} = \frac{n_m^{(k)} + \alpha_k}{\sum_{k=1}^{K} n_m^{(k)} + \alpha_k} \tag{36}$$

In the analysis of the main paper the average of the estimated $\hat{\theta}$ and $\hat{\varphi}$ from the 10 last samples of the stored Gibbs simulations are used to construct the daily news topic frequencies. Because of lack of identifiability, the estimates of $\hat{\theta}$ and $\hat{\varphi}$ can not be combined across samples for an analysis that relies on the content of specific topics. However, statistics insensitive to permutation of the underlying topics can be computed by aggregating across samples, see Griffiths and Steyvers (2004).

Before estimation three parameters need to be predefined: the number of topics and the two parameter vectors of the Dirichlet priors, $\alpha$ and $\beta$. Here, symmetric Dirichlet priors, with $\alpha$ and $\beta$ each having a single value, are used. In turn, these are defined as a function of the number of topics and unique words:

$$\alpha = \frac{50}{K}, \quad \text{and} \quad \beta = \frac{200}{N}$$

The choice of $K$ is discussed in Section F.2. In general, lower (higher) values for $\alpha$ and $\beta$ will result in more (less) decisive topic associations. The values for the Dirichlet hyper-parameters also reflect a clear compromise between having few topics per document and having few words per topic. In essence, the prior specification used here is the same as the one advocated by Griffiths and Steyvers (2004).

Using the posterior estimates from the LDA model, the frequency with which each topic is represented in the newspaper for a specific day is computed. This is done by first collapsing all the articles in the newspaper for one specific day into one document. Following Heinrich (2009) and Hansen et al. (2018), a procedure for querying documents outside the set on which the LDA is estimated is then implemented. In short, this corresponds to using the same Gibbs simulations as described above, but with the difference that the sampler is run with the estimated parameters $\boldsymbol{\Phi} = \{\varphi_k\}_{k=1}^{K}$ and hyper-parameter $\alpha$ held constant.

Denote by $\tilde{W}$ the vector of words in the newly formed document. Topic assignments, $\tilde{Z}$, for this document can then be estimated by first initializing the algorithm by randomly assigning topics to words and then performing a number of Gibbs iterations using:

$$P(\tilde{Z}_i = k \mid \tilde{\boldsymbol{Z}}_{-(i)}, \tilde{\boldsymbol{W}}; \alpha, \beta) \propto (n_{\tilde{m}, -i}^{(k)} + \alpha_k)\hat{\varphi}_{k,t} \tag{37}$$

Since $\hat{\varphi}_{k,t}$ does not need to be estimated when sampling from (37), fewer iterations are needed to form the topic assignment index for the new document than when learning both the topic and word distributions. Here 2000 iterations are performed, and only the average of every 10th draw is used for the final inference. After sampling, the topic distribution can be estimated as before:

$$\hat{\tilde{\theta}}_{\tilde{m},k} = \frac{n_{\tilde{m}}^{(k)} + \alpha_k}{\sum_{k=1}^{K} n_{\tilde{m}}^{(k)} + \alpha_k} \tag{38}$$

## G.3 Time-varying Autoregressive models

To model the time-varying Autoregressive processes we follow the general setup used in Primiceri (2005), where the parameters follow independent random walks. Letting $y_t$ denote the dependent variable, the model structure can be written as:

$$y_t = \rho_t y_{t-1} + \sigma_t w_t \qquad\qquad w_t \sim i.i.d. N(0,1) \tag{39a}$$

$$\rho_t = \rho_{t-1} + v_t^\rho \qquad\qquad v_t^\rho \sim i.i.d. N(0, \sigma_\rho^2) \tag{39b}$$

$$log(\sigma_t) = log(\sigma_{t-1}) + v_t^\sigma \qquad\qquad v_t^\sigma \sim i.i.d. N(0, \sigma_\sigma^2) \tag{39c}$$

This system contains two time-varying state variables, $\rho_t$ and $log(\sigma_t)$, and two hyper-parameters $\sigma_\rho^2$ and $\sigma_\sigma^2$. The system is estimated using Carter and Kohn's multimove Gibbs sampling approach (Carter and Kohn (1994)) together with the mixture proposal by Kim et al. (1998) to estimate the stochastic volatility part of the model.

We specify the initial values $\rho_0 \sim N(0,1)$, $y_0 \sim N(0,1)$, and $log(\sigma_0) \sim N(0,1)$. The priors for the hyper-parameters are all from the inverse-Wishart distribution:

$$\underline{\sigma}_\rho^2 \sim IW(\underline{T}^{\sigma_\rho^2}, \underline{T}^{\sigma_\rho^2} \cdot \kappa_{\sigma_\rho^2}^2) \quad \underline{T}^{\sigma_\rho^2} = 10, \kappa_{\sigma_\rho^2} = 0.01$$

$$\underline{\sigma}_\sigma^2 \sim IW(\underline{T}^{\sigma_\sigma^2}, \underline{T}^{\sigma_\sigma^2} \cdot \kappa_{\sigma_\sigma^2}^2) \quad \underline{T}^{\sigma_\sigma^2} = 10, \kappa_{\sigma_\sigma^2} = 0.03$$

where the first element in each prior distribution is the degrees of freedom parameter, and the second the scale parameter. We note that for the inverse-Wishart distribution the prior scale matrix has the interpretation of the prior sum of squared residuals. Therefore, each scale matrix is multiplied by the degrees of freedom parameter. Also, for the inverse-Wishart prior to be proper, the degrees of freedom parameter must be larger than the dimension of the scale matrix. This is the case in all our prior specifications.

To implement the simulations, we break the problem into three blocks, and follow the now standard method first proposed in Primiceri (2005). In short, letting $\boldsymbol{\rho^T} = [\rho_1, ..., \rho_T]'$, $\boldsymbol{\sigma^T} = [\sigma_1, ..., \sigma_T]'$, and $\boldsymbol{y^T} = [y_1, ..., y_T]'$, we first sample:

$$p(\boldsymbol{\rho^T}|\boldsymbol{y^T}, \boldsymbol{\sigma^T}, \sigma_\rho^2) = p(\rho_T|\boldsymbol{y^T}, \boldsymbol{\sigma^T}, \sigma_\rho^2) \prod_{t=1}^{T-1} p(\rho_t|\boldsymbol{y^T}, \boldsymbol{\sigma^T}, \sigma_\rho^2). \tag{41}$$

Since (39), conditional on $\boldsymbol{y^T}$, $\boldsymbol{\sigma^T}$, and $\sigma_\rho^2$, constitute a conditional linear state space system, $\boldsymbol{\rho^T}$ can be drawn using Carter and Kohn's multimove Gibbs sampling approach (Carter and Kohn (1994)).

Second, conditional on $\boldsymbol{\rho^T}$, we can construct the following transformation:

$$y_t - \rho_t y_{t-1} = y_t^* = \sigma_t w_t, \tag{42}$$

which, after squaring and taking logarithms, can be written as:

$$log((y_t^*)^2) = 2log(\sigma_t) + e_t, \tag{43}$$

where $e_t = log(w_t^2)$. Now, equation (43) together with the law-of-motion for $log(\sigma_t)$ in (39), is a new state space system, albeit non Gaussian. Accordingly, we use the mixture approximation proposed in Kim et al. (1998), together with Carter and Kohn's multimove Gibbs sampling approach, to sample $\boldsymbol{\sigma^T}$.

Finally, conditional on $\boldsymbol{y^T}$, $\boldsymbol{\sigma^T}$ and $\boldsymbol{\rho^T}$, the hyper-parameters $\sigma_\sigma^2$ and $\sigma_\rho^2$ are drawn from the inverse-Wishart distribution.