Attention Cycles

Joel P. Flynn*  Karthik A. Sastry†
MIT  MIT

May 16, 2020

Abstract

We document that attention to macroeconomic conditions is highly counter-cyclical using a natural-language-processing method that analyzes the text of US public firms’ descriptions of key risks in regulatory filings. We develop a theoretical framework that rationalizes this phenomenon as an optimal stochastic choice pattern in response to higher stakes for tracking macroeconomic developments in recessions, during which the cost of mistakes is high. In general equilibrium, elevated stakes in recessions generate attention cycles in which attention to fundamentals and endogenous volatility form a positive feedback loop. In a calibrated, one-parameter extension of the real-business-cycle model with stochastic choice, attention cycles generate asymmetrically large amplification of negative shocks, greater amplification of shocks when output is low, and endogenous stochastic volatility of output growth.

*MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. Email: jpflynn@mit.edu.
†MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. Email: ksastry@mit.edu.

We thank Daron Acemoglu, George-Marios Angeletos, Ricardo Caballero, Oguzhan Celebi, Roberto Corrao, Michele Fornino, Drew Fudenberg, Chen Lian, Stephen Morris, Jacob Moscona, Georgios Nikolakoudis, Lawrence Schmidt, Alp Simsek, John Sturm, Adrien Verdelhan, Ivan Werning, Alexander Wolitzky and participants in the MIT Macro Lunch, MIT Theory Lunch, and MIT Sloan Finance Lunch for helpful comments. First draft: May 4, 2020.
1. Introduction

How much are people thinking about the economy? Figure 1 plots the simplest time-series metric thereof, the relative frequency of use for the word “economy,” from three different domains. The first is US Google searches (blue); the second is the full text of quarterly and annual reports of financial performance (10-Q/K) submitted by public firms to the Securities and Exchange Commission (orange); and the third is the full text of US public firms’ quarterly earnings conference calls (green). All three series strikingly jump in the Great Recession, peaking at values that are respectively 2.2, 3.0, and 5.1 times their pre-Recession average. The regulatory filings, which reach furthest back into history, uncover a similar but smaller spike around the dot-com crash, subsequent recession, and the September 11 terrorist attack. And the Google searches, which we can observe with higher frequency and in real-time, are beginning to reach Great Recession levels in the ongoing COVID-19 Pandemic.

These patterns beg for a place in our understanding of economic fluctuations as psychologically driven phenomena, in which reduced activity both shapes and is shaped by individuals’ thinking. The clear spikes in Figure 1 are all, at least anecdotally, times when even “regular conversation” is unusually tilted toward macroeconomic topics, which are strikingly less salient in normal times. This behavior is inconsistent with the assumed behavior of homo economicus in standard full- or constant-attention macroeconomic models, wherein attention cannot vary over the business cycle. Yet if scrutiny toward the macroeconomy can by itself increase agents’ responsiveness to shocks, it is natural to consider the extent to which a self-fulfilling attention cycle can be both a consequence and a cause of the business cycle itself.

This article investigates such a possibility. We begin by systematizing the insights of Figure 1 in a more comprehensive measure of firms’ emphasis on macroeconomic risk in their language. We focus on regulatory filings (SEC Forms 10-Q and 10-K) from the universe of publicly traded US firms from 1998-2017 as a primary source of data on firms’ risk perception. Using a natural-language-processing technique where we compare the words used in these filings to introductory macroeconomics textbooks, we tabulate a metric of each filing’s informativeness about macro developments. We then aggregate these measures into industry-level and economy-wide time-series. This reveals two stylized facts. First, the unemployment rate significantly predicts aggregate attention to the macroeconomy: attention

---

1 Each of the latter two data sources will be described in considerably more detail in Section 2. The Google series is from Google Trends, accessed in April 2020.

2 Our methods are closely related to the pioneering work of Hassan, Hollander, van Lent, and Tahoun (2019) on quantifying firm-level political risk. We discuss our methods in more detail in Section 2.
is counter-cyclical.\footnote{We show that our index features only low correlation with the news and uncertainty measures of Baker, Bloom, and Davis (2016) and the VIX. Thus, firms’ own enumeration of key risks tilts toward the macroeconomy not merely in the times of greatest uncertainty or deviation from the trend, but instead in times of low activity.} Second, attention to the macroeconomy is persistent, approximately to the same or slightly greater extent than aggregate activity, and to a strikingly greater extent than financial-market based risk indices (e.g., VIX) or policy-focused indices (e.g., that constructed by Baker et al., 2016).

Why do attention cycles occur, and how could they matter for the behavior of macroeconomic aggregates? We answer both questions in a Neoclassical market economy with an aggregate demand externality and a \textit{stochastic choice friction}. At the core of the model are firms who choose their total production or capacity as an imperfect function of the macroeconomic state, aggregate productivity. Firms face a cognitive cost in adapting their action toward the state, or precisely responding to shocks.\footnote{The theory for this type of state-dependent stochastic choice is discussed in a more decision-theoretic context by Fudenberg, Iijima, and Strzalecki (2015) and analyzed in games, at different levels of generality, by McKelvey and Palfrey (1995), Morris and Yang (2016), and Flynn and Sastry (2020).}

The core insight from our theory in \textit{partial equilibrium} is that firms facing the cognitive cost to tune their action to the macro state will differentially do so when the cost of making mistakes is high. Formally, the correct metric for such costs is the curvature of agents’ objective functions in their own action. In a Neoclassical theory of the firm, in which firms choose production to maximize risk-adjusted profits, the aforementioned curvature is the

---

\textbf{Figure 1.} \textit{Mentions of “Economy” over Time.}

The y-axis unit is normalized frequency, with each series maximized at 100. The series are respectively maximized in October 2008, Q1 2010, and Q1 2009.
product of two terms: the curvature of firms’ dollar profits, which is highest when output is low in standard parametrizations; and the stochastic discount factor (SDF), which is highest when consumption is lowest owing to the ownership of firms by risk-averse households. This observation provides a parsimonious explanation for attention cycles, with economic driving forces that are robust across a wide array of macroeconomic models as long as they embed one or both of the previous forces.

In equilibrium, our model features counter-cyclical attention when wages are sufficiently rigid and risk aversion is sufficiently high. Flexible and pro-cyclical wages provide a countervailing force of reducing the curvature of profits in downturns that pushes against the demand and SDF forces. But they are unlikely to reverse the attention cycle in reasonable business-cycle calibrations, when pitted against the cyclicity of the risk adjustment.\(^5\)

Instead, when the positive externality of increased demand dominates the negative externality via wage pressure, attention cycles are stronger in general equilibrium than in partial equilibrium. Firms that pay more attention to the macroeconomy respond more to shocks. As a result, aggregate production falls by more in response to a given negative shock when firms pay more attention. As output is then lower, firms have yet greater incentive to track the underlying state closely, pay even more attention, and push down aggregate production further—leading to a partially self-fulfilling attention cycle.

We next illustrate the consequences of attention cycles in our model in closed-form. First, we show that output is the product of the frictionless benchmark with perfect attention and an “attention wedge” that penalizes output. This is because inattention generates erratic choices and hence dispersion in production and the value marginal product of labor across firms. The last translates directly into a reduction of aggregate labor productivity—the Solow residual in this one-factor economy. Owing to the attention cycle, the aforementioned wedge increases in size when productivity is higher, as firms become more lax in their optimization.\(^6\) As a result, misallocation across firms in the model is endogenously higher in booms than recessions, providing an attentive analogue of the macroeconomic literature on the cleansing effect of recessions (Caballero and Hammour, 1994).

From this perspective, our theory of attention cycles flips on its head the durable hypothesis of Pigou (1927) that recessions are the times of especially great errors in forecasting and

\(^5\)Households’ stochastic discount factors are extremely cyclical, as demonstrated by Hansen and Jaganathan (1991). Wages, meanwhile, are known to be highly acyclical over the business cycle (e.g. Solon, Barsky, and Parker, 1994; Grigsby, Hurst, and Yildirimaz, 2019).

\(^6\)The mechanics that map (in)attention to productivity in the model, or the pathway from cross-firm dispersion to aggregate productivity losses, greatly resembles the mechanism of Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). See also Restuccia and Rogerson (2017) for a review of the relevant literature with some emphasis on the implications for comparative economic development rather than advanced-economy business cycles.
judgement but maintains the character of “psychological causes” of the business cycle. In our model, negative shocks crowd in careful thinking about the macroeconomy, which may have been an afterthought in better times—and it is this relatively more precise thinking that generates higher aggregate volatility. Importantly, this mechanism is directional: the equal and opposite will not occur for positive shocks to the system, which instead will be mitigated by their encouraging firms to stop paying attention to economic developments. Taken together, attention cycles in our model generate a host of macro phenomena related to asymmetry; state-dependent shock propagation, stochastic volatility of output and asymmetric shock propagation all emerge owing to the attention cycle. These features emerge as purely endogenous outcomes without adding a “driving process” (e.g., sequence of productivity shocks) with such features.

The remainder of the paper pivots toward assessing the empirical plausibility of our specific theory for attention cycles and a quantification of their possible importance. On the first front we have two strategies. The first is to exploit our rich firm-and-industry-level panel data on attention levels. We document that attention cycles, in accordance with the theory, are more pronounced for industries with highly cyclical demand (e.g., Retail), but present also for acyclical and counter-cyclical industries (e.g., Educational Services and Agriculture). The second strategy leverages the unique firm-level survey data from Coibion, Gorodnichenko, and Kumar (2018) and shows that firms that directly report having more sensitive profits to firm-level decisions also track the macroeconomy much more closely.

Our final section quantifies the macro impact of attention cycles in a calibrated version of our Neoclassical model with flexible wages. We estimate parameters, including the extent of the attention friction, to match the time-series properties of our measured macro attention index. This involves a fundamental leap of faith that time-series patterns in “what firms say” translate quite literally into “what firms do.” But it allows for a powerful thought experiment of letting only our novel data on macro attention guide the most interesting predictions of our model, which can themselves be met with out-of-sample evidence.

Attention cycles have a quantitatively important impact on macroeconomic dynamics in our model. We find in the quantitative model the same patterns suggested by the theory: attention is highest in the states were productivity is the lowest, and more than 87% of this owes to general equilibrium feedback. The model generates quantitatively large asymmetry of shock propagation, state-dependent shock propagation and stochastic volatility. In particular, the shock within our model that replicates the Great Recession’s 4.2% peak-to-trough reduction in output generates also a 153% increase in the conditional volatility of output purely from the endogenous force of shifting attention. Consequently, according to state-of-the-art evidence from Jurado, Ludvigson, and Ng (2015), our model accounts for
the entirety of the time-series stochastic volatility measured in output, a moment not targeted in our calibration. Importantly, none of these features are present in the underlying RBC model: all shocks have a symmetric and state-independent effect on aggregates and the model features no stochastic volatility.

Related Literature This article lies at the intersection of several literatures. On the purely decision-theoretic front, we contribute to a large literature developing models of optimal limited attention. Sims (2003), Gabaix (2014) and the literature on rational inattention emphasize the scope for attention to be tuned to more relevant states and/or attributes of the world.\(^7\)\(^8\) We take more seriously the observable equilibrium implications of these ideas. A literature embedding information acquisition and/or behavioral inattention into macro models (e.g., Woodford, 2003a; Lorenzoni, 2009; Mackowiak and Wiederholt, 2009; Gabaix, 2016) has ignored both state-dependence of attention and its equilibrium consequences. Studies of costly adjustment of prices (e.g., Gorodnichenko, 2008; Alvarez, Lippi, and Paciello, 2011) or infrequent and costly adjustment of investment portfolios (e.g., Abel, Eberly, and Panageas, 2013; Kacperczyk, Van Nieuwerburgh, and Veldkamp, 2016) draw on intuitively similar mechanisms operating through the curvature of payoffs, but consider only partial equilibrium mechanisms.\(^9\)

The following papers are much closer to ours. Mäkinen and Ohl (2015) and Benhabib, Liu, and Wang (2016) come the closest to our focus on firms’ decisions to learn over the business cycle, but localize their findings to very specific models and do not provide direct empirical or quantitative evidence. Nimark (2014) introduces time-varying “newsworthiness” of shocks in a model with normal-mixture fundamentals, and links to macro but not micro evidence. In Caplin and Leahy (1994), reactions to negative shocks release information that intensifies the crash—a similar story to ours that relies on a theoretically very different mechanism.

Zeira (1987, 1994), Rob (1991), and Caplin and Leahy (1993) highlight a different information externality generated by inaction—when no one produces or invests, it is harder to learn about the state of the economy. A modern literature including Van Nieuwerburgh and Veldkamp (2006), Fajgelbaum, Schaal, and Taschereau-Dumouchel (2017), and Straub

---

\(^7\)The former notes, in its discussion of unrestricted information acquisition in the consumption-savings problem of a log utility agent, that information becomes arbitrarily precise for low values of consumption because utility losses from misoptimization limit to infinity.

\(^8\)See Mackowiak, Matejka, and Wiederholt (2018) and Gabaix (2019) for review articles.

\(^9\)Studies by Berger and Vavra (2019) and Alvarez, Beraja, Gonzalez-Rozada, and Neumeyer (2019) on state-dependent responsiveness to shocks in price-setting, framed in the context of the aforementioned menu-cost theory, are also related in spirit to the points raised in this article.
and Ulbricht (2017) returns to this topic with new quantification. At a high level, these theories find equilibrium amplification via constraints on “information supply” rather than “information demand.” Our paper is concerned entirely with the latter force, except for a brief joint analysis in Section 5. A rigorous combination of the two is an interesting avenue for future research.

Finally, an emerging literature studies how firms “think” over the business cycle (Coibion et al., 2018; Altig, Barrero, Bloom, Davis, Meyer, and Parker, 2019; Bachmann and Elstner, 2015). We complement these efforts with novel measurement using publicly available data, taking inspiration from other work using regulatory filings (e.g., Loughran and McDonald, 2011) and firm-leadership conference calls (Hassan et al., 2019) to learn about the risks that firms identify as most important. This is complementary to other efforts in the literature to study identified risks in the media (Baker et al., 2016).

Outline The rest of the paper proceeds as follows. In Section 2, we develop our index of macro attention and show that it is both counter-cyclical and persistent. In Section 3, we augment a Neoclassical RBC model with a stochastic choice friction to understand how attention cycles manifest as both a cause and a consequence of the business cycle. In Section 4, we test the predictions of the theory to evaluate its performance in explaining attention cycles. In Section 5, we calibrate a one-parameter extension of the RBC model with stochastic choice and analyze the impact of attention cycles on macroeconomic dynamics. Section 6 concludes.

2. Attention to the Macroeconomy is Counter-cyclical

We first provide quantitative evidence for the premise of our investigation—that firms’ attention to macroeconomic events is heightened during economic downturns. To do so, we develop a novel index of macro attention at the quarterly frequency in the United States since 1998. The index is based on identifying in public firms’ Forms 10-Q and 10-K the frequency of macro-salient words, which are themselves identified by comparing the corpus of 10-Qs and 10-Ks with macro reference texts (undergraduate textbooks). This builds upon the simple “word-counting” exercise of Figure 1 by both automating the choice of relevant words and providing a weighting scheme used widely in the natural-language-processing literature. Using this measure, we find strong evidence that macro attention is persistent and counter-cyclical, which will motivate our subsequent theoretical analysis.
2.1. Data Sources

Main Sources of Text: 10-K and 10-Q filings Our corpus of text that captures “what firms are thinking” are the quarterly 10-Q and annual 10-K reports that all publicly traded companies in the US submit to the Securities and Exchange Commission (SEC). While not directly designed as direct communication with shareholders, forms 10-Q and 10-K are written knowing that their contents will become public. We use data running from 1998 to 2017 in our main analysis, covering all public firms that appear in Compustat over that time period. Our total sample consists of 479,403 individual documents, or about 6,000 per quarter. Anecdotally, in the modern era, data mining 10-Q and 10-K filings for useful information and/or indicators of firm sentiment is a common practice among equity analysts.

Where does information about the macroeconomy enter an SEC filing? Form 10-K’s generally follow a uniform structure separated into 4 parts and 15 sub-parts (items). Among these, the most relevant is Item 1 which outlines the business of the firm and the risk factors it perceives in the current environment. In other contexts, firms may use the same space to highlight more idiosyncratic and/or industry-specific trends. The form 10-Q, although less uniformly structured, also contains lines for describing prominent company-specific risks as well as (i) adding unstructured notes to the financial statement which offer (possibly macroeconomic) explanations and (ii) adding “other exhibits” of information which may include other shareholder communication that transpired during the quarter.

We index time on the form 10-Q and 10-K using the filing date, not the quarter to which the report necessarily applies. This is a decision for comparability across firms that are reporting in different fiscal calendars. It also means that we are measuring attention at the moment of writing or speaking and not in the context of what time period is being discussed.

Alternative Sources of Text: Firm Conference Calls As an alternative source of text describing “what firms are thinking,” we use the full text of US Public Firms’ sales and earnings conference calls, as recorded by the Fair Disclosure (FD) News Wire. These

---

To give an example, the automaker General Motors in its summary report for 2009 extensively highlighted the threats of the ongoing Great Recession to its line of business:

Our business is highly dependent on sales volume. Global vehicle sales have declined significantly from their peak levels and there is no assurance that the global automobile market will recover in the near future or that it will not suffer a significant further downturn. [...] The deteriorating economic and market conditions that have driven the drop in vehicle sales, including declines in real estate and equity values, rising unemployment, tightened credit markets, depressed consumer confidence and weak housing markets, may not improve significantly during 2010 and may continue past 2010 and could deteriorate further.

Of our sample, about 1/4 (exactly 22.7%) are the form 10-K. The form 10-K is filed at the end of the fiscal year which may not align with the calendar year for all companies. In our sample, 68% of 10-Ks are filed in calendar quarter 1; 14% in calendar quarter 2; and 9% each in calendar quarters 3 and 4.
are the same data analyzed by Hassan et al. (2019) in their study of political risks. We collect our sample by first scraping all calls and then restricting to firms that are listed on a US Stock Exchange (the NYSE, NASDAQ, or NYSE American/MKT). Our conference call sample is 280,074 documents covering a smaller time period from 2004 to 2013. Consistent with the 10Q/Ks, we also index conference calls by the time of the call and not the quarter to which the discussion applies. Appendix D, and especially subsection D.1, contain much more information about these data and how they are analyzed.

Macro Textbooks We use three common beginner and intermediate textbooks for undergraduate macroeconomics: *Macroeconomics and Principles of Macroeconomics* by N. Gregory Mankiw and *Macroeconomics: Principles and Policy* by William J. Baumol and Alan S. Blinder. Hassan et al. (2019) analogously treat undergraduate textbooks (in their case, of government and political science) as an appropriate stand-in for advanced, but not highly technical mastery of the field of interest.

### 2.2. Identifying the Macro Informativeness of Firm Communication

**Notation and formulae** Let us denote by $\mathcal{D}_t$ the set of 10K or 10Q documents reported to the SEC in a given quarter $t$, and $\mathcal{D} := \bigcup_t \mathcal{D}_t$ the set of all such documents over the entire sample. We will think of these sets consisting of individual documents $d$ which are a vector of words $w$ (i.e., the full text stripped of punctuation and structure). Let us analogously denote the set of textbooks as $\mathcal{B}$ consisting of individual books $b$.

At the level of individual documents or books, we define the *term-frequency* as the fraction of the entire document consisting of the term of interest:

$$
tf(w, d) := \frac{1}{|d|} \sum_{w' \in d} \mathbb{1}\{w' = w\} \tag{1}
$$

Next, for a given corpus, we will define the *document-frequency* as the fraction of documents that contain a given term $w$:

$$
df(w, \mathcal{D}) := \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \mathbb{1}\{w \in d\} \tag{2}
$$

Let us finally introduce a combination metric of the previous two: tf-idf, or *term-frequency-inverse-document-frequency*. This is a standard metric in natural-language-processing

---

12 We use the 7th, 3rd, and 12th editions of these books, respectively.
13 For instance, the publisher’s website describes the current edition of Baumol and Blinder’s text as “a ‘robust policy-based approach to teaching introductory economics’ with ‘vivid examples’ relating to ‘contemporary economic problems and policy debates specific to the U.S.’”
that measures whether a word is frequent in a given document, relative to its frequency in the entire corpus:

\[
\text{tf-idf}(w, d, \mathcal{D}) := \text{tf}(w, d) \cdot \log \left( \frac{1}{\text{df}(w, \mathcal{D})} \right)
\]  

(3)

The log functional form for the second term can be thought of either as an heuristic for scaling the relative importance of the two terms, or a way to formalize the connection of tf-idf to information theory (Aizawa, 2003). For the former interpretation, note that the log inverse document frequency is bounded below by 0 when the document frequency is 1. Concretely, if we wanted to know whether documents share the information embedded in the phrase

The quick brown fox jumped over the lazy dog.

we would learn very little from the presence of “the” (log idf \(\approx\) log(1) = 0), but would learn much more from the presence of “fox” (log idf \(\gg\) 0). The exact shape of the log(\(\cdot\)) weighting is interpreted as a heuristic for how to scale less frequent words.

In the information-theoretic interpretation, the average tf-idf across a set of terms \(W\) and corpus of documents \(\mathcal{D}\) is in units of bits of information provided by term by document pairs—that is, how informative is a given document, or are documents on average, about a given set of words \(W\)? This connection is established formally by Aizawa (2003) and will later enable us to connect our measure in the data to the theory when we calibrate the model.

**Identifying macro words** The first step in our methodology for measuring document-level, and later aggregate, macro attention is to identify a set of words \(W_M\) that correspond to the macroeconomy. To do this, we identify words with a high tf-idf score for macroeconomics textbooks when compared against the corpus of 10-Ks and 10-Qs.

We first take all terms \(w\) that appear in a given macro textbook \(b\) and score them by tf-idf\((w, b, \mathcal{D})\). This gives the term frequency in the book scaled by the inverse document frequency in the 10-K/Qs, yielding the words that would distinguish \(b\) (e.g., Mankiw’s Principles) were it an SEC filing. We take the top 200 such words in each of the three textbooks used in the analysis. We then take the intersection to form a list of macro-relevant words. The intersection helps guard against the idiosyncratic choices of certain books. This procedure allows us to identify words that constitute emphasis on the macroeconomy and not merely the myriad financial words that will habitually be used in SEC filings (e.g., “credit” and “interest”).

---

14 For instance, in Mankiw’s textbooks, a parable about supply and demand for “ice cream” is used often enough to make “ice” and “cream” high tf-idf words in our procedure.

15 Appendix Figure 11 prints these words in alphabetical order, along with the time-series properties
The measure: With our list of words $W_M$, we next calculate the document-level tf-idf, summing over terms. Letting $f$ denote a uniquely identified firm, and $d(f,t)$ denote a mapping from firms and quarters to documents, we use this to calculate a firm-level panel of “macro attention”:

$$\text{MacroAttention}(f,t) := \sum_{w \in W_M} \text{tf-idf}(w, d(f,t), D)$$  \hspace{1cm} (4)

To calculate the same statistic in the time-series, we take a simple average across firms or documents. Letting $F_t$ denote the set of firms for which we have documents in a period $t$, or $F_t := \{ f : d(f,t) \in D_t \}$, we calculate:

$$\text{MacroAttention}(t) := \frac{1}{|F_t|} \sum_{f \in F_t} \text{MacroAttention}(f,t)$$  \hspace{1cm} (5)

We calculate also a “seasonally-adjusted” version of the previous that subtracts the average value in that quarter of the year over the whole sample, which we denote by $\text{MacroAttentionA}(t)$.

This removes seasonal fluctuations and, more importantly, the fact that forms 10-K contain disproportionately more macro information.

Note finally that, as constructed, it is straightforward to calculate versions of the previous that use only subsets of the word dictionary, by replacing $W_M$ in Equation 4, or subsets of the firms (e.g., in a particular industry), by replacing $F_t$ in Equation 5.

2.3. Macro Attention in the Business Cycle

Figure 2 plots our aggregate measure in the time-series with adjustment for quarterly effects. The units can be interpreted as word proportions or relative word counts. From peak to

---

Footnotes:

16 Formally, if $q(t) \in \{1,2,3,4\}$ returns the quarter of the year, we first define quarterly averages $(x(q))_{q=1}^4$:

$$x(q) := \frac{\sum_t \text{MacroAttention}(t) \cdot I\{q(t) = q\}}{\sum_t I\{q(t) = q\}}$$  \hspace{1cm} (6)

and define an “Adjusted” metric as

$$\text{MacroAttentionA}(t) := \text{MacroAttention}(t) - x(q(t)) + x(q(1))$$  \hspace{1cm} (7)

i.e., one that partials out quarterly-frequency fluctuations and adds back the trend of Q1.
trough, the percentage of 10-K and 10-Q forms that are related to the selected macro topics increases about 10 basis points. In percentage terms, this is about a 25% increase in the level of attention from pre-Recession levels.

An obvious pattern in Figure 2 is the enormous spike in macroeconomic attention during the Great Recession, coupled with a much smaller spike in the 2001 Recession and immediate aftermath. Appendix Figure 10 shows the scatterplot of our attention measure against US unemployment. In an OLS regression, the two variables are positively associated with an $R^2$ of 18.3% and a $t$-statistic (with robust standard errors) of 4.17 over 80 quarter-year observations. This relationship fits very heavily on the Great Recession, which is the most prominent macro event in the sample. We summarize this fact as the following:

**Fact 1.** Firms’ attention to the macroeconomy is counter-cyclical.

Another salient feature of Figure 2 is the **persistence** of macro attention well after the peak of the Great Recession. Unemployment reaches approximately pre-recession levels by 2015, though macro attention remains considerably elevated. This yields our second stylized fact:

**Fact 2.** Firms’ attention to the macroeconomy is persistent.

This is in sharp contrast to financial-market measures of volatility like the VIX, which spike more strongly around specific events, and also suggests a theory built around correlating attention with activity rather than specific shocks.
Robustness: Conference Calls Appendix D discusses how we extend our method to measure macro attention using our secondary data source, sales and earnings conference calls. These data cover a smaller time period but uncover, on the shared sample, very similar patterns. In particular, Appendix Figure 8 shows that the conference call based metric is also counter-cyclical, though not perfectly correlated with the 10-Q/K method. We take this as evidence that our broad interpretation (Facts 1 and 2) is consistent across data sources.

2.4. Individual Words

Our MacroAttention index can be described as a weighted sum of individual indices corresponding to individual words. What do these single-word indices look like? Figure 3 shows the time-series plot for six selected words of interest, without seasonal adjustment.

Some patterns are worth noting. First, using only mentions of “unemployment,” we see evidence (albeit slightly more muted) of the aforementioned pattern of excess persistence. Unemployment attention (top left) does not recover to pre-Great-Recession levels even when...
unemployment itself does. Such intense interest in the labor market is both more persistent than direct interest in “Recession[s]” (bottom left), and more typical of the Great Recession than the 2001 recession, which had a muted labor market impact relative to its financial and output impacts.

Second, our methods give broadly similar patterns when we focus on macroeconomic words that are subjectively less “automatically associated” with downturns. The second column of Figure 3 shows the cyclical patterns of mentions of consumption, a more “generally relevant” indicator of economic performance, and just the word “economy,” which one would expect to have more neutral connotations. To the extent that even these terms take on negative connotations, we would argue, is more of a result of the forces we are documenting than a cause thereof—precisely the fact that economic issues seem more relevant during unpleasant times may cause such connotations.¹⁷

Finally, attention to policy issues (like “Fed” for monetary policy or “multiplier” for fiscal policy) is less obviously cyclical and more prone to idiosyncratic spikes. The “Fed” series, for instance, spikes noticeably around the collapse of Lehman brothers and the 2013 “taper tantrum,” but has a less systematic pattern around recessions. This basic observation foregrows a point that we will discuss more very shortly—that our measure of macro attention is not completely co-linear with the Baker et al. (2016) measure of policy uncertainty, nor with the market-based measures with which the previous authors show policy uncertainty is highly correlated.

Appendix Figure 11 prints the time-series graphs, in a much smaller format, for all of the words in our sample. A more concise summary of the word-by-word cyclicality is Appendix Figure 12, which shows the distribution of word-by-word term-frequency correlations with unemployment. In our sample, 61 of the 89 words (or 69%) of the words are counter-cyclical, or positively correlated with the unemployment rate.

2.5. Macro Attention is Distinct from News Indices

Our MacroAttention index has been constructed to capture attention to the macroeconomy overall. However, this is plausibly related to other uncertainty measures focused on closely related issues, and/or measured in different domains (financial markets, news media, etc.). To rule out this possibility, we compare the time-series behavior of our index with three alternatives. The first two derive from the related work of Baker et al. (2016) on measuring economic policy uncertainty, which combines information from newspaper reports about

¹⁷But note, to the opposite point, that only seven words on our list (argue, cut, problem, question, unemployed, unemployment, and recession) make the original Loughran and McDonald (2011) list of negative sentiment words in 10Q/Ks.
economic policy with information on tax code redundancy and disagreement among professional forecasters about government spending variables. The third is the VIX index of implied volatility for the S&P 500.

Appendix Table 6 shows the correlation of each measure with the others from 1998 to 2007. We find positive correlations of our measure with both Baker et al. (2016) measures and a negative correlation with the VIX. But, most importantly, we find significant independent variation in our measure, suggesting that we have captured a novel dimension of macroeconomic “salience” that is not purely subsumed in existing measures (and associated theory) for understanding policy and financial risk. Moreover, it is one that by construction relates explicitly to what firms are themselves thinking rather than the media or financial market participants.

3. A Macroeconomic Theory of Attention Cycles

To both explain the attention cycle we observe in the data and explore its macroeconomic implications, we augment a standard Neoclassical RBC model in the most parsimonious possible manner: we add a single parameter to allow for stochastic choice by firms. The model features a risk-averse representative household, heterogeneous firms, and aggregate demand externalities a la Blanchard and Kiyotaki (1987). Firms make decisions about how much quantity to produce under uncertainty. In recessions, this decision is naturally “higher stakes” in utility units for two reasons: (i) the firm’s risk averse owners are more concerned about dollar profits in those states and (ii) dollar profits themselves are more sensitive to choices in the low-demand region of the demand curve. Each of these two forces naturally contributes toward attention cycles in both partial and general equilibrium. We finally show how attention cycles influence business cycles and lead to endogenously higher misallocation in booms than recessions, asymmetric and state-dependent shock propagation, and endogenous stochastic volatility of output.

3.1. Primitives: Consumers and Final Goods

There are countably infinite time periods indexed by $t \in \mathbb{N}$. There is a continuum of goods indexed by $i \in [0, 1]$, which are imperfect substitutes for one another in the production of a final good. There is a single factor of production, labor.

A representative household has constant relative risk-aversion (CRRA) preferences over final-good consumption $C_t$. Payoffs take the following expected-discounted-utility form:

$$U(\{C_{t+j}\}_{j \in \mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \frac{C_{t+j}^{1-\gamma}}{1-\gamma}$$  \hspace{1cm} (8)
where $\beta \in [0, 1)$ is the discount factor. Looking ahead, note that the model has no capital or investment choices, so the parameter $\gamma$ embodies only risk aversion and not intertemporal substitution if we are focused on the model’s predictions for quantities and not prices.\footnote{Of course the value of $\gamma$ matters in an intertemporal substitution capacity for determining the model’s equilibrium interest rate. But we will not focus on those predictions in our analysis.}

The household supplies labor at a wage $w_t$ and owns equity in firms that produce intermediate goods, thus receiving profits $\{\Pi_{it}\}_{i \in [0,1]}$. They can trade a bond in zero net supply at interest rate $r_t$ and have asset position $A_t$ at time $t$. The flow budget constraint of the representative household is therefore given by:

$$C_t + A_{t+1} \leq w_t L_t + \int_{[0,1]} \Pi_{it} di + (1 + r_t) A_t \quad (9)$$

The aggregate final good is produced by a representative, perfectly competitive firm. Its production function takes the following constant-elasticity-of-substitution (CES) form, with elasticity $\epsilon > 1$:

$$X_t = X(\{x_{it}\}_{i \in [0,1]}) = \left( \int_{[0,1]} x_{it} \frac{\epsilon - 1}{\epsilon} \right)^{\frac{\epsilon}{\epsilon - 1}} \quad (10)$$

The final goods firm buys its inputs at prices $\{p_{it}\}_{i \in [0,1]}$ and sells its output at a normalized price of one. Because the final-goods firm is perfectly competitive, it earns no profit.

We assume a non-competitive mechanism for labor supply. Wages are set by the following reduced-form “wage rule”:

$$w_t = \bar{w} \cdot \left( \frac{X_t}{\bar{X}} \right)^{\chi} \quad (11)$$

where $(\bar{w}, \bar{X})$ are constants, and $\chi \geq 0$ is a “pass-through elasticity.” Assuming a wage rule instead of a labor market equilibrium has both technical and economic content for our analysis. On the technical side it greatly simplifies our equilibrium analysis and allows us to describe the economy via a single, easily characterized fixed-point equation. On the economic side, it easily parameterizes the singular force (wage pressure, or labor scarcity) that would push against our main results about attention cycles, in a way we will demonstrate shortly. Such a wage rule is very similar to a wage Phillips curve and, accordingly, a parameterization with low $\chi$ would account well for the low observed pro-cyclicality wages over the business cycle (Solon et al., 1994; Grigsby et al., 2019). In our quantitative analysis, we will show that all of our results are robust to considering wages that are set by more standard market clearing conditions.
3.2. Primitives: Intermediate Goods, Stochastic Choice, and Attention

The key actors in our model are producers of intermediate goods. These firms hire a labor quantity $L_{it}$, take the wage $w_t$ as given, and produce with the following linear technology:

$$x_{it} = \theta_t L_{it}$$

(12)

where $\theta_t$ is an aggregate-level productivity shifter which will be the main “driving force” in the model.

We assume that $\theta_t$ lies in a compact set $\Theta \subset \mathbb{R}$ and is determined exogenously by a Markov process with transition probability given by $\pi : \Theta \rightarrow \Delta(\Theta)$, with $\pi(\theta_t | \theta_{t-1})$ yielding the density of $\theta_t$ conditional on last period’s productivity being $\theta_{t-1}$. At time $t$, all firms know the sequence of previous productivity realizations $\{\theta_s\}_{s<t}$, but not the current period’s productivity.

3.2.1. The Friction

Intermediate goods choose a production level $x_{it}$ in the feasible set $\mathcal{X}$. But they struggle to match that choice to the state of the world $\theta_t$ without making idiosyncratic mistakes. We model this by having them choose stochastic choice rules at a cost. This formulation is sufficiently flexible to capture models of information acquisition and mistake-making in a unified and parsimonious manner, as Flynn and Sastry (2020) describes in more detail.

Formally, each firm chooses a stochastic choice rule $p : \Theta \rightarrow \Delta(\mathcal{X})$ in set $\mathcal{P}$, or a mapping from states of the world to distributions of actions described by probability density (mass) function (PDF) $p(\cdot | \theta)$. We model the cost of information acquisition and/or mistake-making via a cost functional $c : \mathcal{P} \rightarrow \mathbb{R}$ which returns how costly any given stochastic choice rule is to implement in units of utility or cognitive strain. The basic idea we wish to embody within such a functional is that playing actions that are more precise in any given state is more costly.

To make this tension more clear, and also to make the analysis more tractable, let us specialize to the following stochastic choice functional:

$$c(p; \theta_{t-1}) = \lambda \int_{\Theta} \int_{\mathcal{X}} p(x | \theta) \log(p(x | \theta)) \, dx \, \pi(\theta | \theta_{t-1}) \, d\theta$$

(13)

for some scalar $\lambda > 0$. This is an example of a likelihood-separable cost functional as discussed in more detail in Flynn and Sastry (2020). Such cost functionals capture the idea that it is costly for agents to avoid “mistakes” or misoptimizations. Indeed, as formalized by Fudenberg et al. (2015), a formulation with likelihood-separable stochastic choice is equiv-
alent to an additive random utility model of the sort often employed in the literatures on trade and industrial organization. The formulation (13) is exactly isomorphic to a “logit demand model,” or additive random utility model with Gumbel distributed perturbations, and embodies the familiar associated axioms including independence of irrelevant alternatives (IIA).

Naturally, such a formulation lacks a natural intuition in terms of signal processing, or rational inattention, which has become a dominant metaphor for imperfect optimization in economics (see, e.g., Sims, 2003). More substantially, the above model cannot capture potentially interesting behavioral phenomena such as anchoring of the action distribution to unconditionally more common actions. In Appendix B, we show how all of our results generalize to a model with mutual information as the cost functional under a particular condition on the priors held by the agents.

3.2.2. The Firm’s Problem

The firms commit to delivering the realized quantity $x_{it} \in X$ to the market, selling it at the (maximum) price $p_{it}$ at which the final goods firm is willing to buy, and to hiring sufficient labor in production. Intermediate goods firms are owned by the representative household and hence seek to maximize the product of their dollar profit and the household’s marginal utility. They treat the cognitive cost as an additive “psychological cost” also in utility units.

Let us define the “risk-adjusted profits” part of the firm’s payoffs is:

$$\Pi(x_{it}; (X_t, w_t), \theta_t) := X_t^{-\gamma} \cdot \left( x_{it}^{1-\frac{1}{\gamma}} X_t^{\frac{1}{\gamma}} - x_{it} \frac{w_t}{\theta_t} \right)$$

(14)

which substitutes in the form of the household’s marginal utility (the SDF), the final goods firms’ demand for intermediates, and the intermediate goods firm’s cost structure.

Because decisions are separable across time, and $\theta_{t-1}$ is an observed sufficient statistic for history of state realizations, the firm can be thought to solve a series of one-shot problems of choosing a stochastic choice rule in period $t$, conditional on the realization $\theta_{t-1}$. The firm has a conjecture for how the tuple of aggregate output (and, by implication, the wages from the wage rule) move with the state. Denote this conjecture as $\Lambda(\theta_t; \theta_{t-1})$. Given this conjecture, they play a best reply by solving the following program:

$$p^*(\Lambda, \theta_{t-1}) \in \arg \max_{p \in P} \int_{\Theta} \int_X \Pi(x, \Lambda(\theta_t; \theta_{t-1}), \theta) p(x \mid \theta) \ d x \ p(\theta \mid \theta_{t-1}) \ d \theta - c(p)$$

(15)

which is maximization of expected utility, averaged over risk in the state $\theta$ and stochastic action $x$, net of the cost of the chosen stochastic choice rule.
3.3. Equilibrium

Our model is a real business cycle model without capital. Output in each period is supply-determined and the inter-temporal Euler equation determines only the (otherwise unrestricted) interest rate, about which we are not interested in making positive predictions. As such, we can define a very “compact” notion of equilibrium that focuses only on the choices of intermediate goods firms and aggregate output $X_t$. Let us define this formally:

**Definition 1** (Wage-Rule Equilibrium). A wage-rule equilibrium is a set of stochastic choice rules, \( \{p^*(\theta_{t-1})\}_{\theta_{t-1} \in \Theta} \), and a law of motion for aggregate output \( \{X(\theta, \theta_{t-1})\}_{\theta, \theta_{t-1} \in \Theta} \) such that:

1. Intermediate goods firms solve program (15) with conjecture \( \Lambda = X \).
2. The wage is determined by (11).
3. The final goods firm produces with production function (10).
4. The conjectured law of motion is correct, or \( \Lambda = X \).

To tractably study equilibrium, we now simplify the intermediate goods firms’ objective and the final goods firm’s production with quadratic approximations. Let us first define the unrestricted optimum choice for the firm, a production level \( x^*(\Lambda, \theta) \) which solves:

\[
x^*(\Lambda, \theta) := \arg \max_{x \in \mathcal{X}} \Pi(x; \Lambda, \theta)
\]

and is the action they would choose when \( \lambda = 0 \) and there is no cost of playing precise choices. Define \( \bar{\Pi}(\Lambda, \theta) \) as the maximized objective. Now let \( \Pi_{xx}(\Lambda, \theta) \) denote the second derivative of the profits function in \( x \), evaluated at \( x^*(\Lambda, \theta) \):

\[
\Pi_{xx}(\Lambda, \theta) := \frac{\partial^2 \Pi}{\partial x^2} \bigg|_{x^*(\Lambda, \theta)}
\]

This object measures the state-dependent cost of deviations from the optimal production plan and will be central to our analysis. The approximate objective of the intermediate goods firm is:

\[
\bar{\Pi}(x; \Lambda, \theta) := \bar{\Pi}(\Lambda, \theta) + \frac{1}{2} \Pi_{xx}(\Lambda, \theta)(x - x^*(\Lambda, \theta))^2
\]

Note finally that this is well-defined also for \( x < 0 \). Thus, for greater tractability, we can also consider cases in which we remove the constraint of positive production, or \( x > 0 \).

Next, we take a quadratic approximation of the final goods firm’s production function (10). To this end, we approximate around the ideal production points \( x^* \) to the second order,
and write:

\[ X_t = \int_{[0,1]} x_{it} \, di - \frac{1}{2\epsilon \cdot x^*(\Lambda, \theta)} \int_{[0,1]} (x_{it} - x^*(\Lambda, \theta))^2 \, di \]  

(19)

Note that this has a “mean-variance form” that reveals the key economics of interest. Because the aggregator is concave (and more so for lower \( \epsilon \)), it is systematically lower when the dispersion of production decisions (here, around the unconstrained optimal point) is higher.\(^{19}\)

A very similar expression is familiar from the literature on the negative effects of price dispersion in New Keynesian models (e.g., Woodford, 2003b) and of dispersion in value marginal products across manufacturing establishments (e.g., Hsieh and Klenow, 2009).\(^{20}\)

Let us now add these two approximation steps to the previous equilibrium concepts to define the following:

**Definition 2 (Linear-Quadratic Equilibrium).** A linear-quadratic equilibrium is a tuple as defined in Definition 1 that replaces \( \Pi \) in program (15) with \( \tilde{\Pi} \), defined in (18), and production function (10) with (19).

In this theoretical analysis, we will provide all results in the linear-quadratic equilibrium. In the quantitative analysis that we later perform, we will solve the model exactly and with no approximations: all of the qualitative results from this analysis remain intact.

Given the aforementioned approximations, and an additional technical assumption that takes the support of the action space \( \mathcal{X} = \mathbb{R} \), we can characterize the equilibrium of the economy up to a single fixed-point equation. One can thereby establish simple sufficient conditions for equilibrium existence, uniqueness and monotonicity of output in productivity:

**Proposition 1 (Existence, Uniqueness, and Monotonicity).** For any \( \chi > 0 \) and \( \lambda \geq 0 \), an equilibrium in the sense of Definition 2 exists. If \( \gamma > \chi + 1 \) and \( \chi \epsilon < 1 \), there is a unique equilibrium with positive output \( X \) and output is strictly increasing in \( \theta \) in this equilibrium.

**Proof.** See Appendix A.1. \( \square \)

It is finally worth noting that this result can be established at a higher level of generality in a more abstract class of games, nesting the model considered in this paper. See Flynn and Sastry (2020) for these results.

\(^{19}\)Note also that this aggregator remains well-defined for negative production, allowing it to co-exist with an expanded choice set for firms.

\(^{20}\)In the former context, see also Sections 6.1.3 and 7.1.2 of Goodfriend and King (1997) for an earlier reference to the same mechanism, but without the accompanying (approximate) mathematics.
3.4. Attention Cycles in the Model

We now turn to the key step in the theoretical analysis and demonstrate that this economy exhibits attention cycles. This result is unambiguous in the rigid-wage, linear-quadratic economy described by Definition 2 when risk-pricing is sufficiently counter-cyclical and wages are not too pro-cyclical and, we will argue, economically plausible via the same mechanisms in other variants.

First, we require a notion of what it means for firms to pay more attention. We will adopt a simple non-parametric notion of attention that corresponds to firms paying more attention if they play a more precise action distribution. Formally, we will say that firms pay more attention in a given state if the entropy of the cross-sectional distribution of firm production is lower. This means also that agents are paying more, in utility units, for their stochastic choice patterns, and is thus a natural unit for our problem. We will say that the model features an attention cycle if firms pay more attention when aggregate output is lower.\footnote{Again, all of the results are robust to a more general non-parametric notion of attention as defined in Flynn and Sastry (2020).}

**Definition 3** (Attention and Attention Cycles). Consider two cross-sectional distributions of firm actions \( g, g' \in \Delta(X) \). Firms pay more attention under \( g \) relative to \( g' \) if:

\[
\int_X g(x) \log g(x) dx \leq \int_X g'(x) \log g'(x) dx
\]  

(20)

Given an equilibrium with aggregate output \( X \) and stochastic choice \( p \), the model features attention cycles if \( p(\theta) \) is monotone decreasing in the sense of attention in \( X \).

Toward proving that the model features attention cycles, we first characterize the optimal stochastic choice behavior by firms. To this end, the following result characterizes exactly what intermediate goods firms do in our model’s unique equilibrium:

**Proposition 2** (Stochastic Choice by Intermediate Goods Firms). Given aggregate output \( X \), in each state \( \theta \in \Theta \) firms play an action \( x_{it} \) distributed as

\[
N\left(x^*(X(\theta), \theta), \frac{\lambda}{|\Pi_{xx}(X(\theta), \theta)|}\right)
\]

where \( x^*(X(\theta), \theta) \) is the unconstrained optimal action,

\[
x^*(X(\theta), \theta) := \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} \bar{w}^{\epsilon} X^{\epsilon} \cdot X(\theta)^{1-\epsilon} \theta^\epsilon
\]  

(21)
and $|\Pi_{xx}(X(\theta), \theta)|$ is the magnitude of curvature for the firms’ objective function, given by

$$|\Pi_{xx}(X(\theta), \theta)| := (\epsilon - 1)^{-\epsilon} \cdot \bar{\omega}^{1+\epsilon} \cdot X^{-\chi(1+\epsilon)} \cdot \theta^{-1-\epsilon} \cdot X(\theta)^{1-\gamma + \chi(1+\epsilon)}$$

**Proof.** See Appendix A.2.

Our model admits a simple and easy-to-interpret stochastic choice pattern. Firms center their action around the full-information optimum $x^*(X, \theta)$. This choice always increases in the state, and also increases in the aggregate output $X$ if $\chi$ is sufficiently small, or $\chi < 1/\epsilon$. In this case, the “good news” of increased demand offsets the “bad news” of higher wages for the firm. More technically, this condition makes the underlying game one of strategic complements rather than strategic substitutes.

Firms’ random choice pattern is a normal distribution, a convenient consequence of our quadratic form and assumed form of the functional (13). Thus, the distribution of firms’ actions across states follows a normal-mixture model. The variance of this distribution increases when the cost of paying attention $\lambda$ increases or the second-order cost of mistakes $|\Pi_{xx}|$ decreases. Observe that $|\Pi_{xx}|$ increases in $\theta$ for any values of the parameters. It further increases in $X$ as long as $\chi < (1 + \gamma)/(1 + \epsilon)$, or wages do not increase too steeply in output. This condition is in fact satisfied for any $\chi < 1/\epsilon$, or wage rule consistent with strategic complementarity in the economy, and $\gamma \geq 1$, which is steeper risk aversion than log preferences. Thus we view it as a fairly uncontroversial assumption. Finally recall that, in our model’s unique equilibrium, $X$ is increasing in the state (Proposition 1).

We combine these observations, plus a translation of variance to the aforementioned entropy units of attention to state the following summarizing result:

**Proposition 3 (Attention Cycles).** Assume $\gamma > \chi + 1$ and $\chi \epsilon < 1$. Intermediate goods firms pay more attention in lower states of the world and the model features attention cycles, where both concepts are as in Definition 3.

**Proof.** See Appendix A.3.

Note that this result stems from the basic fact that the equilibrium mapping $\theta \mapsto |\Pi_{xx}(\theta)|$ is monotone decreasing in the state of the world. As a result, firms play more precise actions in equilibrium in exactly the lowest productivity states of the world. In these states, as output is lowest, the model then features attention cycles. To understand this result, there are three relevant forces which determine the monotonicity of $|\Pi_{xx}|$, which we will now review in turn. These mechanisms hint at the robustness of our result in alternative settings to the explicit model we have developed.
First, and most importantly, the household’s marginal utility or SDF, $X^{-\gamma}$, is higher when output is lower. This is of course a general function of concave utility functions and does not require the CRRA form. But in the CRRA case we can provide an illustrative calculation of the magnitude. An $x\%$ reduction in total output in a recession is associated with an approximately $\gamma \cdot x\%$ increase in marginal utility, or the “stakes” for getting decisions correct. For high risk aversions (or “prices of risk,” in a more reduced-form interpretation) this force can be magnified tremendously. This force is present in any Neoclassical theory of the firm whose profits are priced by a representative household, or more generally, a SDF. The relevant quantity in a more general model of the SDF is then its elasticity to aggregate production, which we know empirically to be extremely high (see, for example, Hansen and Jagannathan, 1991). Thus, not only is this risk-pricing force present in our model, we argue that it is present in any Neoclassical model; casual empiricism moreover suggests that the magnitude of this force must be large.

A second force comes from the aggregate demand externality. The curvature of a firm’s revenues depends inversely on aggregate demand, holding fixed marginal utility and marginal costs, owing to the aggregate demand externality. Low aggregate demand is associated with operating on the more steeply curved part of the individual-level demand curve which is less forgiving to productive mistakes.

Finally, there is a direct dependence on wages, which is translated into output units by the wage rule. For this to over-power the previous two considerations, we require a sufficiently high pass-through of output to wages: $\chi > (1 + \gamma)/(1 + \epsilon)$. We argue that this lies in the empirically irrelevant part of the parameter space as $\gamma$ which indexes the elasticity of the SDF would far exceed $\epsilon$ in any reasonable calibration and $\chi$ is very small, owing to the relative acyclicality of wages over the business cycle (Solon et al., 1994; Grigsby et al., 2019).

A reasonable extrapolation of these ideas is that we might expect to see the same pattern in any market economy in which the first force, owing to the SDF, dominates—regardless of the extent of aggregate demand externalities, or even if the game of interest had strategic complementarity. Moreover, while the cyclical of factor prices poses some threat to our argument, the empirical observation that wages are not particularly cyclical suggests that our wage-rule economy for low $\chi$ is not an unreasonable approximation. We therefore argue that the basic forces that generate attention cycles in our model are likely to be robust in other models.

Finally, having discussed the robustness of the economic mechanisms, it is natural to inspect the robustness of our results to the particular way in which we have modelled stochastic choice. In Flynn and Sastry (2020), we show a strictly more general version of Propositions
1 and 3 in abstract games. The core of the argument remains the monotonicity of “stakes” in both the state and output. But additional generality comes from (i) considering a larger class of stochastic choice functionals and (ii) clarifying the exact kind of supermodularity (macro complementarity) and discounting (concavity of the aggregate production function) required to achieve monotone precision in a game’s unique equilibrium.\(^{22}\) Finally, we show in Appendix B how all of these results survive in a special case of the same model with mutual information costs, the functional considered in the literature on rational inattention.

3.5. The Qualitative Macroeconomic Consequences of Attention Cycles

Our goal in this section was transparently to explain attention cycles as a consequence of business cycles, or rationalize the basic patterns from Section 2. But we are also interested in understanding the other direction of causality: how do attention cycles affect macro dynamics? In Section 5 we will take a much more detailed look at quantitative implications for business cycles in a calibrated model. Here, we will show on a more qualitative level how attention cycles affect business cycles.

Let us define \(X(\theta, \lambda)\) as a mapping from the state and the scaling parameter for the stochastic choice functional to output in the economy, holding fixed all other parameters. Note that \(X(\cdot, 0)\) returns the mapping from the state to output in the “vanilla RBC model” with no costs of stochastic choice. This is a natural benchmark that shuts down both inattention and the cyclicality thereof.

The following proposition characterizes output in log units as the sum of this RBC core and a wedge \(\log W(\theta, \lambda)\):

**Proposition 4 (Output in the Attention Cycles Model).** *Output can be written in the following way:*

\[
\log X(\log \theta, \lambda) = X_0 + \chi^{-1} \log \theta + \log W(\log \theta, \lambda)
\]

*where \(\log W(\log \theta, \lambda) \leq 0\). When \(\gamma > \chi + 1\) and \(\chi \epsilon < 1\), the wedge has the following properties:*

1. \(\log W(\log \theta, 0) \equiv 0\) for all \(\theta\), or there is no wedge when the friction disappears.
2. \(\partial \log W/\partial \lambda < 0\), or the wedge increases with the friction.
3. \(\partial \log W/\partial \log \theta < 0\) for \(\lambda > 0\), or the wedge is increasing in the state.

*Proof. See Appendix A.4.*

\(^{22}\)In fact, a slightly generalized version of Propositions 1 and 3 that dispenses with the approximated CES aggregator could be proven exactly using the main results of Flynn and Sastry (2020).
Let us analyze this result piece-by-piece. First, output is log-linear in the state absent the friction. Second, output is depressed by the presence of stochastic choice in a way that smoothly increases in the extent of the friction. This has both a partial and general equilibrium component. The partial equilibrium component comes from the fact that higher \( \lambda \) makes firms play more dispersed actions (Proposition 2), which has a cost to output coming from the dispersion penalty in (19). The general equilibrium component comes from iterating this logic until convergence, in terms of both shifting down the mean action and reducing the variance of actions as firms pay more attention. The result verifies that this fixed-point operation converges on a lower value of output.

This result echoes the general principle that inattention dampens the effects of shocks on the macroeconomy. In our model the manifestation is exactly through an increase in the dispersion of actions, which results in lower production by the final goods firm with concave technology.

Point (3) combines this “dampening” observation with a (GE fixed-point) version of Proposition 2. In low states of the world, action dispersion is lower and hence so too is the wedge. In high states of the world, action dispersion is higher and the wedge is higher, but not enough to make output non-monotone in the state (Proposition 1).

3.5.1. A Productivity Interpretation

The wedge \( W(\log \theta, \lambda) \) is akin to an efficiency wedge in an imperfect-reset price or quantity setting model. To make this connection more precise, consider the following simple Corollary that re-casts the wedge in terms of labor productivity, which is also the Solow residual in our one-factor economy:

**Corollary 1** (The Productivity Wedge). Let \( A := X/L \) be the measured productivity of our economy. Productivity can be written as

\[
\log A(\log \theta, \lambda) = A_0 + \log \theta + \chi \epsilon \log W(\log \theta, \lambda)
\]  

(23)

where \( \log W(\cdot) \leq 0 \) is as defined in Proposition 4.

**Proof.** See Appendix A.5.

So the aforementioned wedge, up to scale, also parametrizes the loss of efficiency in this economy due to misoptimization error. Moreover, Proposition 4 shows that these errors should be lower in low states of the world and so productivity should be closer to the frictionless benchmark in these states. Owing to attention cycles, misallocation across firms in the model is endogenously higher in booms than recessions.
The productivity wedge representation allows for two useful parallels between our paper’s mechanism and classic arguments in the macroeconomics literature. Consider first the literature on the cleansing effect of recessions following Caballero and Hammour (1994). Our mechanism is like an attentional, intensive margin version of the same effect: conditional on a given firm operating, it is more focused on making precise and accurate choices in recessions, and this on average reduces the wedge and raises aggregate TFP.

Consider next the large literature on the welfare costs of inflation following Goodfriend and King (1997, 2001). The fundamental, dominant argument in this literature relates to the productivity consequence of imperfect optimization via price dispersion or, equivalently, value marginal product dispersion among firms. Our model is one in which that dispersion is endogenously lower in recessions, due to the economic forces outlined above. This state-variation in the wedge contributes to business-cycle dynamics of output, as indicated in Corollary 1.

3.5.2. Shock Propagation and Volatility

The fact that agents are differentially attentive to shocks across states of the world naturally leads one to expect that the economy is differentially sensitive to shocks across states of the world. This observation is formalized in the following Corollary which shows how the model with attention cycles generates endogenous stochastic volatility, state-dependent propagation of shocks and asymmetric shock propagation—features which are all absent in the frictionless benchmark with fully attentive firms, nested by $\lambda = 0$.

**Corollary 2** (Endogenous Stochastic Volatility, State-Dependent Propagation, and Asymmetric Propagation). *Suppose that productivity follows the process* $\theta_t = \rho_\theta \log \theta_{t-1} + (1 - \rho_\theta) \log \bar{\theta} + \nu_t$ *where* $\text{Var}(\nu_t) = \sigma^2_\nu$. *The model generates endogenous stochastic volatility, state-dependent shock propagation and asymmetric shock propagation:*

1. **Endogenous stochastic volatility:** to a first-order approximation in $\nu_t$, the variance of output conditional on last period’s productivity $\theta_{t-1}$ is given by

   $$\text{Var}(\log X_t|\theta_{t-1}) = \left(\chi^{-1} + \frac{\partial \log W(\log \theta)}{\partial \log \theta}|_{\theta=\theta_{t-1}}\right)^2 \sigma^2_\eta (24)$$

   *Moreover, in the frictionless benchmark* $\lambda = 0$ *volatility is independent of* $\theta_{t-1}$.

2. **State-dependent shock propagation:** the impact on output from a small shock $\nu_t$ in state $\theta_{t-1}$ is given by:

   $$\frac{\partial \log X(\log \theta)}{\partial \log \theta}|_{\theta=\theta_{t-1}} = \chi^{-1} + \frac{\partial \log W(\log \theta)}{\partial \log \theta}|_{\theta=\theta_{t-1}} (25)$$
Moreover, in the frictionless benchmark $\lambda = 0$ the impact of a shock to productivity is independent of $\theta_{t-1}$.

3. Asymmetric shock propagation: the impact of a shock to second-order in $\nu_t$ in state $\theta_{t-1}$ is given by

$$
\frac{\partial \log X(\log \theta)}{\partial \log \theta}|_{\theta = \theta_{t-1}} = \chi^{-1} + \frac{\partial \log W(\log \theta)}{\partial \log \theta}|_{\theta = \theta_{t-1}} + \frac{\partial^2 \log W(\log \theta)}{\partial \log \theta^2}|_{\theta = \theta_{t-1}} \nu_t
$$

(26)

where the sign of second derivative of the wedge $\frac{\partial^2 \log W(\log \theta)}{\partial \log \theta^2}|_{\theta = \theta_{t-1}}$ determines the direction of the asymmetry. In the frictionless benchmark $\lambda = 0$, shocks have a symmetric impact on output.

**Proof.** See Appendix A.6.

To understand this result, see that the sensitivity of the economy to shocks is simply the sum of the frictionless economy’s response to shocks, which is always $\chi^{-1}$, and the response of the attention wedge to shocks, which is always negative and depends on the state. Thus, the economy is mechanically less responsive to shocks than the full attention benchmark as firms do not perfectly track changes in the fundamental. This is a familiar point in the literature with cognitively constrained agents (e.g. Sims, 1998, 2003; Gabaix, 2014), drawn out here in a general equilibrium context.

Novel to our analysis is the fact that the response of the attention wedge to shocks depends on the state (the attention wedge is not linear in $\log \theta$). This generates a host of interesting phenomena. First, when the economy is more sensitive to shocks we see both higher output volatility and greater responses of output to shocks of the same size are an inevitable consequence. Second, as the attention wedge moves non-linearly in the state, asymmetric amplification is a natural consequence, the magnitude and sign of which is driven by the degree of concavity or convexity of the attention wedge.

The ultimate macroeconomic implications of this result hinge on the concavity or convexity of the attention wedge in the state. When the attention wedge is convex (resp. concave), the economy generates greater (smaller) volatility in low states, a larger (smaller) impact of shocks and features larger (smaller) impact of negative than positive shocks. Owing to the intuition that when $\theta$ is very small, it is as-if the economy is operating in its RBC core, with “full pass-through” of fundamental shocks to measured productivity and output, the natural case appears to be a convex wedge. Nevertheless, this cannot be established globally. As a result, it is ultimately a quantitative question whether or not the model generates these properties and how significant they are for business cycle dynamics.

Our quantitative analysis in Section 5 will deliver such a convex wedge, implying that the
economy displays more “stickiness” (in the language of Sims, 1998) in high-output states. In that section we will review more the contributing factors toward and key implications of this finding.

3.6. Extensions: Information Acquisition and Efficiency

In Appendix B, we study the robustness of our results to more canonical information acquisition by examining our model with rational inattention a la Sims (2003). We leverage results from Matějka and McKay (2015) to show that all of the results in this section hold exactly as written with mutual information cost instead of entropic likelihood-separable costs in the case where all actions are exchangeable in the prior. We moreover prove a comparative statics result, that holds for all priors, showing that increases in curvature of firms’ payoffs in any state increase attention in that state. We therefore argue that the conclusions of this section are robust to considering canonical information acquisition.

In Appendix C, we study the efficiency properties of the model. In a similar vein to Angeletos and Pavan (2007), Angeletos and Sastry (2019) and Hébert and La’O (2020), we provide necessary conditions for efficiency. In our economy, efficiency is knife-edge and generally fails owing to cross-firm production externalities and their interaction with stochastic choice.

4. Testing the Theory

Having outlined a theory for how and why attention cycles manifest, and before fully exploring its quantitative potential, we now go back to the data to test the theory’s additional predictions for attention as a function of other environmental shifters.

Let us first outline three such predictions. The first concerns the reliance of countercyclical attention on different macro mechanisms. Given the primacy of the risk aversion force in the expression of Subsection 3.4, we expect attention cycles to be prevalent even in settings in which the cyclicity of demand is low:

Prediction 1. Measured firm-level macro attention should be counter-cyclical even if demand and output are not.

This differentiates our theory from an alternative in which the patterns of attention reflected only relative relevance of macro information for predicting firm-level sales. In this alternative, we may not expect acyclical industries also to exhibit attention cycles.

Nonetheless, the role of the curvature of firms’ revenue suggests that demand cyclical should also matter and crowd in more attention to macro aggregates:
Prediction 2. Measured firm-level macro attention should be more counter-cyclical in contexts where firms face more pro-cyclical demand.

Finally, at the most abstract level possible, the general theory predicts that firms with greater curvature of profits should pay more attention to macroeconomic variables. Thus, should there be a way to measure the curvature of firms’ profits, we should observe that firms with more curvature do a better job of reporting contemporaneous macroeconomic conditions, as formalized below:

Prediction 3. Firms should make smaller mistakes in reporting contemporaneous macroeconomic conditions if they have greater absolute curvature of their profits.

In this section, we explicitly test all three of these predictions to evaluate the performance of the model.

4.1. The Cyclicality of Attention by Sector

Predictions 1 and 2 relate to the relative cyclicality of attention by sector. It is straightforward to calculate industry-level MacroAttention by continuing to take simple averages across firms, but subsetting only to firms in a particular industry, with Equation 4. To conduct this analysis, we partition our sample into 45 different industries. These are based primarily on NAICS2 codes, but we separate manufacturing (NAICS 31-33) and information (NAICS 51) into three-digit categories to maintain comparable numbers of firms in each bin.

For each thusly defined industry, we calculate an output cyclicality measure using BEA data on sectoral GDP since 2005 (linked appropriately to NAICS-definition sectors). We compute our preferred measure as the correlation between log nominal output in a given sector with log nominal output in the entire economy, also sourced from the BEA tables. In the model, this is an imperfect correlate for the primitive shifter of different sectors’ demand cyclicality (and indeed is endogenous to firms’ decisions and attention). Nonetheless, we argue that it captures the reduced-form prediction. We next calculate, separately for each sector \( n \), both a MacroAttention\(_{nt} \) measure as described in Section 2 and the correlation thereof with national-level unemployment. This is essentially repeating the exercise of Figure 2 and creating an equivalent of Fact 1 for each of the 45 sectors.

Figure 4 plots these two objects against one another and reveals a positive relationship. That is, sectors with higher output cyclicality display considerably more cyclicality in their attention. We label, for illustration, some sectors that anecdotally might be associated with high or low cyclicality. The “trend line” in this scatterplot estimates the following regression:

\[
\text{Corr}[\text{MacroAttention}_{nt}, \text{Unemployment}_t] = \alpha + \beta \cdot \text{Corr}[\text{Output}_{nt}, \text{Output}_t] + \epsilon_n \quad (27)
\]
for which estimates are given in Appendix Table 7. Both the slope and intercept are significantly different from 0 at the 5% level. From a back-of-the-envelope extrapolation, which is visible in Figure 4, any sector with an output correlation above $-\hat{\alpha}/\hat{\beta} = -0.398$ is predicted to have counter-cyclical attention. This verifies both Predictions 1 and 2.

4.2. Directly Reported Attention and Curvature

In our next application, we use the survey data collected by Coibion et al. (2018) (hence, CGK) from a representative panel of firms in New Zealand from 2013 to 2016 to speak even more directly to the mechanisms at work in our model.

Reported attention and the business cycle Although the CGK survey took place during relatively tranquil times for the New Zealand economy, it did ask two hypothetical questions directly revealing of the premise for this paper. Each concerned firm’s desire to collect information on the macroeconomy conditional on either good (or poor) conditions:

Suppose that you hear on TV that the economy is doing well [or poorly]. Would it make you more likely to look for more information?

Table 1 reports the percentage of answers in each of five bins, given the conditions of the economy doing “well” or “poorly.” This self-reported demand for information clearly spikes

---

23In Appendix D, we replicate this exercise using data on macro attention from firms’ sales and earnings conference calls. We find similar patterns but have a less precise estimate for Prediction 2 (the “slope” of attention in demand cyclicality).
<table>
<thead>
<tr>
<th>Response</th>
<th>Poorly</th>
<th>Well</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much more likely</td>
<td>44.96</td>
<td>9.77</td>
</tr>
<tr>
<td>Somewhat more likely</td>
<td>30.91</td>
<td>19.42</td>
</tr>
<tr>
<td>No change</td>
<td>12.56</td>
<td>8.67</td>
</tr>
<tr>
<td>Somewhat less likely</td>
<td>7.16</td>
<td>53.35</td>
</tr>
<tr>
<td>Much less likely</td>
<td>4.40</td>
<td>8.79</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Table 1: **Changing Macro Attention in Response to News.**
Data are from the Coibion et al. (2018) survey of firms in New Zealand.

in the context of *bad* news about the macroeconomy and, if anything, contracts with *good* news of the macroeconomy. This supports very precisely our theory that bad conditions increase the stakes for firms’ decisions and hence make keen attention to macroeconomic conditions more important, while good news does not have a symmetric effect.

**Reported objective function curvature and attention** The second possible test in the CGK data relates to Prediction 3 outlined above: that *any shifter* of the stakes of the firm’s objective function (that is, risk-adjusted profits) should scale with attention to macro developments. While the model we have developed does provide an internal explanation for where such variation could come from, one might imagine that it relates to idiosyncratic shapes of firm-level demand curves or other firm-level conditions.

The CGK survey indirectly elicits information on this shifter via questions about purely hypothetical price changes and revenue increases to an “optimal point.” In Appendix E, we show exactly how one can use a pair of linked questions about firms’ hypothetical optimal reset price, and the hypothetical percentage increase in profits that would be associated with that change, to develop an elicited measure of *firm profit curvature* in non-risk-adjusted units. Appendix Table 4 shows that this curvature measure seems to be higher for smaller firms with more within-industry competitors.

As outcomes for macro attention, we can turn to two sources. The first is the absolute-value error in firms’ one-year back-casts for three macro variables: inflation, output growth, and unemployment. The second is firm managers’ reported (binary) interest in *tracking* one of the aforementioned variables. Appendix E shows exactly where these measures come from in the survey.

For each of the aforementioned firm-level outcomes $Y_{it}$, we run the following regression on the firm-level profit curvature variable $\text{ProfitCurv}_{it}$ and a vector of controls $Z_{it}$:

$$ Y_{it} = \alpha + \beta \cdot \text{ProfitCurv}_{it} + \gamma'Z_{it} + \epsilon_{it} $$ (28)
Table 2: **Curvature and Attention.**
Data are from the Coibion et al. (2018) survey in New Zealand.

We control for five bins in the firms’ total reported output and the firms’ 3-digit ANZ-SIC code industries. Finally, we cluster all standard errors by 3-digit industry.

Table 2 shows the results. For inflation we find strong evidence that higher-curvature firms make smaller errors, with some much of the effect being absorbed by control variables when added. For GDP growth we find estimates that are much less precise but have the same signs; and for unemployment, results that are further imprecise and have the wrong signs. We take this as support for Prediction 3 and the exact mechanism that our theory proposes: that the differential stakes of making mistakes is a contributing factor to macro attention.

### 5. Quantifying the Macroeconomic Consequences

Having now tested the key predictions of the theory, and thereby evidenced its key mechanisms, we explore its quantitative bite and macroeconomic implications in a one-parameter extension of the canonical RBC model. In particular, we calibrate the stochastic choice extension of the RBC model from Section 3 to match our facts regarding the elasticity of measured MacroAttention to the business cycle. As indicated by the theory, this parsimonious extension of the RBC model, which we call the real attention cycles (RAC) model, generates four important phenomena not present in the RBC model: asymmetrically large
amplification of negative shocks; greater amplification of shocks when output is low; endoge-
nous stochastic volatility of output growth, whereby volatility of output growth is highest
when output is lowest; and fast crashes and slow recoveries.

5.1. Calibration and Numerical Solution

Recall the RBC environment from Section 3. We make one change for the quantitative
section, to explore quantitatively the robustness to allowing flexible wages that are set via
a more canonical labor market clearing mechanism rather than a reduced form rule. We
assume the household has the inter-temporal preferences of Greenwood, Hercowitz, and
Huffman (1988), henceforth GHH:

\[ U(C_t, L_t) = \left( \frac{C_t - \frac{L_t^{1+\phi}}{1+\phi}}{1-\gamma} \right)^{1-\gamma} \]  

(29)

This admits a simple labor supply expression that replaces our previous wage-rule:

\[ w_t = L_t^\phi \]  

(30)

where \( \phi \), in slightly different units than the previous \( \chi \), parametrizes a similar idea of how
cyclical are wages in the economy. The GHH specification moreover prevents our experimen-
tation with the risk-aversion parameter \( \gamma \) from having any direct effects on the cyclicality of
labor supply and/or wages.

We approximate the action set \( \mathcal{X} \) using an evenly spaced grid of points. This encodes
both a positivity constraint and a maximum action, the latter of which is necessary for
computational feasibility though not a direct implication of the model.\(^{24}\)

Next, to close the model, we assume that the productivity of the intermediate goods
firms \( \theta_t \) follows a Gaussian AR(1) process in logs:

\[ \log \theta_t = (1 - \rho_\theta) \log \bar{\theta} + \rho_\theta \log \theta_{t-1} + \nu_t \]  

(31)

where \( \nu_t \sim^{iid} N(0, \sigma_\theta^2) \) and \( \bar{\theta} \) is a scaling constant. For numerical implementation, we es-
timate the transition process using a discrete approximation of the previous on an evenly
spaced grid between \([\log \bar{\theta} - 0.10, \log \bar{\theta} + 0.10]\) (i.e., allowing for 10% fluctuations in produc-
tivity).\(^{25}\)

\(^{24}\)We choose a grid of 5000 points between 0 and 12.

\(^{25}\)We use 100 grid points and “snap” the predicted process to the endpoints if the continuous approximation
would predict over-shooting. We set \( \bar{\theta} = 3 \).
### Table 3: Parameters.
We select the fixed parameters. Free parameters are estimated by simulated method of moments.

<table>
<thead>
<tr>
<th>Geometrically Fixed</th>
<th>CRRA</th>
<th>Geometrically Free</th>
<th>Weight on entropy penalty</th>
<th>Variance of the productivity innovation</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ</td>
<td>5</td>
<td>λ</td>
<td>exp(−6.29)</td>
<td>0.0055</td>
</tr>
<tr>
<td>ψ</td>
<td>1</td>
<td>σ_θ^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ϵ</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ_θ</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ϵ_W</td>
<td>0.353</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.1.1. Parameters and Moment Matching

The RAC model has six parameters (ignoring the scaling by \( \bar{\theta} \)). As the curvature of utility \( \gamma \) serves no inter-temporal purpose in our model with no capital, the parameter serves only as the coefficient of relative risk aversion and shapes the stochastic discount factor. We impose a value \( \gamma = 5 \) which is midway between standard macro and finance calibrations.\(^{26}\) This reflects our desire to have a model with sensible risk-pricing characteristics as the stochastic discount factor is a key determinant of the greater curvature of firms’ profit functions in downturns. We choose standard values of \( \epsilon = 3 \) for the elasticity of substitution, \( \psi = 1 \) for the curvature of labor disutility, and \( \rho_\theta = 0.90 \) for the persistence of the exogenous productivity process \( \rho_\theta \). The parameters governing attention cost \( \lambda \) and the volatility of the productivity process \( \sigma_\theta^2 \) are free and we describe their estimation shortly.

Given a vector of these parameters \( (\gamma, \psi, \epsilon, \rho_\theta, \lambda, \sigma_\theta^2) \), we solve the model exactly up to approximating the action and state spaces on uniform grids. We do not invoke the quadratic approximation used in the theoretical analysis.\(^{27}\)

We then calibrate both the scaling of attention costs \( \lambda \) and the variance of the productivity innovation \( \sigma_\theta^2 \) using the simulated method of moments. We target two moments: the elasticity of MacroAttention to employment, \( \epsilon_W \); and the variance of quarterly output growth, \( \sigma_Y^2 \). We choose these moments so that our quantitative model matches both our

---

\(^{26}\)We have experimented with different values of \( \gamma \) to explore the robustness of our results. Broadly speaking, we can fit the data with both higher and lower \( \gamma \), and the higher \( \gamma \) calibrations generate more pronounced asymmetry in business cycle dynamics. We find \( \gamma \lesssim 2 \) as a (here, approximate) cut-off for seeing strong business cycle asymmetries. Appendix Figure 14 highlights one salient measure for this—the extent of stochastic volatility, which is like the slope of the line plotted in Figure 7.

\(^{27}\)Formally, we solve for the optimal action distribution \( p^*(X_t, w_t, \theta_t) \) by firms as a function of aggregate production \( X_t \), wages \( w_t \) and productivity \( \theta_t \). We then numerically find the fixed point \( X_t(\theta_t) = X(p^*(X_t, w_t, \theta_t)) \) and \( w_t(\theta_t) = W(p^*(X_t, w_t, \theta_t)) \) where the \( W \) function is the market clearing wage given a distribution of production. The equilibrium action distribution is then given by the action distribution evaluated at this numerical fixed point.
stylized facts regarding MacroAttention and generates a sensible mean level of volatility for output growth. Both targeted values are reported in Table 3.

5.1.2. Attention in the Theory and the Data

Using these moments critically requires a mapping between our empirical measure of macro attention and attention within the model. To this end, we take:

\[
\log \text{MacroAttention}_t = -\text{ActionEntropy}(\theta_t) := \int_{X} p^*(x|\theta_t) \log p^*(x|\theta_t) \, dx
\]

(32)

The justification for this mapping comes from our underlying theory and two key assumptions. First, through the lens of our approximated model introduced earlier, this formulation makes MacroAttention directly proportional to the precision of agents’ actions. Hence our empirical finding that attention increases 1.34 percent per 1 percent change in economic activity \(e_W = -1.34\) should approximately translate in the model as the same percent-to-percent passthrough of employment to precision. This is ultimately the joint of two assumptions, both of which have economic content. The first, more fundamentally, is that what firms say maps to what they do in a meaningful way. The second is that we have captured the right units for this relationship. Our approach is ultimately only a first pass at answering both questions and we think that fully experimental work that explores each of these angles would be exciting and extremely informative about the questions at hand here.

5.1.3. Objective and Results

We choose the free parameters \(\lambda\) and \(\sigma^2_\theta\) to minimize the sum of squared differences between the target empirical moments and the moments estimated from the model:

\[
\{\lambda^*, \sigma^2_\theta^*\} = \arg\min_{\lambda, \sigma^2_\theta} \sum_{m \in \{e_W, \sigma^2_Y\}} \left(\frac{m^M(\lambda, \sigma^2_\theta)}{m} - 1\right)^2
\]

(33)

The results of this estimation are given alongside all targeted moments and fixed parameters in Table 3. Importantly, our in-model estimates of the two moments are -1.31 and 0.351, which are almost exactly correct.

Appendix Figure 13 shows graphically how we meet the main calibration target, which is the cyclicality of attention. In the left panel we show the agents’ action distributions, which become more dispersed in larger states of the world. In the right panel we show the utility

\[^{28}\text{In the first regression, we add quarterly fixed effects in logs to take care of the seasonal and document-compositional adjustments. See Section 2.2 for more discussion.}\]
cost of attention, which increases in the state as we may have predicted in an economy with moderate wage cyclicity.

An auxiliary calculation reveals how important general equilibrium forces are in explaining attention cycles. To this end, we consider a partial equilibrium benchmark in which firms play a best response to the conjecture that output and wages are as in the frictionless benchmark with \( \lambda = 0 \). The elasticity of these firms’ attention to employment in the counterfactual economy, exactly the moment we are trying to match, is only 12.9\% of the same value calculated in our calibrated economy (and, almost equivalently, in the data). Hence, in a realistic calibration, GE forces are more than just a theoretical curiosity—they greatly affect the model’s ability to fit the data and explain attention cycles.

### 5.2. The Attention Wedge

Recall our theoretical decomposition in Proposition 4. Figure 5 shows the analogue within our model. In the dotted line of the left panel, we plot a counterfactual world with the same structural parameters but \( \lambda = 0 \), or no stochastic choice friction. The frictionless output is exactly log-linear in the state, just like in the wage-rule example.\(^{29}\) In the black line of

\[ \log Y' = \frac{1}{\psi} \log \left( 1 - \frac{1}{\epsilon} \right) + \left( 1 + \frac{1}{\psi} \right) \log \theta \]
Figure 6. **State-Dependent Impulse Response.**

Impulse response function of output (left pane) and entropy of the action distribution (right pane) to a 2% productivity shock from steady-state productivity. The blue lines correspond to the (negative) IRF to a positive productivity shock. The orange lines correspond to the IRF to a negative productivity shock. The red lines correspond to the IRF to a negative productivity shock starting 2% below the steady state.

In the left panel, we show log output in the attention cycles model. This is almost everywhere lower, and especially so in higher states of the world in which agents choose quantities more imprecisely. The numerical reason for a slightly negative inattention wedge for low states relates to the non-negativity constraint for firm-level output that was ignored in Section 3.1. At some point, more random production begins to increase mean output purely because there are asymmetric opportunities to miss the optimal target.

The right panel of Figure 5 isolates the difference between the two lines, which is the analogue to the “wedge” $-\log W$ from Proposition 4. This wedge is an increasing function, as suggested by the theoretical results. It is also convex in the quantitative application, which as our theoretical results showed means that the slope of $\log X$ in $\log \theta$ decreases in the state. The next subsection will show in more illustrative ways how this basic pattern generates asymmetric propagation of shocks, state-dependent propagation of shocks and endogenous stochastic volatility.

### 5.3. Asymmetric Business Cycle Behavior

Let us now show two key features of the model’s predictions that follow from the basic pattern noted in the previous subsection, which was the convexity of the attention wedge or concavity of log output as a function of the state.
Asymmetric and State-Dependent Amplification of Shocks  The left pane of Figure 6 shows that a 2% negative shock to productivity features greater propagation relative to 2% positive shock to productivity: on impact, the negative shock causes a 21.9% greater change in output. Because of aggregate demand externalities and the stochastic discount factor, the stakes for making mistakes are larger in the lower state, so agents respond more to the shock. This is shown quantitatively in the right pane of Figure 6 which shows exactly this: the entropy of agents’ actions responds 93.4% more on impact to a negative shock than a positive shock. Attention cycles therefore cause the economy to respond more to negative shocks than positive shocks, despite no inherent non-linearity in the frictionless model.

The calibrated model further generates state-dependent propagation of shocks. The left pane of Figure 6 shows also the impact of a 2% negative productivity shock starting 2% below the steady state. This “double dip” shock is the most severe of the three plotted, because now both the initial level of objective curvature and the response thereof are magnified. The right pane of Figure 6 formalizes this intuition by showing how the entropy of agents actions responds much more to the shock starting below the steady state. In particular, a 2% negative shock starting 2% below the steady state causes, on impact, a 19.0% greater change in output than a 2% negative shock starting at the steady state, and a 45.1% greater change in output than a 2% positive shock starting at the steady state.

To put the previous results in a different language: even when fundamentals (here, productivity) evolve in a symmetric way, output will be more prone to “double dips” than “double increases” because of the aforementioned amplification and dampening roles of endogenous attention. The most fragile moment for the macroeconomy is when it is already in a downturn, because the key agents (here, intermediate goods producers) are responding aggressively to changes in fundamentals. Consequently, the RAC model can generate fast crashes and slow recoveries.

Endogenous stochastic volatility  An additional prediction which distills the lessons of the previous impulse response analysis is the following: the model endogenously generates higher volatility of output growth rates in lower states of the world. Such endogenous stochastic volatility is missing (up to the Jensen’s inequality correction) in the core RBC model, which features a log-linear relationship between output and the state.

Figure 7 plots the conditional volatility of GDP growth as a function of the underlying state in the calibrated model. One sees two important facts. First, the model hits the calibrated unconditional volatility “on average” across states. Second, there is a substantial amount of stochastic volatility in the model. A back-of-the-envelope calculation then reveals that if the model is to generate the 4.2% reduction in output consistent with the Great Recession, then volatility in the model would be up to 153% greater in the Great Recession.
than the boom preceding it. GARCH evidence for industrial production from the post-War era from Jurado et al. (2015) reveals that output volatility was roughly 125% greater in the Great Recession than normal times. As a result, the model with attention cycles that features endogenous stochastic volatility alone can account for the entirety of the stochastic volatility we observe in the time-series, despite the fact that this a non-targeted moment of our calibration.

5.4. **Attention Cycles Meet Uncertainty Shocks**

This article has discussed how demand for more accurately tracking the macro state should be higher in recessions. A large literature emphasizes that “information supply” may also dry up during recessions, primarily because information is postulated to be an external product of goods production or investment.\(^{30}\) Moreover, empirical evidence from Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2018) shows that the cross-sectional dispersion of firms’ productivity and many attributes widens during economic downturns, while in our baseline model dispersion shrinks.

A full combination of the two models is beyond the scope of our analysis. But we highlight here one immediate way in which these the “attention cycles” and “uncertainty cycles” stories interact with one another.

Random costs of attention  Let us assume that $\lambda$ itself is stochastic but common knowledge to agents. Propositions 2 and 4 in the approximated wage-rule economy already suggest how a positive “uncertainty shock,” or increase in the cost of paying attention $\lambda$, would behave in this model: it would increase the dispersion of actions around the optimum and depress output. Appendix Figure 15 verifies this numerically around the steady-state of our model.31

Conditional versus unconditional predictions  Imagine now the following caricature of the quantitative finding in Bloom et al. (2018) holds true: recessions are times of high uncertainty and low productivity. Concretely have $\lambda$ be a random variable that is negatively correlated with $\theta$, but otherwise common knowledge, to maintain the simplicity of the stochastic choice program.

Such a model would attenuate our unconditional predictions of higher aggregate volatility as a function of low output. In the new model, part of the explanation for low output would be that the cost of paying attention is high. Holding fixed productivity $\theta$, higher costs of paying attention would lower the volatility of aggregates. Moreover, if we had the tools for direct measurement of such an object, we may see a large “wedge” between productivity and the measured technology residual (as defined in Proposition 4 and shown in Figure 5) in an uncertainty-driven downturn.

But the model would generate the same conditional predictions as a function of fundamental productivity. Concretely, firms would respond more to negative than positive productivity shocks, and the former would induce higher instantaneous volatility in output, holding fixed the level of attention. An empirical test along these lines, separately identifying via sign restrictions this two-shock model, is a fascinating topic for future research.

6. Conclusion

This paper has investigated the interaction between the attention cycle and the business cycle. We first developed a new measure of firms’ attention to the macroeconomy and documented that attention is counter-cyclical. We built a macroeconomic theory to understand why cyclical attention should manifest in both partial and general equilibrium by adding stochastic choice to an otherwise standard Neoclassical RBC model. We first showed theoretically exactly how high curvature of profits leads to higher attention by firms and therefore that if the model’s equilibrium mapping from productivity to the curvature of profits is monotone decreasing that attention cycles should occur. In partial equilibrium we uncovered two persuasive reasons to suppose that curvature of firms’ profits in their own production should

31The only possible complication is the aforementioned discussion of the zero production constraint. In that context, higher dispersion could increase output.
be lowest when the state is lowest: aggregate demand externalities and risk-pricing via the SDF. We therefore argue that basic asset pricing forces and aggregate demand externalities explain the observed attention cycle. Importantly, we highlighted a general equilibrium force whereby precise attention to fundamentals and declining aggregates form a positive feedback loop.

Calibrating to our new facts a parsimonious, one-parameter extension of the RBC model to include stochastic choice, we uncover a quantitatively important role of cyclical attention on macroeconomic dynamics. In particular, cyclical attention generates asymmetrically larger propagation of negative shocks, larger propagation of all shocks in lower states of the world and endogenous stochastic volatility, whereby volatility of GDP growth is highest when productivity and output are lowest.

Our aim in this paper, from the start, has been to fully explicate a single pathway in the economy to explore the causes and consequences of attention cycles. But there remains much to explore in linking this mechanism with other sources of positive and negative feedback in the economy. One angle, which has already been commented upon in the quantitative section, is integrating together this model of “information demand” with a plausible, and micro-calibrated, model of “information supply” that captures the possibility that precise information about the macroeconomy is simply harder to generate in low-activity periods.

A second, more speculative, point relates to the extent to which narratives may beget themselves via non-economic pathways: in different terms, looking for externalities outside of goods production. In the context of the economics narratives literature owing to Shiller (2017, 2019), the “macro narrative” identified here is fairly coarse—indeed, it is probably an umbrella terminology for a myriad more specific stories for how the macroeconomy will evolve. A theory of cyclical attention begs for a module exploring the extent to which macro attention is self-fulfilling purely as a story in the social consciousness. More formally this could involve a marriage of the stochastic choice model here, or an information acquisition enrichment thereof, with classic models of herding and social learning. This article’s framework suggests that the economic forces pushing individuals to pay attention to the macroeconomy during a downturn may greatly accelerate the virality of narratives because they become more essential for everyone to understand. This is a promising frontier for future work.

\[\text{32} \text{One such model that marries social learning with rational inattention is Caplin, Leahy, and Matějka (2015).}\]
References


Appendices

A. Omitted Proofs

A.1. Proof of Proposition 1

Proof. To prove existence, we first study the problem of a single firm $i$ who is best replying to the conjecture that the law of motion of the aggregate is $\Lambda : \Theta \rightarrow \mathbb{R}$ for fixed $\theta_{t-1}$. See that this agent faces the problem:

$$p^*(\Lambda, \theta_{t-1}) \in \arg \max_{p \in \mathcal{P}} \int_{\Theta} \int_{\mathcal{X}} \bar{\Pi}(x, \Lambda(\theta; \theta_{t-1}), \theta) p(x | \theta) \, dx \, \pi(\theta | \theta_{t-1}) \, d\theta - c(p)$$

(34)

First, let us examine the set of stochastic choice rules:

$$\mathcal{P} = \{p : \Theta \rightarrow \Delta(\mathcal{X})\} = \prod_{\theta \in \Theta} \Delta(\mathcal{X})$$

(35)

See that $\Delta(\mathcal{X})$ is compact as $\mathcal{X}$ is compact. It therefore follows by Tychonoff’s theorem that $\mathcal{P}$ is compact. Define $k : \mathcal{P} \times \mathcal{B} \rightarrow \bar{\mathbb{R}}$, where $\mathcal{B} = \{\hat{\Lambda} : \Theta \rightarrow \mathbb{R}\}$ as:

$$k(p, \hat{\Lambda}) = \int_{\Theta} \int_{\mathcal{X}} \bar{\Pi}(x, \Lambda(\theta; \theta_{t-1}), \theta) p(x | \theta) \, dx \, \pi(\theta | \theta_{t-1}) \, d\theta - c(p)$$

(36)

See that $k$ is jointly lower-semicontinuous in $(p, \Lambda)$ by owing to continuity of $\bar{\Pi}$ and lower semicontinuity of $c$. Thus, by Berge’s theorem, the maximizing correspondence $\mathcal{P}^*(\Lambda)$ is upper hemicontinuous in $\Lambda$. Moreover, we have that $c$ is strictly convex and therefore that $\mathcal{P}^*(\hat{\Lambda})$ is both single-valued and continuous. It immediately follows that in any equilibrium $P_i^* = P^* = \mathcal{P}^*(\hat{\Lambda})$ for all $i$ and thus that there cannot exist asymmetric equilibria.

To show existence of an equilibrium it now suffices to show that there exists a $\Lambda$ such that:

$$\Lambda = X \circ \mathcal{P}^*(\Lambda)$$

(37)

To this end define the operator $T : \mathcal{B} \rightarrow \mathcal{B}$ such that:

$$T(\Lambda) = X \circ \mathcal{P}^*(\Lambda)$$

(38)

We wish to show that $T$ has a fixed point. First, see that $\mathcal{B}$ is the space of bounded functions from a compact set to the reals and is therefore both convex and compact. Thus $\mathcal{B}$ is a compact subset of itself. Second, we have established that $\mathcal{P}^*$ is continuous and $X$ is
continuous by assumption. Thus, $T$ is a continuous map from a compact, convex set to a compact set contained in its domain. By Schauder’s fixed point theorem, there exists a fixed point $\hat{X} = T(\hat{X})$. Thus an equilibrium exits. Moreover, we already established that there cannot exist an asymmetric equilibrium.

To show uniqueness and monotonicity, we now use the approximations and optimal policy derived in Proposition 2. In particular, in any state $\theta$, we know that aggregate output must solve the fixed point equation above, which can be expressed as:

$$X(\theta) = \mathbb{E}_{p^*}(X(\theta), \theta)[x] - \frac{1}{2\epsilon x^*(X(\theta), \theta)} \nabla_{p^*}(X(\theta), \theta)[x]$$  \hspace{1cm} (39)

Recall that the optimal policy is moreover given by:

$$N \left( x^*(X(\theta), \theta), \frac{\lambda}{|\Pi_{xx}(X(\theta), \theta)|} \right)$$  \hspace{1cm} (40)

where:

$$x^*(X, \theta) = v_x X^{1-\chi} \theta^\epsilon$$

$$|\Pi_{xx}(X, \theta)| = v_\Pi X^{-1-\gamma+\chi(1+\epsilon)} \theta^{-1-\epsilon}$$  \hspace{1cm} (41)

for constants $v_x, v_\Pi > 0$ given by:

$$v_x := \left( 1 - \frac{1}{\epsilon} \right)^\epsilon \bar{\omega}^{-\epsilon} \bar{X}^\chi$$

$$v_\Pi := (\epsilon - 1)^{-\epsilon} (1+\epsilon)^{-1} \bar{\omega}^{1+\epsilon} \bar{X}^{-\chi(1+\epsilon)}$$  \hspace{1cm} (42)

It follows that equilibrium output solves:

$$X(\theta) = v_x X(\theta)^{1-\chi} \theta^\epsilon - \frac{\lambda v_x v_\Pi}{2\epsilon} X(\theta)^{\gamma-\chi} \theta$$  \hspace{1cm} (43)

Note that there is always a trivial equilibrium $X = 0$ arising from our approximations. Toward proving uniqueness of non-trivial equilibrium, define:

$$g(X, \theta) = a_0 X^{1-\chi} \theta^\epsilon - a_1 X^{\gamma-\chi} \theta$$  \hspace{1cm} (44)

where $a_0 = v_x > 0$ and $a_1 = \frac{\lambda v_x v_\Pi}{2\epsilon} > 0$. We now compute the derivatives of this function in
\[ g_X(X, \theta) = a_0 (1 - \chi \epsilon) X^{-\chi' \theta} - a_1 (\gamma - \chi) X^{\gamma - \chi - 1} \theta \]
\[ g_{XX}(X, \theta) = -a_0 (1 - \chi \epsilon) \chi \epsilon X^{-\chi' \theta} - a_1 (\gamma - \chi) (\gamma - \chi - 1) X^{\gamma - \chi - 2} \theta \]

Note that if \( 1 - \chi \epsilon > 0 \) and \( \gamma > 1 + \chi \) that:
\[ \lim_{X \to 0} g_X(X, \theta) = +\infty \quad \lim_{X \to \infty} g_X(X, \theta) = -\infty \]

Moreover, if \( \gamma > \chi + 1 \) we have that \( g_{XX}(X, \theta) < 0 \) on \((0, \infty)\). Thus, when \( \gamma > \chi + 1 \) and \( \chi \epsilon < 1 \), \( g(X, \theta) \) crosses \( X \) from above and there exists a unique fixed point.

We now show monotonicity of the fixed point. To this end, we implicitly differentiate the fixed point condition:
\[ \frac{dX(\theta)}{d\theta} = \left[ a_0 (1 - \chi \epsilon) X(\theta)^{-\chi \epsilon \theta} - a_1 (\gamma - \chi) X(\theta)^{\gamma - \chi - 1} \theta \right] \frac{dX(\theta)}{d\theta} + \left[ a_0 \chi \epsilon X(\theta) - a_1 X(\theta)^{\gamma - \chi} \right] \]

Yielding:
\[ \frac{dX(\theta)}{d\theta} = \frac{a_0 \epsilon X(\theta)^{1 - \chi \epsilon \theta} - a_1 X(\theta)^{\gamma - \chi}}{1 - \left[ a_0 (1 - \chi \epsilon) X(\theta)^{-\chi \epsilon \theta} - a_1 (\gamma - \chi) X(\theta)^{\gamma - \chi - 1} \theta \right]} \]

Multiplying both sides by a factor of \( \frac{\theta}{X} \):
\[ \frac{d \log X(\theta)}{d \log \theta} = \frac{a_0 \epsilon X(\theta)^{1 - \chi \epsilon \theta} - a_1 X(\theta)^{\gamma - \chi}}{X(\theta) - \left[ a_0 (1 - \chi \epsilon) X(\theta)^{-\chi \epsilon \theta} - a_1 (\gamma - \chi) X(\theta)^{\gamma - \chi - 1} \theta \right]} \]

Consider first the numerator (as \( \epsilon > 1 \)):
\[ a_0 \epsilon X(\theta)^{1 - \chi \epsilon \theta} - a_1 X(\theta)^{\gamma - \chi} = a_0 (\epsilon - 1) X(\theta)^{1 - \chi \epsilon \theta} + X(\theta) > 0 \]

To show \( \frac{d \log X(\theta)}{d \log \theta} > 0 \) it now suffices to show that the denominator is positive:
\[ X(\theta) = a_0 X(\theta)^{1 - \chi \epsilon \theta} - a_1 X(\theta)^{\gamma - \chi} \theta \geq \begin{cases} \ a_0 (1 - \chi \epsilon) X(\theta)^{1 - \chi \epsilon \theta} - a_1 (\gamma - \chi) X(\theta)^{\gamma - \chi} \theta & \chi \epsilon \in (0,1) \\ \ a_0 (1 - \chi \epsilon) X(\theta)^{1 - \chi \epsilon \theta} - a_1 (\gamma - \chi) X(\theta)^{\gamma - \chi} \theta & \gamma - \chi > 1 \end{cases} \]

Completing the proof. \( \square \)
A.2. Proof of Proposition 2

Proof. We first derive the quadratic approximation of both the firm’s objective and the production function of the final goods firm. The firms’ profits in utility units for the representative household are given by:

$$\Pi(x_{it}; (X_t, w_t), \theta_t) := X_t^{-\gamma} \left( x_{it}^{1-\frac{1}{\epsilon}} X_t^{\frac{1}{\epsilon}} - x_{it} w_t \theta_t \right)$$  \hspace{1cm} (52)

The optimal action in the absence of stochastic choice solves the FOC:

$$\left(1 - \frac{1}{\epsilon}\right) x_{it}^{\frac{1}{\epsilon}} X_t^{\frac{1}{\epsilon}} = \frac{w_t}{\theta_t}$$  \hspace{1cm} (53)

yielding:

$$x_{it}^* = \left(1 - \frac{1}{\epsilon}\right)^{-\epsilon} X_t \left(\frac{w_t}{\theta_t}\right)^{-\epsilon} \hspace{1cm} (54)$$

We now approximate the firm’s profit function to second order in $x_{it}$ around $x_{it}^*$:

$$\Pi(x_{it}; (X_t, w_t), \theta_t) = \Pi(x_{it}^*; (X_t, w_t), \theta_t) + \Pi_x(x_{it}; (X_t, w_t), \theta_t)|_{x_{it}=x_{it}^*}(x - x_{it}^*)$$

$$+ \frac{1}{2} \Pi_{xx}(x_{it}; (X_t, w_t), \theta_t)|_{x_{it}=x_{it}^*}(x - x_{it}^*)^2 + O^3(x_{it}) \hspace{1cm} (55)$$

Application of the envelope theorem implies that:

$$\Pi_x(x_{it}; (X_t, w_t), \theta_t)|_{x_{it}=x_{it}^*} = 0 \hspace{1cm} (56)$$

Thus, our approximation reduces to the quadratic utility function in the Linear-Quadratic equilibrium:

$$\Pi(x_{it}^*; (X_t, w_t), \theta_t) = \Pi(x_{it}^*; (X_t, w_t), \theta_t)$$

$$+ \frac{1}{2} \Pi_{xx}(x_{it}^*; (X_t, w_t), \theta_t)|_{x_{it}=x_{it}^*}(x - x_{it}^*)^2 \hspace{1cm} (57)$$

It remains to characterize the intercept and curvature. We first characterize the intercept:

$$\Pi(x_{it}^*; (X_t, w_t), \theta_t) = X_t^{-\gamma} \left( \frac{w_t}{\theta_t} \right)^{1-\epsilon} \left( \left[ 1 - \frac{1}{\epsilon} \right]^{(1-\frac{1}{\epsilon})} - \left[ 1 - \frac{1}{\epsilon} \right]^{\epsilon} \right)$$

$$= X_t^{-\gamma} \left( \frac{w_t}{\theta_t} \right)^{1-\epsilon} \epsilon^{-\epsilon} (\epsilon - 1)^{\epsilon-1} \hspace{1cm} (58)$$
We now characterize the curvature:

\[ \Pi_{xx}(x_{it}; (X_t, w_t), \theta_t)|_{x_{it}=x_{it}^*} = X_t^{-\gamma} \cdot (B_{xx}(x_{it}^*, X_t) - C_{xx}(x_{it}^*, w_t, \theta_t)) \] (59)

where we have that:

\[ B_{xx}(x_{it}^*, X_t) = -\frac{1}{\epsilon} \left(1 - \frac{1}{\epsilon}\right) \left(x_{it}^* - \frac{1}{\epsilon} X_t^\frac{1}{\epsilon^2} \right) \]

\[ = -\frac{1}{\epsilon} \left(1 - \frac{1}{\epsilon}\right) \left(1 - \frac{1}{\epsilon}\right)^{-\frac{1+\epsilon}{\epsilon}} X_t^{-\frac{1}{\epsilon} - \frac{1}{\epsilon^2}} \left(\frac{w_t}{\theta_t}\right)^{\frac{1}{1+\epsilon}} \] (60)

\[ C_{xx} \equiv 0. \]

and \( C_{xx} \equiv 0. \) We now substitute for the wage rule:

\[ w_t = \bar{w} \cdot \left(\frac{X_t}{X}\right)^{\chi} \] (61)

This yields the following equations for the optimal action and curvature:

\[ x_{it}^* = \left(1 - \frac{1}{\epsilon}\right) \bar{w}^{-\epsilon} X^\chi X_t^{1-\chi^\epsilon} \theta_t^\epsilon \] (62)

\[ \Pi_{xx}(x_{it}; (X_t, w_t), \theta_t)|_{x_{it}=x_{it}^*} = -(\epsilon - 1)^{-\epsilon} \bar{w}^{1+\epsilon} X_t^{(1+\epsilon)^{1-\epsilon}} X_t^{1-\gamma + \chi(1+\epsilon)} \]

Turning to the aggregator we have that:

\[ X_t = X(\{x_{it}\}_{i\in[0,1]}) = \left(\int_{[0,1]} \frac{x_{it} - 1}{x_{it} \cdot dx} \right)^\frac{1}{\epsilon^2} \] (63)

Taking a quadratic approximation around \( x_{it}^* = x_{it}^* \) yields:

\[ X(\{x_{it}\}_{i\in[0,1]}) \approx x_t^* + DX(x^*) \cdot (x_t - x_t^*) + \frac{1}{2} (x_t - x_t^*)' HX(x^*) (x_t - x_t^*) \] (64)

where \( D \) is the Jacobian operator and \( H \) Hessian operator. We moreover obtain that:

\[ DX(x^*)(i) = (x_t^*)^{-\frac{1}{\epsilon}} X_t^\frac{1}{\epsilon} = 1 \] (65)

and:

\[ HX(x^*)(i, j) = \begin{cases} -\frac{1}{\epsilon x_t^\epsilon}, & i = j, \\ \frac{1}{\epsilon x_t^\epsilon}, & i \neq j. \end{cases} \] (66)
Noting finally that firms’ actions are independent conditional on \( \theta \) reveals that the cross terms in the above expansion average to zero as all covariances are identically zero. Thus:

\[
X(\{x_{it}\}_{i \in [0,1]}) \approx x^*_t + \mathbf{\bar{\Pi}} \cdot (x_{it} - x^*_t) - \frac{1}{2 \epsilon x^*_t} (x_{it} - x^*_t)'I(x_{it} - x^*_t)
\]  

(67)

Yielding the following mean-variance form:

\[
X_t \approx \int_{[0,1]} x_{it} \, di - \frac{1}{2 \epsilon x^*_t} \int_{[0,1]} (x_{it} - x^*_t)^2 \, di
\]  

(68)

We now characterize the optimal policy for firms in any equilibrium. To this end, note that for an arbitrary function \( u : \mathcal{X} \times \Theta \rightarrow \mathbb{R} \) and prior \( \pi \in \Delta(\Theta) \) that the stochastic choice problem is given by:

\[
\max_{p \in \mathcal{P}} \int_{\mathcal{X}} \int_{\Theta} u(x, \theta) p(x|\theta) \, dx \, d\pi(\theta) - \lambda \int_{\Theta} \int_{\mathcal{X}} p(x|\theta) \log p(x|\theta) \, dx \, d\pi(\theta)
\]  

(69)

This can be formulated as a Lagrangian with the following form: (\( \kappa(x, \theta) \) are the non-negativity constraints and \( \gamma(\theta) \) are the constraints that all action distributions integrate to unity)

\[
\mathcal{L}(\{p(x|\theta), \kappa(x, \theta)\}_{x \in \mathcal{X}, \theta \in \Theta}, \{\gamma(\theta)\}_{\theta \in \Theta}) = \int_{\Theta} \int_{\mathcal{X}} u(x, \theta) p(x|\theta) \, dx \, d\pi(\theta)
\]

\[
- \lambda \int_{\Theta} \int_{\mathcal{X}} p(x|\theta) \log p(x|\theta) \, dx \, d\pi(\theta)
\]

\[
+ \kappa(x, \theta)p(x|\theta) + \gamma(\theta) \left( \int_{\mathcal{X}} p(x|\theta) \, dx - 1 \right)
\]  

(70)

Taking the first-order condition of this program with respect to \( p(x|\theta) \) yields:

\[
u(x, \theta) - \lambda(\log p(x|\theta) + 1) + \kappa(x, \theta) + \gamma(\theta) = 0
\]  

(71)

Which can be re-arranged as:

\[
p(x|\theta) \propto \exp\{\lambda^{-1} u(x, \theta)\}
\]  

(72)

In the linear quadratic equilibrium, we can simply replace \( u \) in the above with \( \bar{\Pi} \) evaluated at equilibrium aggregate output. Thus:

\[
p(x|\theta) \propto \exp \left\{ - \frac{1}{2} \frac{\|\Pi_{xx}(X(\theta), \theta)\|}{\lambda} (x - x^*(X(\theta), \theta))^2 \right\}
\]  

(73)
Which can be rewritten as:

\[
p(x|\theta) \propto \exp \left\{ -\frac{1}{2} \left( \frac{x - x^*(X(\theta), \theta)}{\sqrt{|\Pi_{xx}(X(\theta), \theta)|}} \right)^2 \right\}
\]

(74)

Taking \( X = \mathbb{R} \), it is then immediate that \( p(x|\theta) \) is given by the following Gaussian form:

\[
p(x|\theta) = \frac{1}{\sqrt{2\pi |\Pi_{xx}(X(\theta), \theta)|}} \exp \left\{ -\frac{1}{2} \left( \frac{x - x^*(X(\theta), \theta)}{\sqrt{|\Pi_{xx}(X(\theta), \theta)|}} \right)^2 \right\}
\]

(75)

Which can be expressed more compactly as that actions are distributed:

\[
N \left( x^*(X(\theta), \theta), \frac{\lambda}{|\Pi_{xx}(X(\theta), \theta)|} \right)
\]

(76)

Thus completing the proof.

\[\square\]

A.3. Proof of Proposition 3

Proof. Recall that the firm action distribution in any state \( \theta \) has PDF given by:

\[
p(x|\theta) = \frac{1}{\sqrt{2\pi |\Pi_{xx}(X(\theta), \theta)|}} \exp \left\{ -\frac{1}{2} \left( \frac{x - x^*(X(\theta), \theta)}{\sqrt{|\Pi_{xx}(X(\theta), \theta)|}} \right)^2 \right\}
\]

(77)

It is well known that a Gaussian random variable with any mean and variance \( \sigma^2 \) has entropy \( h(\sigma^2) = \frac{1}{2} \ln(2\pi e \sigma^2) \). Thus the entropy in state \( \theta \) is given by:

\[
h(X(\theta), \theta) = \frac{1}{2} \ln \left( 2\pi e \frac{\lambda}{|\Pi_{xx}(X(\theta), \theta)|} \right)
\]

(78)

Attention is then monotone decreasing in \( \theta \) in equilibrium if and only \( h(X(\theta), \theta) \) is increasing in \( \theta \). To this end, it suffices to show that \( |\Pi_{xx}(X(\theta), \theta)| \) is decreasing in \( \theta \). Recall that this is given by:

\[
|\Pi_{xx}(X(\theta), \theta)| = v_{\Pi} X(\theta)^{-1-\gamma+\chi(1+\epsilon)} \theta^{-1-\epsilon}
\]

(79)

where \( v_{\Pi} > 0 \). By proposition 1, we have that \( X(\theta) \) is monotone increasing in \( \theta \). It therefore suffices to show that:

\[
\gamma > \chi(1+\epsilon) - 1
\]

(80)
We know that $\gamma > \chi + 1$ and $\chi \epsilon < 1$ by hypothesis. These together imply the above, completing the proof.

### A.4. Proof of Proposition 4

**Proof.** We first derive output in the fully attentive $\lambda = 0$ limit. Recall the fixed-point equation for output from Proposition 1:

$$X(\theta; \lambda) = a_0 X(\theta; \lambda)^{1-\chi \epsilon} - a_1(\lambda) X(\theta; \lambda)^{\gamma-\chi \epsilon}$$  \hspace{1cm} (81)

When $\lambda = 0$, we have that $a_1(0) = 0$. Thus:

$$X(\theta; 0) = a_0 X(\theta; \lambda)^{1-\chi \epsilon}$$  \hspace{1cm} (82)

Or simply:

$$X(\theta; 0) = a_0^{\frac{1}{1-\chi \epsilon}}$$  \hspace{1cm} (83)

Now define the proportional wedge between equilibrium output and output without the attention friction as:

$$X(\theta; \lambda) := a_0^{\frac{1}{1-\chi \epsilon}} W(\theta; \lambda)$$  \hspace{1cm} (84)

We can rewrite the fixed point equation in terms of the wedge:

$$a_0^{\frac{1}{1-\chi \epsilon}} W(\theta; \lambda) W(\theta; \lambda)^{\gamma-\chi \epsilon} = a_0^{\frac{1}{1-\chi \epsilon}} W(\theta; \lambda)^{1-\chi \epsilon} - a_1(\lambda) a_0^{\frac{\gamma-\chi \epsilon}{1-\chi \epsilon}} W(\theta; \lambda)^{\gamma-\chi \epsilon}$$  \hspace{1cm} (85)

Rearranging we obtain:

$$W(\theta; \lambda) = W(\theta; \lambda)^{1-\chi \epsilon} - a_1(\lambda) a_0^{\frac{\gamma-\chi \epsilon}{1-\chi \epsilon}} W(\theta; \lambda)^{\gamma-\chi \epsilon}$$  \hspace{1cm} (86)

Toward the claims in the proposition, first observe that output can be rewritten in the claimed form:

$$\log X(\log \theta, \lambda) = X_0 + \chi^{-1} \log \theta + \log W(\log \theta, \lambda)$$  \hspace{1cm} (87)

where $X_0 := \frac{1}{\chi \epsilon} \log a_0$. We first prove that the wedge is positive. To this end, see that the wedge is positive and unique under the exact same conditions that $X(\theta)$ is positive and unique: $\chi \epsilon < 1$ and $\gamma > \chi + 1$. Moreover, $W(\theta; \lambda)$ crosses the 45 degree line from above. To show that $W(\theta; \lambda) \leq 1$, it then suffices to show that the RHS of the fixed point equation is less than unity when evaluated at $W(\theta; \lambda) = 1$. As $a_1(\lambda), a_0 > 0$, this is immediate. Thus $\log W(\theta; \lambda) \leq 0$, as claimed. By construction, we have that $W(\theta; 0) = 1$ so that
log \(W(\theta; 0) = 0\). It is moreover immediate from the fact that \(a_1'(\lambda) > 1\) that \(W_\lambda(\theta; \lambda) < 0\).

Toward the final claim, we show that \(\log W(\lambda; \theta)\) is monotone decreasing in \(\log \theta\). First, we implicitly differentiate the fixed point condition:

\[
\frac{dW}{d\theta} = \left[(1 - \chi\epsilon) W(\theta; \lambda)^{-\chi\epsilon} - a_1(\lambda) a_0 \frac{\gamma - 1}{\chi} W(\theta; \lambda)^{\gamma - 1} \theta^{\frac{\gamma - 1}{\chi}}\right] \frac{dW}{d\theta} \tag{88}
\]

Or:

\[
\frac{dW}{d\theta} = \frac{-a_1(\lambda) a_0 \frac{\gamma - 1}{\chi} W(\theta; \lambda)^{\gamma - 1} \theta^{\frac{\gamma - 1}{\chi}}}{1 - \left[(1 - \chi\epsilon) W(\theta; \lambda)^{-\chi\epsilon} - a_1(\lambda) a_0 \frac{\gamma - 1}{\chi} W(\theta; \lambda)^{\gamma - 1} \theta^{\frac{\gamma - 1}{\chi}}\right]} \tag{89}
\]

Which we can rewrite as:

\[
\frac{d \log W}{d \log \theta} = \frac{-a_1(\lambda) a_0 \frac{\gamma - 1}{\chi} W(\theta; \lambda)^{\gamma - 1} \theta^{\frac{\gamma - 1}{\chi}}}{W(\theta; \lambda) - \left[(1 - \chi\epsilon) W(\theta; \lambda)^{1 - \chi\epsilon} - a_1(\lambda) a_0 \frac{\gamma - 1}{\chi} W(\theta; \lambda)^{\gamma - 1} \theta^{\frac{\gamma - 1}{\chi}}\right]} \tag{90}
\]

By positivity of \(a_1(\lambda)\) and \(a_0\) and the hypothesis that \(\gamma > \chi + 1\), the numerator of this expression is negative. To show that the wedge is monotone decreasing, we need to show that the denominator is positive. To this end, we see that:

\[
W(\theta; \lambda) = W(\theta; \lambda)^{1 - \chi\epsilon} - a_1(\lambda) a_0 \frac{\gamma - 1}{\chi} W(\theta; \lambda)^{\gamma - 1} \theta^{\frac{\gamma - 1}{\chi}}
\geq \begin{cases} \chi\epsilon \in (0, 1) & W(\theta; \lambda)^{1 - \chi\epsilon} - a_1(\lambda) a_0 \frac{\gamma - 1}{\chi} W(\theta; \lambda)^{\gamma - 1} \theta^{\frac{\gamma - 1}{\chi}} \\ \gamma > \chi + 1 & \end{cases} \tag{91}
\]

Completing the proof.

\[\square\]

A.5. Proof of Corollary 1

Proof. Observe that the labor demand of any given firm \(i\) is given by:

\[
l_{it} = \frac{x_{it}}{\theta_t} \tag{92}
\]

Total labor demand in the economy is then given by:

\[
L_t = \int_{[0,1]} l_{it} \, di = \frac{1}{\theta_t} \int_{[0,1]} x_{it} \, di \tag{93}
\]
From Proposition 2, we know that:

\[ \int_{[0,1]} x_i d\theta = x^\epsilon(X(\theta_t), \theta_t) = v_x X(\theta_t)^{1-\chi \epsilon} \theta_t^\gamma \]  

(94)

Thus:

\[ L_t = v_x X(\theta_t)^{1-\chi \epsilon} \theta_t^{\gamma-1} \]  

(95)

Aggregate productivity is simply the Solow residual in this economy and is straightforwardly given by:

\[ \log A(\theta) = \log X(\theta) - \log L(\theta) \]
\[ = \log X(\theta) - \log v_x - (1 - \chi \epsilon) \log X(\theta) + (\epsilon - 1) \log \theta \]
\[ = -\log v_x + \chi \epsilon \log X(\theta) + (\epsilon - 1) \log \theta \]  

(96)

Plugging in our representation of aggregate output from Proposition 4 we obtain:

\[ \log A(\theta; \lambda) = (\chi \epsilon X_0 - \log v_x) + \log \theta + \chi \epsilon \log W(\log \theta, \lambda) \]  

(97)

where \( W(\cdot) \) inherits all of the properties proved in Proposition 4. Defining \( A_0 := (\chi \epsilon X_0 - \log v_x) \) completes the proof.

A.6. Proof of Corollary 2

Proof. First, recall from Proposition 4 the expression for output:

\[ \log X(\log \theta, \lambda) = X_0 + \chi^{-1} \log \theta + \log W(\log \theta, \lambda) \]  

(98)

Recall also that when \( \lambda = 0 \) that \( \log W(\log \theta, \lambda) = 0 \). Thus the model in the absence of the friction generates state-independent amplification and volatility of output. First, consider the response of output to a small shock \( \nu_t \) starting from \( \theta_{t-1} \). We have immediately that:

\[ \frac{\partial \log X(\log \theta)}{\partial \log \theta} \bigg|_{\theta=\theta_{t-1}} = \chi^{-1} + \frac{\partial \log W(\log \theta)}{\partial \log \theta} \bigg|_{\theta=\theta_{t-1}} \]  

(99)

Where we moreover note that \( \frac{\partial \log W(\log \theta)}{\partial \log \theta} \bigg|_{\theta} \) is a non-linear function of \( \theta \):

\[ \frac{d \log W}{d \log \theta} = \frac{-a_1(\lambda) a_0 \frac{2-\chi}{\chi} \gamma \epsilon W(\theta; \lambda) W(\theta; \lambda)^{\gamma-1} \theta^{\gamma-1}}{W(\theta; \lambda) - (1 - \chi \epsilon) W(\theta; \lambda)^{1-\chi \epsilon} - a_1(\lambda) a_0 \frac{2-\chi}{\chi} \gamma \epsilon \theta^{\gamma-1} \theta^{\gamma-1}} \]  

(100)
Toward showing endogenous stochastic volatility, simply approximate to first order in \( \eta \) output:

\[
\log X(\nu_t; \theta_{t-1}) \approx \tilde{X}(\theta_{t-1}) + \frac{\log \partial X(\log \theta)}{\partial \log \theta}|_{\theta=\theta_{t-1}} \nu_t
\]  

(101)

Taking the variance of this expression conditional on \( \theta_{t-1} \) yields:

\[
\text{Var}(\log X_t|\theta_{t-1}) = \left( \chi^{-1} + \frac{\partial \log W(\log \theta)}{\partial \log \theta}|_{\theta=\theta_{t-1}} \right)^2 \sigma^2_{\theta}
\]  

(102)

Taking a second-order approximation in \( \nu_t \) and differentiating yields the asymmetric shock propagation equation:

\[
\frac{\partial \log X(\log \theta)}{\partial \log \theta}|_{\theta=\theta_{t-1}} = \chi^{-1} + \frac{\partial \log W(\log \theta)}{\partial \log \theta}|_{\theta=\theta_{t-1}} + \frac{\partial^2 \log W(\log \theta)}{\partial \log^2 \theta}|_{\theta=\theta_{t-1}} \nu_t
\]  

(103)

Completing the proof.

\[\square\]

**B. Attention Cycles with Rational Inattention**

In this section, we show how the core logic of attention cycles carries over to a setting with information acquisition and optimal signal processing. To do this, we consider the class of posterior-separable cost functionals. Denti (2020) provides this formulation as a representation theorem in stochastic choice space of the usual posterior-based definition of Caplin and Dean (2013):

**Definition 4 (Posterior-Separable Cost Functionals).** A cost functional \( c \) has a posterior-separable representation if and only if there exists a convex and continuous \( \hat{\phi} \) such that:

\[
c(p) = \int_X \hat{\phi}(\{p(x|\theta)\}_{\theta \in \Theta}) \, dx
\]  

(104)

where:

\[
\hat{\phi}(\{p(x|\theta)\}_{\theta \in \Theta}) = p(x)\phi \left( \left\{ \frac{p(x|\theta)\pi(\theta)}{p(x)} \right\} \right)_{\theta \in \Theta}
\]  

(105)

whenever \( p(x) > 0 \) and \( \hat{\phi} = 0 \) otherwise.

Intuitively, such a cost functional considers the cost to the agent of arriving at any given posterior and adds that up over the distribution of posteriors that are realized. Important cost functionals such as the mutual information cost functional considered in the literature on rational inattention are members of this class. Indeed, mutual information is the special case of the above where \( \phi \) returns the entropy of the distribution that is its argument.
The mathematical structure of posterior-separable cost functionals does not admit the same prior-independence property as likelihood-separable cost functionals. As a result, we will not be able to carry all of our results over to this setting. Nevertheless, as we will argue, the key qualitative forces apply.

In the setting with likelihood separable choice in the single-agent context, we showed that greater curvature of payoffs leads to more precise actions (Proposition 2). With posterior-separable choice, the above result does not hold in general. This is because the prior also influences the states in which the agent would like to learn precisely. In particular, even if a state features high curvature, if it is unlikely to arise, the agent may not care to acquire precise information in that state. A particular case where this complication can be bypassed is when costs are given by mutual information and all actions are exchangeable in the prior in the sense that all actions are ex ante equally attractive (Matějka and McKay, 2015). This is a natural case to consider and yields a particularly revealing structure to the optimal policy: the agent’s actions in state \( \theta \) are given by a normal distribution centered on the objective optimum and with variance inversely proportional to the curvature of their objective in that state – a normal mixture model.

**Proposition 5.** Suppose that \( u(x, \theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2 \) and costs are posterior separable with entropy kernel \( \lambda \phi(\cdot) \) for some \( \lambda > 0 \). If all actions are exchangeable in the prior, then in the limit of the support of the action set to infinity, \( \hat{x} \to \infty \) for \( \pi = -\bar{x} = \hat{x} \), the optimal stochastic choice rule is given by:

\[
p(x|\theta) = \frac{1}{\sqrt{\pi \lambda \beta(\theta)}} \exp \left\{ -\frac{1}{2} \left( \frac{x - \gamma(\theta)}{\sqrt{\lambda \beta(\theta)}} \right)^2 \right\}
\]

(107)

Which is to say that the agent’s actions follow a normal mixture model with conditional action density given by:

\[
x|\theta \sim N(\gamma(\theta), \frac{\lambda}{2\beta(\theta)})
\]

(108)

**Proof.** We first show that mutual information can be written in the claimed stochastic choice form. These arguments follow closely Matějka and McKay (2015) and Denti (2020). The agent can design an arbitrary signal space \( S \) and choose a joint distribution between signals and states \( g \in \Delta(S \times \Theta) \). As in Sims (2003), the mutual information is the reduction in
entropy from having access to this signal relative to the prior:

\[ I(g) = \int_S \int_\Theta g(s, \theta) \log \left( \frac{g(s, \theta)}{\pi(\theta) \int_\Theta g(s, \theta) d\theta} \right) d\theta ds \tag{109} \]

We now argue that it is without loss to consider a choice over stochastic choice rules \( p : \Theta \to \Delta(X) \). Suppose \( x \) is an optimal action conditional on receiving any \( s \in S_x \). Suppose that there exist \( S^1_x, S^2_x \subseteq S_x \) of positive measure such that \( g(\theta|s_1) \neq g(\theta|s_2) \) for all \( s_1 \in S^1_x, s_2 \in S^2_x \). Now generate a new signal structure \( g' \) such \( \tilde{s} \in S^1_x \cup S^2_x \) is sent whenever any \( s \in S^1_x \cup S^2_x \) was sent under \( g \). Clearly, \( x \) is optimal conditional on receiving \( \tilde{s} \). Thus, expected payoffs under \( g' \) are the same as those under \( g \). Moreover, \( g' \) is simply a garbling of \( g \) in the sense of Blackwell. Thus \( C(g') < C(g) \) for any convex cost functional \( C \). As \( I \) is convex, this is a contradiction. Thus, there must be at most one posterior (realized with positive density) associated with each action. As \( g(s, \theta) = g(s|\theta)\pi(\theta) \), the choice of \( g(s, \theta) \in \Delta(S \times \Theta) \) is a choice over \( g(\cdot|\cdot) : \Theta \to \Delta(S) \). Moreover, there is a unique posterior \( \mu(\theta|s) \) associated with each (non-dominated) action which is determined exactly by \( g(\cdot|\cdot) \). Hence, the agent directly chooses a mapping \( p(\cdot|\cdot) : \Theta \to \Delta(X) \). The agent’s problem can then be directly re-written in the claimed stochastic choice form for some \( c_I \):

\[ \max_{P \in \mathcal{P}} \int_\Theta \int_X u(x, \theta) dP(x|\theta) d\pi(\theta) - c_I(P) \tag{110} \]

Moreover, separating terms, one achieves the following representation of \( c_I \):

\[ c_I(p) = \int_\Theta \int_X p(x|\theta) \log p(x|\theta) dx d\pi(\theta) - \int_X p(x) \log p(x) dx \tag{111} \]

where:

\[ p(x) = \int_\Theta p(x|\theta) d\pi(\theta) \tag{112} \]

The stochastic choice problem can now be expressed by the Lagrangian: \( (\kappa(x, \theta) \) are the non-negativity constraints and \( \gamma(\theta) \) are the constraints that all action distributions integrate to unity)

\[
\mathcal{L}(\{ p(x|\theta), \kappa(x, \theta) \}_{x \in X, \theta \in \Theta}, \{ \gamma(\theta) \}_{\theta \in \Theta}) = \int_\Theta \int_X u(x, \theta)p(x|\theta) dx d\pi(\theta) \\
- \lambda \left( - \int_X p(x) \log p(x) dx + \int_\Theta \int_X p(x|\theta) \log p(x|\theta) dx d\pi(\theta) \right) + \kappa(x, \theta)p(x|\theta) + \gamma(\theta) \left( \int_X p(x|\theta) dx - 1 \right) \tag{113}
\]
Any time that \( p(x|\theta) > 0 \), taking the FOC pointwise with respect to \( p(x|\theta) \) and rearranging we have that:

\[
p(x|\theta) = \frac{p(x) \exp\{u(x, \theta)\}}{\int_{\mathcal{X}} p(\tilde{x}) \exp\{u(\tilde{x}, \theta)\} d\tilde{x}}
\]

(114)

Moreover, we can plug the above back into the general problem and take the FOC. Rearranging we have that for all \( x \) such that \( p(x) > 0 \):

\[
\int_\Theta \int_{\mathcal{X}} \frac{\exp\{u(x, \theta)\}}{p(\tilde{x}) \exp\{u(\tilde{x}, \theta)\} d\tilde{x}} d\pi(\theta) = 1
\]

(115)

Up to now we have applied standard techniques from Matějka and McKay (2015). We now use our utility function and exchangability assumption to derive our novel result. In particular, we take the utility function as:

\[
u(x, \theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2
\]

(116)

And assume exchangability in the prior such that all actions are \textit{ex-ante} equally attractive in the limit:

\[
\int_\Theta \int_{\mathcal{X}} \frac{\exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\}}{\exp\{-\beta(\theta)\lambda^{-1}(\tilde{x} - \gamma(\theta))^2\} d\tilde{x}} \pi(\theta) d\theta = 1 \quad \forall x \in \mathcal{X}
\]

(117)

Under this condition, in the limit of the support to infinity, the unconditional action distribution converges to the improper uniform distribution \( p(x) = p(x') \) for all \( x \in \mathcal{X} \). The conditional action distribution then becomes:

\[
p(x|\theta) = \frac{\exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\}}{\int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(x' - \gamma(\theta))^2\} d\tilde{x}}
\]

(118)

The denominator of this expression can be computed:

\[
\int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\} dx = \int_{\mathcal{X}} \sqrt{\frac{2\pi \frac{\lambda}{2\beta(\theta)}}{2\pi \frac{\lambda}{2\beta(\theta)}}} \exp \left\{-\frac{1}{2} \left( \frac{x - \gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}} \right)^2 \right\} dx
\]

\[
= \sqrt{2\pi \frac{\lambda}{2\beta(\theta)}} \int_{\mathcal{X}} \frac{1}{2\pi \frac{\lambda}{2\beta(\theta)}} \exp \left\{-\frac{1}{2} \left( \frac{x - \gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}} \right)^2 \right\} dx
\]

\[
= \sqrt{2\pi \frac{\lambda}{2\beta(\theta)}}
\]

(119)
It follows that:

\[
p(x|\theta) = \frac{1}{\sqrt{\pi \lambda \beta(\theta)}} \exp \left\{ -\frac{1}{2} \left( \frac{x - \gamma(\theta)}{\sqrt{\frac{\lambda}{2 \beta(\theta)}}} \right)^2 \right\} \tag{120}
\]

Which is to say that \(X|\theta\) is a Gaussian random variable with mean \(\gamma(\theta)\) and variance \(\frac{\lambda}{2 \beta(\theta)}\).

This result extends the known results on Gaussian optimality of stochastic choice with mutual information (Sims, 2003) to a domain with a stochastic weight on the deviation from optimality. For our purposes, the novel and interesting feature is that the variance of the action distribution in any given state is inversely-proportional to curvature. It follows that if all actions are exchangeable in the prior when:

\[
\gamma(\theta) = x^*(X(\theta), \theta) \\
\beta(\theta) = \frac{1}{2} |\Pi_{xx}(X(\theta), \theta)|
\tag{121}
\]

where \(X(\theta)\) is the unique equilibrium level of aggregate production, then the model with mutual information is exactly equivalent to the model with entropic likelihood-separable cost that we studied. All results from Section 3 then carry directly.

Away from the exchangeability condition on the prior, one can still establish a comparative statics result whereby a small increase in curvature over a small set of states gives rise to an increase in precision in those states. Thus, even if attention is not ranked by curvature, small increases in curvature nevertheless increase attention. This is stated formally as Proposition 6.

**Proposition 6.** Suppose that \(u(x, \theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2\) and costs are posterior separable with differentiable kernel \(\phi = \log\). Now perturb \(\beta(\theta)\) in a neighborhood of width \(\delta > 0\) of some \(\hat{\theta}\) at which \(\pi(\hat{\theta})\) exists and is finite by \(\varepsilon > 0\):

\[
\hat{\beta}(\theta) = \beta(\theta) + \varepsilon \mathbb{I} \left[ \theta \in [\hat{\theta} - \delta, \hat{\theta} + \delta] \right]
\tag{122}
\]

Moreover assume that the optimal policy is differentiable in \(\varepsilon\) at \(\varepsilon = 0\). In the limit of \(\delta \to 0\) the change in the optimal action density at \(\varepsilon = 0\) in state \(\hat{\theta}\) is such that \(p(\hat{\theta})\) is becoming more precise about \(\gamma(\hat{\theta})\) under \(\phi\).

**Proof.** To prove the result, we first derive how the log density changes in \(\varepsilon\). Recall from Proposition 5 that the optimal stochastic choice rule is given by:

\[
p(x|\theta) = \frac{p(x) \exp\{u(x, \theta)\}}{\int_{\tilde{x}} p(\tilde{x}) \exp\{u(\tilde{x}, \theta)\} \, d\tilde{x}}
\tag{123}
\]

59
which under our assumption on payoffs is given by:

\[
p(x|\theta) = \frac{p(x) \exp\{-\beta(\theta)(x - \gamma(\theta))^2\}}{\int_\mathcal{X} p(\tilde{x}) \exp\{-\beta(\tilde{x} - \gamma(\theta))^2\} d\tilde{x}}
\]

Taking the derivative of this expression with respect to \( \varepsilon \), evaluating at \( \varepsilon = 0 \) and rearranging:

\[
\frac{dp(x|\theta)}{d\varepsilon}_{|\varepsilon=0} = p(x|\theta) \left[ -(x - \gamma(\theta))^2 + \int_\mathcal{X} (\bar{x} - \theta)^2 p(\bar{x}|\theta) d\bar{x} \right] \mathbb{I}[|\theta - \hat{\theta}| \leq \delta] + p(x|\theta) \left[ \frac{d \log p(x)}{dx} |_{\varepsilon=0} - \int_\mathcal{X} \frac{d \log p(\bar{x})}{dx} |_{\varepsilon=0} p(\bar{x}|\theta) d\bar{x} \right]
\]

Taking the limit \( \delta \to 0 \), whenever \( \theta \neq \hat{\theta} \):

\[
\lim_{\delta \to 0} \frac{dp(x|\theta)}{d\varepsilon}_{|\varepsilon=0} = p(x|\theta) \left[ \frac{d \log p(x)}{dx} |_{\varepsilon=0} - \int_\mathcal{X} \frac{d \log p(\bar{x})}{dx} |_{\varepsilon=0} p(\bar{x}|\theta) d\bar{x} \right] \quad \forall \theta \neq \hat{\theta}
\]

Now note that \( p(x) = \int_{\Theta} p(x|\theta) \pi(\theta) d\theta \). Thus by the dominated convergence theorem and the fact that the conditional density is always bounded:

\[
\lim_{\delta \to 0} \frac{dp(x)}{d\varepsilon}_{|\varepsilon=0} = \int_{\Theta} \lim_{\delta \to 0} \frac{dp(x|\theta)}{d\varepsilon}_{|\varepsilon=0} \pi(\theta) d\theta
\]

Noting now that \( \pi(\hat{\theta}) \) exists and is finite, we can compute this integral by ignoring \( \hat{\theta} \). Thus:

\[
\lim_{\delta \to 0} \frac{dp(x)}{d\varepsilon}_{|\varepsilon=0} = 0
\]

It follows that:

\[
\lim_{\delta \to 0} \frac{d \log p(x|\theta)}{d\varepsilon}_{|\varepsilon=0} = \left[ -(x - \gamma(\theta))^2 + \int_\mathcal{X} (\bar{x} - \theta)^2 p(\bar{x}|\theta) d\bar{x} \right] \mathbb{I}[|\theta - \hat{\theta}| \leq \delta]
\]

where \( \int_\mathcal{X} (\bar{x} - \theta)^2 p(\bar{x}|\theta) d\bar{x} > 0 \). We therefore have exactly that the action distribution in state \( \hat{\theta} \) is becoming more precise about \( \gamma(\hat{\theta}) \) under the metric \( h = \log \).

}\]

Proving such a result globally (i.e. for non-infinitesimal changes in curvature) is challenging. This is for the reason that when we increase the costs of misoptimizing in some states, the shape of the prior distribution has global effects on the stochastic choice so as to render comparisons in terms of precision impossible. Intuitively, if a state is very unlikely, you do not learn about how to play there even if making mistakes in that state is very bad. Nevertheless, the result still isolates the general feature that higher curvature of payoffs for firms will tend to give rise to more precise attention, they key idea in the theory. Moreover,
the same qualitative forces that gave rise to this curvature (the SDF and AD externalities) still apply in this context.

C. (In)efficiency of Attention Cycles

A further question of interest is when equilibria of our model are efficient. This is of relevance as this determines whether attention cycles will have any normative content. As our agents are symmetric, \textit{ex-ante} Pareto efficiency and utilitarian efficiency are equivalent. We therefore say that a stochastic choice rule is efficient if it maximizes utilitarian welfare:

**Definition 5.** A stochastic choice rule \( p^E \in \mathcal{P} \) is efficient if it solves the following program:

\[
p^E \in \arg \max_{p \in \mathcal{P}} \int_{\Theta} \int_{\mathcal{X}} u(x, X(p(\theta)), \theta) \, dP(x|\theta) \, d\pi(\theta) - c(p)
\]  

(130)

Critically, see that an efficient stochastic choice rule both fully internalizes the effect choices have on aggregates and the costs of stochastic choice. We now ask, when will equilibrium be efficient?

**Proposition 7.** Suppose that there exists a unique efficient \( p^E \) obtained as an interior solution of the efficient program. Moreover, suppose that the aggregator is linear:

\[
X(g) = \int_{\mathcal{X}} f(x)g(x) \, dx
\]  

(131)

for some non-constant function \( f \). A necessary condition for efficiency of an equilibrium stochastic choice rule \( p^* \) to be efficient is that:

\[
\int_{\mathcal{X}} u_X(x, X(p^*(\theta)), \theta) \, dP^*(x|\theta) \, dx = 0
\]  

(132)

for almost all \( \theta \in \Theta \).

**Proof.** Recall that the planner’s problem is given by:

\[
p^E \in \arg \max_{p \in \mathcal{P}} \int_{\Theta} \int_{\mathcal{X}} u(x, X(p(\theta)), \theta) \, dP(x|\theta) \, d\pi(\theta) - c(p)
\]  

(133)

and that the aggregator is a linear function of the distribution:

\[
X(g) = \int_{\mathcal{X}} f(x)g(x) \, dx
\]  

(134)

for some non-constant function \( f \). If the efficient allocation is obtained as an interior solution,
then as this program is globally concave, we have that:

\[
\frac{\partial c(p^E)}{\partial p^E(x'|\theta)} - \frac{\partial c(p^E)}{\partial p^E(x|\theta)} = u(x', X(p^E(\theta)), \theta) - u(x, X(p^E(\theta)), \theta)
\]

\[
+ \int_X [f(x') - f(x)] u_X(\bar{x}, X(p^E(\theta)), \theta)p^E(\bar{x}|\theta) \, d\bar{x}
\] (135)

In any equilibrium, there are two possibilities in each state \(\theta\). Either two or more actions are played in that state, or one action is played in that state. Suppose that two or more actions are played with positive density in state \(\theta\) and denote either of these by \(x, x'\). We have that \(p(x|\theta), p(x'|\theta) > 0\). Now consider the perturbation of the equilibrium stochastic choice \(p\) such that:

\[
\tilde{p}(x'|\theta) = p(x'|\theta) + \varepsilon
\]

\[
\tilde{p}(x|\theta) = p(x|\theta) - \varepsilon
\] (136)

where we note that if \(p \in \mathcal{P}\), then we also have that \(\tilde{p} \in \mathcal{P}\). Taking the first-order condition with respect to \(\varepsilon\) and evaluating at \(\varepsilon = 0\), we obtain a necessary condition for equilibrium:

\[
\frac{\partial c(p)}{\partial p(x'|\theta)} - \frac{\partial c(p)}{\partial p(x|\theta)} = u(x', X(p(\theta)), \theta) - u(x, X(p(\theta)), \theta)
\] (137)

By the fact that there exists a unique efficient allocation \(p^E\), we require that \(p = p^E\) for efficiency. Thus, we require that:

\[
\frac{\partial c(p^E)}{\partial p^E(x'|\theta)} - \frac{\partial c(p^E)}{\partial p^E(x|\theta)} = \frac{\partial c(p)}{\partial p(x'|\theta)} - \frac{\partial c(p)}{\partial p(x|\theta)}
\] (138)

This equality implies that:

\[
\int_X [f(x') - f(x)] u_X(\bar{x}, X(p^E(\theta)), \theta)p^E(\bar{x}|\theta) \, d\bar{x} = 0
\] (139)

which can be re-expressed as:

\[
[f(x') - f(x)] \int_X u_X(\bar{x}, X(p^E(\theta)), \theta)p^E(\bar{x}|\theta) \, d\bar{x} = 0
\] (140)

As \(f\) is non-constant, we have that \(f(x') \neq f(x)\) for some \(x, x' \in \mathcal{X}\). Thus, we require for efficiency that:

\[
\int_X u_X(\bar{x}, X(p(\theta)), \theta)p(\bar{x}|\theta) \, d\bar{x} = \int_X u_X(\bar{x}, X(p(\theta)), \theta)p(\bar{x}|\theta) \, d\bar{x} = 0
\] (141)
where this is required for almost all $\theta \in \Theta$ as welfare is equivalent up to outcomes in sets of $\theta$ of measure zero. If one action is played in equilibrium, then as $p^E$ is unique and interior, we know that $p^* \neq p^E$. In this case, the integral condition above is vacuous and therefore remains necessary.

The answer is therefore simple: exactly when average externalities are zero. Efficiency is therefore knife-edge in any game with a non-trivial co-ordination motive. To make this more explicit, we once again return to the quadratic payoff environment from earlier, which generalizes the setting of Angeletos and Pavan (2007) to our class of weighted-quadratic objective functions, and derive a much simpler efficiency condition.

**Proposition 8.** Suppose that utility is quadratic:

$$u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)(x - \gamma(X, \theta))^2$$

(142)

and that there is a unique efficient equilibrium obtained as an interior solution of the efficient program. For a given distribution of actions $p \in \Delta(\mathcal{X})$, define:

$$Bias[p, \theta] = \int_X (x - \gamma(X(p), \theta))p(x)\,dx$$

$$MSE[p, \theta] = \int_X (x - \gamma(X(p), \theta))^2p(x)\,dx$$

(143)

as the bias and mean-squared error induced by an action distribution. A necessary condition for efficiency of an equilibrium with stochastic choice $p^*$:

$$0 = \alpha_X(X(p^*(\theta), \theta) - \beta_X(X(p^*(\theta), \theta)MSE[p^*(\theta), \theta]$$

$$+ 2\gamma_X(X(p^*(\theta), \theta)\beta(X(p^*(\theta), \theta)Bias[p^*(\theta), \theta]$$

(144)

for almost all $\theta \in \Theta$. Moreover, if there is no pure externality $\alpha(X, \theta) = \alpha(\theta)$, then a necessary condition for efficiency of an equilibrium with stochastic choice $p^*$:

$$\underbrace{\frac{d \log \beta(X(p^*(\theta), \theta)}{d \log X}_{\text{Elasticity of Curvature to Aggregate}} = \underbrace{2\gamma_X(X(p^*(\theta), \theta)}_{\text{Movement of Optimal Choice}} \underbrace{\frac{Bias[p^*(\theta), \theta]}{X(p^*(\theta))}}_{\text{Bias-dispersion Ratio}}$$

(145)

for almost all $\theta \in \Theta$.

**Proof.** Recall from Proposition 7 that a necessary condition for efficiency of an equilibrium
\[ \int_{X} u_{X}(\tilde{x}, X(p(\theta)), \theta)p(\tilde{x}|\theta) \, d\tilde{x} = 0 \]  
(146)

for almost all \( \theta \in \Theta \). We now have that utility is given by:

\[ u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)(x - \gamma(X, \theta))^2 \]  
(147)

We therefore have that:

\[ u_{X}(x, X, \theta) = \alpha_{X}(X, \theta) - \beta_{X}(X, \theta)(x - \gamma(X, \theta))^2 + 2\gamma_{X}(X, \theta)\beta(X, \theta)(x - \gamma(X, \theta)) \]  
(148)

Plugging this into the necessary condition and evaluating at the equilibrium aggregate \( \hat{X}(\theta) = X(p(\theta)) \), we obtain:

\[ 0 = \int_{X} \left[ \alpha_{X}(X(p(\theta)), \theta) - \beta_{X}(X(p(\theta)), \theta)(\tilde{x} - \gamma(X(p(\theta)), \theta))^2 \right. \\
\left. + 2\gamma_{X}(X(p(\theta)), \theta)\beta(X(p(\theta)), \theta)(\tilde{x} - \gamma(X(p(\theta)), \theta)) \right] p(\tilde{x}|\theta) \, d\tilde{x} \]  
(149)

Which can be rewritten in terms of the equilibrium bias and variance with respect to \( \gamma \) as:

\[ 0 = \alpha_{X}(X(p^*(\theta)), \theta) - \beta_{X}(X(p^*(\theta)), \theta)MSE[p^*(\theta), \theta] \\
+ 2\gamma_{X}(X(p^*(\theta)), \theta)\beta(X(p^*(\theta)), \theta)Bias[p^*(\theta), \theta] \]  
(150)

yielding the first claim in the proposition. When we additionally impose that \( \alpha(X, \theta) = \alpha(\theta) \), then we can eliminate the first term to obtain:

\[ \beta_{X}(X(p^*(\theta)), \theta)MSE[p^*(\theta), \theta] = 2\gamma_{X}(X(p^*(\theta)), \theta)\beta(X(p^*(\theta)), \theta)Bias[p^*(\theta), \theta] \]  
(151)

Multiplying both sides by \( X(p(\theta)) \) and re-arranging terms we then have:

\[ \frac{\beta_{X}(X(p^*(\theta)), \theta)X(p^*(\theta))}{\beta(X(p^*(\theta)), \theta)} = 2\gamma_{X}(X(p^*(\theta)), \theta) \frac{Bias[p^*(\theta), \theta]}{X(p^*(\theta))} \frac{\left( \frac{MSE[p^*(\theta), \theta]}{X(p^*(\theta))} \right)^2}{2} \]  
(152)

Which is to say that:

\[ \frac{d\log \beta(X(p^*(\theta)), \theta)}{d\log X} = 2\gamma_{X}(X(p^*(\theta)), \theta) \frac{Bias[p^*(\theta), \theta]}{X(p^*(\theta))} \frac{\left( \frac{MSE[p^*(\theta), \theta]}{X(p^*(\theta))} \right)^2}{2} \]  
(153)

completing the proof. \( \square \)
This proposition makes clear that with pure externalities in utility, it is very challenging to achieve efficiency. Moreover, even when there are no pure externalities, efficiency will still only obtain when the elasticity of curvature to aggregates is exactly balanced by movements in the optimum and the bias-dispersion ratio. Consequently, attention cycles generated by the model will typically be inefficient owing to inefficiency of monopolistic production and its interaction with stochastic. Thus, attention cycles will generally be inefficient and there will often be an explicit role for policy intervention.

D. Measuring Attention with Conference Calls

In this Appendix, we describe in more detail our replication of the main data exercises using an independent dataset on sales and earnings conference calls. We find broadly similar patterns of cyclical attention and heterogeneity across sectors.

D.1. Conference Call Data

We obtain data from the Fair Disclosure (FD) Wire service, which records transcripts of sales and earnings conference calls for public companies around the world. We obtain an initial sample of 294,900 calls by scraping, from the Lexis Nexis online API, all Fair Disclosure postings containing the tag “Conference Call - Final” in the document title. Later filters will help extract the (in practice, very small) number of false positive hits.

Although the data in principle cover 2001 to 2020 (all documents available on Lexis Nexis as of February 2020), the data coverage are considerably better for the restricted sample of 2004 to 2014. This subsample includes 261,034 documents, or 89% of the pulled documents, even though they are only 11/20 = 55% of the surveyed time periods.

We next subset to documents that have reported firm names and stock tickers, which are automatically associated with documents by Lexis Nexis. When matches are probabilistic, we use the first (highest probability) match.34 We finally restrict to firms that are listed on one of three US stock exchanges: the NYSE, the NASDAQ, or the NYSE-MKT (Small Cap). We finally connect tickers to firm identifiers (GVKEY) using the master cross-walk available on Wharton Research Data Services (WRDS). These operations together reduce the sample size to 164,805 calls.

We finally restrict to conference calls that are sales or earnings reports. This further reduces the sample to 158,810 total observations, by removing conference calls related to other activities (e.g., mergers). All in all, this sample is about 3,600 firm observations per quarter, or about 60% of the per-quarter observations we obtained via the SEC filings.

34In the essentially zero-measure cases in which there is a tie, we take the alphabetically first ticker.
D.2. Measuring Macro Attention

Software and Methods To tabulate histograms of words within documents, we use the CountVectorizer function in the FeatureExtraction module of the standard Python package Scikit Learn. We apply this algorithm to the entire text of conference calls, including any annotation text and the names of speakers.

Changes to the Words List For consistency, we try to use the same word list that we used for the 10K and 10Q filings. This presents a challenge, however, because some words that are uncommon in the 10K and 10Q corpus are extremely common in the conference call corpus and drown out any useful “signal” in the data. For this reason, we subjectively drop the following words that pose the most difficulty and ex post seem to have low interpretability for our exercise: “people”, “let”, “want”, “point”, “much”, “get”, “question”, “think,” and “percent.” On the remaining set of 80 words, we use the same methods outlined previously. This includes calculating document-specific term frequencies using the fraction of appearances divided by the total number of words; and corpus-appropriate inverse-document-frequencies in the conference call data.

D.3. Results: Time-series Trends

Figure 8 plots the conference-call derived measure, alongside the 10K/Q measure on the common subset of the sample. Theses are positively correlated and both capture our main stylized fact, which is correlation with the business cycle. But the measures do differ sharply in terms of “smoothness of response,” with the conference call measure responding more immediately. A similar observation was possible also from the patterns in Figure 1 based on a single emblematic word, “economy.”

A perhaps more relevant metric for this difference is the behavior of our calibrated moments, reported in Table 3. The elasticity of macro attention to employment is -2.293 and the auto-correlation is 0.881 for conference calls over the restricted sample. In numerical simulations with these parameters (not reported for brevity), the qualitative features of the macro calibration are quite similar, though the magnitudes of equilibrium amplification and asymmetric behavior of business cycles are smaller.

D.4. Results: Industry Heterogeneity

Figure 9 replicates Figure 4 with the conference call sample, showing the heterogeneity in the cyclicality of attention by sector. The regression for the trend line has an intercept of 0.282 and a slope of 0.147, with respective robust standard errors (across 45 industries)
Figure 8. **Macro Attention from Conference Calls versus 10K/Q.**
Both series are seasonally adjusted, as indicated by (7).

equal to 0.070 and 0.146. The former is significantly different from 0, verifying Prediction 1 very strongly. The latter is much less precisely measured, but has a point estimate with is positive (as predicted by the theory) and of comparable magnitude (within error) to the value measured for the 10-K/Q in Table 6.

**E. Coibion et al. (2018) Survey**

In this section, we describe precisely how we incorporate the survey results of Coibion et al. (2018) into our empirical analysis. We make use of the full dataset contained in the replication files posted on the article’s page hosted by the *American Economic Review*: [https://www.aeaweb.org/articles?id=10.1257%2Faer.2011299](https://www.aeaweb.org/articles?id=10.1257%2Faer.2011299). All direct references to survey questions by wave or number match the “Appendix 5: Selected Survey Questions” in the online appendix available at the same link.

**E.1. Profit Function Curvature**

We draw our measure of profit function curvature from the answers to two survey questions about hypothetical price changes. These are jointly asked as Question 17 of Wave 5, Part B:
If this firm was free to change its price (i.e. suppose there was no cost to renegotiating contracts with clients, no costs of reprinting catalogues, etc.) right now, by how much would it change its price? Please provide a percentage answer.

By how much do you think profits would change as a share of revenues? Please provide a numerical answer in percent.

Denote the answer for prices as $\Delta p_i$ and the answer for profits as $\Delta \Pi_i$. Under the assumption that the following second-order approximation holds for the deviation of profits from their frictionless optimum (e.g., a version of (18), in percentage units for the outcome and the choice variable), the following relationship holds between the measurable quantities and the profit function curvature $\text{ProfitCurv}_i$:

$$\Delta \Pi_i = \text{ProfitCurv}_i \cdot \Delta p_i^2$$  \hspace{1cm} (154)

We use this expression to calculate an empirical analogue of profit curvature. The top panel of Table 4 provides summary statistics of measured profit curvatures among the 3,153 firms for which we can measure it. The median reported curvature is 0.12, which means that
Table 4: **Profit Curvature in the Data.**
The top panel gives summary statistic. The bottom panel gives normalized regression coefficients for a number of possible correlates.

A one-percentage-point deviation from the optimal price for such a firm corresponds to a 0.12-percentage-point deviation from optimal profits as a fraction of revenue.

The bottom panel of Figure 4 shows firm and manager-level correlates for our measure in the CGK data. The table reports coefficients of the following regression:

\[
\hat{\text{ProfitCurv}}_i = \beta \cdot \hat{X}_i + e_{it} \tag{155}
\]

where the hat denotes that both variables have been normalized to z-score units (i.e., with means subtracted and standard deviation divided out), so the coefficient \( \beta \) is a “normalized” metric of the standard-deviation-to-standard-deviation effect. We find strong evidence that the firms with higher profit function curvature are smaller and have more competitors. There is only weaker evidence that the associated managers are more skilled and/or better rewarded. We interpret this cautiously as evidence that likely confounds via manager skill and firm sophistication (i.e., better managers grow firms larger, and make better forecasts) are going the “wrong direction” to explain our reduced-form correlations between profit curvature and forecasting accuracy.
Table 5: Curvature and Inflation Attention in Waves 1 versus 4.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Absolute inflation BCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave</td>
<td>1</td>
</tr>
<tr>
<td>ProfitCurv_{it}</td>
<td>-1.172</td>
</tr>
<tr>
<td></td>
<td>(0.195)</td>
</tr>
<tr>
<td>Controls?</td>
<td>✓</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.024</td>
</tr>
<tr>
<td>Within $R^2$</td>
<td>0.024</td>
</tr>
<tr>
<td>$N$</td>
<td>3,153</td>
</tr>
</tbody>
</table>

E.2. Outcomes: Back-cast Errors

For back-cast errors, we use the following questions that are split among waves of the survey. In survey wave 1, firms are asked the following question:

During the last twelve months, by how much do you think prices changed overall in the economy?

Although the wording of the question is not entirely clear about what indicator is being referred to, we follow CGK and interpret this as the annual percent change in CPI, with realized value 1.6%. Firms are asked a similar question in wave 4, but we prefer the wave 1 version because the sample size is slightly larger. Table 5 recreates Table 2 from the main text, first for the wave 1 back-cast of inflation (reported for the main text) and next for the wave 4 back-cast of inflation (not reported in the main text, but quantitatively very similar).

For GDP growth, we use the following question from wave 4:

What do you think the real GDP growth rate has been in New Zealand during the last 12 months? Please provide a precise quantitative answer in percentage terms.

and compare with a realized value of 2.5%. Finally, for unemployment, we use the following question also from wave 4:

What do you think the unemployment rate currently is in New Zealand? Please provide a precise quantitative answer in percentage terms.

and compare with a realized value of 5.7%. All realized values are taken from the replication files of CGK, to deal with any ambiguity about statistical releases, and ensure comparability with that study.
E.3. Outcomes: Tracking Indicators

We finally use, for the lower panel of Table 2, the following questions from wave 4 about tracking different variables:

Which macroeconomic variables do you keep track of? Check each variable that you keep track of.

1. Unemployment rate
2. GDP
3. Inflation
4. None of these is important to my decisions

We code for each variable a binary indicator of whether the firm lists the variable of interest. We lump together GDP in this question (by implication, in levels) with quantitative forecasts of GDP Growth in Table 2.
F. Additional Tables and Figures

Figure 10. Macro Attention and Unemployment.
The fitted line is from OLS regression and the t-statistic is heteroskedasticity-robust.

<table>
<thead>
<tr>
<th></th>
<th>MacroAttentionA</th>
<th>BBD (Total)</th>
<th>BBD (News)</th>
<th>VIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>MacroAttentionA</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>BBD (Total)</td>
<td>0.409</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>BBD (News)</td>
<td>0.300</td>
<td>0.870</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>VIX</td>
<td>-0.238</td>
<td>0.435</td>
<td>0.449</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 6: Quarterly-frequency Correlations.
The sample is 1998-Q1 to 2017-Q4. Macro Attention is measured with quarterly fixed-effects partialed out. VIX is the average of daily observations.
Figure 11. Text Frequency for Every Word.
Figure 12. Correlations with Unemployment by Word.
Of these, 61 of 89 (or 69%) are positive.

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Coefficient</th>
<th>SE</th>
<th>$R^2$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant ($\alpha$)</td>
<td>0.100</td>
<td>(0.048)</td>
<td>0.111</td>
<td>45</td>
</tr>
<tr>
<td>Corr[Output$_{nt}$, Output$_t$] ($\beta$)</td>
<td>0.251</td>
<td>(0.100)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Industry Output and Attention Cyclicality.
This regression corresponds to the trend-line in Figure 4.
Figure 13. **Action Distributions by State.**
The left panel shows action distributions for 10 equally spaced states between the minimum and maximum. The right shows the entropy as a function of the state, in log deviations from the long-run average.

Figure 14. **Sensitivity to $\gamma$.**
We re-calibrate the model for different assumed values of the coefficient of relative risk aversion. The left panel plots the estimated value of $\lambda$ and the right panel reports a summary statistic for the extent of stochastic volatility generated by the model: the relative conditional variance of output growth starting at $\theta = 0.98$ to $\theta = 1.02$. Our baseline calibration of $\gamma = 5$ is denoted in orange. Note that wide variance in the precise value of $\lambda$ is partially an artifact of “rescaling” the profits function, and the quantitative variance of moments of interest (like the right panel) is much smaller and more interpretable.
Figure 15. **Effects of Heightened Uncertainty.**

This figure fixes $\theta = \bar{\theta}$, varies $\lambda$ from the calibrated value $\bar{\lambda}$, and plots the resulting log output in deviation from the reference level $\bar{X}$. In particular, fluctuations plus or minus 10 percent in the attention cost (horizontal axis) result in output moving plus or minus 2 tenths of a percent or 20 basis points (vertical axis).