Attention Cycles

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

May 4, 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’ description 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 macro 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. A calibrated, one-parameter extension of the real-business-cycle model with stochastic choice features asymmetrically large amplification of negative shocks, greater amplification of shocks when output is low, and endogenous stochastic volatility of output growth—ingredients for fast crashes and slow recoveries.

∗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, 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.
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). \(^1\) 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 how individuals think. The visible spikes in attention in Figure 1 are all, at least anecdotally, times when usual conversation is unusually attuned toward scrutinizing macro news—an activity that is 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.

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) filed by the universe of publicly traded US firms from 1998-2017 as a primary source of data on firms’ perceived risk sources. Using a natural-language-processing technique where we compare the words used in these filings to introductory macroeconomics textbooks to identify key “macro words,” we tabulate a metric of each filing’s informativeness about macro developments, which we aggregate into industry-level and aggregate time-series. \(^2\) This reveals two stylized facts. First, the unemployment rate significantly predicts aggregate attention to the

\(^1\) 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.
macroeconomy: attention is counter-cyclical. Second, attention to the macroeconomy is persistent, remaining high even after unemployment and output have recovered.

Why do attention cycles occur, and how could they matter for the behavior of macroeconomic aggregates? We answer both questions in a theory of state-dependent stochastic choice. Our initial, abstract model features a continuum of agents who care about an exogenous state and an aggregate of the cross-sectional distribution of others' actions. Rather than simply choosing an action to best reply to conjectures regarding the aggregate, agents choose a stochastic choice rule describing the distribution of actions they play in each state of the world. They have incentive to do so because they face a cost functional defined over stochastic choice rules: playing more precise action distributions is more expensive. We restrict attention to a highly tractable set of cost functionals that are likelihood-separable, wherein the agent chooses state-by-state which action distribution to play and pays in proportion to cost of so doing, averaged over states. This formalism captures, in reduced form, phenomena such as costly control, difficulty in planning and information acquisition.

The core insight from our theory in partial equilibrium is that agents facing some cost to tune their action to the macro state (e.g., firm managers who need to actively research and plan for macro contingencies) will differentially do so when the cost of making mistakes

---

3We 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.
is high. Formally, in states where agents have an objective function with greater absolute curvature, they will play more precise actions. Importantly, the curvature of firms’ profits in the Neoclassical theory of the firm is the product of two terms: the curvature of firms’ dollar profits, and the stochastic discount factor (SDF) which prices these profits. Thus, all else equal, when the SDF is high or their profits are more responsive to their own behavior, firms should pay more attention.

This simple observation provides a parsimonious explanation for attention cycles, with economic driving forces that are robust across a wide array of macroeconomic models. In particular, recessions are precisely the periods in which the SDF is high, owing to the fact that consumption is low, marginal utility is high and dollar profits are therefore more valuable. This force is further intensified for firms that also face cyclical product demand, which occurs naturally in models with aggregate demand externalities, as the curvature of revenues is greatest when firms face the most cyclical demand. However, even a firm whose business has no link with the aggregate (or has zero beta in the language of finance) should feel pressure to perform well in recessions by precisely responding to the idiosyncratic risks it does face.

But the story does not end in partial equilibrium. Because of our theory’s analytical tractability, we can prove how attention evolves in equilibrium. In this context, we provide novel results on uniqueness of equilibrium and monotonicity of the precision of agents’ equilibrium actions in our class of large games with stochastic choice. Most importantly, our results show that models with strategic complementarity (that is not so strong as to allow multiple equilibria) and curvature of payoff functions that is monotone decreasing in both exogenous and endogenous states will feature the most precise actions in exactly the worst exogenous states of the world. Thus, our theoretical results show how the initial partial equilibrium impetus toward cyclical attention feeds upon itself in general equilibrium in a world with macroeconomic strategic complementarity (e.g., a Keynesian cross). 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.

This reasoning flips on its head the durable hypothesis of Pigou (1927) that recessions are times of great errors in forecasting and judgement but maintains the character of “psychological causes.” In our model, negative shocks crowd in careful thinking about the macroeconomy, which may have been an afterthought in better times. 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, the theory of attention cycles has the potential to explain a
host of macro phenomena related to asymmetry—disproportionately large shocks and high volatility in low states of the world, and a tendency to fall faster downward than to climb back up—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 simple, one-parameter extension of the standard RBC model that incorporates stochastic choice. We calibrate the model to match the time-series properties of our measured macro attention index. We map these data to the model via a formal information-theoretic justification to measure a model-derived object, the entropy of firms’ action distribution. This involves a fundamental leap of faith that time-series patterns in “what firms say” translate quite literally into “what firms do.” This 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 65% of this owes to general equilibrium feedback. Critically, 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 135% 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.\textsuperscript{4,5} 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, 2003; 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.\textsuperscript{6}

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 \textit{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 and Ulbricht (2017) returns to the 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 6. A rigorous combination of the two is an interesting avenue for future research.

Our abstract model of stochastic choice in large games is most related to the concept

\textsuperscript{4}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.

\textsuperscript{5}See Mackowiak, Matejka, and Wiederholt (2018) and Gabaix (2019) for review articles.

\textsuperscript{6}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.
of quantal response equilibrium introduced by McKelvey and Palfrey (1995), where agents play a logit best reply to the actions played by others. Our model differs in that it features a continuum of players, dependence of payoffs on aggregators of the cross-sectional action distribution, and more general cost functionals of the form studied by Fudenberg, Iijima, and Strzalecki (2015) in the context of stochastic choice data. Our results on equilibrium uniqueness and monotonicity of equilibrium action precision therefore contribute to the recent literature understanding uniqueness and comparative statics in games with stochastic choice (Yang, 2015; Morris and Yang, 2016).

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 build a general model of stochastic choice in large games to understand attention cycles at a deep level. In Section 4, we map a standard macroeconomic model to the general model to understand why attention cycles should manifest. In Section 5, we test the predictions of the theory to evaluate its performance in explaining attention cycles. In Section 6, we calibrate a one-parameter extension of the RBC model with stochastic choice and analyze the impact of attention cycles on macroeconomic dynamics. Section 7 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 both by 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. 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.\(^7\) 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.\(^8\)

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 are the

---

\(^7\)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.

\(^8\)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.
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 Section E, and especially subsection E.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.9 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.10

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:

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

(1)

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

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

(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 process-

---

9We use the 7th, 3rd, and 12th editions of these books, respectively.
10For 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.”
ing that measures whether a word is frequent in a given document, \textit{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

\[
\text{The quick brown fox jumped over the lazy dog.}
\]

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 \textit{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 \( \text{tf-idf}(w, b, \mathcal{D}) \). 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 \textit{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 \textit{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”).

\footnote{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.}

\footnote{Appendix Figure 13 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{3cm} (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{3cm} (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 $\text{MacroAttentionA}(t)$.$^{13}$ 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}

\hspace{3cm}

\begin{align}
\text{Formally, if } q(t) &\in \{1, 2, 3, 4\} \text{ 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{3cm} (6)
\end{align}

and define an “Adjusted” metric as

$$\text{MacroAttentionA}(t) := \text{MacroAttention}(t) - x(q(t)) + x(q(1))$$  \hspace{3cm} (7)

i.e., one that partials out quarterly-frequency fluctuations and adds back the trend of Q1.
through, 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 the 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. Figure 12 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.

**Robustness: Conference Calls** Appendix E 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, Figure 8 shows that the conference call based metric is also...
quite counter-cyclical, though not perfectly correlated with the 10-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 with 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) never recovers 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 relative to 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.

Figure 3. Specific Components of Macro Attention.
See Appendix Figure 13 for all 89 words in the same format.
than a cause thereof—precisely the fact that economic issues seem more relevant during unpleasant times may cause such connotations.\footnote{But 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.}

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 foreshadows 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 13 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 14, 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 our 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 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 General Theory of Attention Cycles

In this section, we develop a simple abstract theory to explain counter-cyclical and persistent attention (Facts 1 and 2). Formally, we build an equilibrium theory of stochastic choice and study how the precision of agents’ actions varies across states. We begin by showing how the desire to make greater or smaller mistakes in any given state manifests in an abstract individual decision problem. The key insight is that the curvature of an agent’s objective determines an agent’s incentives to make mistakes: in states with absolutely high curvature, mistakes are costly and so actions are precise. Moving the analysis to equilibrium, we provide existence and uniqueness results, and explicitly model general equilibrium feedback that shapes the curvature of utility. We provide general conditions under which the precision of actions is monotone in the state and attention cycles arise in equilibrium. The key conditions to generate this result are strategic complementarity and curvature of payoffs that is monotone decreasing in both exogenous and endogenous states.

Critically, for both the proof and the interpretation, the analysis features a general equilibrium feedback whereby more precise actions beget yet more precise actions. Strategic complementarity—akin to the general equilibrium feedback we would expect in a Keynesian, demand-driven model—amplifies attention cycles.

3.1. The Single-Agent Problem with Costly Stochastic Choice

There is a single agent who chooses an action \(x \in \mathcal{X}\), which is a closed interval. There is an underlying and payoff-relevant state of the world \(\theta \in \Theta\), which is either a closed interval or a finite set, over which the agent has prior \(\pi \in \Delta(\Theta)\). Agents have a payoff function \(u: \mathcal{X} \times \Theta \to \mathbb{R}\) with \(u(x, \theta)\) yielding the agent’s payoff from taking action \(x\) in state \(\theta\). For tractability, we assume that the utility function is twice continuously differentiable and strictly concave in its first argument.

Agents are either unsure about the underlying state of the world or find it costly to adapt their action to the state. We model this uncertainty or cost of adaptation by having agents choose a stochastic choice rule at cost. This formulation is sufficiently flexible to capture models of information acquisition and mistake-making in a unified and parsimonious manner, as we will shortly make clear. Formally, the agent chooses a stochastic choice rule \(P: \Theta \to \Delta(\mathcal{X})\) with \(P(x|\theta)\) describing the cumulative distribution of actions \(x\) taken by the agent in state \(\theta\). We call the set of measurable stochastic choice rules \(\mathcal{P}\). When this admits a density function, or we wish to refer to the mass function, we denote a stochastic choice rule by \(p(x|\theta)\). We model the cost of information acquisition, paying attention or controlling mistakes via a cost functional \(c: \mathcal{P} \to \mathbb{R}\) which simply returns how costly any
given stochastic choice rule is for the agent.

The agent’s problem is then to choose a stochastic choice rule to maximize their expected utility under that stochastic choice rule net of the costs of employing that rule. Hence, agents solve the following program:

$$\max_{P \in \mathcal{P}} \int_\Theta \int_\mathcal{X} u(x, \theta) \, dP(x|\theta) \, d\pi(\theta) - c(P)$$

(8)

To make clear the nature of such a stochastic choice formulation and introduce the key classes of problem we will later consider in the analysis, we define also the following class of cost functionals:

**Definition 1** (Likelihood-Separable Cost Functional). A cost functional $c$ has a likelihood-separable representation if there exists a strictly increasing and strictly convex function $\phi : \mathbb{R}_+ \to \mathbb{R}$ such that for any stochastic choice rule $P$ with corresponding density $p$:

$$c(P) = \int_\Theta \int_\mathcal{X} \phi(p(x|\theta)) \, dx \, d\pi(\theta)$$

(9)

Likelihood-separable 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 equivalent to an additive random utility model of the sort often employed in the literatures on trade and industrial organization. Critically, under such a formulation, the agent’s first-order condition among any two choices $x, x' \in \mathcal{X}$ in state $\theta$ where $p(x|\theta), p(x'|\theta) > 0$ is given simply by:

$$u(x', \theta) - u(x, \theta) = \phi'(p(x'|\theta)) - \phi'(p(x|\theta))$$

(10)

which has no dependence on the agent’s prior. Employing such a formulation therefore allows us to isolate the component of stochastic choice that stems from agents finding it challenging to adapt their action to the state of the world, conditional on being aware of that state.

Naturally, such a formulation lacks a natural intuition in terms of signal processing 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 our single-agent result generalizes to the case with mutual information cost, the cost functional studied in the literature on rational inattention.
3.2. Curvature and State-Dependent Attention

We now study state-dependence in an agent’s “attention.” We map this concept to the precision with which agents hone their stochastic choice rule upon the unconstrained, state-contingent optimum. That is, an agent is paying attention in state realization \( \theta = \theta_0 \) if they are always close to playing the course of action that is best adapted toward \( \theta_0 \)—they have planned very precisely for this contingency.

Let us proceed toward this result in steps. We require first a notion of the unconstrained optimum as a function of the state:

\[
x^*(\theta) = \arg \max_{x \in X} u(x, \theta)
\]  \hspace{1cm} (11)

Now consider a second-order approximation of the agent’s objective function in the presence of stochastic choice where, state-by-state, we approximate the agent’s payoff around \( x^*(\theta) \). Applying a standard envelope theorem to remove the first-order term, we obtain the following approximation:

\[
u(x, \theta) \approx u(x^*(\theta), \theta) + \frac{1}{2} u_{xx}(x^*(\theta), \theta)(x - x^*(\theta))^2
\]  \hspace{1cm} (12)

where \( \bar{u}(\theta) \equiv u(x^*(\theta), \theta) \) is an action-independent constant. Intuitively, to second order, the impact of the agent’s action on their payoff depends only on how poorly they track their optimal action \( x^*(\theta) \) and the curvature of their payoff in their own action around the full-information optimum \( |u_{xx}(x^*(\theta), \theta)| \). We therefore interpret this curvature as the unit cost of making mistakes. Critical to our analysis is how this unit cost varies with the state—that is, how the “stakes for being accurate” change across situations.

It would seem natural that in states with high absolute curvature, the agent’s action distribution would concentrate closely around \( x^* \). This is indeed exactly the case with likelihood-separable stochastic choice. We first formalize a notion of precision that can be defined for arbitrary probability distributions:

**Definition 2** (Precision). Given a function \( h \), a symmetric distribution \( g \) is more precise about a point \( x^* \) than \( g' \) about \( x^{*'} \) under \( h \) if \( h \circ g(|x - x^*|) \) is faster decreasing in \( |x - x^*| \) than is \( h \circ g'(|x' - x^{*'}|) \) in \( |x' - x^{*'}| \).\(^{15}\)

Informally, this definition requires that a distribution is more precise than another if its density is more rapidly decreasing away from the point about which precision is being considered. This definition generalizes the property that Gaussian distributions are more

\(^{15}\)On an asymmetric support, we call a distribution \( g \) symmetric if \( g(x) = g(-x) \) whenever both \( g(x) \) and \( g(-x) \) are defined.
precise about their mean when they have a lower standard deviation to cases with non-Gaussian densities by exactly capturing the idea that a distribution is more precise if its tails decay faster from the point about which a distribution is centered.\footnote{To see this, recall that a Gaussian random variable with mean $\mu$ and standard deviation $\sigma$ has pdf:}

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ - \frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}$$

Thus, for two gaussian distributions with means $\mu, \mu'$ and standard deviations $\sigma, \sigma'$ such that $\sigma < \sigma'$, we have that $h \circ g(|x - \mu|)$ is faster decreasing than $h \circ g'(|x - \mu'|)$ whenever $h$ is monotone. Thus, under monotone $h$, we have that Gaussian distributions with lower standard deviations are more precise about their mean under our definition of precision.

We can now formalize the idea that in states with higher curvature, the action distribution of the agent features more precision for arbitrary likelihood-separable cost functionals:

**Proposition 1.** Suppose that $u(x, \theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2$ and costs are likelihood separable with differentiable kernel $\phi$. Take any $\theta, \theta' \in \Theta$ such that $\beta(\theta') > \beta(\theta)$, the optimal stochastic choice rule is such that $p^*(\theta')$ is more precise about $\gamma(\theta')$ under $\phi'$ than is $p^*(\theta)$ about $\gamma(\theta)$.

**Proof.** See Appendix A.1.

In view of the approximation argument used earlier, we can further argue that to a second order approximation of any objective, precision about the full-information optimal action can be ranked by the curvature of the agent’s objective at the full-information optimal action. The requirement for precision to be monotone under $\phi'$ is natural: this is the unit in which the marginal costs of precision are measured for the agent. In Appendix B, we provide generalizations of this result to consider mutual information cost, the cost functional studied in the literature on rational inattention, that have the same character but include also a “second-order” effect of concentrating the action distribution in one state on the unconditional “default action distribution” toward all actions are anchored.

### 3.3. State-Dependent Attention in General Equilibrium

We have now understood how state-dependent attention manifests in an individual’s decision problem via the curvature of an agent’s objective. We now turn to the more natural setting for macro modeling: an equilibrium environment in which the fundamental is endogenous to other players’ actions. Our key results will provide conditions under which a continuum of agents like the last section’s can interact strategically, and the resulting equilibrium is both unique and features monotonicity of attention.
A continuum of identical agents is indexed by $i \in [0, 1]$. They take actions $x_i \in \mathcal{X}$ which are aggregated by an aggregator $X : \Delta(\mathcal{X}) \rightarrow \mathbb{R}$ that takes a cross-sectional distribution of actions to an aggregate. One natural interpretation of the aggregator, which will resurface in our later application to market economies, is that $X$ is a production function mapping each firm $i$’s production choice into a final good. Agents have identical utility functions $u : \mathcal{X} \times \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ where $u(x, X, \theta)$ is an agent’s utility from playing $x$ when the aggregate is $X$ and the state is $\theta$. An equilibrium is the following:

**Definition 3 (Equilibrium).** An equilibrium $\Omega$ is a tuple comprising a collection of stochastic choice rules $\{P^*_i\}_{i \in [0,1]}$ and an equilibrium law of motion for aggregates $\hat{X} : \Theta \rightarrow \mathbb{R}$ such that:

1. Each agent’s stochastic choice rule is consistent with the equilibrium law of motion:

$$P^*_i \in \arg \max_{P \in \mathcal{P}} \int_{\Theta} \int_{\mathcal{X}} u(x, \hat{X}(\theta), \theta) \, dP(x|\theta) \, d\pi(\theta) - c(P) \quad \forall i \in [0,1]$$

2. The equilibrium law of motion is consistent with the stochastic choice rules played by agents:

$$\hat{X} = X \circ \int_{[0,1]} P^*_i \, di$$

An equilibrium is symmetric if $P^*_i = P^*$ for all $i \in [0,1]$.

To ensure equilibrium existence and to rule out asymmetric equilibria, we make the following two technical assumptions.

**Assumption 1 (Bounded and Continuous Payoffs).** The utility function $u$ is bounded and uniformly continuous in $(x, X, \theta)$. The aggregator $X$ is bounded and the cost functional $c$ is lower semi-continuous.

**Assumption 2 (Convex Cost Functional).** The cost functional $c$ is strictly convex.

We can now state and prove the following result:

**Theorem 1.** Under Assumptions 1 and 2, there exists an equilibrium $\Omega$. Moreover, there does not exist an asymmetric equilibrium.

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

Having shown existence and established that the analysis is well-posed, we also wish to study when there is a unique equilibrium in this setting so that our study of cross-state comparisons is well-posed. To this end, we identify conditions on payoffs, aggregators and stochastic choice functionals sufficient to guarantee uniqueness. For payoffs, we first require complementarities in the underlying game in the form of supermodularity in payoffs between an agent’s own action and the state:
Assumption 3 (Supermodularity). Payoffs $u$ are supermodular in $(x, X)$:

$$u(x', X', \theta) - u(x, X', \theta) \geq u(x', X, \theta) - u(x, X, \theta)$$  (16)

For all $x' \geq x, X' \geq X, \theta$.

We moreover require that these complementarities are not too strong in the sense that payoffs are sufficiently concave to prevent the effect of complementarities between the aggregate and the agent's own action from becoming too large.\textsuperscript{17}

Assumption 4 (Sufficient Concavity). Payoffs satisfy the following condition for all $\alpha \in \mathbb{R}_+$:

$$u(x, X + \alpha, \theta) - u(x - \alpha, X, \theta) \geq u(x', X + \alpha, \theta) - u(x' - \alpha, X, \theta)$$  (17)

For all $x' \geq x, X, \theta$.

Having identified conditions on payoffs, we now turn to finding conditions on the aggregator of the cross-sectional distribution of actions. Naturally, we require that the aggregator is monotone in the sense of first-order stochastic dominance:

Assumption 5 (Monotone Aggregator). For all $g, g' \in \Delta(X)$:

$$g' \succeq_{FOSD} g \implies X(g') \geq X(g)$$  (18)

We moreover require that the aggregator satisfies discounting, which is to say that it is concave in simple shifts of the cross-sectional action distribution:

Assumption 6 (Discounted Aggregator). There exists $\beta \in (0, 1)$ such that for any distribution $g \in \Delta(X)$ and any $\alpha \in \mathbb{R}_+$:

$$X(\{g(x - \alpha)\}_{x \in X}) \leq X(\{g(x)\}_{x \in X}) + \beta \alpha$$  (19)

Finally, we need to identify conditions on the underlying stochastic choice environment that are sufficient to guarantee uniqueness. To this end, we first define a new statistical property of a function that we label quasi-MLRP. This condition allows us to relate the underlying cost functional to the distribution of actions induced by optimality.

\textsuperscript{17}The condition we identify also appears in the literature on uniqueness in games. See Weinstein and Yildiz (2007) for an example.
Definition 4 (Quasi-MLRP). A function \( f : \mathbb{R}_+ \rightarrow \mathbb{R} \) satisfies quasi-MLRP if for any two distributions \( g', g \in \Delta(\mathcal{X}) \):

\[
\left( f(g'(x')) - f(g'(x)) \geq f(g(x')) - f(g(x)) \quad \forall x' \geq x \right) \implies g' \succeq_{FOSD} g \quad (20)
\]

It is moreover worthwhile to note that important functions satisfy quasi-MLRP and that the class of \( f \) satisfying the quasi-MLRP property \( \mathcal{F} \) is non-empty. Indeed, as the name suggests, quasi-MLRP is a strict weakening of MLRP. These facts are shown formally in Lemma 1.\(^{18}\)

Lemma 1. \( \mathcal{F} \) is non-empty. Moreover, the functions (i) \( f(x) = \log(x) \) and (ii) \( f(x) = x \) satisfy quasi-MLRP. Thus, quasi-MLRP is a strict weakening of MLRP.

Proof. See Appendix A.3.

With this definition in hand, we can now state our final assumption on stochastic choice functionals, which will ensure we can always translate dominance in payoff units to dominance in terms of distributions:

Assumption 7 (Attention Costs Are Not Too Steep). Costs have a likelihood-separable representation with differentiable kernel \( \phi \) such that \( \phi' \) satisfies quasi-MLRP.

We can now state and prove our uniqueness theorem:

Theorem 2. Under Assumptions 1 - 7, there exists a unique equilibrium \( \Omega \).

Proof. See Appendix A.4.

We show this result by defining an equilibrium operator that maps the law of motion of the aggregate in the state to the resulting optimal stochastic choice rule and then maps this back to a law of motion of the aggregate, and then determining that said operator is a contraction map.\(^{19}\) This result extends classic uniqueness results to the realm of stochastic choice and

\[^{18}\text{A complete characterization of } \mathcal{F} \text{ is beyond this paper. Nevertheless, see Appendix A.3 for a result that provides a more verifiable sufficient condition for a function to be a member of } \mathcal{F}.\]

\[^{19}\text{More formally, let } \mathcal{B} = \{ \hat{X} : \Theta \rightarrow \Theta \} \text{ and define the operator } T : \mathcal{B} \rightarrow \mathcal{B}: \]

\[
T\hat{X} = X \circ p^*(\hat{X}) \quad (21)
\]

One notes that \( \mathcal{B} \) is a subset of the space of bounded functions and we endow it with the sup norm. To show uniqueness of the equilibrium law of motion of aggregates, it then suffices to prove that \( T \) is a contraction map. We prove this by showing that under the given assumptions, \( T \) satisfies both of Blackwell’s conditions of monotonicity and discounting. Given the unique equilibrium-consistent law of motion \( T\hat{X} = X \), the equilibrium stochastic choice rule is then the unique solution of the stochastic choice problem given that law of motion \( p^*(\hat{X}) \).
may be of independent interest to researchers looking to understand the relationship between stochastic choice and multiplicity.\textsuperscript{20}

We now wish to derive results describing how the equilibrium aggregate $\hat{X} : \Theta \rightarrow \mathbb{R}$ moves with the underlying state and how the dispersion of the cross-sectional action distribution $p^* : \Theta \rightarrow \Delta(X)$ depends on the state. To show monotonicity of aggregates, we simply require a stronger supermodularity assumption that not only are individual actions and aggregate actions complements, but so too is the underlying state itself a complement to both individual actions and aggregates in utility:

**Assumption 8.** The payoff function $u$ is supermodular in $(x, X, \theta)$:

$$u(x', X', \theta') - u(x, X, \theta) \geq u(x, X', \theta') - u(x, X, \theta) \quad \forall \theta' \geq \theta, X' \geq X, x' \geq x \quad (25)$$

With this assumption in hand, we can formalize that the unique equilibrium aggregate is monotone in the state:

**Proposition 2.** Under Assumptions 1 - 8, the unique equilibrium $\hat{X}$ is monotonically increasing in $\theta$.

*Proof.* See Appendix A.5. $\square$

The intuition for this result is simple: higher $\theta$ makes higher actions more desirable so that the distribution of actions in higher states dominates the distribution in lower states. The proof strategy makes use of the contraction mapping property used in the uniqueness proof, shows that monotonicity is preserved by the fixed point operator and therefore, by the properties theorem, that the fixed point must itself be monotone.

\textsuperscript{20}As many assumptions go into this result, it is heartening to know that a large class of economies (including those that we will later study) satisfy the conditions of the uniqueness theorem. An important example of an economy that satisfies Assumptions 1 - 7 is one with:

1. Weighted-quadratic payoffs:
   $$u(x, X, \theta) = \alpha(X, \theta) - \beta(\theta)(x - (\alpha \gamma(\theta) + (1 - \alpha)X))^2 \quad (22)$$

2. Constant elasticity of substitution aggregators (for $\sigma > 1$):
   $$X = \left( \int_X x_i^{1-\sigma^{-1}} \, di \right)^{\frac{\sigma}{\sigma - 1}} \quad (23)$$

3. Likelihood separable and entropic stochastic choice functionals (for $\lambda > 0$):
   $$\phi(p) = \lambda p \log p \quad (24)$$
3.4. **Attention Cycles in Equilibrium**

We now turn to understanding when the cross-sectional action distribution features monotone precision in the underlying state. To show this result, we return to the quadratic environment from earlier:

**Assumption 9.** The utility function is given by:

\[
u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)(x - \gamma(X, \theta))^2 \tag{26}\]

where \( \beta(X, \theta) \) is positive and monotonically decreasing in \((X, \theta)\) and \( \gamma(X, \theta) \) is monotonically increasing in \((X, \theta)\).

Once again, such a utility function can be justified via a second-order approximation of any utility function, where the assumptions on \( \beta \) and \( \gamma \) map to saying that curvature \( u_{xx} \) is monotonically increasing in \((x^*, X, \theta)\) and that the full-information optimal action \( x^* \) is monotonically increasing in \((X, \theta)\).

In this context, we can now build on our uniqueness and monotonicity results to show that the equilibrium action distribution features monotone decreasing precision in the state:\(^{21,22}\)

**Theorem 3.** Under Assumptions 1 - 9, the unique equilibrium \( p^*(\theta) \in \Delta(X) \) is is more precise about \( \gamma(\hat{X}(\theta), \theta) \) than \( p^*(\theta') \) about \( \gamma(\hat{X}(\theta'), \theta') \) under \( \phi' \) for any \( \theta' \geq \theta \).

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

This result provides a general equilibrium analogue of the precision results in the previous section (Proposition 1) and thereby a set of sufficient conditions for equilibrium action precision to be monotone in the state. The intuition for this result is that when the state increases, the endogenous aggregate also increases, and so the absolute curvature of utility falls. As a result, it is less costly to misoptimize. Agents therefore put less effort into playing precise actions and precision of actions is decreasing in the state. As a result of this lower precision in high states, agents more precisely track the underlying fundamental in lower

\(^{21}\)Assumption 9 can be relaxed to consider objective functions of the form:

\[
u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)\Gamma(|x - \gamma(X, \theta)|) \tag{27}\]

for monotone increasing \( \Gamma \) with the normalization \( \Gamma(0) = 0 \). The proof of Theorem 3 then carries exactly as written under this slightly more general specification. We specialize to the quadratic form for expositional purposes given the justification of quadratic \( \Gamma \) by an approximation argument.

\(^{22}\)Once again, the set of models satisfying Assumptions 1 - 9 is not empty, and contains many models of interest. In particular, taking the example from Footnote 20 and making \( \beta \) log-linear (Cobb-Douglas) in \((X, \theta)\) is consistent with the assumptions for some specifications of parameters, and features GE amplification of attention cycles. See Appendix D for an example of attention cycles in GE in closed-form.
states, accentuating further the partial equilibrium desire to pay attention in these states. This causes attention cycles to be partially self-fulfilling.

This general equilibrium feedback loop, which allows us to determine the effect of moving from a single-agent to a strategic world, accentuates attention cycles under our given assumptions. This allows us both to state a sharp general equilibrium result and to appreciate the potential for “macro mechanisms” to both cause and be caused by the dynamics of attention. With these results in hand, we now have a theory for how attention cycles arise in general equilibrium. What remains is to map the reduced-form forces in the strategic environment to tangible (and quantifiable) macro mechanisms.

4. Attention Cycles in Macroeconomic Models

We now explore how attention cycles manifest in a standard RBC model and its rigid-wage, Keynesian limit, mapping this model to the abstract framework we have developed. 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. Each firm’s decision problem about how much to produce naturally has more curvature in their chosen action when the state of the world is poor, in particular because the firm’s risk averse owners are more concerned about dollar profits in those states, and dollar revenues are themselves most sharply curved in these states owing to aggregate demand externalities. The conjunction of these two forces will map naturally to the “state-dependent stakes” from the abstract model, and lead to attention cycles.

4.1. Primitives

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. Let the quantities of the intermediate good and final good be denoted by \( x_{it} \) and \( X_t \), respectively. There is a single factor of production, labor, denoted by \( L_t \).

A representative household has preferences over final-good consumption and labor, \( \{C_t, L_t\}_{t \in \mathbb{N}} \). Payoffs take the following expected-discounted-utility form:

\[
U(\{C_{t+j}, L_{t+j}\}_{j \in \mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j U(C_{t+j}, L_{t+j})
\]  

(28)

where \( \beta \in [0, 1) \) is the discount factor, \( U \) is increasing and concave in its first argument, decreasing and convex in its second, and differentiable.
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:

$$C_t + A_{t+1} \leq w_t L_t + \int_{[0,1]} \Pi_{it} \, di + (1 + r_t) A_t \quad (29)$$

The representative household maximizes its utility (28) subject to the aforementioned sequence of constraints, taking all prices as given.

The aggregate final good is produced by a representative, perfectly competitive firm. Its production technology is represented by the constant returns to scale functional $X : \Delta(\mathcal{X}) \to \mathbb{R}_+$ for some compact $\mathcal{X} \subset \mathbb{R}$. 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.

The key actors in our model are producers of intermediate goods. These firms hire a labor quantity $L_{it}$ and produce with the following technology

$$x_{it} = \theta_t f(L_{it}) \quad (30)$$

where $\theta_t$ is an aggregate-level productivity shifter and $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a constant or decreasing returns-to-scale function, with $f' > 0$ and $f'' \leq 0$.

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 $\pi : \Theta \to \Delta(\Theta)$, 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 $t$, a firm knows the sequence of previous productivity realizations $\{\theta_s\}_{s < t}$, but not the current period’s productivity.

Intermediate goods firms, like the agents in the previous section’s abstract problem, struggle to adapt their action $x_{it}$ to match the state of the world $\theta_t$. We model this by having them choose stochastic choice rules at a cost. Each firm chooses rule $P : \Theta \to \Delta(\mathcal{X})$ in the set of stochastic choice rules $\mathcal{P}$. The firms commit to delivering the realized quantity $x_{it} \in \mathcal{X}$ 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 seek to maximize the product of their dollar profit $\Pi_{it}(\cdot)$ and the representative household’s marginal utility, $U_C(C_t, L_t)$, net of their non-pecuniary cost.
of attention $c : \mathcal{P} \rightarrow \mathbb{R}$. Formally, the firm’s payoff is:

$$u(x_{it}, X_{t}, w_{t}, L_{t}, \theta_{t}) = U_{C}(X_{t}, L_{t})\Pi_{it}(x_{it}, X_{t}, w_{t}, \theta_{t}) \quad (31)$$

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, labor and wages $\Lambda_{t} \equiv (X_{t}, w_{t}, L_{t})$ moves with the state, which we will denote by $\hat{\Lambda}$. Given this conjecture, they play a best reply by solving the following program:

$$p^{*}(\hat{\Lambda}; \theta_{t-1}) \in \arg\max_{p \in \mathcal{P}} \int_{\Theta} \int_{X} u(x, \hat{\Lambda}(\theta, \theta_{t-1}), \theta)p(x \mid \theta; \theta_{t-1}) \, dx \pi(\theta \mid \theta_{t-1}) \, d\theta - c(p) \quad (32)$$

**Equilibrium** An equilibrium of the model can then be described by laws of motion for all endogenous variables:

$$\hat{\Omega} = \{C(\theta_{t}; \theta_{t-1}), L(\theta_{t}; \theta_{t-1}), A(\theta_{t}; \theta_{t-1}), w(\theta_{t}; \theta_{t-1}), X(\theta_{t}; \theta_{t-1}), p^{*}(\theta_{t}; \theta_{t-1})\}_{\theta_{t}, \theta_{t-1} \in \Theta} \quad (33)$$

that are consistent with optimality according to the household’s problem (28), the firms’ problem (32) and market clearing for final goods, intermediate goods and labor. Using market clearing and the optimality conditions for the representative household, one notes that aggregate consumption is always simply output and, as bonds are in zero net supply, $A \equiv 0$. Thus, an equilibrium can always be summarized more compactly by:

$$\Omega = \{\hat{\Lambda}(\theta_{t}, \theta_{t-1}), p^{*}(\theta_{t}; \theta_{t-1})\}_{\theta_{t}, \theta_{t-1} \in \Theta} \quad (34)$$

which is simply, for each possible state of yesterday’s productivity, the equilibrium concept according to the game-theoretic analysis given in Definition 3, with the complication that $\hat{\Lambda}$ is multidimensional.

If we take $U$ to be separable in consumption and labor and suppose that wages are *exogenous*, then the model collapses exactly back to that studied in the abstract analysis in Section 3, as $\hat{\Lambda} = \hat{X}$. Theorem 3 applies exactly (up to approximation) conditional on our verifying that the curvature of firms’ objective functions satisfies the required regularity conditions and has monotone decreasing curvature in the state $\theta$ and aggregate $X$. Thus, in the rigid-wage Keynesian limit of this model, which is similar to a stochastic choice version of the liquidity trap models of Eggertsson and Krugman (2012) and Korinek and Simsek (2016), our abstract theoretical results all apply.
Away from the case with rigid wages and separable utility, our theoretical results do not apply exactly but can still be used as a guide. The idea remains to verify that firms’ profits “on average,” considering both productivity and wage effects, become more sharply curved when productivity and production are low. Formally, in view of Proposition 1, we know that this will always give rise to a partial equilibrium desire to pay more attention in low states. We moreover know that if the equilibrium map from states to curvature of payoffs is monotone decreasing, that Proposition 1 implies that attention cycles occur. As we will soon argue, any sensibly calibrated RBC model (including our later quantitative model) must have this feature, allowing us to understand why an RBC model is likely to feature attention cycles.

4.2. Understanding the Firm’s Problem

Let us now consider the firm’s problem in more detail to understand the forces that shape the curvature of their profits. Under the assumption that the demand function for any firm’s good depends only on their own product and the total level of product $d(x_{iit}, X_t)$, the firm’s payoff, not including the cost of stochastic choice, can be written in the following three terms:

$$u(x_{iit}, X_t, w_t, L_t, \theta_t) = U_C(X_t, L_t) \cdot (B(x_{iit}, X_t) - C(x_{iit}, w_t, \theta_t)) \quad (35)$$

The first term, outside the parentheses, is the marginal utility of the representative household or SDF, which prices the profits of the firm. The two terms in parentheses are defined as the firm’s “benefits,” or revenues, and “costs.”

Let us now consider the following state-by-state approximation to the firm’s objective as in Equation 12. The key quantity, the state-dependent curvature of payoffs, is the following:

$$u_{xx}(x_{iit}, X_t, w_t, L_t, \theta_t) = U_C(X_t, L_t) \cdot (B_{xx}(x_{iit}, X_t) - C_{xx}(x_{iit}, w_t, \theta_t)) \quad (36)$$

Note that the household’s marginal utility, and variation therein, is critical to quantify the payoff cost of the randomness in production induced by imperfect planning.

To be more specific, we specialize to the environment we will consider for the later quantitative study of the RBC model with attention cycles. First, assume that the final goods firm has a CES production technology, or:

$$X_t = X([x_{iit}]_{i \in [0,1]}) = \left( \int_{[0,1]} x_{iit}^\varepsilon \, di \right)^{\frac{1}{\varepsilon}} \quad (37)$$

23This does not appear in the standard RBC model as, if the firm could costlessly condition its action upon the state $\theta$ (i.e., $c(\cdot) \equiv 0$), then the solution is invariant to risk adjustment.
where $\varepsilon > 1$ is the elasticity of substitution between varieties of the intermediate good. Next, let us specialize to a linear production for the intermediates good producer, or $f(L) = L$. With these specific functional forms, we can provide an exact expression for the monopolist’s approximated objective:

**Proposition 3 (State-dependent curvature).** The second-order approximation of the intermediate firm’s objective function is given in the form of Equation 12 with the following functions:

\[
\bar{u}(X_t, w_t, L_t, \theta_t) = v_0(\varepsilon) \cdot \left(\frac{w_t}{\theta_t}\right)^{1-\varepsilon} \cdot X_t \cdot U_C(X_t, L_t) \\

u_{xx}(x_{it}, X_t, w_t, L_t, \theta_t)|_{x_{it}=x^*_{it}} = -v_2(\varepsilon) \cdot \left(\frac{w_t}{\theta_t}\right)^{1+\varepsilon} \cdot X_t^{-1} \cdot U_C(X_t, L_t)
\]

where the constants $v_0(\varepsilon), v_2(\varepsilon) > 0$ are given by:

\[
v_0(\varepsilon) = (\varepsilon - 1)\varepsilon^{-1} \varepsilon^{-\varepsilon} \\
v_2(\varepsilon) = (\varepsilon - 1)^{-\varepsilon} \varepsilon^{\varepsilon - 1}
\]

*Proof.* See Appendix A.7.

Let us focus on curvature as this is the object for understanding attention cycles.\(^{24}\) First and most importantly, we observe that aggregate demand affects the curvature in two ways. Critically, the SDF $U_C(X_t, L_t)$ which will be higher when output is lower by concavity of $u$ in its first argument, elevates the curvature of the firm’s objective function in low states.

To enable a back-of-the-envelope calculation for the severity of this force, let us assume a separable CRRA form for utility and write $U_C(X_t, L_t) = X_t^{-\gamma}$. A 5% reduction in total output in a recession is associated with an approximately $\gamma \cdot 5\%$ 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.

A second force is the following: 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. Finally, the dependence on wages and productivity comes entirely through the second, benefits-curve force. When marginal costs are high, or $w/\theta$ is high, then firms are operating in the steeper, low-demand portion of the revenues curve.\(^{25}\) Thus,\(^\text{24}\) The state-dependent mean pay-off, evaluated at the optimal policy, is higher when wages are lower, productivity is higher, or aggregate demand is higher.

\(^{25}\)Note that there is no curvature in the total cost curve $C(\cdot)$, so its second derivative in the representation given in Equation 36 is identically zero.
the state itself always acts in the direction of attention cycles, but the wage serves as a countervailing force, decreasing in curvature in downturns.

4.3. **Summary and Predictions**

We have shown how to map the RBC economy to the abstract game theoretic model studied extensively in Section 3. Crucially, we have seen how the curvature of firms’ profit functions depends on both endogenous and exogenous states. When wages are fixed in the Keynesian limit, all of our theoretical results apply, and Theorem 3 guarantees that attention cycles occur as the curvature of profits is diminishing in both productivity and output. When wages are flexible, we cannot directly apply our theoretical results as the endogenous state is multi-dimensional. Nevertheless, the occurrence of attention cycles comes down to a horse race between cyclicality of marginal costs $\frac{w_t}{\theta_t}$ and both the effects of the SDF and aggregate demand externalities. Indeed, we know empirically that the SDF is extremely elastic to the state of the economy (see, for example, Hansen and Jagannathan (1991)) and that wages are relatively acyclical (Solon, Barsky, and Parker, 1994; Grigsby, Hurst, and Yildirmaz, 2019). Any empirically relevant model must therefore have an equilibrium map from the state to the curvature of profits that is decreasing in the state; Proposition 1 then guarantees the occurrence of attention cycles:

**Prediction 1.** Measured firm-level macro attention should be counter-cyclical.

The role of the curvature of firms’ revenue (benefits) functions moreover suggests an auxiliary prediction. When demand is more elastic to aggregate demand, or the Keynesian output-demand feedback is steeper, curvature should be more cyclical. We can translate this into a *cross-sectional* prediction about when our theory will be most applicable:

**Prediction 2.** Measured firm-level macro attention should be more counter-cyclical in contexts where firms face more pro-cyclical demand.

Owing to the fact that wages always act as a countervailing force to both curvature and strategic complementarity, the rigid-wage (or zero lower bound), Keynesian limit of the model features larger attention cycles.\(^{26}\)

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

---

\(^{26}\)The model therefore also makes a prediction that the presence of the zero lower bound should intensify attention cycles, though this is hard to test empirically owing to the fact that the deepest recessions are most associated with the zero lower bound.
firms with more curvature do a better job of reporting contemporaneous macroeconomic conditions.

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

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

## 5. Testing the Theory of Attention Cycles

Having outlined a theory for how and why attention cycles manifest, we now go back to the data to test its additional predictions for both attention and behavior. We will do so in two steps. First, we will return to the original motivating evidence in US public firms’ attention to macro phenomena and show that our model can help interpret *sectoral* heterogeneity in the extent of attention cyclicality. Second, we will turn to more precise, though much smaller-scale, evidence collected by Coibion et al. (2018) for manufacturing firms in New Zealand to directly test the mechanism for our model. These authors’ survey includes a unique question that *indirectly elicits* the curvature of firms’ profit functions in a specific choice variable (here, the optimal reset price). We find that firms that report having more steeply curved profit functions, or higher sensitivity of dollar profits to “mistakes,” are more likely to accurately report or track contemporaneous macroeconomic conditions.

### 5.1. The Cyclicality of Attention by Sector

Predictions 1 and 2 from Section 4 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:

\[
\text{Corr}[\text{MacroAttention}_{nt}, \text{Unemployment}_t] = \alpha + \beta \cdot \text{Corr}[\text{Output}_{nt}, \text{Output}_t] + \epsilon_n
\]  

for which estimates are given in 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 \(-\alpha/\beta = -0.398\) is predicted to have counter-cyclical attention.\(^{27}\)

5.2. 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

\(^{27}\)In Appendix E, 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).
Response | Poorly | Well  
---|---|---
Much more likely | 44.96 | 9.77 
Somewhat more likely | 30.91 | 19.42 
No change | 12.56 | 8.67 
Somewhat less likely | 7.16 | 53.35 
Much less likely | 4.40 | 8.79 
Total | 100.00 | 100.00

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

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 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 written down does that 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 F, 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 F 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 $Z_{it}$:

$$Y_{it} = \alpha + \beta \cdot \text{ProfitCurv}_{it} + \gamma \cdot Z_{it} + \epsilon_{it}$$

(41)

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, though much of these effects is 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 the exact mechanism that our theory proposes: that the
differential stakes of making mistakes is a contributing factor to macro attention.

6. Quantifying Attention Cycles

Having now tested the key predictions of the theory, and thereby evidenced its key mechanism, 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 4 to match our facts regarding the elasticity of measured MacroAttention to the business cycle (Fact 1) and its persistence (Fact 2). This parsimonious extension of the RBC model, which we call RBC-AC, generates four important phenomena not present in the RBC model: asymmetrically large amplification of negative shocks; greater amplification of shocks when output is low; endogenous stochastic volatility of output growth, whereby volatility of output growth is highest when output is lowest; and fast crashes and slow recoveries.

6.1. Primitives

Recall the RBC environment from section 4. We now impose some additional structure for this quantitative application. The representative household has GHH preferences:

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

which we adopt to prevent large income effects in labor supply and ensure that our quantitative model indeed features complementarity in production.\(^{28}\)

The productivity of the intermediate goods firms \( \theta_t \) follows a Gaussian AR(1) process in logs:

\[ \log \theta_t = \rho \log \theta_{t-1} + \nu_t \]

where \( \nu_t \sim IID \ N(0, \sigma_\theta^2) \). The stochastic choice cost functional for the intermediates goods firms is likelihood separable with an entropy kernel. That is, their cost functional for a stochastic choice rule \( P \) is:

\[ c(P) = \lambda \int_\Theta \int_X \phi(p(x|\theta)) \, dx \, d\pi(\theta) \]

with kernel \( \phi(x) = x \log(x) \). That is, the cost of an action distribution in any given state is

---

\(^{28}\)In the presence of large income effects, others producing more could bid up wages by so much to outweigh the aggregate demand externality. In this case, firms’ first-best action would be to produce less when aggregate production is higher and actions would be substitutes.
### Table 3: Parameters.
We select the fixed parameters. Free parameters are estimated by simulated method of moments.

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
</table>
| Fixed | $\gamma$  
|       | CRRA Disutility of labor                        | 5        |
|       | $\psi$  
|       | Elasticity of substitution                       | 1        |
|       | $\varepsilon$  
|       | Elasticity of substitution                       | 3        |
|       | $\rho_\theta$  
|       | Persistence of productivity                      | 0.95     |
| Free  | $\lambda$  
|       | Weight on entropy penalty                        | 13.5     |
|       | $\sigma^2_\theta$  
|       | Variance of the productivity innovation          | 0.00427  |
| Matched | $\sigma^2_Y$  
|       | Variance of quarterly output growth              | 0.353    |
|       | $e_W$  
|       | Elasticity of MacroAttention to employment (Fact 1) | -1.34   |
|       | $\rho_W$  
|       | Persistence of log MacroAttention (Fact 2)       | 0.849    |

6.2. Calibration and Numerical Solution

The RBC-AC model has six parameters. 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.\(^{29}\) 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 $\varepsilon = 3$ for the elasticity of substitution, $\psi = 1$ for the curvature of labor disutility, and $\rho_\theta = 0.95$ for the persistence of the exogenous productivity process $\rho_\theta$. The parameters governing attention cost $\lambda$ and the volatility of the productivity process $\sigma^2_\theta$ are free and we describe their estimation shortly.

Given a vector of these parameters $(\gamma, \psi, \varepsilon, \rho_\theta, \lambda, \sigma^2_\theta)$, we solve the model exactly up to approximating the action and state spaces on uniform grids. Importantly, we do not invoke the quadratic approximation used in the theoretical analysis.\(^{30}\)

We then calibrate both the scaling of attention costs $\lambda$ and the variance of the productivity innovation $\sigma^2_\theta$ using the simulated method of moments. We target three moments: the

---

\(^{29}\)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. Appendix Figure 15 highlights one salient measure for this—the extent of stochastic volatility, which is like the slope of the line plotted in Figure 7.

\(^{30}\)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.
elasticity of MacroAttention to employment, $e_W$ (Fact 1); the persistence of MacroAttention, $\rho_W$ (Fact 2); and the variance of quarterly output growth, $\sigma^2_Y$. We choose these moments so that our quantitative model matches both our stylized facts regarding MacroAttention and generates a sensible mean level of volatility for output growth.

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

$$\log \text{MacroAttention}^M_t = \text{ActionEntropy}(\theta_t) \equiv \int_X \phi(p^*(x|\theta_t)) \, dx \quad (45)$$

The justification for such a mapping comes from the information theoretic interpretation of the tf-idf statistic from which our MacroAttention statistic is derived. In particular, Aizawa (2003) shows that the tf-idf statistic can be interpreted as the reduction in entropy regarding a given set of terms (our macro words) from observing a set of documents (that period’s universe of firm 10-K or 10-Q forms). As a result, our MacroAttention index provides a measure of the information regarding the macroeconomy embedded within firms’ language. As we have used entropy as the cost functional, the endogenous entropy of the action distribution is both the non-pecuniary cost of attention and a measure of the information about the macroeconomy in what firms do. This mapping therefore assumes that there is exactly as much information about the macroeconomy in what firms say and do. With this mapping in hand, given data from the model, we estimate the slope of action entropy to log employment, the persistence of action entropy and the variance of output growth to obtain the model analogues for $e_W$, $\rho_W$ and $\sigma^2_Y$, respectively.$^{31}$

Finally, 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:

$$\{\lambda^*, \sigma^2_{\theta^*}\} = \arg \min_{\lambda, \sigma^2_{\theta}} \sum_{m \in \{e_W, \rho_W, \sigma^2_Y\}} \left( \frac{m^M(\lambda, \sigma^2_\theta)}{m} - 1 \right)^2 \quad (46)$$

The results of this estimation are given alongside all targeted moments and fixed parameters in Table 3.

6.3. Attention Cycles and General Equilibrium Feedback

We now first examine the nature of attention cycles in the calibrated model. To this end, the left pane of Figure 5 plots the production distribution by firms as the state varies on

$^{31}$In both regressions, we add quarterly fixed effects in logs to take care of the seasonal and document-compositional adjustments. See Section 2.2 for more discussion.
Figure 5. **State-dependent Action Distributions and Action Entropy.**
Left pane: equilibrium production distribution by state in the calibrated RBC-AC model. Right pane: entropy of the equilibrium production distribution in the calibrated RBC-AC model (solid) and entropy of the production distribution when firms best reply to the conjecture that they are in the baseline RBC economy (dashed line).

the interval [0.95, 1.05]. Actions are more “sharply peaked” in low states. This is further demonstrated in the right pane of Figure 5 which shows that the entropy of the equilibrium action distribution is monotonically increasing in the state. This figure therefore serves as the quantitative analogue of our Theorem 3: because curvature is monotone decreasing in the state in equilibrium, so too is attention.\(^{32}\)

Attention cycles in the model are partially self-fulfilling. More precise actions cause the economy to better track the underlying state, which causes output to fall by more in lower states. When output falls by more, because of complementarities in production, it is better for agents to produce less and so the aggregate falls. As curvature is higher when the aggregate is lower, this causes an increase in the curvature of firms’ objective functions, leading them to take more precise actions. Consequently, owing to these general equilibrium forces, higher precision begets higher precision. We demonstrate this effect quantitatively in the right pane of Figure 5. In particular, the dashed line shows the relationship between the entropy of the action distribution and the state where firms to best reply to the (mistaken) conjecture that they are in the baseline RBC economy (No GE). This isolates one (in our view, intuitive) measure of the partial equilibrium effect of attention cycles. Concretely, computing the elasticity of macro attention to unemployment in No GE reveals that 65.4% of the elasticity of attention to unemployment in our model stems from GE forces. Therefore, general equilibrium forces do not appear to be a theoretical curiosity: they are quantitatively

\(^{32}\)Curvature is itself decreasing in the state in equilibrium because the stochastic discount factor and aggregate demand externalities both increase the curvature of profits in low states sufficiently to outweigh any effect on curvature from the cyclicality of marginal costs (the presence of which, as mentioned in Section 4.1, technically made the main attention cycles results theoretically ambiguous).
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 dashed lines correspond to the (negative) IRF to a positive productivity shock. The black 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.

relevant for understanding attention cycles.

6.4. The Macroeconomic Implications of Attention Cycles

Having understood the key forces underlying attention cycles, we now explore their significant macroeconomic implications.

Impulse responses. We first explore how the RBC-AC model gives rise to asymmetric propagation of positive and negative shocks, features absent from the vanilla RBC model. In particular, the left pane of Figure 5 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 20.9% 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 6 which shows exactly this: the entropy of agents actions responds 66.4% 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 underlying model.

Second, the RBC-AC model 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 17.7% greater change in output than a 2% negative shock starting at the steady state, and a 42.3% 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. And, in some broader sense, the “Keynesian narrative” for self-fulfilling fluctuations is more true in the downturn even though the primitive properties of the environment (formally, all fixed determinants of strategic interaction) are completely unchanged.

**Endogenous stochastic volatility** An additional prediction which distills the lessons of the previous impulse response analysis is the following: the RBC-AC model endogenously generates higher volatility in lower states of the world. Figure 7 plots the conditional volatility of GDP growth as a function of the underlying state. One sees three important facts. First, the RBC-AC model generates uniformly less volatility than the RBC model given the same driving process for productivity. The reason for this is more familiar: agents are inattentive to shocks, so the economy responds less to them than in the fully attentive benchmark.

---

Figure 7. **Endogenous Stochastic Volatility.**
Conditional volatility of output growth in the RBC-AC model (orange line) and the RBC model (blue line). The dashed orange line is unconditional output volatility in the RBC-AC model, which was a moment to which we calibrated.
(as in Sims (2003) and Gabaix (2016)). Second, the volatility of GDP growth in the RBC model is flat as there is no state-dependence of shock propagation and the driving process for technology features constant volatility. As a result, all of the stochastic volatility generated by the model is endogenous. Third, 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 136% 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.

6.5. Discussion and Extensions

Analytical analogues of the numerical results  In this section we have shown numerically how attention cycles give rise to asymmetric propagation of shocks, state-dependent propagation of shocks and endogenous stochastic volatility. However, we stress that these are not some artifact of our particular calibration, but general features likely to be obtained in any RBC model with attention cycles. We show this in Appendix D, which solves in closed-form a model with attention cycles which generates each of the aforementioned phenomena. Our closed-form example further illustrates the analytical tractability of the theory and scope for researchers to use similar models to explore other issues in the class of games we study.

(In)efficiency of Attention Cycles  Stochastic volatility in our model is not the efficient response to a shock process that gets more vicious in recessions, but instead is entirely driven by cross-firm externalities. In Appendix C, we study the normative properties of the general class of games studied in Section 3 and show that, except under knife-edge conditions unlikely to be satisfied, models with attention cycles will be inefficient. Our efficiency results complement those of the literature on welfare in games with exogenous information Angeletos and Pavan (2007) and with endogenous information acquisition (Angeletos and Sastry, 2019; Hébert and La’O, 2020). In such a world it is tempting to think about corrective policy that insures firms against failure in low states of the world (e.g., lenient bankruptcy law or loan forgiveness) that could fight the fundamental force in our model and draw back the feedback loop contributing to high volatility. We leave a full exposition of such matters, which may require a substantially less stylized model to comment on fully, to future research.
Taking information acquisition seriously  Our main analysis used *likelihood-separable stochastic choice* as a paradigm for imperfect optimization. This tractably captured the key features of interest. However, the main ideas of our analysis extend readily to more standard models of parametric or unrestricted information acquisition, like Sims (2003); models of “sparse” optimization as a function of only a few relevant features of the world, like Gabaix (2014); or other theories of an imperfectly specified decision rule, like Ilut and Valchev (2017). To illustrate this point, we show in Appendix B robustness of our main qualitative results to considering unrestricted information acquisition with mutual information costs. This is an attractive comparison to our basic case because it contributes one additional additive term to the cost functional, which heuristically captures the force of anchoring toward a prior distribution. Because information acquisition can both condition on the conditional prior (e.g., previous state of the world), and asymmetrically uncover the truth of different possible states of the world, agents will try both to learn more in low states and differentially learn about the possibility that the state becomes even lower. This leads to the same attention-severity feedback loop: as long as the proxy for greater attention (sharper information, more considered features, closer approximation of the policy rule) reduces dampened response to the shock of interest, we will obtain the same business cycle properties we have emphasized.

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.33 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. To speak to this force within our framework, in Appendix G we show how the model responds to a reduced-form shock to the cost of paying attention $\lambda$. As one might expect, negative attention shocks or positive uncertainty shocks (formally, positive $\lambda$ shocks) serve the role of dampening the IRF of GDP to productivity shocks. However, in response to a 2% productivity shock, a contemporaneous increase in attention costs by approximately 100% is required to numerically cancel the attention cycle. We therefore argue the quantitative impact of attention cycles is likely to survive in any sensible model that features both forces. A full model that “horse races” these two externalities, each disciplined by direct firm-level measurement, is a fascinating topic that we leave to future study.

---

7. 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 developed an abstract theory to understand why cyclical attention should manifest in both partial and general equilibrium. We further highlighted an important general equilibrium force whereby precise attention to fundamentals and declining aggregates form a positive feedback loop. Mapping the macroeconomic models to the abstract theory, we uncovered two persuasive reasons to suppose that curvature of firms’ profits in their own production should 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. 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.\footnote{One such model that marries social learning with rational inattention is Caplin, Leahy, and Matějka (2015).} 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.

References


Appendices

A. Omitted Proofs

A.1. Proof of Proposition 1

Proof. Consider the stochastic choice problem:

$$\max_{P \in \mathcal{P}} \int_{\Theta} \int_{X} u(x, \theta) \, dP(x|\theta) \, d\pi(\theta) - c(P)$$  \hspace{1cm} (47)$$

When costs are likelihood separable with kernel $\phi$ we have that this problem is given by:

$$\max_{P \in \mathcal{P}} \int_{\Theta} \int_{X} u(x, \theta) p(x|\theta) \, dx \, d\pi(\theta) \, - \int_{\Theta} \int_{X} \phi(p(x|\theta)) \, dx \, d\pi(\theta)$$  \hspace{1cm} (48)$$

Consider now the perturbation of a policy $p$ for a given $\theta$ and $x$ such that $p(x|\theta) > 0$:

$$\tilde{p}(\gamma(\theta)|\theta) = p(\gamma(\theta)|\theta) + \varepsilon$$

$$\tilde{p}(x|\theta) = p(x|\theta) - \varepsilon$$  \hspace{1cm} (49)$$

See that if $p \in \mathcal{P}$, then $\tilde{p} \in \mathcal{P}$. Taking the first-order condition with respect to $\varepsilon$ reveals that a necessary condition for optimality of $p$ is that:

$$u(\gamma(\theta), \theta) - u(x, \theta) = \phi'(p(\gamma(\theta))|\theta) - \phi'(p(x|\theta))$$  \hspace{1cm} (50)$$

Under the assumption that utility is quadratic we have that:

$$u(x, \theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2$$  \hspace{1cm} (51)$$

This first-order condition therefore reduces to:

$$\beta(\theta)(x - \gamma(\theta))^2 = \phi'(p(\gamma(\theta))|\theta) - \phi'(p(x|\theta))$$  \hspace{1cm} (52)$$

Now consider $\theta' > \theta$ such that $\beta(\theta') > \beta(\theta)$ and take $x', x$ such that:

$$|x - \gamma(\theta)| = |x' - \gamma(\theta')|$$  \hspace{1cm} (53)$$
Continue to assume that there is positive density at all points. We therefore have that:

$$
\phi'(p(\gamma(\theta')|\theta')) - \phi'(p(x'|\theta')) = \beta(\theta')(x' - \gamma(\theta'))^2 \\
> \beta(\theta)(x - \gamma(\theta))^2 \\
= \phi'(p(\gamma(\theta)|\theta)) - \phi'(p(x|\theta))
$$

(54)

We now need to consider those points that have no density. Considering the first-order condition for \(p(x|\theta)\) yields:

$$
u(x,\theta) - \phi'(p(x|\theta)) - \lambda(\theta) - \kappa(x,\theta) = 0
$$

(55)

where \(\lambda(\theta)\) is the Lagrange multiplier on the constraint that \(\int_X p(x|\theta) \, dx = 1\) and \(\kappa(x,\theta)\) is the Lagrange multiplier on the constraint that \(p(x|\theta) \geq 0\). When \(p(x|\theta) = 0\), we have that \(\kappa(x,\theta) < 0\). We therefore have that:

$$
\kappa(x,\theta) = u(x,\theta) - \lambda(\theta)
$$

(56)

When utility is quadratic, this is given by:

$$
\kappa(x,\theta) = -\beta(\theta)(x - \gamma(\theta))^2 + \alpha(\theta) - \lambda(\theta)
$$

(57)

which is monotonically decreasing in \(|x - \gamma(\theta)|\). Thus, if there is an \(x\) such that \(p(x|\theta) = 0\), then there exists a \(\bar{x}(\theta)\) such that \(p(x|\theta) = 0\) if and only if \(|x - \gamma(\theta)| \geq |\bar{x}(\theta) - \gamma(\theta)|\). Moreover, we have that \(\bar{x}(\theta)\) can be ranked according to \(\beta(\theta)\). Thus, if \(p(x|\theta) = 0\), then we know that \(p(x'|\theta') = 0\).

Putting everything together we have that:

$$
\phi'(p(\gamma(\theta')|\theta')) - \phi'(p(x'|\theta')) \geq \phi'(p(\gamma(\theta)|\theta)) - \phi'(p(x|\theta))
$$

(58)

for all \(x \in X\). It then follows by the definition of precision that \(p(\theta')\) is more precise about \(\gamma(\theta')\) than \(p(\theta)\) about \(\gamma(\theta)\) under \(\phi'\).

\(\square\)

A.2. Proof of Theorem 1

Proof. To prove existence, we first study the problem of a single agent \(i\) who is best replying to the conjecture that the law of motion of the aggregate is \(\hat{X}: \Theta \rightarrow \mathbb{R}\). See that this agent
faces the problem:

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

(59)

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

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

(60)

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 \( \mathcal{B} = \{ \hat{X} : \Theta \rightarrow \mathbb{R} \} \) as:

\[ k(P, \hat{X}) = \int_{\Theta} \int_{\mathcal{X}} u(x, \hat{X}(\theta), \theta) \, dP(x|\theta) \, d\pi(\theta) - c(P) \]  

(61)

See that \( k \) is jointly lower-semicontinuous in \((P, \hat{X})\) by Assumption 1 owing to continuity of \( u \) and lower semicontinuity of \( c \). Thus, by Berge’s theorem, the maximizing correspondence \( \mathcal{P}^*(\hat{X}) \) is upper hemicontinuous in \( \hat{X} \). Moreover, by Assumption 2 we have that \( c \) is strictly convex and therefore that \( \mathcal{P}^*(\hat{X}) \) is both single-valued and continuous. It immediately follows that in any equilibrium \( P_i^* = P^* = \mathcal{P}^*(\hat{X}) \) 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 \( \hat{X} \) such that:

\[ \hat{X} = X \circ \mathcal{P}^*(\hat{X}) \]  

(62)

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

\[ T(\hat{X}) = X \circ \mathcal{P}^*(\hat{X}) \]  

(63)

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 equilibria. This completes the proof.  

\[ \square \]
A.3. Proof of Lemma 1

Proof. We wish to show that \( f(x) = \log(x) \) and \( f(x) = x \) satisfy quasi-MLRP. Non-emptiness of \( \mathcal{F} \) then immediately follows. To see that quasi-MLRP is satisfied for \( f(x) = \log x \) is immediate as the quasi-MLRP property is just MLRP. Indeed for quasi-MLRP we require for any two distributions \( g', g \in \Delta(\mathcal{X}) \):

\[
\left( f(g'(x')) - f(g'(x)) \geq f(g(x')) - f(g(x)) \quad \forall x' \geq x \right) \implies g' \succeq_{\text{FOSD}} g
\]  

(64)

With \( f(x) = \log(x) \) this becomes:

\[
\left( \frac{g'(x')}{g'(x)} \geq \frac{g(x')}{g(x)} \quad \forall x' \geq x \right) \implies g' \succeq_{\text{FOSD}} g
\]

(65)

The left hand side of this implication is simply the MLRP property. MLRP implies FOSD. Thus, one sees immediately that quasi-MLRP is a weakening of MLRP.

That this weakening is strict is shown by proving that \( f(x) = x \) satisfies quasi-MLRP. This requires us to prove that for any two distributions \( g', g \in \Delta(\mathcal{X}) \):

\[
\left( g'(x') - g'(x) \geq g(x') - g(x) \quad \forall x' \geq x \right) \implies g' \succeq_{\text{FOSD}} g
\]

(66)

To do this, we first prove a technical lemma which may be of use for characterizing other functions that satisfy quasi-MLRP:

Lemma 2. For any two distributions \( g', g \in \Delta(\mathcal{X}) \), the following holds:

\[
\left( f(g'(x')) - f(g'(x)) \geq f(g(x')) - f(g(x)) \quad \forall x' \geq x \right) \implies
\]

\[
\left( \frac{\int_x^x [f(g'(\tilde{x})) - f(g(\tilde{x}))] \, d\tilde{x}}{x - x} \geq \frac{\int_x^x [f(g'(\tilde{x})) - f(g(\tilde{x}))] \, d\tilde{x}}{x - x} \quad \forall x \in \mathcal{X} \right)
\]

(67)

Proof. To prove the required implication, we begin with the hypothesis:

\[
f(g'(x')) - f(g'(x)) \geq f(g(x')) - f(g(x)) \quad \forall x' \geq x
\]

(68)

Which can be rewritten as:

\[
f(g'(x')) + f(g(x)) \geq f(g(x')) + f(g'(x)) \quad \forall x' \geq x
\]

(69)
We now integrate from \( x \) to \( x' \) with respect to \( x \) to obtain the inequality:

\[
(x' - x) f(g'(x')) + \int_x^{x'} f(g(x)) \, dx \geq (x' - x) f(g(x')) + \int_x^{x'} f(g'(x')) \, dx
\]  

(70)

Imposing \( x' = x \) we obtain:

\[
(x - x) [f(g'(x)) - f(g(x))] \geq \int_x^{x} [f(g'(...)) - f(g(...))] \, dx
\]  

(71)

Applying the same procedure but this time integrating from \( x \) to \( \bar{x} \) with respect to \( x' \) and evaluate at \( x' = x \) to obtain this inequality:

\[
\int_x^{\bar{x}} [f(g'(\bar{x})) - f(g(\bar{x}))] \, d\bar{x} \geq (\bar{x} - x) [f(g'(x)) - f(g(x))]
\]  

(72)

Combining our two inequalities we obtain the required one:

\[
\frac{\int_x^{\bar{x}} [f(g'(\bar{x})) - f(g(\bar{x}))] \, d\bar{x}}{\bar{x} - x} \geq \frac{\int_x^{x} [f(g'(...)) - f(...)] \, dx}{x - x}
\]  

\( \forall x \in \mathcal{X} \)  

(73)

Which completes the proof. \( \square \)

If it can be established that:

\[
\left( \frac{\int_x^{\bar{x}} [f(g'(\bar{x})) - f(g(\bar{x}))] \, d\bar{x}}{\bar{x} - x} \geq \frac{\int_x^{x} [f(g'(...)) - f(...)] \, dx}{x - x} \right) \quad \forall x \in \mathcal{X}
\]  

(74)

\[ \implies g' \geq_{\text{FOSD}} g \]

then we will have established that function \( f \) satisfies quasi-MLRP.

We now use this to prove that \( f(x) = x \) satifies quasi-MLRP. Plugging in to the derived integral condition, we obtain:

\[
\frac{G(x) - G'(x)}{\bar{x} - x} \geq \frac{G'(x) - G(x)}{x - x} \quad \forall x \in \mathcal{X}
\]  

(75)

Re-arranging this:

\[
G(x) \geq G'(x) \quad \forall x \in \mathcal{X}
\]  

(76)

which is the definition that \( g' \geq_{\text{FOSD}} g \). This completes the proof and establishes that quasi-MLRP is a strict weakening of MLRP. \( \square \)
A.4. Proof of Theorem 2

Proof. In the proof of Theorem 1, we showed that the set of equilibrium laws of motion for the aggregate $\hat{X} : \Theta \to \mathbb{R}$ are the fixed points of the operator $T : \mathcal{B} \to \mathcal{B}$:

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

(77)

where $\mathcal{B}$ is the space of bounded functions from $\Theta$ to the reals, which we endow with the sup norm. To prove uniqueness, we will use the stated assumptions to prove that $T$ is a contraction map. Once it is established that $T$ is a contraction, we have by the Banach fixed point theorem that there is a unique $\hat{X}$ such that $\hat{X} = T(\hat{X})$ and therefore that there is a unique law of motion of aggregates in equilibrium. By the fact that $\mathcal{P}^*(\hat{X})$ is single-valued (see Theorem 1), we then have that there is a unique $\Omega = (P^*, \hat{X})$.

To show that $T$ is a contraction, we wish to apply Blackwell’s sufficient conditions for an operator to be a contraction. More specifically, if $T$ operates on the space of bounded functions and is endowed with the sup norm, then the following are sufficient for $T$ to be a contraction:

1. Monotonicity: $\hat{X}' \geq \hat{X} \implies T(\hat{X}') \geq T(\hat{X})$ for any $\hat{X}', \hat{X} \in \mathcal{B}$
2. Discounting: there exists $\beta \in (0, 1)$ such that $T(\hat{X} + \alpha) \leq T(\hat{X}) + \beta \alpha$ for all $\alpha \in \mathbb{R}_+$ and any $\hat{X} \in \mathcal{B}$

Toward proving these properties, we first derive some necessary conditions for optimal stochastic choice. These are similar to the conditions used to prove monotone precision in the proof of Proposition 1. To this end, see that the stochastic choice program under an equilibrium conjecture $\hat{X}$ is given by:

$$\max_{p \in \mathcal{P}} \int_{\Theta} \int_X u(x, \hat{X}(\theta), \theta) \, dP(x|\theta) \, d\pi(\theta) - c(P)$$

(78)

By Assumption 7 costs are likelihood separable and this program is given by:

$$\max_{p \in \mathcal{P}} \int_{\Theta} \int_X u(x, \hat{X}(\theta), \theta) \, dP(x|\theta) \, d\pi(\theta) - \int_{\Theta} \int_X \phi(p(x|\theta)) \, dx \, d\pi(\theta)$$

(79)

Take the optimal policy $p$ and now consider a perturbation for actions $x, x' \in \mathcal{X}$ in state $\theta \in \Theta$ such that $p(x|\theta; \hat{X}), p(x'|\theta; \hat{X}) > 0$ by $\varepsilon > 0$:

$$\tilde{p}(x'|\theta; \hat{X}) = p(x'|\theta; \hat{X}) + \varepsilon$$
$$\tilde{p}(x|\theta; \hat{X}) = p(x|\theta; \hat{X}) - \varepsilon$$

(80)
Taking the FOC with respect to $\varepsilon$ and evaluating at $\varepsilon = 0$ yields the following necessary optimality condition:

$$u(x', \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) = \phi'(p(x'|\theta; \hat{X})) - \phi'(p(x|\theta; \hat{X}))$$ (81)

By the previous necessary conditions and the supermodularity assumption (Assumption 3) we have that (for all $x' \geq x$ in the support of both stochastic choice rules and all $\theta$):

$$\phi'(p(x'|\theta; \hat{X}')) - \phi'(p(x|\theta; \hat{X}')) \geq u(x', \hat{X}'(\theta), \theta) - u(x, \hat{X}'(\theta), \theta)$$

$$\geq u(x', \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta)$$ (82)

$$= \phi'(p(x'|\theta; \hat{X})) - \phi'(p(x|\theta; \hat{X}))$$

We now need to check the case where the stochastic choice rules do not have full support. Moreover, if $p(x'|\theta) > 0$ and $p(x|\theta) = 0$ then:

$$u(x', \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) \leq \phi'(p(x'|\theta; \hat{X})) - \phi'(p(x|\theta; \hat{X})) = \phi'(p(x'|\theta; \hat{X})) - \phi'(0)$$ (83)

Now take any $x, x'$ in the support of $p(\theta; \hat{X})$ such that $x' \geq x$. By the complementary slackness conditions, and supermodularity assumption, one of two cases holds: either both $x$ and $x'$ are in the support of $p(\theta; \hat{X}')$; or $x$ is not in the support of $p(\theta; \hat{X}')$ and $x'$ is.

Under the first case, the given argument goes through exactly. In the second case, simply supplement Equation 82 with Equation 83.

In either of the above cases, if $\phi'$ satisfies quasi-MLRP (Assumption 7), then we have that $p(\theta; \hat{X}') \succeq_{FOSD} p(\theta; \hat{X})$ for all $\theta$. It then follows by the monotonicity property of the aggregator (Assumption 5) that $X(p(\theta; \hat{X}')) \geq X(p(\theta; \hat{X}))$ for all $\theta$ and therefore that $T(\hat{X}') \geq T(\hat{X})$, which establishes the required monotonicity property.

We now prove discounting. To this end, we will show that when we take $\hat{X}' = \hat{X} + \alpha$ for $\alpha \in \mathbb{R}_+$ that the resulting stochastic choice is dominated by an $\alpha$ right translation of the original stochastic choice under $\hat{X}$. Where we accumulate all mass at the endpoint $\bar{x}$ if $x - \alpha$ would lie outside of the support. Under this transformation, observe by the necessary condition for optimality and the sufficient concavity condition on utility (Assumption 4) that:

$$\phi'(p(x' - \alpha|\theta; \hat{X})) - \phi'(p(x - \alpha|\theta; \hat{X})) = u(x' - \alpha, \hat{X}(\theta), \theta) - u(x - \alpha, \hat{X}(\theta), \theta)$$

$$\geq u(x', \hat{X}(\theta) + \alpha, \theta) - u(x, \hat{X}(\theta) + \alpha, \theta)$$ (84)

$$= \phi'(p(x'|\theta; \hat{X} + \alpha)) - \phi'(p(x|\theta; \hat{X} + \alpha))$$
Moreover, by quasi-MLRP of \( \phi' \) (Assumption 7), we have that \( p_{-\alpha}(\theta, \hat{X}) \succeq_{\text{FOSD}} p(\theta, \hat{X} + \alpha) \) where \( p_{-\alpha} \) is the described right translation by \( \alpha \) of \( p \). Moreover, by the discounting property of the aggregator (Assumption 6), we then have that:

\[
T(\hat{X} + \alpha) \leq X \circ p_{-\alpha}(\hat{X}) \leq T(\hat{X}) + \beta \alpha
\]  

which establishes the discounting property.

We have now shown that \( T \) satisfies Blackwell’s sufficient conditions and is a contraction map. By the Banach fixed point theorem, there then exists a unique equilibrium \( \Omega \). \( \square \)

A.5. Proof of Proposition 2

Proof. To show that the unique equilibrium aggregate law of motion of monotone in \( \theta \), we use the properties theorem. More formally, Define the set of monotone increasing and bounded functions \( \mathcal{M} = \{ \hat{X} \in \mathcal{B} | \hat{X}(\theta') \geq \hat{X}(\theta) \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta \} \). If we can show that \( T(\hat{X}) \in \mathcal{M} \) for any \( \hat{X} \in \mathcal{M} \), then we know that the unique fixed point of \( T \) is in \( \mathcal{M} \) and therefore that the unique equilibrium law of motion is in \( \mathcal{M} \). To this end, we wish to show that:

\[
\hat{X}(\theta') \geq \hat{X}(\theta) \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta \implies T(\hat{X})(\theta') \geq T(\hat{X})(\theta) \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta
\]

This follows immediately from the necessary condition used in the proof of Theorem 2. More precisely, we use the necessary optimality condition and Assumption 8 by taking \( X' = \hat{X}(\theta') \geq X(\theta) = X \):

\[
\phi'(p(x'|\theta', \hat{X})) - \phi'(p(x|\theta', \hat{X})) = u(x', \hat{X}(\theta'), \theta') - u(x, \hat{X}(\theta'), \theta') \\
\geq u(x', \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) \\
= \phi'(p(x'|\theta, \hat{X})) - \phi'(p(x|\theta, \hat{X}))
\]

By the quasi-MLRP property of \( \phi' \) (Assumption 7) we then have that \( p(\theta'; \hat{X}) \succeq_{\text{FOSD}} p(\theta; \hat{X}) \) and thus by the monotonicity of the aggregator (Assumption 5) that \( T(\hat{X})(\theta') \geq T(\hat{X})(\theta) \). By application of the standard properties theorem, we then have that the unique fixed point \( \hat{X} \in \mathcal{M} \), proving the result. \( \square \)

A.6. Proof of Theorem 3

Proof. This proof follows from an appropriately modified general equilibrium version of the proof of Proposition 1. Under Assumptions 1 - 7, by Theorem 2, we have that there exists a unique equilibrium law of motion \( \hat{X} \). Moreover, additionally under Assumption 8, by
Proposition 3, we have that \( \hat{X} \) is monotone increasing in \( \theta \). Recall also by Theorem 1, that the unique symmetric stochastic choice rule consistent with the unique equilibrium \( \hat{X} \) solves the following program:

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

(88)

Where under assumption 7, we have that this program is given by:

\[
p \in \arg \max_{p \in \mathcal{P}} \int_{\mathcal{X}} \int_{\mathcal{Y}} u(x, \hat{X}(\theta), \theta) dP(x|\theta) d\pi(\theta) - \int_{\mathcal{X}} \int_{\mathcal{Y}} \phi(p(x|\theta)) dx d\pi(\theta)
\]  

(89)

where we will suppress the dependence of the optimal policy on \( \hat{X} \) as it is unique. Consider now a perturbation of the equilibrium optimal policy \( p \) for a given \( \theta \) and \( x \) such that \( p(x|\theta) > 0 \):

\[
\tilde{p}(\gamma(\hat{X}(\theta), \theta)|\theta) = p(\gamma(\hat{X}(\theta), \theta)|\theta) + \varepsilon
\]

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

(90)

See that if \( p \in \mathcal{P} \), then \( \tilde{p} \in \mathcal{P} \). Taking the first-order condition with respect to \( \varepsilon \) and evaluating at \( \varepsilon = 0 \) reveals that a necessary condition for optimality of \( p \) is that:

\[
u(\gamma(\hat{X}(\theta), \theta), \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) = \phi'(p(\gamma(\hat{X}(\theta), \theta)|\theta)) - \phi'(p(x|\theta))
\]  

(91)

Under Assumption 9, we moreover, have that

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

(92)

Thus our necessary condition simplifies to:

\[
\beta(\hat{X}(\theta), \theta)(x - \gamma(\hat{X}(\theta), \theta))^2 = \phi'(p(\gamma(\hat{X}(\theta), \theta)|\theta)) - \phi'(p(x|\theta))
\]  

(93)

Now consider \( \theta' \geq \theta \). We have by monotonicity of \( \hat{X} \) that \( \hat{X}(\theta') \geq \hat{X}(\theta) \). It follows that \( \beta(\hat{X}(\theta'), \theta') \leq \beta(\hat{X}(\theta), \theta) \) given our assumption that \( \beta \) is monotonically decreasing in both arguments. Now take \( x, x' \) such that:

\[
|x - \gamma(\hat{X}(\theta), \theta)| = |x' - \gamma(\hat{X}(\theta'), \theta')|
\]  

(94)
It follows that:

\[
\phi'(p(\gamma(\hat{X}(\theta),\theta)|\theta)) - \phi'(p(x|\theta)) = \beta(\hat{X}(\theta),\theta)(x - \gamma(\hat{X}(\theta),\theta))^2 \\
\geq \beta(\hat{X}(\theta'),\theta')(x' - \gamma(\hat{X}(\theta'),\theta'))^2 \\
= \phi'(p(\gamma(\hat{X}(\theta'),\theta')|\theta')) - \phi'(p(x'|\theta))
\] (95)

We now that to take care of those points that have no density. To this end consider the first-order condition for \(p(x|\theta)\):

\[
u(x,\hat{X}(\theta),\theta) - \phi'(p(x|\theta)) - \lambda(\theta) - \kappa(x,\theta) = 0 (96)
\]

where \(\lambda(\theta)\) is the Lagrange multiplier on the constraint that \(\int_{\mathcal{X}} p(x|\theta) = 1\) and \(\kappa(x,\theta)\) is the Lagrange multiplier on the constraint that \(p(x|\theta) \geq 0\). When \(p(x|\theta) = 0\), we have that \(\kappa(x,\theta) \leq 0\). Given our assumption on utility, this is given by:

\[
\kappa(x,\theta) = -\beta(\hat{X}(\theta),\theta)(x - \gamma(\hat{X}(\theta),\theta))^2 + \alpha(\theta) - \lambda(\theta) (97)
\]

which is monotonically decreasing in \(|x - \gamma(\hat{X}(\theta),\theta)|\). Thus, if there is an \(x\) such that \(p(x|\theta) = 0\), then there exists an \(\bar{x}(\theta)\) such that \(p(x|\theta) = 0\) if and only if \(|x - \gamma(\hat{X}(\theta),\theta)| \geq |\bar{x}(\theta) - \gamma(\hat{X}(\theta),\theta)|\). Moreover, by monotonicity of \(\beta(\hat{X}(\theta),\theta)\) is \(\theta\), we have that \(p(x|\theta) = 0\) implies that \(p(x'|\theta') = 0\). Hence, we have always that:

\[
\phi'(p(\gamma(\hat{X}(\theta),\theta)|\theta)) - \phi'(p(x|\theta)) \geq \phi'(p(\gamma(\hat{X}(\theta'),\theta')|\theta')) - \phi'(p(x'|\theta))
\] (98)

for all \(x \in \mathcal{X}\). It follows then by the definition of precision that \(p(\theta)\) is more precise about \(\gamma(\hat{X}(\theta),\theta)\) than \(p(\theta')\) about \(\gamma(\hat{X}(\theta'),\theta')\) under \(\phi'\).

\[A.7. \ \text{Proof of Proposition 3}\]

\[\text{Proof.}\ \text{Recall the form of the firm’s profit function:}\]

\[
u(x_{it}, X_t, w_t, L_t, \theta_t) = U_C(X_t, L_t) \cdot (B(x_{it}, X_t) - C(x_{it}, w_t, \theta_t))
\] (99)

where benefits under the CES demand assumption reduce to:

\[
B(x_{it}; X_t) = x_{it}^{1 - \frac{\epsilon}{2}} X_t^{\frac{\epsilon}{2}}
\] (100)
and reduce under the linear production assumption to:

\[ C(x_{it}; \theta_t, w_t) = x_{it} \frac{w_t}{\theta_t} \] (101)

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

\[ \left(1 - \frac{1}{\varepsilon}\right) x_{it}^{\varepsilon - \frac{1}{2}} X_t^{\frac{1}{2}} = \frac{w_t}{\theta_t} \] (102)

yielding:

\[ x_{it}^* = \left(1 - \frac{1}{\varepsilon}\right) \varepsilon X_t \left(\frac{w_t}{\theta_t}\right)^{-\varepsilon} \] (103)

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

\[ u(x_{it}, X_t, w_t, L_t, \theta_t) = u(x_{it}^*, X_t, w_t, L_t, \theta_t) + u_x(x_{it}, X_t, w_t, L_t, \theta_t)|_{x_{it}=x_{it}^*}(x - x_{it}^*) \]
\[ + \frac{1}{2} u_{xx}(x_{it}, X_t, w_t, L_t, \theta_t)|_{x_{it}=x_{it}^*}(x - x_{it}^*)^2 + O^3(x_{it}) \] (104)

Application of the envelope theorem implies that:

\[ u_x(x_{it}, X_t, w_t, L_t, \theta_t)|_{x_{it}=x_{it}^*} = 0 \] (105)

Thus, our approximation reduces to the quadratic form:

\[ u(x_{it}, X_t, w_t, L_t, \theta_t) = u(x_{it}^*, X_t, w_t, L_t, \theta_t) \]
\[ + \frac{1}{2} u_{xx}(x_{it}, X_t, w_t, L_t, \theta_t)|_{x_{it}=x_{it}^*}(x - x_{it}^*)^2 + O^3(x_{it}) \] (106)

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

\[ u(x_{it}^*, X_t, w_t, L_t, \theta_t) = U_C(X_t, L_t) \cdot (B(x_{it}^*, X_t) - C(x_{it}^*, w_t, \theta_t)) \]
\[ = U_C(X_t, L_t) \left(X_t \left(\frac{w_t}{\theta_t}\right)^{1-\varepsilon}\right) \left([1 - \frac{1}{\varepsilon}]^{\varepsilon(1-\frac{1}{2})} - [1 - \frac{1}{\varepsilon}]^\varepsilon\right) \]
\[ = U_C(X_t, L_t) \left(X_t \left(\frac{w_t}{\theta_t}\right)^{1-\varepsilon}\right) \varepsilon^{-\varepsilon} (\varepsilon - 1)^{\varepsilon-1} \] (107)

We now characterize the curvature:

\[ u_{xx}(x_{it}, X_t, w_t, L_t, \theta_t)|_{x_{it}=x_{it}^*} = U_{C}(X_t, L_t) \cdot (B_{xx}(x_{it}^*, X_t) - C_{xx}(x_{it}^*, w_t, \theta_t)) \] (108)
where we have that:

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

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

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

(109)

and \( C_{xx} \equiv 0 \). Plugging the above in yields the claimed curvature and completes the proof.  

\[ \Box \]

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. We take the following as a definition of posterior-separable cost functionals. Denti (2019) provides this formulation as a representation theorem in stochastic choice space of the usual posterior-based definition of Caplin and Dean (2013):

**Definition 5** (Posterior-Separable Cost Functionals). A cost functional \( c \) has a posterior-separable representation if and only if there exists a convex and continuous \( \phi \) such that:

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

(110)

where:

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

(111)

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 1). 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 4.** 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 $\bar{x} = -\bar{x} = \hat{x}$, the optimal stochastic choice rule is given by:

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

(113)

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

$$x|\theta \sim N \left( \gamma(\theta), \frac{\lambda}{2\beta(\theta)} \right)$$

(114)

**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 (2019). 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:

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

(115)

---

35Formally, all actions are exchangeable in the prior if:

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

(112)
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 $S_x^1, S_x^2 \subseteq S_x$ of positive measure such that $g(\theta|s_1) \neq g(\theta|s_2)$ for all $s_1 \in S_x^1, s_2 \in S_x^2$. Now generate a new signal structure $g'$ such $\tilde{s} \in S_x^1$. Suppose $\tilde{s}$ was sent whenever any $s \in S_x^1 \cup S_x^2$ 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 $c_I$:

$$
\max_{P \in \mathcal{P}} \int_\Theta \int_X u(x, \theta) \, dP(x|\theta) \, d\pi(\theta) - c_I(P) \tag{116}
$$

Moreover, separating terms, one achieves the following representation of $c_I$:

$$
c_I(p) = \int_\Theta \int_X p(x|\theta) \log p(x|\theta) \, dx \, d\pi(\theta) - \int_X p(x) \log p(x) \, dx \tag{117}
$$

where:

$$
p(x) = \int_\Theta p(x|\theta) \, d\pi(\theta) \tag{118}
$$

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)

$$
\mathcal{L}(\{p(x|\theta), \kappa(x, \theta)\}_{x \in X, \theta \in \Theta}, \{\gamma(\theta)\}_{\theta \in \Theta}) = \int_\Theta \int_X u(x, \theta) p(x|\theta) \, dx \, d\pi(\theta) - \lambda \left( -\int_X p(x) \log p(x) \, dx + \int_\Theta \int_X p(x|\theta) \log p(x|\theta) \, dx \, d\pi(\theta) \right) + \kappa(x, \theta) p(x|\theta) + \gamma(\theta) \left( \int_X p(x|\theta) \, dx - 1 \right) \tag{119}
$$

Any time that $p(x|\theta) > 0$, taking the FOC pointwise with respect to $p(x|\theta)$ and rearranging we have that:

$$
p(x|\theta) = \frac{p(x) \exp\{u(x, \theta)\}}{\int_X p(\tilde{x}) \exp\{u(\tilde{x}, \theta)\} \, d\tilde{x}} \tag{120}
$$
Moreover, we can plug the above back into the general problem and take the FOC. Re-
arranging we have that for all $x$ such that $p(x) > 0$:

$$\int_\Theta \int_X p(\tilde{x}) \exp\{u(\tilde{x}, \theta)\} \, d\tilde{x} \, d\pi(\theta) = 1$$

(121)

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:

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

(122)

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

$$\int_\Theta \int_X \exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\} \pi(\theta) d\theta = 1 \quad \forall x \in X$$

(123)

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

$$p(x|\theta) = \frac{\exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\}}{\int_X \exp\{-\beta(\theta)\lambda^{-1}(\tilde{x} - \gamma(\theta))^2\} \, d\tilde{x}}$$

(124)

The denominator of this expression can be computed:

$$\int_X \exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\} \, dx = \int_X \sqrt{\frac{2\pi}{2\beta(\theta)}} \exp\left\{-\frac{1}{2} \left( \frac{x - \gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}} \right)^2 \right\} \, dx$$

$$= \sqrt{\frac{2\pi}{2\beta(\theta)}} \int_X \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left( \frac{x - \gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}} \right)^2 \right\}$$

$$= \sqrt{\frac{2\pi}{2\beta(\theta)}}$$

(125)

It follows that:

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

(126)

Which is to say that $X|\theta$ is a Gaussian random variable with mean $\gamma(\theta)$ and variance
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. The following corollary is immediate but conceptually important:

**Corollary 1.** In the environment of Proposition 4, if \( \beta(\theta) \) is decreasing (increasing) in \( \theta \), then \( \text{Var}(X|\theta) \) is increasing (decreasing) in \( \theta \).

**Proof.** From Proposition 4, we have that:

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

(127)

If \( \beta(\theta) \) is monotone increasing (decreasing) in \( \theta \) it follows immediately that \( \text{Var}(X|\theta) \) is monotone decreasing (increasing) in \( \theta \).

Thus, if the curvature of payoffs is monotone in the state then the agent’s actions will also feature monotone variance in the underlying state of the world. As we show in Section 4, many macro-finance models when approximated have the exact feature that curvature is state-monotone. Direct application of this result then yields the important prediction that attention, precision and the propensity to make mistakes will all also be monotone in the state, generating attention cycles.

Away from the restrictive 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 5.

**Proposition 5.** 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 \( \varepsilon > 0 \):

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

(128)

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 4 that the optimal stochastic choice rule is given by:

\[
p(x|\theta) = \frac{p(x) \exp\{u(x, \theta)\}}{\int_{X} p(\tilde{x}) \exp\{u(\tilde{x}, \theta)\} d\tilde{x}}
\] (129)

which under our assumption on payoffs is given by:

\[
p(x|\theta) = \frac{p(x) \exp\{-\hat{\beta}(\theta)(x - \gamma(\theta))^2\}}{\int_{X} p(\tilde{x}) \exp\{-\hat{\beta}(\theta)(\tilde{x} - \gamma(\theta))^2\} d\tilde{x}}
\] (130)

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

\[
\frac{dp(x|\theta)}{d\varepsilon} \bigg|_{\varepsilon=0} = p(x|\theta) \left[-(x - \gamma(\theta))^2 + \int_{X} (\tilde{x} - \theta)^2 p(\tilde{x}|\theta) d\tilde{x}\right] \text{I}[|\theta - \hat{\theta}| \leq \delta] + p(x|\theta) \left[\frac{d\log p(x)}{dx} \bigg|_{\varepsilon=0} - \int_{X} \frac{d\log p(x)}{dx} \bigg|_{\varepsilon=0} p(\tilde{x}|\theta) d\tilde{x}\right]
\] (131)

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

\[
\lim_{\delta \to 0} \frac{dp(x|\theta)}{d\varepsilon} \bigg|_{\varepsilon=0} = p(x|\theta) \left[\frac{d\log p(x)}{dx} \bigg|_{\varepsilon=0} - \int_{X} \frac{d\log p(x)}{dx} \bigg|_{\varepsilon=0} p(\tilde{x}|\theta) d\tilde{x}\right] \quad \forall \theta \neq \hat{\theta}
\] (132)

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:

\[
\lim_{\delta \to 0} \frac{dp(x)}{d\varepsilon} \bigg|_{\varepsilon=0} = \int_{\Theta} \lim_{\delta \to 0} \frac{dp(x|\theta)}{d\varepsilon} \bigg|_{\varepsilon=0} \pi(\theta) d\theta
\] (133)

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

\[
\lim_{\delta \to 0} \frac{dp(x)}{d\varepsilon} \bigg|_{\varepsilon=0} = 0
\] (134)

It follows that:

\[
\lim_{\delta \to 0} \frac{d\log p(x|\theta)}{d\varepsilon} \bigg|_{\varepsilon=0} = \left[-(x - \gamma(\theta))^2 + \int_{X} (\tilde{x} - \theta)^2 p(\tilde{x}|\theta) d\tilde{x}\right] \text{I}[|\theta - \hat{\theta}| \leq \delta]
\] (135)

where \( \int_{X} (\tilde{x} - \theta)^2 p(\tilde{x}|\theta) d\tilde{x} > 0 \). 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 will
tend to give rise to more precise attention, they key idea in the theory.

C. (In)efficiency of Attention Cycles

A further question of interest is when equilibria of our model are efficient. This is of rele-
vance as this determines whether attention cycles will have any normative content. As our
agents are symmetric, *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 6.** A stochastic choice rule \( p^E \in \mathcal{P} \) is efficient if it solves the following program:

\[
p^E \in \arg \max_{p \in \mathcal{P}} \int_{\Theta} \int_{X} u(x, X(p(\theta)), \theta) \, dP(x|\theta) \, d\pi(\theta) - c(p) \tag{136}
\]

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 equi-
librium be efficient?

**Proposition 6.** Suppose that there exists a unique efficient \( p^E \) obtained as an interior so-
lution of the efficient program. Moreover, suppose that the aggregator is linear:

\[
X(g) = \int_{X} f(x)g(x) \, dx \tag{137}
\]

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

\[
\int_{X} u_X (x, X(p^*(\theta)), \theta) \, dP^*(x|\theta) \, dx = 0 \tag{138}
\]

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

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

\[
p^E \in \arg \max_{p \in \mathcal{P}} \int_{\Theta} \int_{X} u(x, X(p(\theta)), \theta) \, dP(x|\theta) \, d\pi(\theta) - c(p) \tag{139}
\]

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

\[
X(g) = \int_{X} f(x)g(x) \, dx \tag{140}
\]
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:

\[
\frac{\partial c(p^E)}{\partial p^E(x'|\theta)} - \frac{\partial c(p^E)}{\partial p^E(x|\theta)} = u(x', X(p^E(\theta)), \theta) - u(x, X(p^E(\theta)), \theta) \\
+ \int_{X} [f(x') - f(x)] u_X(\tilde{x}, X(p^E(\theta)), \theta) p^E(\tilde{x}|\theta) \, d\tilde{x}
\]

(141)

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:

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

(142)

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:

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

(143)

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

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

(144)

This equality implies that:

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

(145)

which can be re-expressed as:

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

(146)

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:

\[
\int_X u_X(\tilde{x}, X(p^E(\theta)), \theta)p^E(\tilde{x}|\theta) \, d\tilde{x} = \int_X u_X(\tilde{x}, X(p(\theta)), \theta)p(\tilde{x}|\theta) \, d\tilde{x} = 0 \tag{147}
\]

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 7.** Suppose that utility is quadratic:

\[
u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)(x - \gamma(X, \theta))^2 \tag{148}
\]

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

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

\[
\text{MSE}[p, \theta] = \int_X (x - \gamma(X(p), \theta))^2p(x) \, dx \tag{149}
\]

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

\[
0 = \alpha_X(X(p^*(\theta)), \theta) - \beta_X(X(p^*(\theta)), \theta)\text{MSE}[p^*(\theta), \theta]
+ 2\gamma_X(X(p^*(\theta)), \theta)\beta_X(X(p^*(\theta)), \theta)\text{Bias}[p^*(\theta), \theta] \tag{150}
\]

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 \( p^* \):

\[
\frac{d \log \beta(X(p^*(\theta)), \theta)}{d \log X} = 2\gamma_X(X(p^*(\theta)), \theta) \frac{\text{Bias}[p^*(\theta), \theta]}{X(p^*(\theta))} \text{Movement of Optimal Choice} \left( \frac{\text{MSE}[p^*(\theta), \theta]}{X(p^*(\theta))} \right)^2 \text{Bias-dispersion Ratio} \tag{151}
\]
for almost all $\theta \in \Theta$.

**Proof.** Recall from Proposition 6 that a necessary condition for efficiency of an equilibrium $p$ is that:

$$
\int_X u_X(\tilde{x}, X(p(\theta)), \theta) p(\tilde{x}|\theta) d\tilde{x} = 0
$$

(152)

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

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

(153)

We therefore have that:

$$
u_X(x, X, \theta) = \alpha_X(X, \theta) - \beta_X(X, \theta)(x - \gamma(X, \theta))^2 + 2\gamma_X(X, \theta)\beta(X, \theta)(x - \gamma(X, \theta))
$$

(154)

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

$$
0 = \int_X \left[ \alpha_X(X(p(\theta)), \theta) - \beta_X(X(p(\theta)), \theta)(\tilde{x} - \gamma(X(p(\theta)), \theta))^2 + 2\gamma_X(X(p(\theta)), \theta)\beta(X(p(\theta)), \theta)(\tilde{x} - \gamma(X(p(\theta)), \theta)) \right] p(\tilde{x}|\theta) d\tilde{x}
$$

(155)

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

$$
0 = \alpha_X(X(p^*(\theta)), \theta) - \beta_X(X(p^*(\theta)), \theta) \text{MSE}[p^*(\theta), \theta] + 2\gamma_X(X(p^*(\theta)), \theta) \beta(X(p^*(\theta)), \theta) \text{Bias}[p^*(\theta), \theta]
$$

(156)

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:

$$
\beta_X(X(p^*(\theta)), \theta) \text{MSE}[p^*(\theta), \theta] = 2\gamma_X(X(p^*(\theta)), \theta) \beta(X(p^*(\theta)), \theta) \text{Bias}[p^*(\theta), \theta]
$$

(157)

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

$$
\frac{\beta_X(X(p^*(\theta)), \theta) X(p^*(\theta))}{\beta(X(p^*(\theta)), \theta)} = 2\gamma_X(X(p^*(\theta)), \theta) \frac{\text{Bias}[p^*(\theta), \theta]}{X(p^*(\theta))} \left( \frac{\text{MSE}[p^*(\theta), \theta]}{X(p^*(\theta))} \right)^2
$$

(158)
Which is to say that:

\[
\frac{d \log \beta(X(p^*(\theta)), \theta)}{d \log X} = 2 \gamma_X(X(p^*(\theta)), \theta) \frac{\text{Bias}[p^*(\theta), \theta]}{X(p^*(\theta))} \left( \frac{\text{MSE}[p^*(\theta), \theta]}{X(p^*(\theta))} \right)^2
\] (159)

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 and there will often be an explicit role for policy intervention.

\section*{D. Attention Cycles in Closed-Form}

In this section we develop a simple-closed form example to exemplify how attention cycles arise in general equilibrium and their implications. We show explicitly how all of the quantitative phenomena in the calibrated RBC-AC model arise qualitatively in this simple example. In particular, the simple example features asymmetric shock propagation, state dependent shock propagation and endogenous stochastic volatility.

To this end, we specialize the general environment from the previous section by imposing more structure on payoffs, cost functionals, and aggregators. We assume that that payoff function is a weighted-quadratic one, with weighting function and tracking functions being quadratic and linear combinations of the aggregate and the state respectively:

\[
u(x, X, \theta) = h(X, \theta) + \frac{1}{\delta X + \varepsilon \theta + \omega X^2} (x - (\beta X + \gamma \theta))^2
\] (160)

which can naturally be justified through an appropriate quadratic approximation of utility functions. We further specify that the cost functional is the average entropy of the stochastic choice rule employed by the agent:

\[
c(p) = \lambda \int_\Theta \int_X p(x|\theta) \log p(x|\theta) \, dx \, d\pi(\theta)
\] (161)

which generates logistic choice. Finally, we take the aggregator as a mean-variance aggregator for \(a \in \mathbb{R}\) which can be justified as an approximation of a CES aggregator:

\[
X(g) = \mathbb{E}_g[x] - \alpha \text{Var}_g[x]
\] (162)
We will also take the action set \( \mathcal{X} = \mathbb{R} \).

In this environment, the equilibrium can be solved for in closed-form. This is stated formally in the following proposition.

**Proposition 8.** The optimal stochastic choice rule is given by:

\[
x|\theta \sim N\left(\beta \hat{X}(\theta) + \gamma \theta, \lambda \left(\delta \hat{X}(\theta) + \varepsilon \theta + \omega \hat{X}(\theta)^2\right)\right)
\]

(163)

When \( \gamma - \alpha \lambda \varepsilon > 0 \) and \( \alpha \lambda \omega > 0 \), there is a unique equilibrium. This unique equilibrium is inefficient when \( h(X, \theta) = h(\theta) \). Moreover, in the unique equilibrium, the aggregate is a monotonically increasing and concave function of the state and given in closed-form by:

\[
\hat{X}(\theta) = -\frac{(1 - \beta + \alpha \lambda \delta) + \sqrt{(1 - \beta + \alpha \lambda \delta)^2 + 4 \alpha \lambda \omega (\gamma - \alpha \lambda \varepsilon) \theta}}{2 \alpha \lambda \omega}
\]

(164)

The equilibrium cross-sectional variance \( \text{Var}(X|\theta) \) is monotonically increasing in the state. Finally, in the limit of \( \omega \to 0 \), the unique equilibrium is linear in the state and given by:

\[
\hat{X}(\theta) = \frac{\gamma - \alpha \lambda \varepsilon}{1 - \beta + \alpha \lambda \delta} \theta
\]

(165)

**Proof.** We first derive the optimal stochastic choice rule given any equilibrium law of motion \( \hat{X} \). To this end, see that the agents all face the problem:

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

(166)

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)

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

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

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

(167)

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

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

(168)
Which can be re-arranged as:

\[ p(x|\theta) \propto \exp\{\lambda^{-1}u(x, \hat{X}(\theta), \theta)\} \quad (169) \]

where the constant of proportionality ensures that \( \int_X p(x|\theta) \, dx = 1 \) for all \( \theta \in \Theta \). Given that utility has the form:

\[ u(x, X, \theta) = h(X, \theta) + \frac{1}{\delta X + \varepsilon \theta + \omega X^2} (x - (\beta X + \gamma \theta))^2 \quad (170) \]

We then have that the conditional density satisfies:

\[ p(x|\theta) \propto \exp\left\{-\lambda^{-1} \frac{1}{\delta X + \varepsilon \theta + \omega X^2} (x - (\beta X + \gamma \theta))^2\right\} \quad (171) \]

which is the kernel of a Gaussian random variable with variance \( \lambda(\delta \hat{X}(\theta) + \varepsilon \theta + \omega \hat{X}(\theta)^2) \) and mean \( \beta \hat{X}(\theta) + \gamma \theta \). This yields the first claim:

\[ X|\theta \sim N \left(\beta \hat{X}(\theta) + \gamma \theta, \lambda \left(\delta \hat{X}(\theta) + \varepsilon \theta + \omega \hat{X}(\theta)^2\right)\right) \quad (172) \]

We moreover have that the aggregator is given by:

\[ X(g) = \mathbb{E}_g[x] - \alpha \text{Var}_g[x] \quad (173) \]

Plugging the obtained mean and variance into our aggregator we obtain the following fixed-point equation for the equilibrium aggregate:

\[ \hat{X}(\theta) = \beta \hat{X}(\theta) + \gamma \theta - \alpha \lambda(\delta \hat{X}(\theta) + \varepsilon \theta + \omega \hat{X}(\theta)^2) \quad (174) \]

This is simply a quadratic equation:

\[ \alpha \lambda \omega \hat{X}(\omega)^2 + (1 - \beta + \alpha \lambda \delta) \hat{X}(\theta) - (\gamma - \alpha \lambda \varepsilon) \theta = 0 \quad (175) \]

When we have that \( \gamma - \alpha \lambda \varepsilon > 0 \), then the unique equilibrium is the positive root of this equation:

\[ \hat{X}(\theta) = \frac{-(1 - \beta + \alpha \lambda \delta) + \sqrt{(1 - \beta + \alpha \lambda \delta)^2 + 4 \alpha \lambda \omega (\gamma - \alpha \lambda \varepsilon) \theta}}{2 \alpha \lambda \omega} \quad (176) \]

We now verify that this is increasing and concave. To this end, see that we can write:

\[ \hat{X}(\theta) = a_0 + \sqrt{a_1 + a_2 \theta} \quad (177) \]
where \(a_0 < 0, a_1 > 0\) and \(a_2 > 0\). Thus:

\[
\hat{X}'(\theta) = \frac{a_2}{\sqrt{a_1 + a_2 \theta}} > 0
\]

\[
\hat{X}''(\theta) = -\frac{1}{2}a_2^2(a_1 + a_2 \theta)^{-\frac{3}{2}} < 0
\]

(178)

We now prove the claim that \(Var(X|\theta)\) is monotone increasing in the state. To this end, see that:

\[
Var(X|\theta) = \lambda \left( \delta \hat{X}(\theta) + \varepsilon \theta + \omega \hat{X}(\theta)^2 \right)
\]

(179)

Taking the derivative of this with respect to the state yields:

\[
\frac{dVar(X|\theta)}{d\theta} = \lambda \left( \delta \hat{X}'(\theta) + \varepsilon + \omega \hat{X}'(\theta) \hat{X}(\theta) \right) > 0
\]

(180)

by monotonicity of the equilibrium \(\hat{X}\) in \(\theta\).

We now verify the claim that the equilibrium is inefficient when \(h(X, \theta) = h(\theta)\). As this environment satisfies all of the assumptions of Proposition 7, we can directly apply the derived necessary condition for efficiency:

\[
0 = \alpha_X(X(p^*(\theta)), \theta) - \beta_X(X(p^*(\theta)), \theta)MSE[p^*(\theta), \theta]
+ 2\gamma_X(X(p^*(\theta)), \theta)\beta(X(p^*(\theta)), \theta)Bias[p^*(\theta), \theta]
\]

(181)

where we note that \(Bias[p^*(\theta), \theta] = 0\) as \(X|\theta\) has mean given by \(\beta \hat{X}(\theta) + \gamma \theta\) and \(\alpha_X = 0\). We therefore require that:

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

(182)

However, we have that \(MSE[p^*(\theta), \theta] \neq 0\) and \(\beta_X(X(p^*(\theta)), \theta) \neq 0\). It follows that the unique equilibrium is inefficient.

To prove the final part of the proposition, we derive \(\hat{X}(\theta)\) in the limit of \(\omega \to 0\). Applying L’Hopital’s immediately yields:

\[
\lim_{\omega \to 0} \hat{X}(\theta) = \frac{\gamma - \alpha \lambda \varepsilon}{1 - \beta + \alpha \lambda \delta} \theta
\]

(183)

Completing the proof.

This proposition captures in closed-form much of the analysis so far. In particular, it provides us with a concrete model which exemplifies our more general theoretical results: equilibrium exists, equilibrium is unique, equilibrium aggregates are monotone increasing in
the state, the cross-sectional variance of actions is increasing in the state and the equilibrium is inefficient. Crucially, therefore, this example captures exactly the phenomenon of attention cycles. The underlying cause of these attention cycles is the fact that the curvature of the utility function is monotonically decreasing in the state in equilibrium. This causes agents to care less about matching the optimal action in good states of the world and therefore gives rise to counter-cyclical attention. The inefficiency of these cycles stems from the fact that agents do not internalize the impact of their actions on the curvature of others’ utility. This leaves room for policy to improve welfare.

We now develop the macroeconomic implications of the current model. In particular, we show in this simple example how the model generates: state-dependent shock propagation; asymmetric shock propagation; and endogenous stochastic volatility. These analytical results shed light on the numerical results we obtain in the calibration of the RBC model.

To this end, let us now extend the above model and allow for a persistent fundamental that is perfectly revealed at the end of each period. More formally, suppose that time is discrete \( t \in \mathbb{N} \) and the fundamental \( \theta \) follows the \( AR(1) \) process in logs:

\[
\log \theta_t = \rho \log \theta_{t-1} + \varepsilon_t
\]  

where \( \varepsilon_t \sim \Lambda \) and \( \theta_t \geq 0 \) for all \( t \). Moreover, we suppose that each period, the agents play the described one-shot game. It follows that period-by-period, the equilibrium aggregate is given by:

\[
\hat{X}(\varepsilon_t, \theta_{t-1}) = \frac{-(1 - \beta + \alpha \lambda \delta) + \sqrt{(1 - \beta + \alpha \lambda \delta)^2 + 4\alpha \lambda \omega (\gamma - \alpha \lambda \varepsilon) (\theta^\nu_{t-1} \varepsilon_t^\nu)}}{2\alpha \lambda \omega}
\]  

which can be rewritten as:

\[
\hat{X}(\varepsilon_t, \theta_{t-1}) = -a + \sqrt{a^2 + b(\theta^\nu_{t-1} \varepsilon_t^\nu)}
\]  

for constants:

\[
a = \frac{1 - \beta + \alpha \lambda \delta}{2\alpha \lambda \omega} > 0
\]

\[
b = \frac{\gamma - \alpha \lambda \varepsilon}{\alpha \lambda \omega} > 0
\]  

Note additionally in the \( \lambda \to 0 \) limit where there is no stochastic choice that the law of motion is:

\[
\hat{X}^{NSC}(\varepsilon_t, \theta_{t-1}) = \frac{\gamma}{1 - \beta} (\theta^\nu_{t-1} \varepsilon_t^\nu)
\]
which is linear in $e_t$ and log-linear in $\epsilon_t$.

We now show how the model with stochastic choice generates state dependent shock propagation, asymmetric shock propagation and endogenous stochastic volatility. Note first that none of these features are present in the model without stochastic choice: the law of motion $\hat{X}^{NSC}$ is always linear in the shock.

Toward seeing that the model features asymmetric shock propagation, see that log $\hat{X}$ is concave in $\epsilon_t$. Thus for any shock of a given size $\epsilon_t > 0$:

$$| \log \hat{X}(\epsilon_t, \theta_{t-1}) - \log \hat{X}(0, \theta_{t-1})| = | \log \hat{X}(-\epsilon_t, \theta_{t-1}) - \log \hat{X}(0, \theta_{t-1})| \tag{189}$$

To see that the model generates state-dependent amplification, with greater amplification in lower states, observe again by the concavity of $\hat{X}$ that whenever $\theta_{t-1}' > \theta_{t-1}$:

$$| \log \hat{X}(\epsilon, \theta_{t-1}) - \log \hat{X}(0, \theta_{t-1}')| = | \log \hat{X}(\epsilon, \theta_{t-1}) - \log \hat{X}(0, \theta_{t-1})| \tag{190}$$

To see how the model generates endogenous stochastic volatility, we approximate the law of motion $\hat{X}$ to first-order in $\epsilon_t$ around $\epsilon_t = 0$:

$$\log \hat{X}(\epsilon_t, \theta_{t-1}) \approx \log \hat{X}(0, \theta_{t-1}) + \partial_\epsilon \log \hat{X}(0, \theta_{t-1}) \epsilon_t \tag{191}$$

Thus:

$$Var(\log \hat{X}(\epsilon_t, \theta_{t-1})|\theta_{t-1}) \approx \partial_\epsilon \log \hat{X}(0, \theta_{t-1})^2 \sigma_\epsilon^2 \tag{192}$$

Moreover, as $\hat{X}$ is concave, the coefficient above $\hat{X}'(0, \theta_{t-1})^2$ is highest when the state is lowest. Thus, stochastic volatility is highest in the lowest states and lowest in the highest states.

We therefore see that this example not only captures the essence of the more general theoretical analysis, but also generates interesting behavior for aggregates. Namely, the existence of endogenous attention cycles leads to: asymmetrically greater propagation of negative shocks to fundamentals; fast crashes and slow recoveries; and greater volatility in bad states of the world. This example importantly provides intuition for the qualitative features of the numerical exercise.
E. 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.

E.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. 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.

E.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.

36In the essentially zero-measure cases in which there is a tie, we take the alphabetically first ticker.
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.

E.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

---

Figure 8. **Macro Attention from Conference Calls versus 10K/Q.**

Both series are seasonally adjusted, as indicated by (7).
Figure 9. **Macro Attention Cyclicality versus Output Cyclicality, with Conference Calls.**
The sample for attention cyclicality is 2004-Q1 to 2014-Q4. The sample for output cyclicality is 2005-Q1 to 2017-Q4. This is the same exercise plotted in Figure 4.

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.

### E.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) 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.

**F. Coibiaon et al. (2018) Survey**

In this section, we describe precisely how we incorporate the survey results of Coibiaon 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.

**F.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 (12), 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} (193)

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 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:

$$\text{ProfitCurv}_i = \beta \cdot \hat{X}_i + e_{it}$$  \hspace{1cm} (194)

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.

### Summary Statistics

| Mean | Quantiles 5 | 25 | 50 | 75 | 95 | 0.280 | 0.020 | 0.05 | 0.12 | 0.28 | 1.00 |
|------|-------------|----|----|----|----|-------|-------|------|------|------|------|------|

### Correlates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Norm. coef.</th>
<th>t-stat</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of price review</td>
<td>-0.106</td>
<td>-7.80</td>
<td>0.011</td>
</tr>
<tr>
<td>log output</td>
<td>-0.066</td>
<td>-9.43</td>
<td>0.015</td>
</tr>
<tr>
<td>Firm age</td>
<td>-0.117</td>
<td>-10.17</td>
<td>0.014</td>
</tr>
<tr>
<td>Employment</td>
<td>-0.122</td>
<td>-7.19</td>
<td>0.015</td>
</tr>
<tr>
<td>Labor share</td>
<td>-0.138</td>
<td>-7.98</td>
<td>0.020</td>
</tr>
<tr>
<td>Number of competitors</td>
<td>0.130</td>
<td>6.81</td>
<td>0.017</td>
</tr>
<tr>
<td>log income</td>
<td>0.015</td>
<td>0.55</td>
<td>0.000</td>
</tr>
<tr>
<td>Some or more college</td>
<td>0.043</td>
<td>1.87</td>
<td>0.002</td>
</tr>
<tr>
<td>Tenure at firm</td>
<td>-0.117</td>
<td>-5.73</td>
<td>0.014</td>
</tr>
<tr>
<td>Tenure in industry</td>
<td>-0.058</td>
<td>-2.33</td>
<td>0.003</td>
</tr>
<tr>
<td>Manager age</td>
<td>-0.091</td>
<td>-3.25</td>
<td>0.008</td>
</tr>
</tbody>
</table>

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.
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>

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.
F.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.

G. Uncertainty Shocks in the RBC-AC Model

As discussed in Section 6, we here study the effect of uncertainty shocks in the RBC-AC model. Formally, we model such shocks by imagining that the multiplicative scaling of the cost of attention were itself a random process, with mean $\bar{\lambda}$, persistence $\rho_\lambda$, and the possibility for idiosyncratic shocks $\nu^\lambda_t$:

$$\lambda_t = \rho_\lambda \lambda_{t-1} + (1 - \rho_\lambda) \bar{\lambda} + \nu^\lambda_t$$ (195)

We might interpret a positive realization of $\nu^\lambda_t$, or shock to the cost of attention, as a reduced-form equivalent of the difficulty of acquiring information or planning actions in a low-information environment. Let us take $\rho_\lambda = 0.95$ to match the persistence of productivity shocks and $\bar{\lambda}$ equal to our calibrated value from earlier for quantification.

Figure 10 shows the impulse response of output to a shock that increases the cost of attention by 10% of its steady-state value, for two possible values of productivity $\theta$ (held constant throughout). Shocks to the cost of paying attention increase output in our model because they fight the mechanism we have in mind—that more “tuned-in” firms will respond more aggressively to productivity shocks. This force is stronger in a relatively low productivity state in which all reactions and stakes are magnified.

What happens if try to capture the idea that recessions involve simultaneous shocks to fundamentals and the cost of acquiring information? To this end, we repeat the impulse response exercise plotted in the black solid line of Figure 6—a 2% reduction in productivity
starting from steady-state—and now add on top a positive shock to $\lambda_t$ with initial magnitude equal to various percentages of $\bar{\lambda}$. Larger positive shocks dampen the negative effect of the productivity shock on output (left panel), and in particular dampen the amplification coming from firms’ playing more precise actions (right panel). To cancel out this effect completely on impact requires an approximately 100% increase in the per-unit cost of paying attention. To first approximation, the corresponding “benefits percent change” coming from higher marginal utility is $\gamma \cdot \{\text{Percent GDP Reduction}\} = 5 \cdot 4\% = 20\%$. So, quantitatively, very large increases in the cost of acquiring information seem necessary to fully cancel out the amplification force we have identified in this paper.

Benchmarking these numbers against plausible, micro-calibrated estimates for the effects of low production on the difficulty of planning or acquiring information is among the tasks we leave for a more structured future study of both mechanisms side-by-side.
Figure 11. Attention and productivity shocks, Combined.
H. Additional Tables and Figures

Figure 12. Macro Attention and Unemployment.
The fitted line is from OLS regression and the t-statistic is heteroskedasticity-robust.

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 13. Text Frequency for Every Word.
Figure 14. **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></td>
<td>45</td>
</tr>
<tr>
<td>Corr[Output$_{nt}$, Output$_t$] ($\beta$)</td>
<td>0.251</td>
<td>(0.100)</td>
<td>0.111</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: **Industry Output and Attention Cyclicality.**  
This regression corresponds to the trend-line in Figure 4.
Figure 15. 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 $\theta = 0.02$ to $\theta = -0.02$. 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.