Attention Cycles

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

November 9, 2021

Abstract

Using data from US public firms' regulatory filings and financial statements, we document that firms' attention to macroeconomic conditions rises in downturns and that their propensity to make input-choice mistakes rises in booms. We explain these phenomena with a business-cycle model in which firms face a cognitive cost of making precise decisions. Because firms are owned by risk-averse households, there are greater incentives to deliver profits when aggregate consumption is low. Thus, firms exert more cognitive effort and make smaller input-choice mistakes in aggregate downturns. In the data, consistent with our model, financial markets punish mistakes more in downturns and macroeconomically attentive firms make smaller mistakes. When calibrated to match our firm-level evidence, *attention cycles* generate quantitatively significant asymmetric, state-dependent shock propagation and stochastic volatility of output growth.

*MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. Email: jpflynn@mit.edu.
†MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. Email: ksastry@mit.edu.
We thank Daron Acemoglu, George-Marios Angeletos, Martin Beraja, Ricardo Caballero, Oğuzhan Çelebi, Lawrence Christiano, Roberto Corrao, Ian Dew-Becker, Martin Eichenbaum, Michele Fornino, Drew Fudenberg, Antoine Levy, Chen Lian, Jeremy Majerovitz, Andrea Manera, Nicolas Minudri, Stephen Morris, Jacob Moscona, George Nikolakoudis, Christina Patterson, Giorgio Primiceri, Parinitha Sastry, Lawrence Schmidt, Christropher Sims, Alp Simsek, John Sturm, Adrien Verdelhan, Iván Werning, Christian Wolf, Alexander Wolitzky and participants in the MIT Macro Lunch, MIT Theory Lunch, MIT Sloan Finance Lunch, the 2021 NBER Summer Institute Impulse and Propagation Mechanisms Meeting, and Northwestern University Macroeconomics Seminar for helpful comments. First posted version: May 4, 2020.

1 Introduction

Firms often make decisions which, from an observer's *ex post* perspective, are inconsistent with profit maximization. An influential explanation for this behavior, prominently articulated by Simon (1947), is that firms' managers face constraints on their attention and decisionmaking capacity. Under this view, changing microeconomic and macroeconomic conditions shape firm-level incentives in the allocation of both physical and cognitive resources. A direct implication is that the state of the economy as a whole may be central for determining firms' apparent "bounded rationality."

In this paper, we study the joint determination of *business cycles*, aggregate fluctuations in economic production, and *attention cycles*, aggregate fluctuations in cognitive effort and mistakes. Our analysis has three parts. First, we introduce two strategies to measure the attention cycle, which capture attention to the macroeconomy in "what firms say" and precision in "what firms do." We find that firms speak more about the macroeconomy and make smaller input-choice mistakes, defined relative to a model benchmark, in downturns. Second, to interpret this evidence, we develop a macroeconomic model in which firms face a cognitive cost of making precise decisions. Because firms are owned by risk-averse investors, incentives for precise profit maximization are higher when aggregate consumption is low. This *risk-pricing channel* rationalizes our finding of smaller mistakes during downturns. Third, we combine our model and evidence to quantify the effects of attention cycles on the dynamic properties of aggregate output and labor productivity. We find that state-dependent attention explains a quantitatively significant fraction of the observed asymmetry and state-dependence in the responses of these macroeconomic aggregates to shocks.

Motivating Evidence To establish the premise that firms' allocation of attention varies with the business cycle, we begin our analysis with a model-free measurement of attention to the macroeconomy in firms' language. Our dataset is the full text of all US public firms' end-of-quarter and end-of-year financial performance reports (Forms 10-Q and 10-K) from 1995-2017. We measure each filing's attention to macroeconomic developments using a natural-language-processing technique that compares the filing's word choice to that of macroeconomics references, which we take to be introductory college-level textbooks. This method, which builds on the approach used by Hassan et al. (2019) to study firm-level attention to political risks, is designed to isolate attention to the macroeconomy from the standard vocabulary of firm communication. We find that aggregate macroeconomic attention in language is counter-cyclical, increasing in years of high unemployment or below-trend performance of the S&P 500.

This finding suggests that firms pay more attention to external conditions during downturns. Later, we relate our measure of attention in firms' language with a measure of mistakes in firms' input choices. To build toward that goal, we first introduce a model of firms' state-dependent attention, production and "mistake-making."

Theoretical Analysis Our model describes firms' state-dependent choices of attention and production in equilibrium. Firms face a cognitive cost in making their input choices contingent on the microeconomic and macroeconomic state, or "steadying their hand" to precisely respond to shocks in productivity, demand, or input prices. They choose state-contingent, and potentially imperfect, plans to maximize risk-adjusted profits net of this cognitive cost. Both the premise of costly planning and the emphasis on state-dependence contrasts our approach with that of the existing macroeconomic literature on decision frictions (e.g., Woodford, 2003; Maćkowiak and Wiederholt, 2009; Angeletos and La'O, 2010; Gabaix, 2020). The environment is Neoclassical, with aggregate demand externalities (Blanchard and Kiyotaki, 1987) and an aggregate shock that shifts the productivity distribution to generate a business cycle. We show that equilibrium analysis in this setting is well-posed and tractable, proving equilibrium existence, uniqueness, and monotonicity of output in the aggregate shock. This allows us to jointly study the causes and consequences of attention cycles in the model's unique equilibrium.

In partial equilibrium, firms pay more attention, measured by the extent of their cognitive effort, and make smaller mistakes, measured by the variance of their actions around the *ex post* ideal point, when the cost of making mistakes is high. Formally, the correct metric for such costs is the curvature of agents' objective functions in their own action. When firms choose production to maximize risk-adjusted profits, the aforementioned curvature is the product of two terms: the curvature of firms' profits, which is highest when aggregate output and firm-specific productivity are low in standard parameterizations; and the stochastic discount factor (SDF), which is highest when aggregate consumption is low because risk-averse households own the firms. We use these observations to characterize patterns of choice, attention, and misoptimization among firms.

In general equilibrium, the cyclicality of the aforementioned incentives is determined by household risk aversion, the extent of aggregate demand externalities, and the elasticity of wages to real output. We show that, when household relative risk aversion exceeds an empirically modest lower bound of one plus the elasticity of real wages to output, equilibrium cognitive effort decreases in output and the average size of agents' mistakes increases in output. Put differently, market incentives push firms toward paying more attention to decisions and precisely maximizing current profits when the aggregate economy is doing poorly.

We next use the model to investigate the macroeconomic consequences of this mechanism. We show that output is the product of the counterfactual output under the full-attention benchmark with an *attention wedge* that is less than one. The wedge arises because inattentive, stochastic choices translate into dispersion of value marginal products ("misallocation") across firms and reduced aggregate total factor productivity (TFP). Due to counter-cyclical attention, the wedge widens when the economy is booming and firms are more lax in their profit maximization. As a result, misallocation across firms in the model is endogenously higher in booms than recessions, providing a rational-inattention analogue of the cleansing effect of recessions (as defined, for instance, by Caballero and Hammour, 1994). We show that, dynamically, the amplification of negative shocks via increased attention leads to asymmetric, state-dependent shock propagation and endogenous stochastic volatility.

Testing the Theory: The Misoptimization Cycle Our model's predictions for misoptimization in input choices can be taken directly to the data. In the next part of the paper, we develop and implement tests of these predictions. We use data on firm production and input choices from public firms' financial statements from 1986-2018, collected in Compustat Annual Fundamentals. We estimate firm-level TFP using conventional methods and estimate empirical policy functions for labor choice conditional on these TFP estimates, time-invariant firm-level fixed effects, and sector-by-time fixed effects. We show that, in our model, the empirical policy function estimates the counterfactual "unconstrained optimal" input choice of firms and that its residuals estimate the *ex post* misoptimizations. We validate that these measured misoptimizations have negative effects on firms' stock returns and profitability in the data, consistent with our interpretation.

Our main aggregate-level finding is that the variance of firms' misoptimizations is procyclical, as predicted by the main case of our model. In a linear regression of our *Misoptimization Dispersion* measure on the unemployment rate, a five percentage point increase in the latter predicts a 53% decrease in the former. While the pro-cyclicality of Misoptimization Dispersion contrasts with existing evidence that firm fundamentals like TFP have counter-cyclical variance (e.g., Kehrig, 2015; Bloom et al., 2018), we underscore that our finding is empirically and theoretically compatible with a story in which dispersion *around* fundamentals decreases in downturns, reducing misallocation, while the fundamentals themselves become more volatile. We conduct a battery of robustness checks to build confidence that our finding is not driven by cyclicality in unmodeled features, like adjustment costs or financial frictions, or in measurement error and misspecification.

We next test the main model mechanism for counter-cyclical attention, markets' greater punishment of mistakes of the same size during downturns. Specifically, we investigate whether the measured negative relationship of misoptimizations with stock returns and profitability steepens in recessions. We find strong evidence in the case of returns and weak evidence in terms of profitability. We interpret this evidence as validation of our macro-financial story of attention cycles: markets particularly reward firms for delivering profits in recessions, which incentivizes precise, attentive responses to changing business conditions in these states.

We also link back to our motivating evidence and document that macroeconomic attention in language is associated with smaller misoptimizations at the firm level. This result suggests that our initial, model-free finding of counter-cyclical "macroeconomic attention" is closely related to our subsequent, model-implied finding of pro-cyclical misoptimization.

Quantification In a final section, we assess our findings' quantitative importance in a numerical calibration of the model. We calibrate the cognitive friction and coefficient of relative risk aversion to match the level and cyclicality of misoptimizations in the data, and use external calibrations for the substitutability between goods and elasticity of real wages to output. This approach lets the data speak directly toward our model's novel mechanism, while using standard calibrations to discipline that mechanism's interaction with other forces.

We find that negative productivity shocks have larger effects than equal-sized positive ones; all shocks have larger effects when the aggregate state is low; and macroeconomic volatility is highest in low states. These findings can rationalize an economically significant fraction of observed non-linearities and state-dependence of macroeconomic dynamics as consequences of state-dependent attention. In the model, negative productivity shocks have a 7% larger effect on aggregate output (5% larger on employment) than positive shocks of the same size. This amounts to 25% of the asymmetry estimated by Ilut et al. (2018) using industry-level data on productivity and employment. Similarly, in the model, an aggregate shock that replicates a 5% peak-to-trough reduction in output from steady-state, comparable to what was experienced during the Great Recession, generates also a 11% increase in the conditional volatility of output. This is 19% of the increase in statistical uncertainty about output growth between the trough of the Great Recession and the preceding peak as measured by Jurado et al. (2015).

Related Literature Our modeling of cognitive frictions via stochastic choice is a departure from more standard techniques of modeling information acquisition and/or processing in the macroeconomics literature. This choice is motivated in our context by an interest in directly studying firms' choices. A byproduct is that we can analytically study the effects of firms' "mistakes" in equilibrium.[1] The distinction between stochastic choice and information acquisition for equilibrium predictions is discussed at length on a more theoretical level

[1] Our interest in firm choices and "mistakes" also relates to literature that has studied various forms of cyclical cleansing and restructuring without emphasizing cognitive frictions, such as Caballero and Hammour (1994, 1996), Koenders and Rogerson (2005), and Berger (2012).

in Morris and Yang (2021) and in our own work, Flynn and Sastry (2021).[2] Complementary recent work by Ilut and Valchev (2021) models cognitive constraints as learning about an unknown mapping from states to actions, focusing on different applications and using numerical analysis of aggregative equilibrium.

Our study, despite its methodological differences, nonetheless contributes to a large literature on cognitive frictions and behavioral inattention in macro models. Most such models (e.g., Woodford, 2003; Maćkowiak and Wiederholt, 2009; Angeletos and La'O, 2010; Gabaix, 2020) have ignored both state-dependence of attention and its equilibrium consequences.[3] Sims (2003) and Gabaix (2014) show, in different models sharing a "rational attention" premise, that optimal costly attention should be tuned to the most payoff-relevant attributes of the decision problem. In this vein, Mäkinen and Ohl (2015), Chiang (2021), and Benhabib et al. (2016) share our focus on the cyclicality of attention due to firms' changing incentives.[4] We, in contrast to all three studies, motivate our analysis with and quantitatively benchmark our findings against empirical evidence on cyclical misoptimization and risk-pricing incentives. Our approach of modeling stochastic choice also allows us to develop this mapping to the data and to study macroeconomic implications for state-dependent dynamics.

Our empirical findings relate to a large literature studying the cyclicality of microeconomic dispersion. Our measurement of pro-cyclical misoptimization is consistent with work by Kehrig (2015) and Bloom et al. (2018) documenting that dispersion in firm-level productivity rises in recessions in the US manufacturing sector. Concretely, *fundamentals* are more volatile in downturns, while misoptimizations around them are less volatile. Our findings are also qualitatively consistent with those of Eisfeldt and Rampini (2006) and Bachmann and Bayer (2014), who document pro-cyclical investment-rate dispersion; Dew-Becker and Giglio (2020), who document acyclicality of implied volatility in the cross-section of firm returns; Ilut et al. (2018), who document larger response of employment to negative versus positive productivity shocks; and Berger and Vavra (2019), who document counter-cyclical responsiveness of price-setters to nominal shocks.

[2]In older literature on game theory, Harsanyi (1973) Selten (1975), and Myerson (1978) study the implications of related models of "trembling hands" or control costs for equilibrium predictions. In decision theory, Fudenberg et al. (2015) study the axiomatization of a control-cost formulation like our own.

[3]Some studies of costly adjustment of prices, including Gorodnichenko (2008) and Alvarez et al. (2011), or of infrequent and costly adjustment of investment portfolios, including Abel et al. (2013) and Kacperczyk et al. (2016), draw on intuitively similar mechanisms operating through the curvature of payoffs, but consider only partial equilibrium mechanisms.

[4]In Mäkinen and Ohl (2015), decreasing returns to scale lead firms to demand more information when aggregate productivity is low, while we emphasize a novel risk-pricing mechanism. Contemporaneous work by Chiang (2021) has considered a complementary mechanism in the context of a model where firms acquire Gaussian signals about the state. Benhabib et al. (2016) predict that firms should have less demand for information during recessions, a prediction at odds with our empirical findings.

Our model provides a micro-foundation for cyclical allocative inefficiency. Our focus on cyclicality distinguishes our work from that of David et al. (2016), David and Venkateswaran (2019), and Ma et al. (2020), who study the steady-state effects of cognitive frictions on inefficiency. Our approach to quantify misallocation via a structural interpretation of *choice dispersion* differs from the existing literature quantifying misallocation using structurally estimated marginal products (Hsieh and Klenow, 2009).[5]

Finally, a recent literature studies how firms "think" using their language choice in regulatory filings (Loughran and McDonald, 2011; Song and Stern, 2021) and earnings calls (Hassan et al., 2019, 2020). We complement these efforts with novel measurement of the macroeconomic content of firms' language. While not intrinsically related to firms' attention, papers that study the macroeconomic content of news coverage (Baker et al., 2016; Bybee et al., 2021) and monetary policy communication (Lucca and Trebbi, 2009; Hansen and McMahon, 2016; Handlan, 2020) share a similar approach.

Overview The rest of the paper proceeds as follows. In Section 2, we present our motivating evidence on macroeconomic attention. In Section 3, we introduce our model. In Section 4, we present our theoretical results on the cyclical allocation of attention and its macroeconomic implications. In Section 5, we build our empirical measure of misoptimizations and test the model's predictions. In Section 6, we calibrate our model and analyze the impact of attention cycles on macroeconomic dynamics. Section 7 concludes.

2 Motivating Evidence: The Macroeconomic Attention Cycle

In this section, we describe our strategy to measure macroeconomic attention in firms' language and present our finding that firms discuss macroeconomic topics more in downturns.

2.1 Data and Measurement

Our data source is the full text of the quarterly 10-Q and annual 10-K reports submitted by all US public firms to the Securities and Exchange Commission (SEC). We use data from 1995 to 2018 in our main analysis.[6] Our total sample consists of 479,403 individual documents, or about 5,000 per quarter, which we index by their date of filing.[7]

[5] Macaulay (2020) similarly treats choice dispersion as evidence of misoptimization in a study of households' cyclical attention toward savings choices.

[6] The relevant digitized documents are hosted by the Security and Exchange Commission's EDGAR (Electronic Data Gathering, Analysis, and Retrieval), which began operation in 1994. We choose 1995 as a starting point at which a nearly comprehensive sample of firms' reports are available in the system.

[7] This timing convention allows 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, not the period being discussed.

Our premise is that firms devote a significant proportion of their filings to enumerating key risks for their business, and via their choice of language reveal a proxy of attention to specific domains.[8] The key challenge for identifying attention toward macroeconomic risks is to differentiate characteristic language of macroeconomics from the intrinsically economic and financial vocabulary of standard firm activities (e.g., "credit" and "costs"). To address this, we apply a simple natural language processing technique that identifies specific documents as "attentive to the macroeconomy" if their word choice is both different from the standard word choice in regulatory filings and similar to the word choice of macroeconomics-focused references. Following the method pioneered by Hassan et al. (2019) to study firm attention to political risks, we use introductory college-level textbooks as reference texts. This choice balances our considerations of keeping the relevant macroeconomic vocabulary mostly non-technical (e.g., "unemployment" instead of "tightness"), but still specific (e.g., "inflation" instead of "price").

To operationalize this method, we first define $\text{tf}(w)_{it}$ as the *term frequency* for a word w in the communication of firm i at time t, measured as the proportion of total English-language words; and $\text{df}(w)$ as the *document frequency* of a given word w among all observed (i, t) regulatory filings, measured as a proportion of total documents that use the word at least once. We summarize the frequency of a word w in document (i, t), relative to its overall frequency among our studied documents, via the "term frequency inverse document frequency," or tf-idf, defined in the following expression:

$$\text{tf-idf}(w)_{it} := \text{tf}(w)_{it} \cdot \log\left(\frac{1}{\text{df}(w)}\right) \tag{1}$$

The log functional form is a heuristic in natural language processing for scaling the relative importance of each term, and it is bounded below by 0 when a word appears in all documents.[9] Thus $\text{tf-idf}(w)_{it}$ is high when the word w is particularly *characteristic* of the (i, t) document relative to others.

We next measure whether this characteristic language compares with the language of three undergraduate macroeconomic textbooks: *Macroeconomics* and *Principles of Macroeconomics* by N. Gregory Mankiw and *Macroeconomics: Principles and Policy* by William J. Baumol and Alan S. Blinder.[10] In each book, we calculate the tf-idf of each individual word

[8] To give an example, the automaker General Motors in its summary report for 2009 highlighted the threats of the ongoing Great Recession to its business: "[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."

[9] One justification for the functional form is derived from information theory by Aizawa (2003).

[10] We use electronic copies of the 7th, 3rd, and 12th editions of these books, respectively.

using term frequencies in the book and (inverse) document frequencies among regulatory filings, to find words that are common in the textbooks but relatively out of place in standard firm communication.[11] We rank the top 200 words by this metric in each textbook, and take the intersection among the three books as a final set of 89 words.[12] Appendix Figure A1 prints these words in alphabetical order, and plots their time-series frequency. Many of the words relate to common macro indicators ("unemployment","inflation"); some to the topic or profession itself ("macroeconomics","economist"); and some to policy ("Fed," "multiplier"). There are also "false positive" words that are related to pedagogy, like "question" and "equation." To allow the method to be fully devoid of direct researcher manipulation, we do not remove such words from the main analysis.

We then use our set of macroeconomic words, denoted by $\mathcal{W}_M$, to calculate our firm-by-time measures of attention as the sum of the (idf-weighted) macroeconomic word frequency:

$$\text{MacroAttention}_{it} = \sum_{w \in \mathcal{W}_M} \text{tf-idf}(w)_{it} \tag{2}$$

To generate an measure MacroAttention_t, we average $\text{MacroAttention}_{it}$ across firms. In our main aggregate results, we remove seasonal trends modeled as quarter-of-the-year means.

2.2 Attention to the Macroeconomy is Counter-Cyclical

Figure 1 plots the time series of our (log) macroeconomic attention metric, at the quarterly frequency and net of seasonal trends, against two macroeconomic time series: the US unemployment rate (Unemployment_t) and the linearly detrended log price of the S&P 500 ($\log \text{SPDetrend}_t$).[13] We find that macroeconomic attention persistently rises when the macroeconomy and financial market are distressed, most tellingly around the 2008 financial crisis and subsequent recession.

To measure the cyclicality of MacroAttention, we estimate a linear regression of MacroAttention on each macroeconomic variable, or

$$\log \text{MacroAttention}_t = \alpha + \beta_Z \cdot Z_t + \epsilon_t \tag{3}$$

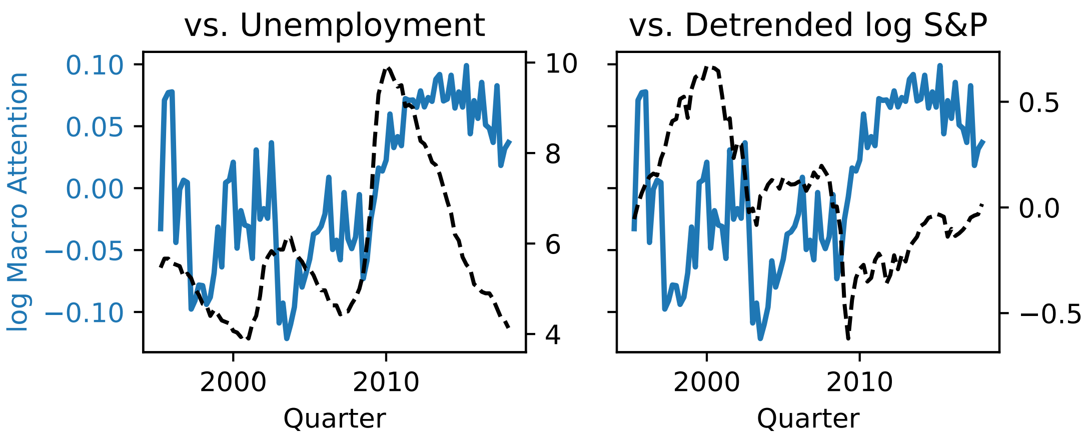
for $Z_t \in \{\text{Unemployment}_t/100, \log \text{SPDetrend}_t\}$. Our coefficient estimates are, respectively,

[11] To reduce the number of required computations, we calculate document frequencies using the Master Dictionary of words in 10-Ks on Bill McDonald's website (McDonald, 2021).

[12] Taking the intersection helps guard against the idiosyncratic language of certain books. For instance, in *Principles of Macroeconomics* by N. Gregory Mankiw, 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.

[13] We remove seasonal trends in log Macro Attention by subtracting averages in each quarter of the year. For this reason, the measure has zero mean by construction.

Figure 1: Macro Attention is Counter-Cyclical



Notes: The blue line, measured by the left axis of each plot, is the log of Macro Attention, adjusted for seasonality. The black dashed lines, measured by the right axis of each plot, are respectively the US unemployment rate and linearly detrended S&P 500 price.

1.529 (SE: 0.405) and -0.104 (SE: 0.029), with R^2 values of 0.180 and 0.237.[14] The former estimate conveys that a five percentage point swing in unemployment from peak to trough of the business cycle predicts an approximately 7.6% increase in macroeconomic attention.

To measure the persistence of Macro Attention, we estimate the following AR(1) model:

$$\log \text{MacroAttention}_t = \alpha + \rho \cdot \log \text{MacroAttention}_{t-1} + \epsilon_t \tag{4}$$

The point-estimate auto-regressive coefficient is $\hat{\rho} = 0.820$ (SE: 0.050), implying a shock half-life of 3.5 quarters.

2.3 Interpretation and Discussion

We interpret the above findings as suggestive evidence that firms' attention allocation varies with the business cycle and, in particular, focuses more intensely on macroeconomic risks during downturns. But, by itself and without a model interpretation, the evidence above does not *directly* demonstrate any change in the process or outcome of firm decisionmaking over the business cycle. In Sections 3 and 4, we will write a macroeconomic model that structures the translation from "attention" to "actions," describes the macroeconomic causes and consequences of a model-consistent "Attention Cycle," and generates predictions that can be directly tested in firm decisions. Before proceeding to this macroeconomic model, we

[14]Standard errors for these results and the next result are heteroskedasticity and auto-correlation (HAC) robust with a four-quarter Bartlett kernel bandwidth. Each regression has 80 observations.

briefly summarize robustness checks and additional exercises based on our Macro Attention measure.

Comparison with the News In Appendix Figure A3, we compare our Macro Attention series at the quarterly frequency with the News Index of Baker et al. (2016), which captures newspaper discussion of economic policy, and the "macroeconomic" sub-topics of Bybee et al. (2021), who use machine learning to flexibly categorize the text of *Wall Street Journal* by topic.[15] Macro Attention has, respectively, correlations of 0.21 and 0.17 with each. The news series are much more sharply peaked around turning points, while the firm attention measure is considerably more persistent. We take this as evidence that firm-level attention requires a substantially different theoretical interpretation which emphasizes more persistent incentives for attention.[16] Our model, in turn, will predict these persistent incentives.

Industry-Level Patterns We might expect the cyclicality of macroeconomic attention to differ across differently cyclical industries. To study this heterogeneity, we partition our sample into 44 different industries.[17] For each industry, we calculate "output cyclicality" as the correlation between sectoral GDP growth, calculated using quarterly BEA data since 2005 linked appropriately to NAICS-definition sectors, with aggregate nominal GDP growth. In Appendix Figure A4, we plot in the cross section of industries the relationship between this output cyclicality and the coefficient of sector-level Macro Attention on the US unemployment rate. Of the 42 industries, 36 feature counter-cyclical macroeconomic attention. The extent of counter-cyclicality increases with the industry's output cyclicality, but almost acyclical industries also have counter-cyclical attention. The model, consistent with this observation, will focus on a risk-pricing mechanism that gives firms a uniform incentive to concentrate attention in aggregate downturns regardless of the cyclicality of firms' demand.

Alternative Data Construction Appendix Figure A1 plots the time-series behavior of each word-level component at the quarterly frequency. In our sample, 61 of the 89 words have a positive quarterly correlation with the unemployment series (Appendix Figure A2). Online Appendix E.1 describes how we replicate our procedure using the full text of US Public Firms' sales and earnings conference calls as an alternative dataset. This produces a similar counter-cyclical pattern over a smaller time period (2004-2013). Online Appendix

[15]For the latter, we average over the topic series for: "economic growth," "Federal Reserve," "financial crisis", "recession," and "macroeconomic data."

[16]Nimark (2014) and Chahrour et al. (2021) model media focus as responding to *a priori* unlikely events and draw out implications for business cycles in environments with information frictions.

[17]These classifications 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. The industry categorization is reviewed in more detail in Online Appendix F.1, in the context of our empirical analysis in Section 5.

E.2 describes an alternative procedure for analyzing the 10-Q/Ks which uses the frequency of algorithmically determined word stems rather than full words to build our word set $\mathcal{W}_M$ and MacroAttention measure. This yields essentially identical results to our main procedure.

3 Model

We now describe our model, a Neoclassical Real Business Cycle model augmented with a stochastic choice friction that captures attention and "mistake making." We introduce the model primitives, the structure of stochastic choice, and the equilibrium concept.

3.1 Consumers and Final Goods

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

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

$$\mathcal{U}(\{C_{t+j}, L_{t+j}\}_{j\in\mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \left(\frac{C_{t+j}^{1-\gamma}}{1-\gamma} - v(L_{t+j}) \right) \tag{5}$$

where $\beta \in [0, 1)$ is the discount factor, v is an increasing and convex labor disutility, and $\gamma > 0$ is the coefficient of relative risk aversion. The household supplies labor at a wage w_t and owns equity in firms that produce intermediate goods, thus receiving profits $\{\pi_{it}\}_{i\in[0,1]}$.

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

$$X_t = X(\{x_{it}\}_{i\in[0,1]}) = \left(\int_{[0,1]} x_{it}^{\frac{\epsilon-1}{\epsilon}} \,\mathrm{d}i \right)^{\frac{\epsilon}{\epsilon-1}} \tag{6}$$

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

Wages are set by the following "wage rule":

$$w_t = \bar{w} \cdot \left(\frac{X_t}{\bar{X}} \right)^{\chi} \tag{7}$$

where $\bar{w} > 0$ and $\bar{X} > 0$ are constants, and $\chi \geq 0$ measures the extent of real wage rigidity.

Given wages, firms demand labor and households supply sufficient labor to meet demand. Describing wage dynamics via the wage rule is for technical simplicity as it allows us to study the economy via a scalar fixed-point equation that is simple to characterize.[18] In Online Appendix B, we show how to micro-found such a wage rule when households have the inter-temporal preferences of Greenwood et al. (1988) and markets clear in the standard fashion. We show in Online Appendix D that our later quantitative results of Section 6 are robust to considering wages that are set in this manner.

3.2 Intermediate Goods Firms and Productivity Shocks

Each intermediate goods firm i is a monopoly producer of its own variety and faces a demand curve $d(x_{it}, X_t) = X_t^{\frac{1}{\epsilon}} x_{it}^{-\frac{1}{\epsilon}}$ from the final goods producer. They hire a labor quantity L_{it}, pay wage w_t per worker, and produce with the following linear technology:

$$x_{it} = \theta_{it} L_{it} \tag{8}$$

where θ_{it} is a firm-level shifter of productivity, which lies in a set $\Theta \subset \mathbb{R}_+$.

We parameterize the stochastic process for firm-level productivity in the following way that captures "aggregate productivity shocks" while allowing for rich cross-firm heterogeneity. There is an aggregate productivity state $\theta_t \in \Theta$, which follows a first-order Markov process with transition density given by $h(\theta_t \mid \theta_{t-1})$. The cross-sectional productivity distribution is given in state Θ by the mapping $G : \Theta \to \Delta(\Theta)$, where we denote the productivity distribution in any state θ_t by $G_t = G(\theta_t)$ with corresponding density g_t. To give the parameter θ_t an interpretation as "aggregate productivity," we assume that the total order on θ_t ranks distributions G_t by first-order stochastic dominance, or $\theta \geq \theta'$ implies $G(\theta) \succsim_{FOSD} G(\theta')$. We proceed in the rest of our main analysis under this "Real Business Cycle" interpretation of our model, though we will emphasize when presenting results why neither the "supply-side" view of economic dynamics nor the specific structure for productivity dynamics is not essential for our main conclusions or proposed macroeconomic propagation mechanism.

3.3 Attention and Mistake Making

We define the firm-level *decision state* $z_{it} = (\theta_{it}, X_t, w_t) \in \mathcal{Z}$ as the concatenation of all decision-relevant variables that the firm takes as given.[19] All firms believe that the vector z_{it}

[18] Blanchard and Galí (2010) and Alves et al. (2020) use similar wage-rule formulations to describe real wage rigidity.

[19] As will become clear, $\mathcal{Z} = \Theta \times \mathcal{X} \times \mathcal{W}$, where $\mathcal{X}$ is feasible set of production, and $\mathcal{W}$ is the image of $\mathcal{X}$ via the wage rule (7).

follows a first-order Markov process with transition densities described by $f : \mathcal{Z} \to \Delta(\mathcal{Z})$, with $f(z_{it}|z_{i,t-1})$ being the density of z_{it} conditional on last period's state being $z_{i,t-1}$. We denote the corresponding set of possible transition densities by $\mathcal{F}$. At time t, each firm i knows the sequence of previous $\{z_{is}\}_{s<t}$ but not the contemporaneous value z_{it}.

3.3.1 The Cost of Attention

Intermediate goods firms choose a production level x_{it} in the feasible set $\mathcal{X}$. But, due to cognitive and/or organizational constraints, they struggle to match that choice to the state $z_{it} = (\theta_{it}, X_t, w_t)$ without making idiosyncratic mistakes. We model this by having them choose stochastic choice rules at a cost. This formulation is sufficiently flexible to capture models of information acquisition and mistake-making in a unified and parsimonious manner, as Flynn and Sastry (2021) describes in more detail.

Formally, each firm chooses a stochastic choice rule $p : \mathcal{Z} \to \Delta(\mathcal{X})$ in set $\mathcal{P}$, or a mapping from states of the world to *distributions of actions* described by probability density (mass) function $p(\cdot \mid z_{it})$ when the firm-level state is z_{it}. A firm using rule p commits to delivering the realized quantity $x_{it} \in \mathcal{X}$ to the market, selling it at the (maximum) price $q_{it} = d(x_{it}, X_t)$ at which the final goods firm is willing to buy, and to hiring sufficient labor in production. Helpfully, it is fully equivalent to interpret firms' choices as committing to hire $L_{it} = \frac{x_{it}}{\theta_{it}}$ workers at wage w_t, producing the maximum level $x_{it} = \theta_{it} L_{it}$, and selling at price $q_{it} = d(x_{it}, X_t)$.

We model the cost of information acquisition and/or mistake-making via a cost functional $c : \mathcal{P} \times \Lambda \times \mathcal{Z} \times \mathcal{F} \to \mathbb{R}$ which returns how costly any given stochastic choice rule is to implement in units of utility or cognitive strain. The cost can depend on a firm-specific type $\lambda_i \in \Lambda \subseteq \mathbb{R}_+$, by assumption independent from the decision state and distributed in the cross-section as $L \in \Delta(\Lambda)$, and the previous value of the decision state $z_{i,t-1}$, which under the Markov assumption summarizes the transition probabilities for z_{it}.[20] The basic idea that we wish to embody, consistent with our motivation of studying costly cognition and mistake making, is that playing actions that are more *precise* in any given state is more costly.

To make this tension more clear, and also to make the analysis more tractable, we specialize to the following cost functional, which equals the negative expected entropy of the action distribution multiplied by a scaling $\lambda_i > 0$:

$$c(p, \lambda_i, z_{i,t-1}, f) = \lambda_i \int_{\mathcal{Z}} \int_{\mathcal{X}} p(x \mid z_{it}) \log(p(x \mid z_{it}))\, \mathrm{d}x\, f(z_{it} \mid z_{i,t-1})\, \mathrm{d}z_{it} \tag{9}$$

[20]It is simplifying to assume no relationship between the cost shifter λ_i and the productivity state θ_{it}, which is within z_{it}. All of our results would go through essentially unchanged, but with more cumbersome definitions of equilibrium, in a model with productivity-correlated and/or dynamic λ_i.

This is an example of a *likelihood-separable cost functional* as discussed in more detail in Flynn and Sastry (2021). Such cost functionals capture the idea that it is costly for agents to avoid "mistakes" or misoptimizations, relative to an unrestricted (costless) optimal choice.[21] In Section 4.5.3, we discuss the robustness of our results to considering models of information acquisition.

3.3.2 The Firm's Problem

Intermediate goods firms are owned by the representative household and maximize the product of their dollar profit, which we write as $\pi(x_{it}, z_{it})$, and the household's marginal utility, which we write as $M(X_t)$. We thusly define "risk-adjusted profits" as the product of these terms:

$$\Pi(x_{it}, z_{it}) = \pi(x_{it}, z_{it}) \cdot M(X_t) \tag{10}$$

We observe that, under our assumed structure for the firms' cost and revenue structure and the household's utility function, the profit function and marginal utility are respectively

$$\pi(x_{it}, z_{it}) := x_{it} \left(x_{it}^{-\frac{1}{\epsilon}} X_t^{\frac{1}{\epsilon}} - \frac{w_t}{\theta_{it}} \right) \qquad\qquad M(X_t) = X_t^{-\gamma} \tag{11}$$

Finally, firms treat the cognitive cost as an additive "psychological cost" also in utility units.

Because decisions are separable across time, and $z_{i,t-1}$ 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 $z_{i,t-1}$. The firm has a conjecture for how aggregate output and wages move over time, as embedded in their subjective prior distribution $f(z_{it} \mid z_{i,t-1})$. Given this conjecture, they play a best reply by solving the following program:

$$\max_{p \in \mathcal{P}} \int_{\mathcal{Z}} \int_{\mathcal{X}} \Pi(x, z_{it})\, p(x \mid z_{it})\, \mathrm{d}x \; f(z_{it} \mid z_{i,t-1})\, \mathrm{d}z - c\left(p, \lambda_i, z_{i,t-1}, f\right) \tag{12}$$

which is maximization of expected utility, averaged over risk in the state z_i and stochastic action x, net of the cost of the chosen stochastic choice rule.

[21] Indeed, as formalized by Fudenberg et al. (2015), a formulation with likelihood-separable stochastic choice is often equivalent to an additive random utility model. The formulation (9) is exactly isomorphic to a "logit demand model," or additive random utility model with Gumbel distributed perturbations, and embodies the familiar associated axioms including independence of irrelevant alternatives (IIA).

3.4 Linear-Quadratic Approximation and Equilibrium

To tractably study equilibrium, we simplify the intermediate goods firms' objective and the final goods firm's production with quadratic approximations. Both approximations are derived in Online Appendix A.7.

Toward simplifying the intermediate goods firms' objective, we first define an intermediate firm's *ex post* optimal production level

$$x^*(z_i) := \arg\max_{x\in\mathcal{X}} \ \Pi(x, z_i) \tag{13}$$

We next define $\bar{\Pi}(z_i)$ as risk-adjusted profits evaluated at $x^*(z_i)$ and $\Pi_{xx}(z_i)$ as the function's second derivative in x evaluated at the same point. The latter measures the *state-dependent cost* of misoptimizations relative to $x^*(z_i)$ and will be central to our analysis. The objective of the intermediate goods firm is, to the second order:

$$\tilde{\Pi}(x, z_i) := \bar{\Pi}(z_i) + \frac{1}{2}\Pi_{xx}(z_i)(x - x^*(z_i))^2 \tag{14}$$

So that this is globally defined, we will also apply the simplifying assumption that $\mathcal{X} = \mathbb{R}$.

Next, we approximate the final goods firm's production function (6) around the *ex post* optimal production levels $x^*(z_i)$ to the second order. We first define the aggregate of the *ex post* optimal production levels as

$$X^* = \left(\int_0^1 x^*(z_i)^{1-\frac{1}{\epsilon}}\,\mathrm{d}i\right)^{\frac{\epsilon}{\epsilon-1}} \tag{15}$$

We then write the approximate production function as

$$X = X^* - \frac{1}{2\epsilon}\int_0^1 \frac{(x_i - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}}(x^*(z_i))^{1+\frac{1}{\epsilon}}}\,\mathrm{d}i \tag{16}$$

Importantly, because the aggregator is concave, it is systematically lower when the dispersion of these misoptimizations is higher.

We now define equilibrium, up to these approximations, in terms of the stochastic choices of intermediate goods firms p, and the transition density f:[22]

Definition 1 (Equilibrium)**.** *An equilibrium is a stochastic choice rule $p \in \mathcal{P}$ and a transition density $f \in \mathcal{F}$ such that:*

[22] Point 4 embeds our notion of rational expectations equilibrium (REE) by requiring that firms' subjective prior about endogenous variables is "correct."

1. *Intermediate goods firms' stochastic choice rules p solve program (12) given f, with $\tilde{\Pi}$, defined in (14), replacing Π.*
2. *The transition density f is consistent with p in the sense that:*[23] *(i) the marginal distribution of firm-level productivity is given by G; (ii) aggregate output is given by the aggregator (16) evaluated in the cross-sectional distribution of production implied by p and G; and (iii) the wage is derived from the wage rule (7) evaluated in aggregate output.*

4 Theoretical Results

We now present our main theoretical results, in four parts. First, we characterize firms' attention and misoptimization in partial equilibrium. These results show how firms' attention choice is shaped by two critical mechanisms, a *profit-curvature channel* related to the dollar cost of misoptimization and a *risk-pricing channel* related to the utility translation or risk adjustment of these costs. Second, we derive two conditions—a restriction to aggregate strategic complementarity and a lower bound on risk aversion—under which attention is counter-cyclical (and misoptimization pro-cyclical) in general equilibrium. We argue these conditions are *ex ante* reasonable based on existing macro-financial evidence. Third, we characterize equilibrium output in the economy as the product of the full-attention, frictionless benchmark and an attention wedge that is smaller than one and decreasing in the underlying productivity of the economy. This attention wedge reflects the cyclical nature of misallocation in our economy: when productivity is lower, firms optimize more precisely and productivity is closer to the frictionless benchmark. Finally, we show how endogenous variation in misallocation drives asymmetric and state-dependent shock propagation, and endogenous stochastic volatility of output growth.

4.1 Attention and Misoptimization in Partial Equilibrium

We begin by describing the stochastic choice behavior of firms. We observe that, under the entropic cost, firms' costly control problem is *linearly separable* across state realizations z_{it}. This implies that firms' optimal policies will be independent of the prior distribution f. In this way, consistent with our motivation, our model isolates firms' difficulties in making

[23] Formally, this first requirement means that:

$$g_t(\theta_{it}) = \mathrm{marg}_{\theta_{it}\in\Theta} \int_{\Theta} f(\theta_{it}, X_t, w_t|\tilde{\theta}, X_{t-1}, w_{t-1}) g_{t-1}(\tilde{\theta}) d\tilde{\theta} \tag{17}$$

while the second means that (16) is evaluated with respect to the density of production $p(\cdot, X_t, w_t)g_t(\cdot)$. In particular, this implies that the marginal of f for X_t and w_t are Dirac distributions on their true values.

state-contingent plans. The following result leverages this observation and describes the solution to the firm's problem and allows us to characterize its comparative statics:

Proposition 1 (Firms' Optimal Stochastic Choice Rules)**.** *The production of a type-λ_i firm, conditional on realized state $z_i = (\theta_i, X, w)$, can be written as*

$$x_i = x^*(z_i) + \sqrt{\frac{\lambda_i}{|\pi_{xx}(z_i)|\, M(X)}} \cdot v_i$$

where $x^*(z_i)$ is the unconstrained optimal action, $M(X)$ is the household's marginal utility, $|\pi_{xx}(z_i)|$ is the magnitude of curvature for the firms' profit function, and v_i is an idiosyncratic, standard normal random variable.

Proof. See Appendix A.1. □

Economically, Proposition 1 says that firms center their action around the full-attention optimum $x^*(z_i)$ but, due to costly control, make an idiosyncratic misoptimization. The variance of the misoptimization increases if the marginal cost of precision increases (higher λ_i), and decreases if either of two components of the marginal benefits of precision increases. The first component of this marginal benefit is the state-dependent curvature of the firms' dollar profit function, $|\pi_{xx}(z_i)|$, which translates small misoptimizations into their dollar cost near the optimal production level. How this $|\pi_{xx}|$ moves as a function of the aggregate business cycle hinges on the specific assumed structure of firms' demand curves and cost structure. The second component of this marginal benefit is the household's marginal utility, $M(X)$, which translates dollar losses into utility losses, which can be directly compared to the utility cost of cognition. When households are risk-averse, this marginal utility is a decreasing function of X: the representative household is "hungrier" for a given firm's dollar profits, in utility terms, when the aggregate economy is doing poorly, and therefore less tolerant of misoptimizations.

The previous argument used only the structure of the cost functional and the assumption of a "Neoclassical firm" owned by a representative household. We can use the specific assumed structure of our model to re-state the comparative statics in the discussion above in terms of firms' productivity θ_i and aggregate output X, after substituting in wages as a function of output via the wage rule. In particular, the curvature terms of interest can be written in the following way:[24]

$$|\pi_{xx}(z_i)| := v_\pi(\epsilon, \chi, \bar{X}, \bar{w}) \cdot \theta_i^{-1-\epsilon} X^{\chi(1+\epsilon)-1} \qquad M(X) = X^{-\gamma} \tag{18}$$

[24] The first expression, and the associated constant, are derived in Appendix A.7.

where $v_\pi(\epsilon, \chi, \bar{X}, \bar{w}) > 0$. We observe also that the variance of production conditional on the realized decision state, or $\mathbb{E}[(x_i - x^*(z_i))^2 \mid z_i]$, is a summary statistic for both misoptimization and "attention" measured by realized cognitive costs, which are decreasing in this variance. We summarize the comparative statics of this conditional variance, and by extension of attention and misoptimization, under the assumed payoff structure in the following result:

Corollary 1 (Comparative Statics for Mistakes)**.** *Consider a type-λ_i firm in state $z_i = (\theta_i, X, w)$. The extent of misoptimization, $\mathbb{E}[(x_i - x^*(z_i))^2 \mid z_i]$, increases in θ_i. Moreover, $\mathbb{E}[(x_i - x^*(z_i))^2 \mid z_i]$ strictly increases in X if and only if $\gamma > \chi(1+\epsilon) - 1$.*

Proof. Immediate from combining Proposition 1 with Equation 18. □

The (absolute) curvature of the profit function always increases in marginal costs, and therefore decreases in θ_i, in our model. The monotonicity of curvature in aggregate output depends jointly on the cyclicality of wages, which contributes a term with exponent $\chi(1+\epsilon)$; the aggregate demand externality, which contributes a term with exponent -1; and marginal utility, which contributes an exponent $-\gamma$. In particular, an economy with sufficiently cyclical wages would have misoptimization decrease in aggregate output due to the profit-curvature channel, while an economy with sufficient risk aversion or sufficiently cyclical marginal utility would have misoptimization increase in aggregate output due to the risk-pricing channel. We will discuss the interpretation of this parameter condition "horse race" in the context of our general-equilibrium result in the next subsection.

4.2 Attention and Misoptimization Cycles in Equilibrium

We now translate the partial equilibrium behavior of the firm into general equilibrium and describe the model's aggregate predictions. We first characterize conditions under which equilibrium analysis is well-posed and output is a uniquely determined, monotone function of the underlying productivity state:

Proposition 2 (Existence, Uniqueness, and Monotonicity)**.** *For any $\chi > 0$, an equilibrium in the sense of Definition 1 exists. If $\chi\epsilon < 1$ and $\gamma > \chi + 1$, there is a unique equilibrium. In this equilibrium, output can be expressed via a function $X : \Theta \to \mathbb{R}$ that is strictly positive and strictly increasing.*

Proof. See Appendix A.2. □

To establish these properties, we derive a representation of equilibrium as a fixed point for aggregate output X as a function of the aggregate state θ. To establish uniqueness and monotonicity, we derive conditions under which the fixed-point equation is a contraction map that depends positively on productivity. The condition $\chi\epsilon < 1$ ensures that firms' production plans are on average an increasing function of aggregate output, by bounding wage pressure relative to the aggregate demand externality. The condition $\gamma > \chi+1$ bounds the variance of actions around this optimum and ensures that, even in the presence of endogenous dispersion, there is positive but bounded complementarity.

The latter condition $\gamma > \chi + 1$ is both conservative in the model, as it ensures uniqueness and monotonicity for any possible distribution of λ_i, and highly plausible in practice. The elasticity of detrended real wages to GDP in (detrended) US data since 1987 is 0.095, and micro-level studies find similarly severe wage rigidity (Solon et al., 1994; Grigsby et al., 2021).[25] Moreover, to be consistent with the restrictions $\chi\epsilon < 1$ (upward-sloping best responses) and $\epsilon > 1$ (substitutable goods), χ cannot exceed one. The corresponding conditions $\gamma > 1.097$ or $\gamma > 2$ are likely both slack by one or two orders of magnitude, given abundant evidence in financial economics about the high cyclicality of the stochastic discount factor (e.g., Hansen and Jagannathan, 1991).

We now demonstrate conditions under which this economy exhibits aggregate attention and misoptimization cycles. We say that firms "pay more attention" in a state if, averaging over idiosyncratic states (θ_i, λ_i), they pay a greater attention cost conditional on that state being realized. We say that firms "misoptimize more" in a state if, again averaging over idiosyncratic states (θ_i, λ_i), they have a lower expected mean-squared error around the *ex post* optimal action $x^*(z_i)$.[26] We now show that, under the stated assumptions for tractable equilibrium analysis, the model features counter-cyclical attention and pro-cyclical misoptimization:

Proposition 3 (Attention and Misoptimization Cycles)**.** *Assume* $\chi\epsilon < 1$ *and* $\gamma > \chi + 1$*. Intermediate goods firms pay more attention and misoptimize less in lower-productivity, lower-output states.*

Proof. See Appendix A.3. □

The proof of this result verifies that the stated conditions are sufficient for output X to be monotone increasing in θ, via Proposition 2; observes that the same conditions are sufficient for the idiosyncratic variance of choices to be monotone increasing in θ_i and X, via Corollary 1; and translates these into predictions for cross-firm averages.

Economically, Proposition 3 says that a calibration of the model that is consistent with existing evidence about wage rigidity and the stochastic discount factor predicts that firms

[25] This calculation uses quarterly-frequency, seasonally-adjusted data on real GDP and median, CPI-adjusted wages of all full-time employed wage and salary workers. Both series are linearly detrended.

[26] Mathematical definitions of both notions are included in the proof of Proposition 3 in Appendix A.3.

should pay more attention to their decisions and make smaller misoptimizations, conditional on stochastic fundamentals, in downturns. The model's prediction about attention is qualitatively consistent with our motivating evidence on macroeconomic attention in language, with the heavy caveat that the language-based measurement has no direct analogue in the model. The prediction about misoptimizations is a testable prediction on firm choices, which we will map to the data in Section 5.

The mechanism for these results hinges on the risk-pricing channel described in partial-equilibrium by Proposition 1—in words, firms exert additional effort to avoid profit-reducing errors in downturns because of the market's heightened aversion to low profits in these bad states. This idea is qualitatively consistent with received wisdom in business management that managing performance in downturns is especially critical for firm valuation.[27] We will formulate a test of the model mechanism along these lines also in Section 5.

4.3 Misoptimization, Output, and Productivity

Having established model conditions that generate attention and misoptimization cycles, we now study the effects of these phenomena on output and production. The proofs of our earlier results demonstrate that there are scalar sufficient statistics in equilibrium for both firm-specific productivity and firm-specific costs of attention. The cross-sectional distribution of productivity is summarized by $\hat{\theta}(G) = \left(\mathbb{E}_G[\theta_i^{\epsilon-1}]\right)^{\frac{1}{\epsilon-1}}$. We therefore, without loss of generality, write $\theta_t = \hat{\theta}(G_t)$ for the remainder of the analysis. The cross-sectional distribution of attention costs is summarized by $\lambda := \mathbb{E}[\lambda_i]$.

Let us define $\log X(\log\theta)$ as a mapping from the log state to log output in the economy, holding fixed all other parameters. The following result describes output in log units as the sum of an "RBC core" factor and an *attention wedge* $\log W(\log\theta)$:

Proposition 4 (Equilibrium Output Characterization)**.** *Equilibrium output follows:*

$$\log X(\log\theta) = X_0 + \chi^{-1}\log\theta + \log W(\log\theta) \tag{19}$$

where X_0 is a constant and $\log W(\log\theta) \leq 0$, with equality if and only if $\lambda = 0$. When $\chi\epsilon < 1$, $\gamma > \chi + 1$, and $\lambda > 0$, the wedge has the following properties:

1. $\frac{\partial \log W}{\partial \lambda} < 0$, *or the wedge widens with the average cost of attention.*
2. $\frac{\partial \log W}{\partial \log\theta} < 0$, *or the wedge widens as the state increases.*

[27] This perspective is well illustrated by a report, "Advantage in Adversity: Winning the Next Downturn," prepared by Boston Consulting Group (Reeves et al., 2019a) and a related summary in the Harvard Business Review (Reeves et al., 2019b).

Proof. See Appendix A.4. □

Absent inattention, output is log-linear in aggregate productivity. With inattention and under our stated conditions from Propositions 2 and 3, output is depressed by the presence of stochastic choice in a way that increases in magnitude in the extent of inattention and, for a fixed level of inattention, increases in magnitude in the state. Both results have a partial-equilibrium and general-equilibrium component. In partial equilibrium, both increasing λ and increasing the productivity state θ make firms play more dispersed actions, as shown in Propositions 1 and 3, and this dispersion has a cost to output when the aggregate production function is concave. In general-equilibrium, we iterate this logic until convergence. Our comparative statics results verify that this fixed-point operation converges on a lower value of output.

To understand better the role of the wedge, we can re-cast the wedge in terms of labor productivity or "total factor productivity" in our one-factor economy:

Corollary 2 (The Productivity Wedge)**.** *Let $A := X/L$ be the measured productivity of our economy. Productivity can be written as*

$$\log A(\log\theta) = \log\theta + \chi\epsilon \log W(\log\theta) \tag{20}$$

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

Proof. See Appendix A.5. □

The productivity wedge representation allows for three useful parallels between our paper's mechanism and classic arguments in the macroeconomics literature. The first relates to a literature on the cleansing effect of recessions following Caballero and Hammour (1994). Our mechanism is like an attentional, *intensive margin* version of the same effect: conditional on a given firm operating, it is more focused on making precise and accurate choices in recessions, and this on average reduces the wedge and raises aggregate labor productivity.

The second relates to an empirical literature studying the the dynamics of labor productivity in the United States, and in particular highlighting its unstable and often negative cyclicality (e.g., Galí and Gambetti, 2009; Barnichon, 2010; Galí and Van Rens, 2021). Our model accommodates a non-monotone relationship between aggregate labor productivity and aggregate output, due to the competing forces of increased productivity with increased misallocation.

The third relates to the literature on the aggregate effects of resource misallocation across firms, pioneered by Hsieh and Klenow (2009) and Restuccia and Rogerson (2008), and related to informational and cognitive frictions by David et al. (2016), David and Venkateswaran

(2019), and Ma et al. (2020). The mechanism whereby dispersion in firm-level value marginal products depresses aggregate productivity is shared with these analyses. What is new is our prediction about the cyclicality of this force and the implications of that cyclicality for business cycles.

4.4 Shock Propagation and Volatility

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

Corollary 3 (Endogenous Stochastic Volatility, State-Dependent Propagation, and Asymmetric Propagation)**.** *Suppose that aggregate productivity follows the process* $\log \theta_t = \rho_\theta \log \theta_{t-1} + (1-\rho_\theta) \log \bar{\theta} + \nu_t$ *where* $Var[\nu_t] = \sigma_\theta^2$*. The model generates the following properties:*

1. *Endogenous stochastic volatility: to a first-order approximation in* ν_t*, the variance of output conditional on last period's productivity* θ_{t-1} *is given by*

$$\mathbb{V}[\log X_t | \theta_{t-1}] = \left(\chi^{-1} + \frac{\partial \log W(\log \theta)}{\partial \log \theta}\bigg|_{\theta=\theta_{t-1}} \right)^2 \sigma_\theta^2 \tag{21}$$

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

$$\frac{\partial \log X(\log \theta)}{\partial \log \theta}\bigg|_{\theta=\theta_{t-1}} = \chi^{-1} + \frac{\partial \log W(\log \theta)}{\partial \log \theta}\bigg|_{\theta=\theta_{t-1}} \tag{22}$$

3. *Asymmetric shock propagation: the impact of a shock to second-order in* ν_t *in state* θ_{t-1} *is given by*

$$\frac{\partial \log X(\log \theta)}{\partial \log \theta}\bigg|_{\theta=\theta_{t-1}} = \chi^{-1} + \frac{\partial \log W(\log \theta)}{\partial \log \theta}\bigg|_{\theta=\theta_{t-1}} + \left(\frac{\partial^2 \log W(\log \theta)}{\partial \log \theta^2}\bigg|_{\theta=\theta_{t-1}} \right) \nu_t \tag{23}$$

where the sign of the wedge's second derivative determines the direction of asymmetry.

Proof. See Appendix A.6. □

To understand this result, note that the sensitivity of the economy to shocks is simply the sum of the frictionless economy's response to shocks, which is always χ^{-1}, and the response of the attention wedge to shocks. This latter response is always negative in the case studied by Proposition 4, so the economy is mechanically less responsive to shocks than the full attention benchmark. This dampened response is a familiar prediction in the literature with cognitively constrained agents (e.g. Sims, 1998, 2003; Gabaix, 2014), drawn out here in a general equilibrium context. Novel to our analysis is the fact that the response of the attention wedge to shocks depends on the state, or the fact that the attention wedge need not be linear in $\log \theta$. This is a direct consequence of modeling state-dependent attention and accommodating attention cycles. This fact may generate the asymmetry, state-dependence, and stochastic volatility indicated in Corollary 3.

The ultimate macroeconomic implications of this result hinge on the concavity or convexity of the attention wedge in the state. When the attention wedge is concave (respectively, convex), the economy generates greater (smaller) volatility in low states, a larger (smaller) impact of shocks in low states, and features larger (smaller) impact of negative than positive shocks from any initial state. Owing to the intuition that when θ is very small, it is as-if the economy is operating in its "full pass-through" RBC core, the natural case appears to be a concave wedge. While we cannot establish theoretically that the wedge is globally concave, it cannot be globally convex and therefore must be concave in at least some part of the parameter space. Our quantitative analysis in Section 6 will feature a globally concave wedge. In that section we will review the implications of this finding.

4.5 Discussion and Extensions

In the next parts of the paper, we will directly test the micro and macro predictions of the model (Section 5) and numerically examine the implications of a model that matches the empirical facts (Section 6). Before proceeding, we briefly summarize some extensions of our main analysis.

4.5.1 Wage-Rule Shocks

Our main analysis considered supply shocks operating through the shifts in the productivity distribution. Consider instead an economy with fixed productivities but stochastic $\bar{w}$, the level term in the wage rule. Shocks to $\bar{w}$ can be interpreted as a *labor wedge shock* or *demand shock* with the appropriate micro-foundation. Inspection of the formulas in Proposition 1, reveals that $\bar{w}$ is tantamount to an aggregate shifter of firms' revenue productivity. Propositions 2 and 3 provide conditions for output to be monotone *decreasing* in $\bar{w}$ and for counter-cyclical attention (pro-cyclical misoptimization) in a labor-wedge driven economy: when the labor wedge is high, and the economy is in a recession, attention is high. Proposition 4 and Corollary 3 hold as written, with $\bar{w}^{-1}$ replacing θ.

4.5.2 Multiple Inputs, Classical Labor Supply, Asset Pricing

In Online Appendix B, we extend the model to allow for intermediate inputs, separate capital owners and laborers, and market-clearing wages rather than a wage rule. The first two features will be useful in mapping the model to the data in Section 5, while the third enables a more Neoclassical micro-foundation. We show under general conditions how the main results derived in this section, regarding the cyclicality of attention of misoptimization and the effects on output, continue to hold so long as the extent of the cognitive friction is not too large. Together, these extensions demonstrate the stability of our main model insights to a richer macroeconomic environment.

4.5.3 Other Cost Functionals and Information Acquisition

In Flynn and Sastry (2021), we establish more general analogs of Propositions 2 and 3 in abstract games rather than business cycle models and allow for a more general class of *likelihood-separable* costs of stochastic choice. The core of the argument remains the monotonicity of "stakes" in both the state and output. We clarify how a restriction of cost functionals to a "quasi-MLRP" class, defined in that paper, preserves natural notions of monotone action distributions and action precision as a function of average incentives.[28]

In Online Appendix C.1, we show how our results translate in a slight variant of our model that accommodates cognitive inertia and persistence of misoptimizations. This extension is useful in matching the patterns we uncover in the data in the next section.

In the same Online Appendix, we also study the robustness of our results to more canonical information acquisition by examining our model with Gaussian signal extraction with costly precision (C.2) and mutual-information costs as studied by Sims (2003) (C.3). In both cases we provide conditions under which a greater cost of making mistakes, for instance due to the risk-pricing mechanism we highlight, leads to firms making smaller mistakes. We therefore argue that the conclusions of this section are robust to considering canonical information acquisition. We do, however, note that application of unrestricted rational inattention to

[28]Our analysis in that paper also clarifies the exact kind of supermodularity (macroeconomic complementarity) and discounting (concavity of the aggregate production function) required to achieve monotone precision in a game's unique equilibrium. In fact, a slightly generalized version of Propositions 2 and 3 that dispenses with the approximated CES aggregator could be proven exactly using the main results of Flynn and Sastry (2021).

the model we study is not analytically possible using any known techniques as our firms' have non-Gaussian priors and non-quadratic payoff functions, and aggregation is non-linear.

5 Testing the Model: The Misoptimization Cycle

In this section, we describe a method to test the model's main predictions using structurally estimated "misoptimizations" by firms. We develop empirical tests that verify the two major predictions of model's main case: that firms on average make smaller misoptimizations in downturns, and that firms face sharper market incentives to do so in the same states. We then show how higher levels of macroeconomic attention, as measured in our motivating analysis of Section 2, predict lower misoptimizations at the firm level.

5.1 Measuring Misoptimizations

5.1.1 Data

Our dataset for public firms' production and input choices is Compustat Annual Fundamentals. Compustat compiles information from firms' financial statements on sales, employment, variable input expenses, and capital measured via net and gross values of plants, property and equipment (PPE). It also provides a historically consistent industry classification, based on firms' main operational NAICS codes. We organize firms into 44 industries, which are defined at the NAICS 2-digit level but for Manufacturing (31-33) and Information (51), which we split into the 3-digit level.[29]

We use a sample period from 1986 to 2018, to focus on the modern macroeconomic regime after the Volcker disinflation. We view this relative stability in the macroeconomic environment as particularly important for measuring externally valid patterns in macroeconomic attention and its relationship with firm-level decisionmaking. We apply standard filters to remove firms that are based outside the US, are insufficiently large, are likely to have been involved in a merger or acquisition, or do not report one or more important pieces of data. These filters yield a final sample of 68,198 firm-year observations, or about 2,200 per year. Online Appendix F.1 describes the procedure in detail.

5.1.2 From Theory to Estimation

Our main goal is to derive an empirical proxy for "misoptimizations" in production or input-choice relative to an *ex post* optimal level. Our overall strategy is to measure residuals in

[29]We drop financial firms and utilities due to their markedly different production functions and profit structures.

firms' choices relative to empirically estimated production and policy functions.

We now describe the details. We assume that each firm i at time t, within sector $j(i)$, operates a sector-specific, constant-returns-to-scale, Cobb-Douglas production technology over labor, materials, and capital, with total factor productivity θ_{it}:

$$\log x_{it} = \log\theta_{it} + \alpha_{L,j(i)}\log L_{it} + \alpha_{M,j(i)}\log M_{it} + \alpha_{K,j(i)}\log K_{it} \tag{24}$$

with the restriction, in each sector, that $\alpha_{L,j(i)} + \alpha_{M,j(i)} + \alpha_{K,j(i)} = 1$.[30] This is a natural generalization of our studied model from Section 3, with a bundle of inputs replacing the single theoretical factor of production and heterogeneity in technology across industries. Next, we assume that input choice follows a policy function that is log-linear in individual firm characteristics, industry-by-time trends, individual productivity, and a residual m_{it}:

$$\log L_{it} = \eta_i + \chi_{j(i),t} + \beta\log\theta_{it} + m_{it} \tag{25}$$

This is a generalization of the optimal policy in our model from Section 3, with input costs and demand varying at the industry-by-time and firm level. Online Appendix F.3 derives Equation 25 exactly in such an extended model, provided that firms can cost-minimize over input bundles. In this mapping to the model, we can interpret the residual as the firm's percentage misoptimization from the counterfactual *ex post* optimal level L^*_{it}, or $m_{it} \approx \frac{L_{it}-L^*_{it}}{L^*_{it}}$. As observed in Online Appendix F.3, and readily apparent in our main theoretical model, these are equivalent to percentage misoptimizations in production, or $\frac{L_{it}-L^*_{it}}{L^*_{it}} = \frac{x_{it}-x^*_{it}}{x^*_{it}}$.

In the data, for a myriad reasons unmodeled in our main analysis, these residuals are likely to be persistent. To capture this, we assume that the m_{it} are determined by an AR(1) process with persistence $\rho \in (0,1)$ and scaled innovations u_{it}, or

$$m_{it} = \rho m_{i,t-1} + \left(\sqrt{1-\rho^2}\right)u_{it} \tag{26}$$

The innovations u_{it} are mean zero and have a variance $\mathbb{E}[u^2_{it} \mid i, Z_t]$ which varies a function of fixed individual characteristics and all aggregate state variables Z_t. We provide a formal micro-foundation for this AR(1) structure of persistent mistakes in Online Appendix C.1. As long as σ^2_{it} is sufficiently persistent, $\mathbb{E}[m^2_{it} \mid Z_t] \approx \mathbb{E}[u^2_{it} \mid Z_t]$, where both expectations average over fixed individual characteristics indexed by i. We will use this approximation in practice, as the one-step-ahead variance of Equation 26 is easier to measure with a relatively

[30] The Cobb-Douglas assumption is a convenient and common step to enable production function estimation via input-cost shares (e.g., Foster et al., 2001, 2008; Bloom et al., 2018). Moreover, a number of studies including Basu and Fernald (1997), Foster et al. (2008), and Flynn et al. (2019) argue that constant returns to scale in physical terms is a reasonable approximation for large, US-based firms.

short sample than the stationary variance.

5.1.3 Estimation Procedure

To empirically estimate our model, we proceed as follows. First, we estimate productivity as the residual of the production function, Equation 24, after estimating input elasticities using cost shares in each sector. We use the standard input definitions of Keller and Yeaple (2009) and provide more details about our procedure in Online Appendix F.[31]

Given estimates of productivity, we estimate the system of Equations 25 and 26. We first estimate Equation 25 via ordinary least squares (OLS) and obtain a preliminary estimate $\hat{m}^0_{it}$ of the residual. We next estimate Equation 26 using $\hat{m}^0_{it}$ to obtain an estimate $\hat{\rho}$ of the residual persistence (in our main procedure, 0.70). We finally estimate via OLS the "quasi-differenced" equation for labor choice:

$$\log L_{it} - \hat{\rho} \log L_{i,t-1} = \eta_i + \chi_{j(i),t} + \beta_0 \log \hat{\theta}_{it} + \beta_1 \log \hat{\theta}_{i,t-1} + (m_{it} - \hat{\rho} m_{i,t-1}) \tag{27}$$

where the residual $(m_{it} - \rho m_{i,t-1})$, under correct specification of ρ, is $\left(\sqrt{1-\rho^2}\right) u_{it}$. We estimate Equation 27 via OLS, rather than imposing a model-derived form for (β_0, β_1), to avoid our estimates of the residual reflecting mechanical misspecification of cost and demand curves.

We finally use our estimates of Equation 27, and point estimate of ρ, to generate estimates of $\hat{u}_{it}$.[32] Our measure of aggregate "Misoptimization Dispersion" is an estimate of the variance $\mathbb{E}[m^2_{it} \mid Z_t]$ with weights s^*_{it} proportional to firms' *predicted* sales based on fundamentals:[33]

$$\text{MisoptimizationDispersion}_t = \frac{\sum_{i \in \mathcal{I}_t} s^*_{it} \cdot \hat{u}^2_{it}}{\sum_{i \in \mathcal{I}_t} s^*_{it}} \tag{28}$$

The weights are the appropriate ones for mapping average misoptimization to misallocation

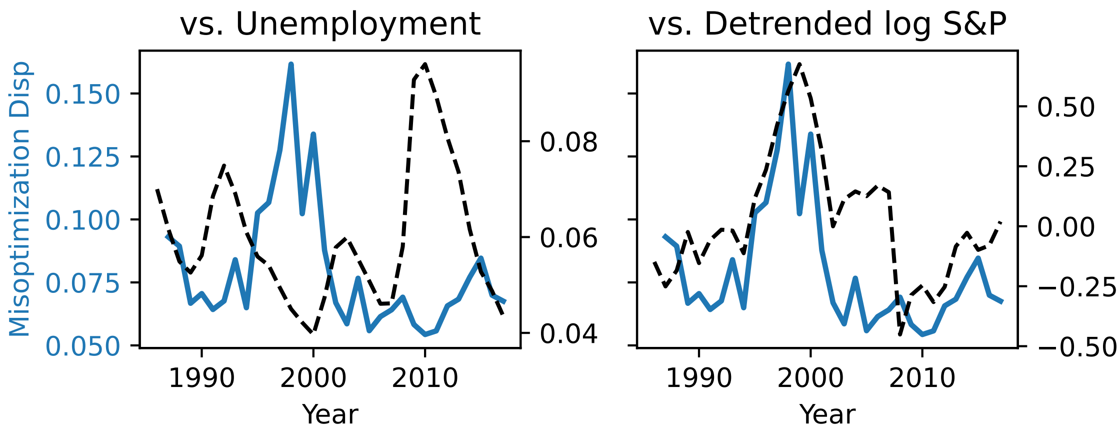
[31] We treat labor expenditures as the product of reported employees and a sector-specific wage calculated from the US County Business Patterns; materials expenditures as the sum of variable costs and administrative expenses (COGS + XSGA) net of depreciation and labor expenditures; and the capital stock as the initial gross level of plant, property, and equipment plus net investment. Instead of imputing rental rates for capital, we impose constant returns to scale and a fixed profit share of 0.75 (e.g., in the model, $\epsilon = 4$).

[32] Appendix Table A7 contains our estimates of Equations 26 and 27 under the baseline procedure outlined in this section, along with several alternative choices used in robustness checks. In all estimations, we drop the top and bottom 1% tails of the TFP distribution to limit the effects of outliers. Results are quantitatively highly similar without this trimming.

[33] In particular, the weights are the exponentiated fitted values $\exp\left(\hat{\beta} \log \theta_{it}\right)$ from the regression:

$$\log \text{Sales}_{it} = \beta \log \theta_{it} + \eta_i + \chi_{j(i),t} + \epsilon_{it}$$

Figure 2: Misoptimization Dispersion is Pro-Cyclical



Notes: The blue line, measured by the left axis of each plot, is Misoptimization Dispersion as defined by Equation 28. The black dashed lines, measured by the right axis of each plot, are respectively the US unemployment rate and linearly detrended S&P 500 price.

in the theory, and will aid in our subsequent structural interpretation of our findings. In a nutshell, these weights give higher influence to larger, more productive firms while not "double-counting" misoptimizations in both input choice and total production. In Appendix A.7.3, we show that pro-cyclical MisoptimizationDispersion by our empirical definition is sufficient for pro-cyclical misoptimization as defined in Proposition 3, and thus constitutes a particularly demanding test of that model prediction.

5.2 Misoptimizations Rise in Booms and Fall in Downturns

Figure 2 plots aggregate Misoptimization Dispersion, measured at the annual frequency, against the annual average unemployment rate and (detrended) end-of-year S&P 500 price. Misoptimization Dispersion rises when the real economy and financial markets are doing well (e.g., the late 1990s), falls during recessionary or financial crisis periods (e.g., 1990, 2001, and 2008), and is approximately as persistent as the overall business cycle.

Mirroring our earlier analysis of Macroeconomic Attention (Section 2), we benchmark the slope of this relationship by estimating a linear regression of Misoptimization Dispersion on each macroeconomic variable. Specifically, we estimate

$$\text{MisoptimizationDispersion}_t = \alpha + \beta_Z \cdot Z_t + \epsilon_t \tag{29}$$

for $Z_t \in \{\text{Unemployment}_t/100, \log \text{SPDetrend}_t\}$, over our 31 annual observations. We estimate a slope of -0.841 (SE: 0.341, $p = 0.02$) with respect to unemployment and 0.064 (SE:

0.017, $p = 0.001$) with respect to the detrended S&P 500.[34] The regression on unemployment implies that a five percentage point point swing in unemployment is associated with an increase of Misoptimization Dispersion by 0.042 log points, or 53% of its sample mean value. Appendix Figure A5 shows that the same counter-cyclical pattern is apparent in other measures of dispersion, the (weighted) mean of $|\hat{u}_{it}|$ and inter-quartile range of $\hat{u}_{it}$.

These findings are consistent with the model prediction for attention and misoptimization in Proposition 3, and by implication with the follow-up results characterizing characterizing how pro-cyclical misallocation affects macroeconomic dynamics (Proposition 4, Corollary 2, and Corollary 3).

5.2.1 Robustness to Incorporating Other Frictions

We have interpreted our results through the lens of our model of imperfect attention. But time-varying misspecification of our model for production and policy functions could also generate the observed pattern, while demanding a different economic interpretation. To guard against this possibility, we first augment our empirical specification to control directly for two leading, but unmodeled, candidates for firm-level deviations from first-order conditions: adjustment costs and financial frictions.

If firms pay significant physical costs to adjust ostensibly "variable" inputs like labor (e.g., as in Hopenhayn and Rogerson, 1993), then "misoptimizations" by our measure may pick up frictional adjustment. To capture such adjustment costs in reduced form, we add the first lag of labor choice to the policy function (25):

$$\log L_{it} = \eta_i + \chi_{j(i),t} + \beta \log \theta_{it} + \tau \log L_{i,t-1} + m_{it} \tag{30}$$

If Equation 30 were the true model of input choice, and we had instead estimated our main model Equation 25, then variation in $\tau \log L_{i,t-1}$ not spanned by current productivity (or the absorbed effects) would be part of the misoptimization m_{it}. Some of this variation would be truly spanned by previous productivity shocks and be mis-represented as a cognitive error, while the remainder of the variation would be a "bad control" that is spanned by persistent lagged cognitive errors. With these caveats in mind, we recalculate Misoptimization Dispersion using the same methods described above starting with the policy function Equation 30.[35] The regression coefficient of this version of Misoptimization Dispersion on unemployment is -0.439 (SE: 0.196), statistically significant at the 5% level but attenuated relative to

[34] The standard errors are HAC robust, using a Bartlett kernel with a width of three years.

[35] The second column of Table A7 shows the estimated quasi-differenced policy function. Adding a direct control for the previous year's input choices essentially removes all autocorrelation in the misoptimization itself, consistent with the "bad control" hypothesis described in the text.

our baseline estimate (-0.841). Thus, a conservative estimate of Misoptimization Dispersion, purged from autocorrelation due to either adjustment costs *or* persistent cognitive frictions, displays the same cyclical trend as our baseline measure.

We next consider financial frictions, broadly defined as state-dependent constraints or costs of financing input purchases. To capture financial frictions, we add a direct control for leverage (total debt over total assets), as constructed by Ottonello and Winberry (2020), and its interaction with TFP, in the policy function (25):[36]

$$\log L_{it} = \eta_i + \chi_{j(i),t} + \beta \log \theta_{it} + \tau \text{Lev}_{it} + \phi \cdot (\text{Lev}_{it} \times \log \theta_{it}) + m_{it} \tag{31}$$

This specification allows for more levered firms to face, effectively, a "TFP adjustment" (direct effect) and have a different responsiveness to TFP shocks (interactive effect). If Equation 31 were the true model of input choice, and we had instead estimated the main model Equation 25, then both the direct effect and changing responsiveness to TFP (if not spanned by fixed effects) would be erroneously attributed to the cognitive friction. When Misoptimization Dispersion is recalculated with the leverage-augmented policy function, its regression coefficient on unemployment is -0.841 (SE: 0.341), almost identical to that of our main model. This suggests, consistent also with the standard narrative that financial frictions create wedges primarily in recessions, that time-varying financial frictions do not explain our results.

5.2.2 Robustness to Alternative Measurement Strategies

We next consider the following set of alternative econometric strategies in estimating the policy and production functions: (i) estimating sector-specific policy functions, to combat against different market conditions and/or measurement error in TFP; (ii) allowing firm responsiveness to TFP to vary by year, to capture time variation in the same;[37] (iii) allowing for TFP to affect the policy function non-linearly, to capture asymmetric hiring and firing rules as found by Ilut et al. (2018); (iv) allowing output elasticities to change over time, to capture changes in production technology like automation; (v) estimating production functions and policy functions in a pre-sample, while estimating misoptimization in a post sample, to avoid mechanical over-fitting; (vi) and calculating productivity using the proxy-variable method of Olley and Pakes (1996), to guard against over-reliance on cost shares and the necessary imputation of returns to scale. We finally consider two alternative specifications of the main regression, the first with linear and quadratic time-trend controls, which can

[36] The third column of Table A7 shows the estimated quasi-differenced policy function.

[37] Decker et al. (2020), in particular, note a secular pattern of "declining business dynamism" or declining responsiveness to TFP, which in principle could contaminate our results.

absorb confounding low-frequency trends, and the second restricting to the manufacturing sector, which has been the focus of much of the literature on misallocation (e.g., Hsieh and Klenow, 2009; Kehrig, 2015). Appendix Table A1 demonstrates the stability of our main finding, counter-cyclical misoptimization, under each scenario.

5.2.3 Misoptimization in Other Input Choices

We focused on misoptimization in labor choice because it is the only input or output measured in quantity units in our dataset. The model implied that we could measure misoptimization in any variable input choice. As a robustness exercise, we recalculate misoptimizations and Misoptimization Dispersion using as the choice variable total variable cost expenditures (i.e., materials plus labor) and investment rates (i.e., log growth rates of the capital stock).[38] We plot the resulting variations of Misoptimization Dispersion in Appendix Figure A6. We find broadly similar patterns, particularly in the spike of the mid 1990s and falls in the 2002 and 2009-10 downturns. The result with investment rates echoes findings by Eisfeldt and Rampini (2006) and Bachmann and Bayer (2014), using comparable data on US public firms, that investment rates are more dispersed in booms than in downturns. Our model's explanation, complementary to that of the literature, is that much of the apparent "reallocation" of capital in booms is unrelated to fundamentals and therefore a cause of *misallocation*, whereas the less dispersed investments in downturns are more related to fundamentals and therefore a cause of *better allocation*.

5.3 Misoptimization is More Costly in Bad States

We have established that, by our measure, firms make smaller misoptimizations in downturns. We now take to the data the model's predicted mechanism: that firms have higher financial incentives to avoid misoptimization in downturns.

5.3.1 Markets Punish Mistakes More When Aggregate Returns Are Low

We study "state-dependent market punishment" in a regression model that relates a firm's valuation, embodied in its stock return $\Delta \log P_{it}$, to the firm's squared misoptimization

[38] We use investment rates rather than the capital stock for comparability to the literature, although the model strictly speaking predicts an analogue of Equation 25 for $\log K_{it}$.

innovation over that year interacted with the aggregate (S&P 500) stock return $\Delta \log P_t$:[39,40]

$$\Delta \log P_{it} = \beta \cdot \hat{u}_{it}^2 + \phi \cdot \left(\hat{u}_{it}^2 \times \Delta \log P_t\right) + \chi_{j(i),t} + \Gamma' X_{it} + \epsilon_{it} \tag{32}$$

Sector-by-time fixed effects partial out industry trends. The vector of control variables X_{it} can include firm fixed effects and the growth rate of firm-level TFP, to partial out other important determinants of returns. The premise consistent with our interpretation of misoptimizations is that the marginal effect of misoptimization on the stock price is negative.[41] The hypothesis that the market punishes misoptimization more severely in times of distress, or low ΔP_t, is captured by $\phi > 0$. Relative to our model in Sections 3 and 4, this prediction is a heuristic interpretation of the elevated "stakes" from the risk-pricing channel in the comparative statics of Proposition 1 and Corollary 1. In the expanded model of Online Appendix B, we directly derive the regression equation and the prediction $\phi \geq 0$, with equality only if investors are risk-neutral and hence there is no risk-pricing channel.

Table 1 shows our estimates. In all four specifications, with and without firm fixed effects and productivity controls, we verify that $\phi > 0$ or that misoptimization is priced more severely in states of low aggregate returns.[42] Our estimates in column (3), in particular, suggest that mistakes have a zero price if the S&P return is 22%, close to its value in the late 1990s or the height of the dot com bubble. By contrast, in the trough of 2008 ($\Delta P_t = -0.52$), the model implies that pricing is 6.2 times more severe than in the "usual" states of $\Delta \log \mathrm{SP}_t \approx 0.10$. Appendix Table A2 shows the stability of our main finding of $\phi > 0$ to the alternative data-construction approaches highlighted in the previous subsection.

5.3.2 *Unpacking the Mechanism: Profitability vs. Returns*

To further explore the mechanism by which misoptimization affects market valuation, we study profits as an intermediate variable. We observe, for each firm, earnings before interest

[39] For this and all subsequent panel regressions, we drop observations in the 1% tail of both the outcome and main regressor to limit the effect of outliers, which are likely driven in this context by measurement errors (both sides for symmetric variables; only the upper tail for positive variables). Results are qualitatively similar on the full sample.

[40] To be consistent with our earlier analysis, we run this regression using re-scaled mistake innovations $\hat{u}_{it}$. In Appendix Table A3, we replicate the analysis using our (first-stage) estimates of the mistake "level" $\hat{m}_{it}$ and find qualitatively similar results.

[41] We verify this directly in the first four columns of Appendix Table A4 by estimating a variation of Equation 32, and its twin with profitability as the outcome, without the aggregate interaction. Using within and across firm variation, the negative effect is strongly statistically significant; using only within-firm variation, the effect is weakly statistically significant ($p < 0.10$). Moreover, in Appendix Figure A7, we show the binned scatter plot of returns and profitability against $\hat{u}_{it}$, net of industry-by-time fixed effects, to verify that variable's quadratic effect on firm performance around $\hat{u}_{it} = 0$.

[42] Similar results are obtained using mistake "levels" as discussed in Footnote 40 (columns 3-6 of Appendix Table A3)

Table 1: Markets Punish Misoptimization Harder in Low-Return States

	(1)	(2)	(3)	(4)
	Outcome: $\Delta \log P_{it}$			
$\hat{u}_{it}^2$	-0.268	-0.262	-0.097	-0.087
	(0.025)	(0.023)	(0.034)	(0.033)
$\hat{u}_{it}^2 \times \Delta \log P_t$	0.376	0.376	0.443	0.431
	(0.123)	(0.124)	(0.171)	(0.167)
Sector x Time FE	✓	✓	✓	✓
Firm FE			✓	✓
TFP Control		✓		✓
N	41,578	41,578	41,206	41,206
R^2	0.239	0.261	0.385	0.403

Notes: Standard errors are double-clustered at the year and firm level.

and taxes (EBIT) or income net of variable costs, and define profitability π_{it} as this year's EBIT divided by the last year's total variable costs.[43] We study the state-dependent effects of misoptimizations on profitability using the following regression model that mirrors our previous analysis of stock returns:[44]

$$\pi_{it} = \beta_\pi \cdot \hat{u}_{it}^2 + \phi_\pi \cdot \left(\hat{u}_{it}^2 \times \Delta \log P_t\right) + \chi_{j(i),t} + \gamma_i + \epsilon_{it} \tag{33}$$

Heuristically, $\phi_\pi > 0$ isolates the *dollar-profit-curvature* channel of Proposition 1 and Corollary 1. In the expanded model of Online Appendix B, we derive the prediction $\phi_\pi \geq 0$ with equality if the profits function, in dollars, has state-independent curvature.

We report the results in the first column of Table 2: a positive, but small and statistically insignificant, ϕ_π. Thus misoptimizations have a constant effect on firms' dollar income up to statistical precision. This suggests that the mechanism for our earlier finding of state-dependent market punishment relates specifically to the market's greater *reaction* to fixed profit effects of misoptimizations.

We next estimate a sequence of models that explore the joint effects of misoptimizations and profitability on stock returns. Our first model regresses firm stock returns on $\hat{u}_{it}^2$ and π_{it}

[43] Variable costs, by our definition, are cost of good sold (COGS) plus administrative expenses (XSGA) net of depreciation (DP). In other words, they are the sum of what we term "materials" and labor expenses in the production function. Normalization by lagged costs, rather than current costs, limits mechanical denominator bias related to the current period's mistakes. Results are similar when normalizing by total sales or costs in the current or previous period.

[44] For simplicity, in this section, we use firm fixed effects as a control. Results are quantitatively similar in other specifications which control for lagged stock returns and/or TFP growth.

Table 2: Misoptimization, Profits, and Pricing

	(1) Outcome: π_{it}	(2) Outcome: $\Delta \log P_{it}$	(3)	(4)
$\hat{u}_{it}^2$	-0.114	-0.021		
	(0.020)	(0.032)		
$\hat{u}_{it}^2 \times \Delta \log P_t$	0.112			
	(0.089)			
π_{it}		0.400	0.421	0.690
		(0.028)	(0.034)	(0.305)
$\pi_{it} \times \Delta \log P_t$			-0.303	-1.642
			(0.166)	(0.632)
Firm FE	✓	✓	✓	✓
Sector x Time FE	✓	✓	✓	✓
N	50,966	40,879	40,879	40,879
R^2	0.663	0.402	0.402	
First-stage F				17.80

Notes: Standard errors are double-clustered at the year and firm level.

conditional on firm and sector by time fixed effects. We find that, conditional on profitability, misoptimizations have a severely attenuated, and statistically indistinguishable from zero, effect on stock returns (column 2 of Table 2). This is consistent with the model interpretation that misoptimizations matter for prices by reducing current profits.

We next explore the market response to profit anomalies, relative to fixed firm profitability and industry trends, without taking a stand structurally on how those anomalies arise. Our estimating equation is the mirror of Equation 32 with profitability in place of misoptimizations:

$$\Delta \log P_{it} = \beta_{\pi \to P} \cdot \pi_{it} + \phi_{\pi \to P} \cdot (\pi_{it} \times \Delta \log P_t) + \chi_{j(i),t} + \gamma_i + \epsilon_{it} \tag{34}$$

again estimated with firm fixed effects as the control variables. We recover the pattern that the market prices more aggressively respond to profitability when aggregate returns are low, or an estimate of $\phi_{\pi \to P} > 0$ (column 4 of Table 2). This validates, without any intermediate structural estimation of production functions or misoptimization, the idea that the market values "firm performance" more acutely in low-return environments.

We finally present an instrumental variables (IV) estimate of cyclical market response to directly quantify the pathway from misoptimizations to state-dependently-priced profitability to stock returns. Specifically, defining Z_{it} as the vector of regressors $(\pi_{it}, \pi_{it} \times \Delta \log P_t)$

and W_{it} as the vector of instruments $(\hat{u}_{it}^2, \hat{u}_{it}^2 \times \Delta \log P_t)$, we augment the now "second-stage" Equation 34 with a "first-stage" equation

$$Z_{it} = W_{it}' A + \chi_{j(i),t} + \gamma_i + \epsilon_{it} \tag{35}$$

The first stage F-statistic is 17.8, owing to the strong relationship between misoptimizations and profitability. Our IV estimates of $(\beta_{\pi \to P}, \pi_{\pi \to P})$, while not very precisely estimated, suggest if anything a greater cyclicality of the market response to misoptimization than to other determinants of profits (column 5 of Table 2). We interpret this result, along with the others in this subsection, as empirical validation of the model's microeconomic mechanism for state-dependent attention and misoptimization.

5.4 Macro Attention Predicts Smaller Misoptimizations

We finally investigate the relationship between misoptimization, which we could link directly to the model, and Macroeconomic Attention from Section 2, which motivated our analysis but did not have a formal analogue in the model. To this end, we estimate the following empirical model of $\hat{u}_{it}^2$, the squared innovation of the firm's model-implied misoptimization, and the log of measured firm-level attention:[45]

$$\hat{u}_{it}^2 = \beta_a \cdot \log \text{MacroAttention}_{it} + \chi_{j(i),t} + \Gamma' X_{it} + \epsilon_{it} \tag{36}$$

Absorbed effects at the sector-by-time level partial out all trends, including the cyclical patterns studied earlier. Additional controls X_{it} can include individual fixed effects, to isolate variation at the firm level; and the growth rates (log differences) of TFP and end-of-year stock prices, to help further isolate variation in attention unrelated to firm-level fundamentals. An estimate of $\beta_a < 0$ would indicate that firms mention the macroeconomy more in periods of lower misoptimization, consistent with measurable macroeconomic attention contributing toward lower misoptimization.

Our findings, reported in Table 3, are consistent with this prediction: higher macroeconomic attention corresponds with smaller production misoptimization. The main specification in column (1) finds a strongly statistically significant effect (p <0.01).[46] The more

[45] As in our estimation of Equation 32, and for the same underlying reasons noted in Footnote 39, we drop observations in the 1% tails of both the outcome and main regressors (both sides for symmetric variables; only the upper tail for positive variables).

[46] The point estimate implies that a one standard deviation variation in Macro Attention across firms and time, 0.41 log points in our sample, can induce a change in the dependent variable equal to 5% of its average value—a small effect, as we might expect given the measurement error in both the dependent and independent variables.

Table 3: Macro-Attentive Firms Make Smaller Misoptimizations

	(1)	(2)	(3)	(4)
	Outcome: $\hat{u}_{it}^2$			
$\log \text{MacroAttention}_{it}$	-0.0081	-0.0052	-0.0058	-0.0056
	(0.0028)	(0.0029)	(0.0044)	(0.0038)
Sector x Time FE	✓	✓	✓	✓
Firm FE			✓	✓
TFP, Price Controls		✓		✓
N	28,279	24,392	27,875	23,930
R^2	0.053	0.067	0.383	0.384

Notes: Standard errors are double-clustered at the year and firm level.

controlled specifications in columns (2) to (4) estimate negative, similarly sized effects, but with considerably less precision.

Appendix Table A5 explores the timing of this relationship. We find only weak evidence of anticipatory effects, or high attention preceding years of low misoptimizations, and strong evidence of persistent effects, or high attention following years of low misoptimizations.[47] Appendix Table A6 shows that results are similar using our conference-call based measure of attention as well as all considered alternative models for the misoptimizations.

5.5 Discussion and Relationship with the Literature

Before proceeding to the quantitative exercise, we summarize an additional empirical exercise in the Online Appendix and the relation of our findings with the literature on uncertainty and productivity.

5.5.1 Survey Evidence

Our model made no specific prediction about the accuracy of firm-level forecasts, for instance of future sales (e.g., $x_{i,t+1}$) or hiring plans, (e.g., $L_{i,t+1}$). Taken literally, our model suggests that in "high attention" periods firms will be more confident about their plans conditional on fundamentals, but they may be arbitrarily or less uncertain about what those fundamentals are. By contrast, firm-level *backcasts* of specific state variables, like external macroeconomic conditions, may have a more direct interpretation as "attentiveness." In Online Appendix G, we bring out two pieces of evidence consistent with our overall analysis in the survey of

[47]These findings also relate to our measuring attention in regulatory documents filed within a given fiscal year, and not those *describing* a given fiscal year. In particular, the form 10-K describing year t is likely to be filed in year $t + 1$.

firms in New Zealand conducted by Coibion et al. (2018a). First, firms report that they would be significantly more likely to seek out news about the macroeconomy if there were an aggregate negative shock, consistent with the sign of our Attention and Misoptimization Cycles. Second firms that report a higher sensitivity of firm profits to their own choices (in this case, posted prices) demonstrate higher awareness of macroeconomic aggregates, consistent with the profit curvature case of our model.

5.5.2 *Dispersion in Fundamentals Versus Misoptimizations*

Kehrig (2015) and Bloom et al. (2018), using micro-data in the manufacturing sector, estimate that the variance in total factor productivity rises in recessions or periods of negative growth.[48] Our analysis, by contrast, studies variance in input choices *conditional* on productivity, and then uses a structural model to interpret the implications for (mis)allocation.

To demonstrate the empirical consistency of these sets of findings, we follow Bloom et al. (2018) and estimate a first-order autoregressive model for TFP with firm and sector-by-time fixed effects:

$$\log \theta_{it} = \gamma_i + \chi_{j(i),t} + \rho_\theta \theta_{i,t-1} + \epsilon_{it} \tag{37}$$

We estimate "TFP Innovation Variance" as the weighted average, $\mathbb{E}[s^*_{it}\epsilon^2_{it}]/\mathbb{E}[s^*_{it}]$.[49]

Appendix Figure A8 plots our estimate against the unemployment rate and detrended stock price. TFP Innovation Variance spikes markedly in the 2002 and 2007-09 recessions, as well as in the 1990s boom. Our measure is weakly and insignificantly counter-cyclical by our measure, with a regression coefficient of 0.051 (SE: 0.153) on unemployment; and it is weakly and insignificantly higher in US recessions, with a regression coefficient on an NBER recession indicator of 0.010 (SE: 0.009).

Our measure of TFP dispersion has a correlation of 0.39 with the equivalent measure from Bloom et al. (2018) over a common sample, and this correlation increases to 0.47 if we restrict in our data to the manufacturing sector.[50] This suggests that our measurement of TFP stochasticity among US public firms is consistent with the measurement of Bloom et al. (2018) using Census data in the manufacturing sector. The Bloom et al. (2018) measure is weakly and insignificantly *pro-cyclical* based on its regression coefficient on unemployment (coefficient: -0.309, SE: 1.209) and, as reported by Bloom et al. (2018), larger in recessions (coefficient: 0.098, SE: 0.022).

[48]Kehrig (2015) also shows a negative correlation between detrended dispersion and detrended *levels* of output, at the sectoral and aggregate levels.

[49]This choice of weights is consistent with our main analysis, but results are similar with the simple average (unreported for brevity).

[50]The common sample is 1987-2010. We use the variance (square of standard deviation) of TFP innovations on the sample of establishments that are in the Bloom et al. (2018) data for 25 years.

6 Quantifying the Consequences of Attention Cycles

The previous section verified the model's microeconomic and macroeconomic predictions for misoptimization, which governed the model's implications for output and productivity dynamics. In this final section, we quantify the consequences of Attention and Misoptimization Cycles in a numerical calibration of our model that matches the empirical findings. Our strategy is to match the aggregate level and cyclicality of misoptimization-induced dispersion in production, which control the size and dynamics of the attention wedge. This allows us to explore the quantitative importance of attention cycles.

In particular, we show the following three important model features: asymmetrically large amplification of negative shocks; greater amplification of shocks when output is low; and endogenously higher volatility when output is low. Each of these features arises from the endogenous adjustment of attention, and would be missing from a full-attention model.

6.1 Calibration

In our calibration, as in our theoretical results of Section 4, the aggregate state variable $\theta_t = \left(\mathbb{E}_{G_t}[\theta_i^{\epsilon-1}]\right)^{\frac{1}{\epsilon-1}}$ is a one-dimensional sufficient statistic for the productivity distribution. We assume that this productivity state θ_t follows a zero-mean, Gaussian AR(1) process in logs:

$$\log\theta_t = \rho_\theta \log\theta_{t-1} + \nu_t \tag{38}$$

where $\nu_t \sim^{IID} N(0, \sigma_\theta^2)$. Note that, in this formulation, the shocks ν_t may reflect changes to *any* moment of the productivity distribution that induce changes in the aggregator θ_t, such as a standard shock to average productivity or an "uncertainty shock" to productivity dispersion as studied by Bloom et al. (2018). Since we will study purely macroeconomic phenomena, we can therefore be agnostic about what shocks to the productivity distribution induce the dynamics of Equation 38.

The model has six parameters.[51] We calibrate two to standard values and match four to measured moments, as reviewed in Table 4 and explained below.

The first calibrated parameter is the elasticity of substitution between products. We set $\epsilon = 4$, which implies an optimal average markup of $\mu = \frac{\epsilon}{\epsilon-1} = \frac{4}{3}$. This is conservative relative to estimates by De Loecker et al. (2020) (1.60) and Demirer (2020) (1.45) for modern markups for US public firms, and slightly larger than the estimate by Edmond et al. (2018)

[51] Two other parameters of less economic relevance are the constants $\bar{w}$ and $\bar{X}$ in the wage rule. These scale overall production in equilibrium (see Proposition 1). We set $\bar{w}$ and $\bar{X}$ to match the wage and output prevailing in a frictionless market economy with Greenwood et al. (1988) preferences over labor and leisure and elasticity of labor supply $\phi = 1$, evaluated at state $\log\theta = 0$.

Table 4: Parameters for Calibration

Fixed	ϵ	Elasticity of substitution	4
	ρ_θ	Persistence of productivity	0.95
Free	χ	Elasticity of real wages to output	0.095
	λ	Average weight on entropy penalty	0.406
	γ	Coefficient of relative risk aversion	11.5
	σ_θ^2	Variance of the productivity innovation	4.82×10^{-7}
Matched	β_U	Slope of Misopt. Dispersion on -$\frac{\text{Unemployment}}{100}$	0.841
	$\bar{\sigma}_M^2$	Average level of Misopt. Dispersion	0.080
	χ	Regression of detrended real wages on detrended output	0.095
	σ_Y^2	Variance of quarterly output growth	0.337

Notes: "Fixed" parameters are externally set. "Free" parameters are chosen to equate the "Matched" parameters in the model and in the data.

on the same dataset (1.25). The elasticity of substitution crucially controls the translation of misoptimization into output and productivity, with a lower value (e.g., higher steady-state markup) translating a fixed level of dispersion of production levels around *ex post* optimal levels into a larger penalty for output and productivity. We next set the persistence of the productivity shock, at the quarterly frequency, to a standard value of $\rho = 0.95$.

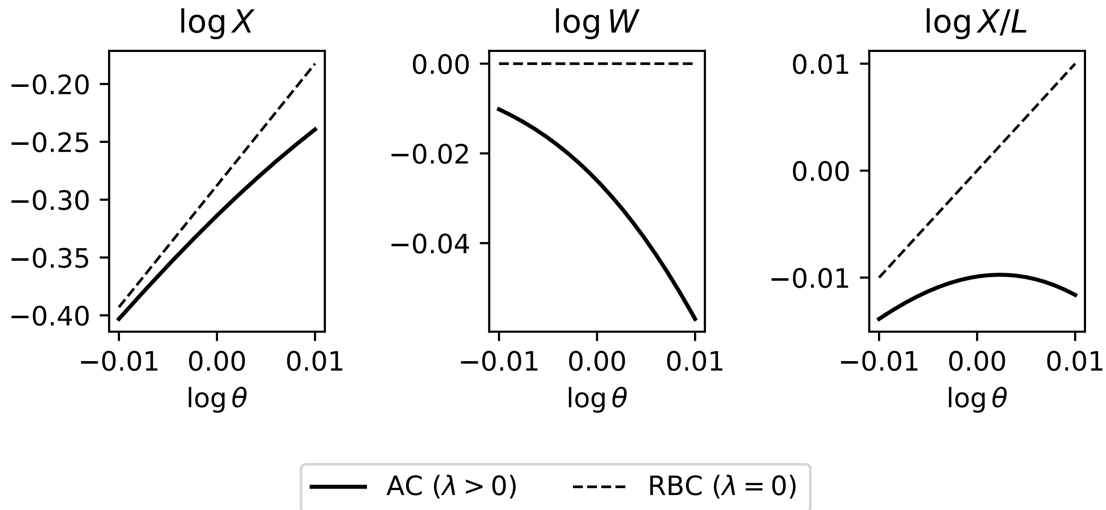
The four remaining free parameters are the elasticity of real wages to output χ, the average attention cost λ, the coefficient of relative risk aversion γ, and the volatility of the productivity process σ_θ^2. We set these four parameters in the model to match four simulated moments with exact analogues in the data, as described in Table 4, and fit the model exactly up to machine precision.

Two of these parameters, χ and σ_θ^2, have a relatively simple interpretation. For the first, we match directly an OLS regression of linearly detrended real wages on linearly detrended GDP, at the quarterly frequency over our studied period 1987-2018.[52] Our estimate of 0.095 lines up with recent evidence on the "flattening" of the wage Phillips curve (Galí and Gambetti, 2019). For the second, we match the variance of quarterly real GDP growth over our sample period. Intuitively, conditional on all other parameters, the variance of output growth is monotone in this variance in the shock process and simple to match exactly.

The key properties of inattention and misallocation are controlled by the remaining two moments. In the model, we calculate the average level of Misoptimization Dispersion, or optimal-sales-weighted dispersion in $\frac{L_{it}-L_{it}^*}{L_{it}^*}$, and the population regression coefficient of

[52] Our real wage series is the median weekly real earnings for wage and salary workers over the age of 16, as reported by the US BLS.

Figure 3: Output, Attention Wedge, and Labor Productivity



Misoptimization Dispersion on log employment.[53] We match these moments respectively to the time-series average of Misoptimization Dispersion, 0.080, and the (negative) slope of Misoptimization Dispersion in unemployment, 0.841. Intuitively, these moments identify the level of misoptimization and the extent of its cyclicality, and in the model, conditional on all other parameters, they identify the average cost of attention λ, itself a sufficient statistic for the cross-sectional distribution, and the coefficient of relative risk aversion γ, which controls the risk-pricing incentive that drives cyclical misoptimization.

Our estimates of γ are thus based entirely on fitting a stochastic discount factor that justifies the observed pattern of misoptimization in our model, rather than incorporating an informed prior from external evidence in financial economics as discussed in Section 4.2. Our finding of $\gamma = 11.5$ (Table 4) is, if anything, slightly conservative relative to the modern asset-pricing literature that estimates, in variations of the consumption capital asset pricing (CCAPM) model, γ of about 15-20 with "unfiltered" measures of consumption (Savov, 2011; Kroencke, 2017) or long-run variation in consumption growth (Parker and Julliard, 2005).

6.2 Output, Productivity, and the Attention Wedge

Figure 3 shows log output, the log attention wedge (as defined in Proposition 4), and log labor productivity in our calibration. In each case, we compare to the predictions of an otherwise identical "pure RBC model" with full attentiveness, or $\lambda = 0$, plotted as a dashed line. As guaranteed by Proposition 4 and Corollary 2, inattention reduces output and

[53] We calculate the "regression coefficient" in the model using numerical integration over the stationary distribution of states. Appendix A.7.3 shows the exact in-model formula for aggregate Misoptimization Dispersion, calculated with optimal-sales-weights, which depends only on the sufficient statistics (θ, λ) of the cross-sectional type distribution.

productivity relative to the fully attentive counterfactual, and this effect grows in larger states. In the mean state, the attention wedge reduces output by 2.6% relative to the fully attentive counterfactual, labor productivity by $\chi\epsilon \times 2.6\% = 1.0\%$, and employment by $(1-\chi\epsilon) \times 2.6\% = 1.6\%$. To decompose further the source of these losses, we calculate also a "partial equilibrium" attention wedge based on firms' inattentive best-responses to the counterfactual RBC dynamics. This partial-equilibrium wedge is compared to the equilibrium wedge in the left panel of Appendix Figure A9. In the mean state, the "partial equilibrium" attention wedge is 1.3% in terms of output, implying that GE interactions account for 1/2 of the losses from inattention.

We observe two further properties of output and productivity dynamics which were *not* immediately clear from the theoretical results and depended on the numerical calibration. First, labor productivity $\log A := \log \frac{X}{L}$ is non-monotone in underlying productivity (rightmost panel) and hence also in aggregate output. As we mentioned in the discussion following Corollary 2, this property results from the dueling forces of productivity shocks and induced misallocation.

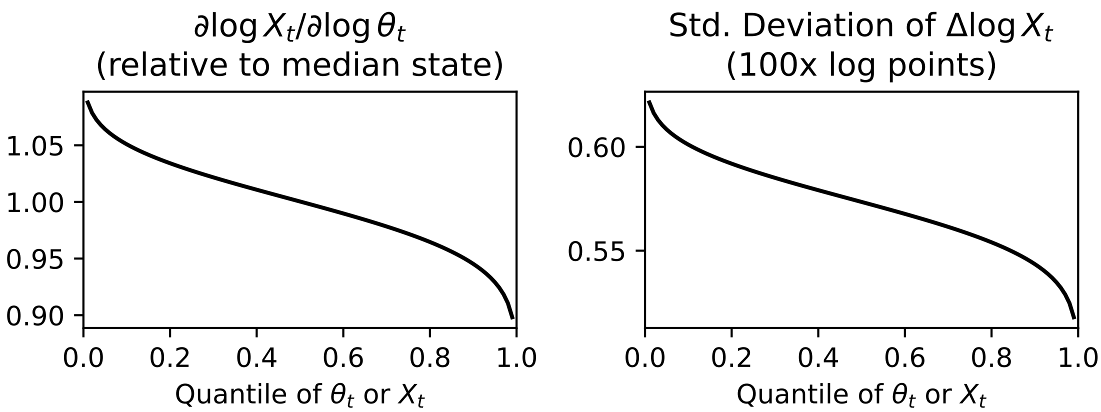
Second, the attention wedge is concave. Appendix Figure A9 shows this concavity directly and provides also a partial-vs.-general-equilibrium decomposition of the finding. Through the lens of Corollary 3, a concave attention wedge implies that, fixing shock sizes, negative shocks have a larger effect on output than positive shocks, and that overall shock response and volatility is higher low-output states. We explore these predictions quantitatively in the next subsection.

6.3 State-Dependence, Asymmetry, and Stochastic Volatility

Figure 4 plots the marginal sensitivity of output to percentage productivity shocks and the standard deviation of output growth as a function of the initial quantile of the stationary distribution for productivity or output. As suggested by the discussion of the concave attention wedge above, both model objects are decreasing functions of the state.

One way to benchmark the extent of asymmetry and state-dependence in shock responses is to consider an "impulse response" thought experiment of a fixed size. Let $\log\hat{\theta}$ be size of fundamental shock that induces a 3% change in output, or solves $\log X(\log\hat{\theta}) - \log X(0) = 0.03$. We compare the normalized effect of this "Positive" shock to the effect of a "Negative" shock from $\log\theta_0 = 0$ to $\log\theta_1 = -\log\hat{\theta}$, and a "Double Dip" shock from $\log\theta_0 = -\log\hat{\theta}$ to $\log\theta_1 = -2\cdot\log\hat{\theta}$. We find that the negative shock has a 7% larger effect on output than the positive shock, and the double dip shock has a 14% larger effect on output than the positive shock. The same results for the response of total labor are 5% and 8%, respectively. The patterns of asymmetry are, in particular, potentially realistic. As a representative example,

Figure 4: Asymmetric Shock Response and Stochastic Volatility



Ilut et al. (2018) estimate that US industries have on average a 20% larger response to negative aggregate productivity shocks than positive shocks.[54] In comparison to this benchmark, our model explains 25% of the empirically realistic shock response via the endogenous reallocation of attention. More generally, to the extent that linear macroeconomic models like our RBC benchmark are plagued with the counterfactual result of left-tailed shock distributions, which cannot even be explained by symmetric, exogenous stochastic volatility, the attention-cycles mechanism offers a partial solution.

To benchmark the extent of stochastic volatility, we observe that a transition from the 90^{th}-percentile to the 10^{th}-percentile productivity state of the reduces output by 4.7%, increases the standard deviation of output growth by 10.6%, and increases the variance of the same by 22.3%. The peak-to-trough fall of output during the Great Recession (e.g., from early 2007 to early 2009) was comparable, and Jurado et al. (2015) estimate that over this episode the forward-looking volatility of industrial production growth over three-month horizons increased by 57%.[55] By this metric, the endogenous re-allocation of attention our model can explain about 19% of empirically reasonable movement in macroeconomic uncertainty, with no underlying stochastic volatility in fundamentals.

6.4 Parameter Robustness and Counterfactual Scenarios

In Online Appendix D, we provide additional results from our numerical exercise. We first explore robustness of our main findings to different external calibrations of wage rigidity χ and substitutability ϵ and to introducing classical labor markets using the preferences

[54]This calculation is based on comparing the "data" estimates in columns 6 and 7 of Table 9 in that paper.

[55]This calculation compares the 3-month "macro uncertainty index," as available at https://www.sydneyludvigson.com/macro-and-financial-uncertainty-indexes, from a low point in April 2007 to a high point in October 2008.

of Greenwood et al. (1988). We also study counterfactual scenarios by altering structural parameters *without* recalibrating the model. To summarize, we find that Attention Cycles generate more pronounced differences from the log-linear, RBC core, and also from the observed equilibrium, in regimes with smaller markups, greater wage rigidity, and higher attention costs.

7 Conclusion

This paper studies cyclical attention to the macroeconomy, *attention cycles*, as both a cause and consequence of macroeconomic dynamics. We develop new measures of firms' attention to the macroeconomy and misoptimization in decision making. We document that macroeconomic attention is counter-cyclical, that the extent of misoptimization is pro-cyclical, that more macroeconomically attentive firms make smaller misoptimizations, and that the market financially punishes firms for misoptimizations more in downturns. We build a macroeconomic theory to understand why cyclical attention should manifest in both partial and general equilibrium by embedding state-dependent stochastic choice in a business cycle model. We derive empirically reasonable conditions under which attention is highest in downturns owing to the higher stakes for making correct decisions, or higher curvature of firms' objectives as a function of choices, in these states. Consistent with our empirical findings, the main contributing macroeconomic force is *risk-pricing* or higher marginal utility in low-production states of the world. Calibrating the model to match our evidence on cyclical misoptimization, we uncover a quantitatively important role for attention cycles in driving macroeconomic dynamics; attention cycles cause asymmetrically larger propagation of negative shocks, larger propagation of all shocks in low-output states and endogenously higher volatility in low-output states.

References

Abel, A. B., Eberly, J. C., and Panageas, S. (2013). Optimal inattention to the stock market with information costs and transactions costs. *Econometrica*, 81(4):1455–1481.

Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. *Information Processing & Management*, 39(1):45–65.

Alvarez, F. E., Lippi, F., and Paciello, L. (2011). Optimal price setting with observation and menu costs. *The Quarterly Journal of Economics*, 126(4):1909–1960.

Alves, F., Kaplan, G., Moll, B., and Violante, G. L. (2020). A further look at the propagation of monetary policy shocks in HANK. *Journal of Money, Credit and Banking*, 52(S2):521–559.

Angeletos, G.-M. and La'O, J. (2010). Noisy business cycles. *NBER Macroeconomics Annual*, 24(1):319–378.

Bachmann, R. and Bayer, C. (2014). Investment dispersion and the business cycle. *American Economic Review*, 104(4):1392–1416.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Barnichon, R. (2010). Productivity and unemployment over the business cycle. *Journal of Monetary Economics*, 57(8):1013–1025.

Basu, S. and Fernald, J. G. (1997). Returns to scale in US production: Estimates and implications. *Journal of Political Economy*, 105(2):249–283.

Benhabib, J., Liu, X., and Wang, P. (2016). Endogenous information acquisition and countercyclical uncertainty. *Journal of Economic Theory*, 165:601–642.

Berger, D. (2012). Countercyclical restructuring and jobless recoveries. *Manuscript, Yale.*

Berger, D. and Vavra, J. (2019). Shocks versus responsiveness: What drives time-varying dispersion? *Journal of Political Economy*, 127(5):2104–2142.

Blanchard, O. and Galí, J. (2010). Labor markets and monetary policy: A new keynesian model with unemployment. *American Economic Journal: Macroeconomics*, 2(2):1–30.

Blanchard, O. J. and Kiyotaki, N. (1987). Monopolistic competition and the effects of aggregate demand. *The American Economic Review*, pages 647–666.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3):623–685.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., and Terry, S. J. (2018). Really uncertain business cycles. *Econometrica*, 86(3):1031–1065.

Bybee, L., Kelly, B. T., Manela, A., and Xiu, D. (2021). Business news and business cycles. Working Paper 29344, National Bureau of Economic Research.

Caballero, R. J. and Hammour, M. L. (1994). The cleansing effect of recessions. *The American Economic Review*, 84(5):1350–1368.

Caballero, R. J. and Hammour, M. L. (1996). On the timing and efficiency of creative destruction. *The Quarterly Journal of Economics*, 111(3):805–852.

Caplin, A. and Dean, M. (2013). Behavioral implications of rational inattention with Shannon entropy. Technical report, National Bureau of Economic Research.

Chahrour, R., Nimark, K. P., and Pitschner, S. (2021). Sectoral media focus and aggregate fluctuations. Mimeo, Cornell University.

Chiang, Y.-T. (2021). Strategic uncertainty over business cycles. Mimeo, University of Chicago.

Coibion, O., Gorodnichenko, Y., and Kumar, S. (2018a). How Do Firms Form Their Expectations? New Survey Evidence. *American Economic Review*, 108(9):2671–2713.

Coibion, O., Gorodnichenko, Y., and Kumar, S. (2018b). Replication data for: How Do Firms Form Their Expectations? New Survey Evidence. American Economic Association (Publisher) and Inter-university Consortium for Political and Social Research (Distributor). Available at: https://doi.org/10.3886/E113095V1.

David, J. M., Hopenhayn, H. A., and Venkateswaran, V. (2016). Information, Misallocation, and Aggregate Productivity. *The Quarterly Journal of Economics*, 131(2):943–1005.

David, J. M. and Venkateswaran, V. (2019). The sources of capital misallocation. *American Economic Review*, 109(7):2531–67.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2020). Changing business dynamism and productivity: Shocks versus responsiveness. *American Economic Review*, 110(12):3952–90.

Demirer, M. (2020). Production function estimation with factor-augmenting technology: An application to markups. Mimeo, MIT.

Denti, T. (2020). Posterior-separable cost of information. Technical report, working paper.

Dew-Becker, I. and Giglio, S. (2020). Cross-sectional uncertainty and the business cycle: evidence from 40 years of options data. Technical Report w27864, National Bureau of Economic Research.

Edmond, C., Midrigan, V., and Xu, D. Y. (2018). How costly are markups? Working Paper 24800, National Bureau of Economic Research.

Eisfeldt, A. L. and Rampini, A. A. (2006). Capital reallocation and liquidity. *Journal of Monetary Economics*, 53(3):369–399.

Flynn, J. P. and Sastry, K. A. (2021). Strategic mistakes. Mimeo, MIT.

Flynn, Z., Traina, J., and Gandhi, A. (2019). Measuring markups with production data. Working Paper 3358472, SSRN.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425.

Foster, L., Haltiwanger, J. C., and Krizan, C. J. (2001). *Aggregate Productivity Growth: Lessons from Microeconomic Evidence*, pages 303–372. University of Chicago Press.

Fudenberg, D., Iijima, R., and Strzalecki, T. (2015). Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Gabaix, X. (2020). A behavioral New Keynesian model. *American Economic Review*, 110(8):2271–2327.

Galí, J. and Gambetti, L. (2009). On the sources of the great moderation. *American Economic Journal: Macroeconomics*, 1(1):26–57.

Galí, J. and Gambetti, L. (2019). Has the US wage Phillips curve flattened? a semi-structural exploration. Working Paper 25476, National Bureau of Economic Research.

Galí, J. and Van Rens, T. (2021). The vanishing procyclicality of labour productivity. *The Economic Journal*, 131(633):302–326.

Gorodnichenko, Y. (2008). Endogenous information, menu costs and inflation persistence. Technical report, National Bureau of Economic Research.

Greenwood, J., Hercowitz, Z., and Huffman, G. W. (1988). Investment, capacity utilization, and the real business cycle. *The American Economic Review*, pages 402–417.

Grigsby, J., Hurst, E., and Yildirmaz, A. (2021). Aggregate nominal wage adjustments: New evidence from administrative payroll data. *American Economic Review*, 111(2):428–71.

Handlan, A. (2020). Text shocks and monetary surprises: Text analysis of fomc statements with machine learning. Mimeo, Brown University.

Hansen, L. P. and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy*, 99(2):225–262.

Hansen, S. and McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99:S114–S133.

Harsanyi, J. C. (1973). Oddness of the number of equilibrium points: a new proof. *International Journal of Game Theory*, 2(1):235–250.

Hassan, T. A., Hollander, S., van Lent, L., Schwedeler, M., and Tahoun, A. (2020). Firm-level exposure to epidemic diseases: Covid-19, sars, and h1n1. Working Paper 26971, National Bureau of Economic Research.

Hassan, T. A., Hollander, S., Van Lent, L., and Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, 134(4):2135–2202.

Hopenhayn, H. and Rogerson, R. (1993). Job turnover and policy evaluation: A general equilibrium analysis. *Journal of political Economy*, 101(5):915–938.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Ilut, C., Kehrig, M., and Schneider, M. (2018). Slow to hire, quick to fire: Employment dynamics with asymmetric responses to news. *Journal of Political Economy*, 126(5):2011–2071.

Ilut, C. and Valchev, R. (2021). Economic agents as imperfect problem solvers. Mimeo, Duke University.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2016). A rational theory of mutual funds' attention allocation. *Econometrica*, 84(2):571–626.

Kehrig, M. (2015). The cyclical nature of the productivity distribution. Working Paper 1854401, SSRN.

Keller, W. and Yeaple, S. R. (2009). Multinational enterprises, international trade, and productivity growth: firm-level evidence from the United States. *The Review of Economics and Statistics*, 91(4):821–831.

Koenders, K. and Rogerson, R. (2005). Organizational dynamics over the business cycle: a view on jobless recoveries. *Federal Reserve Bank of Saint Louis Review*, 87(4):555.

Kroencke, T. A. (2017). Asset pricing without garbage. *The Journal of Finance*, 72(1):47–98.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65.

Lucca, D. O. and Trebbi, F. (2009). Measuring central bank communication: an automated approach with application to fomc statements. Technical report, National Bureau of Economic Research.

Ma, Y., Ropele, T., Sraer, D., and Thesmar, D. (2020). A quantitative analysis of distortions in managerial forecasts. Working Paper 26830, National Bureau of Economic Research.

Macaulay, A. (2020). Cyclical attention to saving. Working Paper 140, UniCredit Foundation.

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

Mäkinen, T. and Ohl, B. (2015). Information acquisition and learning from prices over the business cycle. *Journal of Economic Theory*, 158:585–633.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.

McDonald, B. (2021). Software repository for accounting and finance. Accessed June 2020. Available at: https://sraf.nd.edu/.

Morris, S. and Yang, M. (2021). Coordination and continuous stochastic choice. *The Review of Economic Studies*, Forthcoming.

Myerson, R. B. (1978). Refinements of the nash equilibrium concept. *International Journal of Game Theory*, 7(2):73–80.

Nimark, K. P. (2014). Man-bites-dog business cycles. *American Economic Review*, 104(8):2320–67.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Ottonello, P. and Winberry, T. (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, 88(6):2473–2502.

Parker, J. A. and Julliard, C. (2005). Consumption risk and the cross section of expected returns. *Journal of Political Economy*, 113(1):185–222.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the united kingdom, 1861-1957. *Economica*, 25(100):283–299.

Reeves, M., Rhodes, D., Ketels, C., and Whitaker, K. (2019a). Advantage in adversity: Winning the next downturn. Technical report, Boston Consulting Group Henderson Institute. Retrieved from: https://www.bcg.com/publications/2019/advantage-in-adversity-winning-next-downturn.

Reeves, M., Whitaker, K., and Ketels, C. (2019b). Companies need to prepare for the next economic downturn. *Harvard Business Review*. Retrieved from: https://hbr.org/2019/04/companies-need-to-prepare-for-the-next-economic-downturn.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720.

Savov, A. (2011). Asset pricing with garbage. *The Journal of Finance*, 66(1):177–201.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55.

Simon, H. A. (1947). *Administrative Behavior*. Macmillan, New York.

Sims, C. A. (1998). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, volume 49, pages 317–356. Elsevier.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Solon, G., Barsky, R., and Parker, J. A. (1994). Measuring the cyclicality of real wages: how important is composition bias? *The Quarterly Journal of Economics*, 109(1):1–25.

Song, W. and Stern, S. (2021). Firm inattention and the transmission of monetary policy: A text-based approach. Mimeo, University of Michigan.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps.*

Yasar, M., Raciborski, R., and Poi, B. (2008). Production function estimation in Stata using the Olley and Pakes method. *The Stata Journal*, 8(2):221–231.

A Omitted Proofs

A.1 Proof of Proposition 1

Proof. Consider a firm of type λ_i, with a payoff $u : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ and prior density $\pi \in \Delta(\mathcal{Z})$. The firm's stochastic choice problem can be written as

$$\max_{p \in \mathcal{P}} \int_{\mathcal{X}} \int_{\mathcal{Z}} u(x,z) p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z - \lambda_i \int_{\mathcal{X}} \int_{\mathcal{Z}} p(x|z) \log p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z \tag{39}$$

We can formulate this problem as constrained optimization for choosing $p(x|z)$ pointwise, with constraints embodying non-negativity and the restriction that conditional distributions integrate to one. We can then write a Lagrangian for this problem, giving these constraints multipliers $\kappa(x,z)$ and $\gamma(\theta)$, respectively:

$$\begin{aligned} \mathcal{L}(\{p(x|z), \kappa(x,z)\}, \{\gamma(z)\}) = & \int_{\mathcal{Z}} \int_{\mathcal{X}} u(x,z) p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z \\ & - \lambda_i \int_{\mathcal{Z}} \int_{\mathcal{X}} p(x|z) \log p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z \\ & + \int_{\mathcal{Z}} \int_{\mathcal{X}} \kappa(x,z) p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z \\ & + \int_{\mathcal{Z}} \gamma(z) \left(\int_{\mathcal{X}} p(x|z) \, \mathrm{d}x - 1 \right) \pi(z) \, \mathrm{d}z \end{aligned} \tag{40}$$

The Lagrangian is concave in the collection $\{p(x \mid z)\}$, since the expected utility term and the two constraint terms are linear in these variables, and the control-cost term is convex in these variables. Taking the first-order condition of the Lagrangian with respect to $p(x|z)$ yields the necessary first-order condition

$$u(x,z) - \lambda_i(\log p(x|z) + 1) + \kappa(x,z) + \gamma(z) = 0 \tag{41}$$

Re-arranging this expression and applying the normalization that the density integrates to one, we get the solution

$$p(x|z) = \frac{\exp\left(\lambda_i^{-1} u(x,z)\right)}{\int_{\mathcal{X}} \exp\left(\lambda_i^{-1} u(x',z)\right) \mathrm{d}x'} \tag{42}$$

This solution is invariant to the prior distribution $\pi(z)$, and hence can be indexed solely by the *ex post* realized state z.

To solve our firm's problem, we replace u in the above with $\tilde{\Pi}$ and z with z_i. Performing this substitution, and ignoring the normalizing constant, we get

$$p(x|z_i) \propto \exp\left(-\frac{(x - x^*(z_i))^2}{2\lambda_i|\Pi_{zz}(z_i)|^{-1}}\right) \tag{43}$$

Taking $\mathcal{X} = \mathbb{R}$, it is then immediate that $p(x|z_i)$ is a Gaussian random variable with mean $x^*(z_i)$ and variance $\lambda_i|\Pi_{xx}(z_i)|^{-1}$. Observing that $|\Pi_{xx}(z_i)| = |\pi_{xx}(z_i)|M(X)$, we can re-write the variance as $\lambda_i(|\pi_{xx}(z_i)|M(X))^{-1}$. Finally, we observe that the stochasticity in each firm's action conditional on z_i is independent from z_i and/or any other firm's action. Therefore we obtain the desired representation

$$x_i = x^*(z_i) + \sqrt{\frac{\lambda_i}{|\pi_{xx}(z_i)|M(X)}} \cdot v_i, \qquad v_i \sim \text{Normal}(0,1) \tag{44}$$

□

A.2 Proof of Proposition 2

Proof. We first re-define the state variable as the $\epsilon - 1$ quasi-arithmetic mean of θ_i

$$\theta := \left(\mathbb{E}_{\theta_i}[\theta_i^{\epsilon-1} \mid \theta]\right)^{\frac{1}{\epsilon-1}} \tag{45}$$

The expectation is taken over the distribution $G(\theta)$, which is by assumption increasing in θ via first-order stochastic dominance. Because this redefinition of the state preserves the strict ordering of realizations of θ, and lies within the domain Θ, it is without loss of generality.

To prove existence, we first study the problem of a single firm i who is best replying to the conjecture that the law of motion of the aggregate is $X : \Theta \to \mathbb{R}$. In particular, they believe that output is given by $X(\theta)$ in each state θ.

As established by Proposition 1, the firm's best-response is invariant to the firm's prior state $z_{i,t-1}$ and described by the following random variable conditional on each realization of z_{it}:

$$x_{it} = x^*(z_{it}) + \sqrt{\frac{\lambda_i}{|\Pi_{xx}(z_{it})|}} \cdot v_{it}, \qquad v_{it} \sim \text{Normal}(0,1) \tag{46}$$

As derived in Appendix A.7.1, the mean and variance scalings are the following, after substituting in the equilibrium conjecture $X_t = X(\theta_t)$:

$$\begin{aligned} x^*(z_{it}) &= v_x(\epsilon, \chi, \bar{w}, \bar{X}) \cdot X(\theta_t)^{1-\chi\epsilon}\theta_{it}^{\epsilon} \\ |\Pi_{xx}(z_{it})| &= v_\Pi(\epsilon, \chi, \bar{w}, \bar{X}) \cdot X(\theta_t)^{-1-\gamma+\chi(1+\epsilon)}\theta_{it}^{-1-\epsilon} \end{aligned} \tag{47}$$

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

$$\begin{aligned} v_x &:= \epsilon^{-\epsilon}\left(\epsilon-1\right)^{\epsilon}\bar{w}^{-\epsilon}\bar{X}^{\chi\epsilon} \\ v_\Pi &:= (\epsilon-1)^{-\epsilon}\epsilon^{\epsilon-1}\bar{w}^{1+\epsilon}\bar{X}^{-\chi(1+\epsilon)} \end{aligned} \tag{48}$$

Conditional on the realization of any state θ, aggregate output must solve the fixed point equation defined by combining the aggregate-good production function (16) and the firms' best responses. Applying the law of iterated expectations, we can re-write the aggregate-good production function as

$$X = X^* - \frac{1}{2\epsilon(X^*)^{-\frac{1}{\epsilon}}}\mathbb{E}_{\theta_i}\left[(x^*(z_i))^{-1-\frac{1}{\epsilon}}\mathbb{E}_{\lambda_i,v_i}\left[(x_i - x^*(z_i))^2 \mid \theta_i, \theta\right] \mid \theta\right] \tag{49}$$

where

$$X^* = \left(\mathbb{E}_{\theta_i}[x^*(z_i)^{1-\frac{1}{\epsilon}} \mid \theta]\right)^{\frac{\epsilon}{\epsilon-1}} \tag{50}$$

We now specialize the expressions above using the structure of the best response in Equations 46, 47, and 48. We first compute X^* as

$$X^* = v_x X^{1-\chi\epsilon}\theta^{\epsilon} \tag{51}$$

where θ is the transformation defined in Equation 45. We next calculate the the second, "variance" term. We start with the "misoptimization variance"

$$\mathbb{E}_{\lambda_i,v_i}\left[(x_i - x^*(z_i))^2\right] = \frac{\lambda}{v_\Pi}X^{1+\gamma-\chi(1+\epsilon)}\theta_i^{1+\epsilon} \tag{52}$$

and then calculate the full term

$$(X^*)^{\frac{1}{\epsilon}}\mathbb{E}_{\theta_i}\left[(x^*(z_i))^{-1-\frac{1}{\epsilon}}\mathbb{E}_{\lambda_i,v_i}\left[(x_i - x^*(z_i))^2\right]\right] = \frac{\lambda}{v_\Pi v_x}X^{\gamma-\chi}\theta \tag{53}$$

Substituting in Equations 51 and 53, we derive that equilibrium output solves:

$$X(\theta) = v_x X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} - \frac{\lambda}{2\epsilon v_x v_\Pi}X(\theta)^{\gamma-\chi}\theta \tag{54}$$

There is always a trivial equilibrium $X = 0$ arising from our approximations. Toward proving existence and uniqueness of a non-trivial equilibrium, define:

$$g(X,\theta) = a_0 X^{1-\chi\epsilon}\theta^{\epsilon} - a_1 X^{\gamma-\chi}\theta \tag{55}$$

where $a_0 = v_x > 0$ and $a_1 = \frac{\lambda}{2\epsilon v_x v_\Pi} > 0$. We now compute this function's derivatives in X:

$$
\begin{aligned}
g_X(X,\theta) &= a_0(1-\chi\epsilon)X^{-\chi\epsilon}\theta^\epsilon - a_1(\gamma-\chi)X^{\gamma-\chi-1}\theta \\
g_{XX}(X,\theta) &= -a_0(1-\chi\epsilon)\chi\epsilon X^{-\chi\epsilon-1}\theta^\epsilon - a_1(\gamma-\chi)(\gamma-\chi-1)X^{\gamma-\chi-2}\theta
\end{aligned}
\tag{56}
$$

If $1-\chi\epsilon > 0$ and $\gamma > 1+\chi$, then

$$
\lim_{X\to 0} g_X(X,\theta) = +\infty \qquad \lim_{X\to\infty} g_X(X,\theta) = -\infty \tag{57}
$$

Moreover, if $\gamma > \chi+1$ we have that $g_{XX}(X,\theta) < 0$ on $(0,\infty)$. Thus, when $\gamma > \chi+1$ and $\chi\epsilon < 1$, $g(X,\theta)$ crosses X from above and there exists a unique, positive fixed point for each θ. Iterated for all states $\theta \in \Theta$, this reasoning shows the existence of a unique, positive equilibrium mapping $X : \Theta \to \mathbb{R}_+$.

We now show monotonicity of the fixed point. To this end, we implicitly differentiate the fixed point condition:

$$
\begin{aligned}
\frac{\mathrm{d}X(\theta)}{\mathrm{d}\theta} &= \left[a_0(1-\chi\epsilon)X(\theta)^{-\chi\epsilon}\theta^\epsilon - a_1(\gamma-\chi)X(\theta)^{\gamma-\chi-1}\theta\right]\frac{\mathrm{d}X(\theta)}{\mathrm{d}\theta} \\
&\quad + \left[a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^{\epsilon-1} - a_1X(\theta)^{\gamma-\chi}\right]
\end{aligned}
\tag{58}
$$

Yielding:

$$
\frac{\mathrm{d}X(\theta)}{\mathrm{d}\theta} = \frac{a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^{\epsilon-1} - a_1X(\theta)^{\gamma-\chi}}{1-\left[a_0(1-\chi\epsilon)X(\theta)^{-\chi\epsilon}\theta^\epsilon - a_1(\gamma-\chi)X(\theta)^{\gamma-\chi-1}\theta\right]} \tag{59}
$$

Multiplying both sides by a factor of $\frac{\theta}{X}$:

$$
\frac{\mathrm{d}\log X(\theta)}{\mathrm{d}\log\theta} = \frac{a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} - a_1X(\theta)^{\gamma-\chi}\theta}{X(\theta)-\left[a_0(1-\chi\epsilon)X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1(\gamma-\chi)X(\theta)^{\gamma-\chi}\theta\right]} \tag{60}
$$

We first show that the numerator is positive

$$
\begin{aligned}
a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1X(\theta)^{\gamma-\chi}\theta &= a_0(\epsilon-1)X(\theta)^{1-\chi\epsilon}\theta^\epsilon + X(\theta) \\
&> 0
\end{aligned}
\tag{61}
$$

The first equality substitutes in the original fixed-point equation. The second follows from observing that $\epsilon > 1$, $a_0 > 0$, $X(\theta) > 0$, and $\theta > 0$. To show $\frac{\mathrm{d}\log X(\theta)}{\mathrm{d}\log\theta} > 0$ it now suffices to show that the denominator is positive, which follows from $\chi\epsilon < 1$ and $\gamma > \chi+1 > \chi$:

$$
\begin{aligned}
X(\theta) &= a_0X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1X(\theta)^{\gamma-\chi}\theta \\
&\geq a_0\underbrace{(1-\chi\epsilon)}_{\chi\epsilon<1}X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1\underbrace{(\gamma-\chi)}_{\gamma-\chi>1}X(\theta)^{\gamma-\chi}\theta
\end{aligned}
\tag{62}
$$

This shows that $\frac{\mathrm{d}\log X(\theta)}{\mathrm{d}\log\theta} > 0$ and implies that $X(\theta)$ is an increasing function. $\square$

A.3 Proof of Proposition 3

Proof. Building on the discussion in the main text, we first provide a formal definition of how to rank average attention and misoptimization:

Definition 2 (Aggregate Attention and Misoptimization)**.** *Fix equilibrium laws of motion* $\{X(\theta), w(\theta)\}$*. Firms' aggregate attention in state* θ*,* $a(\theta)$*, is their average realized cognitive cost*

$$a(\theta) := \mathbb{E}_{\theta_i,\lambda_i,v_i}\left[\lambda_i p^*(x|z_i(\theta);\lambda_i)\log p^*(x|z_i(\theta);\lambda_i) \mid \theta\right] \tag{63}$$

where $z_i(\theta) = (\theta_i, X(\theta), w(\theta))$ *and* $p^*(\cdot \mid z_i;\lambda_i)$ *is the uniquely optimal state-contingent plan of a type-*λ_i *firm contingent on realized state* z_i*. Firms aggregate misoptimization is their average mean-squared-error around the ex post optimal action, or*

$$m(\theta) := \mathbb{E}_{\theta_i,\lambda_i,v_i}\left[(x_i - x^*(z_i(\theta)))^2 \mid \theta\right] \tag{64}$$

We now prove the result, starting with the monotonicity of misoptimization. Using the result of Proposition 1, and the substitution of $(x^*(z_i), |\Pi_{xx}(z_i)|)$ as in the proof of Proposition 2, we show that the average "misoptimization variance" of actions conditional on (z_i, λ_i) is

$$m(z_i,\lambda_i,\theta) := \mathbb{E}_{v_i}\left[(x_i - x^*(z_i(\theta)))^2 \mid z_i,\lambda_i,\theta\right] = \frac{\lambda_i X(\theta)^{1+\gamma-\chi(1+\epsilon)}\theta_i^{1+\epsilon}}{v_\Pi(\epsilon,\chi,\bar{w},\bar{X})} \tag{65}$$

where $v_\Pi > 0$ is defined in Equation 48. Using the law of iterated expectations, we can write $m(\theta) = \mathbb{E}_{\theta_i,\lambda_i}[m(z_i,\lambda_i,\theta)]$. Assessing this outer expectation, we derive

$$m(\theta) = \frac{\lambda}{v_\Pi(\epsilon,\chi,\bar{w},\bar{X})} \cdot X(\theta)^{1+\gamma-\chi(1+\epsilon)} \cdot \mathbb{E}_{\theta_i}[\theta_i^{1+\epsilon} \mid \theta] \tag{66}$$

We observe that $\mathbb{E}_{\theta_i}[\theta_i^{1+\epsilon} \mid \theta]$ increases in θ, as $y \mapsto y^{1+\epsilon}$ is an increasing function and $\theta' > \theta \to G(\theta') \succsim_{FOSD} G(\theta)$. We next observe that $X(\theta)^{1+\gamma-\chi(1+\epsilon)}$ increases in X if $\gamma > \chi(1+\epsilon)-1$. This condition is guaranteed by $\gamma > \chi+1$ and $\chi\epsilon < 1$. Moreover, by Proposition 2, the stated conditions ensure that $X(\theta)$ is an increasing function. This proves that $m(\theta)$ increases in θ, or misoptimization is higher in the higher-productivity, higher-output state.

We next consider the monotonicity of attention. The entropy of a Gaussian random variable with variance σ^2 is proportional, up to scaling and constants, to $\log(\sigma^2)$. We therefore derive, up to scaling and constants,

$$a(\theta) = (-1 - \gamma + \chi(1+\epsilon)) \log X(\theta) - (1+\epsilon)\mathbb{E}_{\theta_i}[\log \theta_i \mid \theta] \tag{67}$$

This is monotone decreasing in X if $\gamma > \chi(1+\epsilon) - 1$, as desired. It is monotone decreasing in θ if $\mathbb{E}_{\theta_i}[\log \theta_i \mid \theta]$ increases θ. This is true because $y \mapsto \log y$ is an increasing function and $\theta' > \theta \to G(\theta') \succsim_{FOSD} G(\theta)$. Thus $a(\theta)$ is a decreasing function of θ, and attention is lower in the higher-productivity, higher-output states.

□

A.4 Proof of Proposition 4

Proof. We first derive output in the fully attentive $\lambda = 0$ limit, which we define by some mapping $X_0 : \Theta \to \mathbb{R}$. Recall the fixed-point equation for output from Proposition 2:

$$X(\theta) = a_0 X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} - a_1 X(\theta)^{\gamma-\chi}\theta \tag{68}$$

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

$$X_0(\theta) = a_0 X_0(\theta)^{1-\chi\epsilon}\theta^{\epsilon} \tag{69}$$

Or simply:

$$X_0(\theta) = a_0^{\frac{1}{\chi\epsilon}}\theta^{\frac{1}{\chi}} \tag{70}$$

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

$$W(\theta;\lambda) := \frac{X(\theta)}{X_0(\theta)} = \frac{X(\theta)}{a_0^{\frac{1}{\chi\epsilon}}\theta^{\frac{1}{\chi}}} \tag{71}$$

Via this definition, we re-write output in the form claimed in the Proposition.

$$\log X(\log\theta) = X_0 + \chi^{-1}\log\theta + \log W(\log\theta) \tag{72}$$

where $X_0 := \frac{1}{\chi\epsilon}\log a_0$.

We next prove that the wedge is positive. To prove this and other properties, we write a fixed-point equation for $W(\theta)$. Combining the definition of the wedge with Equation 68, we obtain

$$W(\theta) = W(\theta)^{1-\chi\epsilon} - a_1(\lambda) a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}} W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}} \tag{73}$$

Based on identical arguments to those in the proof of Proposition 2, the wedge is positive and unique under the exact same conditions that $X(\theta)$ is positive and unique: $\chi\epsilon < 1$ and $\gamma > \chi + 1$. Moreover, $W(\theta)$ crosses the 45 degree line from above. To show that $W(\theta) \leq 1$, it then suffices to show that the right-hand-side of the fixed point equation is less than unity when evaluated at $W(\theta) = 1$. As $a_1, a_0 > 0$, this is immediate. Thus $\log W(\theta) \leq 0$, as claimed. Moreover, given that $\frac{\partial a_1}{\partial \lambda} > 1$, it is immediate to show $\frac{\partial W}{\partial \lambda} < 0$.

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

$$\begin{aligned}\frac{\mathrm{d}W}{\mathrm{d}\theta} &= \left[(1-\chi\epsilon)W(\theta)^{-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}(\gamma-\chi)W(\theta)^{\gamma-\chi-1}\theta^{\frac{\gamma-1}{\chi}}\right]\frac{\mathrm{d}W}{\mathrm{d}\theta} \\ &\quad - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}\frac{\gamma-1}{\chi}W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-\chi-1}{\chi}}\end{aligned} \tag{74}$$

or:

$$\frac{\mathrm{d}W}{\mathrm{d}\theta} = \frac{-a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}\frac{\gamma-1}{\chi}W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-\chi-1}{\chi}}}{1-\left[(1-\chi\epsilon)W(\theta)^{-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}(\gamma-\chi)W(\theta)^{\gamma-\chi-1}\theta^{\frac{\gamma-1}{\chi}}\right]} \tag{75}$$

which we can rewrite as, after multiplying by $\frac{\theta}{W}$, as

$$\frac{\mathrm{d}\log W}{\mathrm{d}\log\theta} = \frac{-a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}\frac{\gamma-1}{\chi}W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}}}{W(\theta)-\left[(1-\chi\epsilon)W(\theta)^{1-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}(\gamma-\chi)W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}}\right]} \tag{76}$$

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

$$\begin{aligned}W(\theta) &= W(\theta)^{1-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}} \\ &\geq \underbrace{(1-\chi\epsilon)}_{\chi\epsilon<1}W(\theta)^{1-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}\underbrace{(\gamma-\chi)}_{\gamma>\chi+1}W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}}\end{aligned} \tag{77}$$

This completes the proof. □

A.5 Proof of Corollary 2

Proof. The labor demand of any given firm i is given by $L_i = \frac{x_i}{\theta_i}$. Total labor demand in the economy is then given by:

$$L = \int_{[0,1]} L_i \,\mathrm{d}i \tag{78}$$

Using Proposition 1, the definition of $x^*(z_i)$ in the proof of Proposition 2, and the equilibrium law of motion $X(\theta)$, we write the production of each firm as

$$x_i = x^*(z_i) + \tilde{v}_i = v_x X(\theta)\theta_i^{\epsilon} + \tilde{v}_i \tag{79}$$

where $\tilde{v}_i$ is the misoptimization scaled by its endogenous standard deviation. Plugging this into the expression for L, we derive

$$L = v_x X(\theta) \int_0^1 \theta_i^{\epsilon-1} \, \mathrm{d}i \tag{80}$$

Simplifying and applying a law of large numbers, we write this as

$$L_t = L(\theta) = v_x X(\theta)^{1-\chi\epsilon}\theta^{\epsilon-1} \tag{81}$$

where we define, as in the main text, $\theta := \left(\mathbb{E}_{\theta_i}[\theta_i^{\epsilon-1} \mid \theta_i]\right)^{\frac{1}{\epsilon-1}}$.

Combining the definition $\log A(\theta) = \log X(\theta) - \log L(\theta)$ with Equation 81, we derive

$$\log A(\theta) = -\log v_x + \chi\epsilon \log X(\theta) + (\epsilon - 1)\log\theta \tag{82}$$

Using our representation of aggregate output from Proposition 4, we obtain:

$$\log A(\theta) = (\chi\epsilon X_0 - \log v_x) + \log\theta + \chi\epsilon\log W(\theta) \tag{83}$$

where $W(\cdot)$ inherits all of the properties proved in Proposition 4. We finally observe that $X_0 = \frac{\log v_x}{\chi\epsilon}$, as defined in the proof of Proposition 4, so $\chi\epsilon X_0 - \log v_x = 0$. This completes the proof. $\square$

A.6 Proof of Corollary 3

Proof. From Proposition 4, the expression for output is

$$\log X(\log\theta) = X_0 + \chi^{-1}\log\theta + \log W(\log\theta) \tag{84}$$

Moreover, if $\lambda = 0$, then $\log W(\log\theta) = 0$.

First, consider the response of output to a small shock ν_t starting from θ_{t-1}. We have immediately that:

$$\left.\frac{\partial \log X(\log\theta)}{\partial \log\theta}\right|_{\theta=\theta_{t-1}} = \chi^{-1} + \left.\frac{\partial \log W(\log\theta)}{\partial\log\theta}\right|_{\theta=\theta_{t-1}} \tag{85}$$

When $\lambda = 0$, then $\log W \equiv 0$ according to Proposition 4. Thus $\frac{\partial \log X(\log\theta)}{\partial \log\theta} \equiv \chi^{-1}$ in all states, if inattention is removed from the model. If instead $\lambda > 0$ then $\frac{\partial \log W(\log\theta)}{\partial\log\theta}|_\theta$ is a non-linear function of θ which satisfies

$$\frac{\mathrm{d}\log W}{\mathrm{d}\log\theta} = \frac{-a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}} \frac{\gamma-1}{\chi} W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}}}{W(\theta) - \left[(1-\chi\epsilon)W(\theta)^{1-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}(\gamma-\chi)W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}}\right]} \tag{86}$$

Toward showing endogenous stochastic volatility, we approximate $\log X$ to the first order in the shock ν_t:

$$\log X(\nu_t;\theta_{t-1}) \approx \log X(\theta_{t-1}) + \left.\frac{\partial \log X(\log\theta)}{\partial\log\theta}\right|_{\theta=\theta_{t-1}} \nu_t \tag{87}$$

By the same logic above, this is state independent if $\lambda = 0$ and $\log W \equiv 0$. Taking the variance of this expression conditional on θ_{t-1} yields:

$$\mathrm{Var}[\log X_t \mid \theta_{t-1}] = \left(\chi^{-1} + \left.\frac{\partial \log W(\log\theta)}{\partial\log\theta}\right|_{\theta=\theta_{t-1}}\right)^2 \sigma_\theta^2 \tag{88}$$

Taking a second-order approximation in ν_t and differentiating yields the asymmetric shock propagation equation:

$$\left.\frac{\partial \log X(\log\theta)}{\partial\log\theta}\right|_{\theta=\theta_{t-1}} = \chi^{-1} + \frac{\partial \log W(\log\theta)}{\partial\log\theta}|_{\theta=\theta_{t-1}} + \left(\left.\frac{\partial^2 \log W(\log\theta)}{\partial\log\theta^2}\right|_{\theta=\theta_{t-1}}\right)\nu_t \tag{89}$$

Again, this is state independent if $\lambda = 0$ and $\log W \equiv 0$. □

A.7 Additional Calculations

A.7.1 Quadratic Approximation of Risk-Adjusted Profits

Using the expressions for dollar profits and marginal utility in Equation 11, we can write firms' risk-adjusted profits as the following:

$$\Pi(x, z_i) := X^{-\gamma}\left(x^{1-\frac{1}{\epsilon}}X^{\frac{1}{\epsilon}} - x\frac{w}{\theta_i}\right) \tag{90}$$

where, as throughout, we define the decision state vector $z_i = (\theta_i, X, w)$. The optimal action in the absence of stochastic choice solves the first-order condition

$$\left(1-\frac{1}{\epsilon}\right){x^*}^{-\frac{1}{\epsilon}}X^{\frac{1}{\epsilon}} = \frac{w}{\theta_i} \tag{91}$$

which can be re-arranged to define

$$x^*(z_i) = \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X \left(\frac{w}{\theta_i}\right)^{-\epsilon} \tag{92}$$

We now approximate the firm's profit function to second order in x around $x^*(z_i)$:

$$\begin{aligned}\Pi(x, z_i) &= \Pi(x^*(z_i), z_i) + \Pi_x(z_i)(x - x^*(z_i)) + \frac{1}{2}\Pi_{xx}(z_i)(x - x^*(z_i))^2 + O^3(x) \\ &=: \tilde{\Pi}(x, z_i) + O^3(x)\end{aligned} \tag{93}$$

where $\Pi_x(z_i) := \Pi_x(x, z_i)|_{x=x^*(z_i)}$ and $\Pi_{xx}(z_i) := \Pi_{xx}(x, z_i)|_{x=x^*(z_i)}$. By the envelope theorem, $\Pi_x(z_i) = 0$. Thus, our approximation reduces to the quadratic utility function in the Linear-Quadratic equilibrium:

$$\tilde{\Pi}(x, z_i) = \Pi(x^*(z_i), z_i) + \frac{1}{2}\Pi_{xx}(z_i)(x - x^*(z_i))^2 \tag{94}$$

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

$$\begin{aligned}\Pi(x, z_i) &= X^{-\gamma}\left(X\left(\frac{w}{\theta_i}\right)^{1-\epsilon}\right)\left(\left[1 - \frac{1}{\epsilon}\right]^{\epsilon\left(1-\frac{1}{\epsilon}\right)} - \left[1 - \frac{1}{\epsilon}\right]^{\epsilon}\right) \\ &= X^{-\gamma}\left(X\left(\frac{w}{\theta_i}\right)^{1-\epsilon}\right)\epsilon^{-\epsilon}\left(\epsilon - 1\right)^{\epsilon-1}\end{aligned} \tag{95}$$

We now characterize the curvature, which is the product of marginal utility with the curvature of the dollar-profit function:

$$\Pi_{xx}(z_i) = X^{-\gamma} \cdot \pi_{xx}(z_i) \tag{96}$$

We calculate, using the form of the profit function from Equation 11, the dollar profit function's second derivative:[56]

$$\begin{aligned}\pi_{xx}(x^*(z_i), X) &= -\frac{1}{\epsilon}\left(1 - \frac{1}{\epsilon}\right)(x^*(z_i))^{-1-\frac{1}{\epsilon}}X^{\frac{1}{\epsilon}} \\ &= -\frac{1}{\epsilon}\left(1 - \frac{1}{\epsilon}\right)\left(1 - \frac{1}{\epsilon}\right)^{-\left(1+\frac{1}{\epsilon}\right)\epsilon} X^{-1-\frac{1}{\epsilon}}X^{\frac{1}{\epsilon}}\left(\frac{w}{\theta_i}\right)^{\epsilon\left(1+\frac{1}{\epsilon}\right)} \\ &= -\epsilon^{\epsilon-1}(\epsilon - 1)^{-\epsilon}X^{-1}\left(\frac{w}{\theta_i}\right)^{1+\epsilon}\end{aligned} \tag{97}$$

[56]Because marginal costs are constant, this curvature arises purely from the curvature of the revenue function.

We substitute in the wage rule, Equation 7, to derive

$$\pi_{xx}(z_i) = -\upsilon_\pi(\epsilon, \chi, \bar{X}, \bar{w}) \cdot \theta_i^{-1-\epsilon} X^{\chi(1+\epsilon)-1} \tag{98}$$

as in Equation 18, where the constant is

$$\upsilon_\pi(\epsilon, \chi, \bar{X}, \bar{w}) := (\epsilon - 1)^{-\epsilon} \epsilon^{\epsilon - 1} \bar{w}^{1+\epsilon} \bar{X}^{-\chi(1+\epsilon)} > 0 \tag{99}$$

A.7.2 Quadratic Approximation of Final-Goods Technology

We now consider the second-order approximation of the aggregator, which is re-printed below

$$X(\{x_i\}_{i\in[0,1]}) = \left(\int_0^1 x_i^{\frac{\epsilon-1}{\epsilon}} \, \mathrm{d}i \right)^{\frac{\epsilon}{\epsilon-1}} \tag{100}$$

Technically speaking, we take a quadratic approximation of a discretized version of this aggregator, and then take consider the limit of this approximation. First, we suppose that there are $K \times K' \times K''$ discrete firms. Define the firm-level state for any firm $kk'k''$ as $\omega_{kk'k''} = (\theta_k, \lambda_{k'}, \upsilon_{k''})$ with corresponding production level $x(\omega_{kk'k''})$. Define the CES aggregator in this economy as:

$$X_{KK'K''}(\{x_{kk'k''}\}) = \left(\frac{1}{K} \sum_{k=1}^{K} \frac{1}{K'} \sum_{k'=1}^{K'} \frac{1}{K''} \sum_{k''=1}^{K''} x(\omega_{kk'k''})^{1-\frac{1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{101}$$

Second, we take a quadratic approximation of this function around the firm-level optimal production points $x_{kk'k''} = x^*(\theta_k)$:

$$\begin{aligned} X_{KK'K''} &= X_K^* + \frac{1}{K} \sum_{k=1}^{K} \frac{1}{K'} \sum_{k'=1}^{K'} \frac{1}{K''} \sum_{k''=1}^{K''} D_k (x(\omega_{kk'k''}) - x^*(\theta_k)) \\ &+ \frac{1}{K} \sum_{k=1}^{K} \frac{1}{K'} \sum_{k'=1}^{K'} \frac{1}{K''} \sum_{k''=1}^{K''} \frac{1}{K} \sum_{\tilde{k}=1}^{K} \frac{1}{K'} \sum_{\tilde{k}'=1}^{K'} \frac{1}{K''} \sum_{\tilde{k}''=1}^{K''} \frac{1}{2} D^2_{k\tilde{k}} (x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{\tilde{k}\tilde{k}'\tilde{k}''}) - x^*(\theta_{\tilde{k}})) \end{aligned} \tag{102}$$

where:

$$X_K^* = \left(\frac{1}{K} \sum_{k=1}^{K} x^*(\theta_k)^{1-\frac{1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{103}$$

and:

$$D_k = \frac{\partial X_K^*}{\partial x_{kk'k''}} = (X_K^*)^{\frac{1}{\epsilon}} x^*(\theta_k)^{-\frac{1}{\epsilon}} \tag{104}$$

and:

$$D^2_{k\tilde{k}} = \frac{\partial^2 X_K^*}{\partial x_{kk'k''}\partial x_{\tilde{k}\tilde{k}'\tilde{k}''}} = \begin{cases} -KK'K''\frac{1}{\epsilon}x^*(\theta_k)^{-\frac{1}{\epsilon}-1}(X_K^*)^{\frac{1}{\epsilon}} + \frac{1}{\epsilon}\frac{\partial X_K^*}{\partial x_{kk'k''}}(X_K^*)^{\frac{1}{\epsilon}-1}x^*(\theta_k)^{-\frac{1}{\epsilon}} & \text{if } kk'k'' = \tilde{k}\tilde{k}'\tilde{k}'' \\ \frac{1}{\epsilon}(X_K^*)^{\frac{1}{\epsilon}-1}(x^*(\theta_k))^{-\frac{1}{\epsilon}}(x^*(\theta_{\tilde{k}}))^{-\frac{1}{\epsilon}}D_K(z_{k'}, X_K^*) & \text{if } kk'k'' \neq \tilde{k}\tilde{k}'\tilde{k}''. \end{cases} \tag{105}$$

We now take limits of this approximation in the following order. We first send $K'' \to \infty$. Observe that, for fixed k, k', we have that each k'' firm by Proposition 1 has action distributed as $N(x^*(\theta_k), \sigma^2_{kk'})$. Thus, as $K'' \to \infty$, by the law of large numbers:

$$\frac{1}{K''}\sum_{k''=1}^{K''} D_k(x(\omega_{kk'k''}) - x^*(\theta_k)) \to^{a.s} 0 \tag{106}$$

Thus, the second term in the quadratic expansion above is zero almost surely in the large firm limit.

We can perform the same exercise for the third term in the quadratic expansion, which we can write as:

$$\begin{aligned} Q &= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{K'}\sum_{k'=1}^{K'}\frac{1}{K}\sum_{\tilde{k}=1}^{K}\frac{1}{K'}\sum_{\tilde{k}'=1}^{K'}\left(\frac{1}{K''^2}\sum_{k''=1}^{K''}\sum_{\tilde{k}''=1}^{K''}\frac{1}{2}D^2_{k\tilde{k}}(x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{\tilde{k}\tilde{k}'\tilde{k}''}) - x^*(\theta_{\tilde{k}}))\right) \\ &= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{K'}\sum_{k'=1}^{K'}\frac{1}{K}\sum_{\tilde{k}=1}^{K}\frac{1}{K'}\sum_{\tilde{k}'=1}^{K'}\frac{1}{2}D^2_{k\tilde{k}}\left(\frac{1}{K''^2}\sum_{k''=1}^{K''}\sum_{\tilde{k}''=1}^{K''}(x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{\tilde{k}\tilde{k}'\tilde{k}''}) - x^*(\theta_{\tilde{k}}))\right) \end{aligned} \tag{107}$$

Fix $k = \tilde{k}$, $k' = \tilde{k}'$ and consider the summation in brackets. This has two terms. First, for $\tilde{k}'' = k''$, the summand is simply $(x(\omega_{kk'k''}) - x^*(\theta_k))^2$. Second, $\tilde{k}'' \neq k''$, the summand is the product of two independent normal random variables with common distribution distribution $N(0, \sigma^2_{kk'})$. Thus, in the $K'' \to \infty$ limit we have that:

$$\frac{1}{K''^2}\sum_{k''=1}^{K''}\sum_{\tilde{k}''\neq k''}^{K''}(x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{kk'\tilde{k}''}) - x^*(\theta_k)) \to^{a.s} 0 \tag{108}$$

and:

$$\frac{1}{K''}\sum_{k''=1}^{K''}(x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{\tilde{k}\tilde{k}'k''}) - x^*(\theta_{\tilde{k}})) \to^{a.s} \sigma^2_{kk'} \tag{109}$$

Thus, we can simplify:

$$Q = -\frac{1}{2\epsilon}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{K'}\sum_{k'=1}^{K'}\frac{\sigma_{kk'}^2}{x^*(\theta_k)^{1+\frac{1}{\epsilon}}(X_K^*)^{-\frac{1}{\epsilon}}} \tag{110}$$

We now observe that $\sigma_{kk'}^2 = \frac{\lambda_{k'}}{\lambda}\sigma_k^2$. Thus, taking the $K' \to \infty$ limit we have that:

$$Q \to^{a.s.} -\frac{1}{2\epsilon}\frac{1}{K}\sum_{k=1}^{K}\frac{\lambda\sigma_k^2}{x^*(\theta_k)^{1+\frac{1}{\epsilon}}(X_K^*)^{-\frac{1}{\epsilon}}} \tag{111}$$

Now taking the limit as $K \to \infty$, we can express this as:

$$Q \to^{a.s} -\frac{1}{2\epsilon}\mathbb{E}\left[\frac{\lambda\sigma_k^2}{x^*(\theta_k)^{1+\frac{1}{\epsilon}}(X_K^*)^{-\frac{1}{\epsilon}}}\right] \tag{112}$$

Moreover, in the same limit, by applying the law of large numbers and the continuous mapping theorem we have that:

$$X_K^* \to^{a.s.} \left(\mathbb{E}[x^*(\theta_k)^{1-\frac{1}{\epsilon}}]\right)^{\frac{\epsilon}{\epsilon-1}} \tag{113}$$

Combining all of the above, we have shown that, in the limit, almost surely:

$$X \approx \left(\mathbb{E}[(x^*(\theta_k))^{1-\frac{1}{\epsilon}}]\right)^{\frac{\epsilon}{\epsilon-1}} - \frac{1}{2\epsilon}\mathbb{E}\left[\frac{\lambda\sigma_k^2}{x^*(\theta_k)^{1+\frac{1}{\epsilon}}(X_K^*)^{-\frac{1}{\epsilon}}}\right] \tag{114}$$

Which we denote by the (somewhat imprecise, but standard) integral form over agents:

$$X = \left(\int_0^1 x^*(z_i)^{1-\frac{1}{\epsilon}}\,\mathrm{d}i\right)^{\frac{\epsilon}{\epsilon-1}} - \frac{1}{2\epsilon}\int_0^1 \frac{(x_i - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}}(x^*(z_i))^{1+\frac{1}{\epsilon}}}\,\mathrm{d}i \tag{115}$$

A.7.3 Mapping Misoptimization Dispersion to the Model

Here, we explicitly calculate the within-model analogue to Misoptimization Dispersion. Recall from Definition 2 the definition of aggregate misoptimization,

$$m(\theta) := \mathbb{E}_{\theta_i,\lambda_i,v_i}\left[(x_i - x^*(z_i(\theta)))^2 \mid \theta\right] \tag{116}$$

which, as derived in the proof of Proposition 3, had expression

$$m(\theta) = \frac{\lambda}{v_\Pi(\epsilon, \chi, \bar{w}, \bar{X})} \cdot X(\theta)^{1+\gamma-\chi(1+\epsilon)} \cdot \mathbb{E}_{\theta_i}[\theta_i^{1+\epsilon} \mid \theta] \tag{117}$$

Misoptimization Dispersion is the optimal-sales-weighted population average of the *normalized* mean-squared error of actions. Let us define this model object as

$$\tilde{m}(\theta) = \mathbb{E}_{\theta_i,\lambda_i,v_i}\left[\hat{s}^*(\theta_i)\left(\frac{(x_i - x^*(z_i(\theta)))}{x^*(z_i)}\right)^2 \mid \theta\right] \tag{118}$$

where $s^*(\theta_i)$ are sales weights evaluated at the optimal production levels. We can use the model's structure to simplify these weights:

$$\hat{s}^*(\theta_i) := \frac{q^*(z_i)x^*(z_i)}{\mathbb{E}_{\theta_i}[q^*(z_i)x^*(z_i)]} = \frac{X^{\frac{1}{\epsilon}}(v_x X^{1-\chi\epsilon}\theta_i^\epsilon)^{1-\frac{1}{\epsilon}}}{\mathbb{E}_{\theta_i}[X^{\frac{1}{\epsilon}}(v_x X^{1-\chi\epsilon}\theta_i^\epsilon)^{1-\frac{1}{\epsilon}}]} = \frac{\theta_i^{\epsilon-1}}{\theta^{\epsilon-1}} \tag{119}$$

where, as throughout, $\theta = \left(\mathbb{E}_{\theta_i}[\theta_i^{\epsilon-1}]\right)^{\frac{1}{\epsilon-1}}$. We can therefore write the expected variance of normalized misoptimizations, conditioning on a specific firm, as

$$\tilde{m}(z_i,\lambda_i,\theta) := \mathbb{E}_{v_i}\left[s^*(\theta_i)\left(\frac{x_i - x^*(z_i(\theta))}{x^*(z_i)}\right)^2 \mid z_i,\lambda_i,\theta\right] = \frac{\lambda_i X(\theta)^{\gamma+\chi(\epsilon-1)-1}\theta^{\epsilon-1}}{v_\Pi v_x^2} \tag{120}$$

where $v_\Pi, v_X > 0$ are defined in Equation 48. It is trivial to integrate over (θ_i, λ_i) to derive

$$\tilde{m}(\theta) = \frac{\lambda X(\theta)^{\gamma+\chi(\epsilon-1)-1}\theta^{1-\epsilon}}{v_\Pi v_x^2} \tag{121}$$

We can relate this to $m(\theta)$ by writing

$$\frac{m(\theta)}{\tilde{m}(\theta)} = v_x^2\theta^{\epsilon-1}\mathbb{E}_{\theta_i}[\theta_i^{1+\epsilon} \mid \theta]X^{2(1-\chi\epsilon)} \tag{122}$$

See that, given $\epsilon > 1$ and $\chi\epsilon < 1$, this is an increasing function of both θ and X. Therefore, if $\tilde{m}(\theta)$ is monotone increasing in θ in an equilibrium with monotone $X(\theta)$, then $m(\theta)$ is also monotone increasing in θ.

Online Appendix

for "Attention Cycles" by Flynn and Sastry

Contents

B Extended Model

In this appendix, we formally develop the extension of the baseline model from Section 3 to include multiple inputs and market clearing wages. In the process, we will provide more direct model micro-foundations for the wage rule and the stock return regression analysis in Section 5.3.

B.1 Set-up

Time is discrete, and indexed by $t \in \mathbb{N}$. There are three kinds of firms: perfectly competitive materials firms who use labor to produce materials; intermediate goods producers who differ in their productivity who use labor and materials to produce a monopolistic variety indexed by $i \in [0, 1]$; and final goods firms who produce consumption goods as a constant elasticity of substitution (CES) aggregate of intermediate goods. There are two types of households: capitalists who own the firms in the economy, do not work and have constant relative risk aversion (CRRA) preferences over consumption; workers who supply labor, are hand-to-mouth (consuming all of their labor income in each period), and have Greenwood et al. (1988) (GHH) preferences over consumption and labor. Finally, as in our baseline model, the stochastic choice friction is embedded in the production of intermediate goods: intermediate goods producers perfectly cost-minimize but find it hard to produce the optimal amount.

B.1.1 Firms

Materials are produced by perfectly competitive firms with linear production technology in labor so that aggregate production of materials M_t is given by:

$$M_t = \theta_t^M L_t^M \tag{123}$$

where θ_t^M is the productivity of the materials sector and L_t^M is its labor input.

Intermediate goods producers of variety i are the monopoly producers of that variety. They have firm-specific productivity θ_{it} and use materials m_{it} and labor L_{it} to produce output x_{it} with Cobb-Douglas production technology:

$$x_{it} = \theta_{it} L_{it}^{\alpha} m_{it}^{1-\alpha} \tag{124}$$

where $\alpha \in (0, 1)$. To the extent that other intermediate goods (e.g. capital) exist and are combined in a CRS Cobb-Douglas production function with labor, this is fully general.

The stochastic process of productivity is exactly as described in Section 3.2. There is an aggregate productivity state $\theta_t \in \Theta$, which follows a first-order Markov process with transition density given by $h(\theta_t \mid \theta_{t-1})$. The cross-sectional productivity distribution is given in state Θ by the mapping $G : \Theta \to \Delta(\Theta)$, where we denote the productivity distribution in any state θ_t by $G_t = G(\theta_t)$ with corresponding density g_t. We assume that the total order on θ_t ranks distributions G_t by first-order stochastic dominance, or $\theta \geq \theta'$ implies $G(\theta) \succsim_{FOSD} G(\theta')$. Finally, materials productivity θ_t^M is determined as an increasing function of the overall productivity state θ_t.

Intermediate goods producers perfectly cost-minimize facing wages w_t and intermediate goods prices p_t^M. That is, for given production level x_{it}, they always choose the cost-minimizing input bundle. We define the firm-level *decision state* $z_{it} = (\theta_{it}, X_t, w_t, p_t^M) \in \mathcal{Z}$ as the concatenation of all decision-relevant variables that the firm takes as given; unlike in the baseline model, this definition includes the materials price. All firms believe that the vector z_{it} follows a first-order Markov process with transition densities described by $f : \mathcal{Z} \to \Delta(\mathcal{Z})$, with $f(z_{it}|z_{i,t-1})$ being the density of z_{it} conditional on last period's state being $z_{i,t-1}$. At time t, each firm i knows the sequence of previous $\{z_{is}\}_{s<t}$ but not the contemporaneous value z_{it}.

Given this firms have risk-adjusted profits given by $\Pi(x_{it}, z_{it})$. They then choose stochastic choice rules to maximize expected profits net of control costs, as captured by the following program which is identical to Equation 12 in the main model, with a different definition of the decision state and profits function:

$$\max_{p \in \mathcal{P}} \int_{\mathcal{Z}} \int_{\mathcal{X}} \Pi(x, z_{it})\, p(x \mid z_{it})\, \mathrm{d}x \;\, f(z_{it} \mid z_{i,t-1})\, \mathrm{d}z - c\left(p, \lambda_i, z_{i,t-1}, f\right) \tag{125}$$

Intermediate goods firms generate profits in units of consumption goods given by π_{it}. The firms store these consumption goods and pay them out as dividends d_{it} to their owners in the following period:

$$d_{it+1} = \pi_{it} \tag{126}$$

A unit supply of stock in the firm, which confers the right to the dividend stream, is available at price P_{it}.

The output of intermediate goods firms is combined to produce consumption goods with a CES production technology. Thus, if the intermediate producers produce $\{x_{it}\}_{i\in[0,1]}$, then the aggregate supply of consumption goods is:

$$X_t = X(\{x_{it}\}_{i\in[0,1]}) = \left(\int_{[0,1]} x_{it}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{127}$$

B.1.2 Households

There are two types of households: capitalists and workers. Capitalists own all firms in the economy and workers are hand-to-mouth. Capitalists have preferences over streams of consumption $\{C_{t+j}\}_{j\in\mathbb{N}}$ given by:

$$\mathcal{U}^C(\{C_{t+j}\}_{j\in\mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta_C^j \frac{C_{t+j}^{1-\gamma}}{1-\gamma} \tag{128}$$

where $\beta_C \in [0,1), \gamma \geq 0$. The dynamic budget constraint of capitalists is given by:

$$C_t + A_{t+1} + \int_{[0,1]} P_{it} S_{it+1} \, \mathrm{d}i \leq \int_{[0,1]} d_{it} S_{it} \, \mathrm{d}i + (1+r_t) A_t + \int_{[0,1]} P_{it} S_{it} \, \mathrm{d}i \tag{129}$$

where S_{it} is their stock-holding in intermediate firm i at time t and A_t is their bond-holding at time t.

Workers have preferences over streams of consumption and labor $\{C_{t+j}^W, L_{t+j}\}_{j\in\mathbb{N}}$ given by:

$$\mathcal{U}^W(\{C_{t+j}^W, L_{t+j}\}_{j\in\mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta_W^j U\left(C_t^W - \frac{L_{t+j}^{1+\psi}}{1+\psi}\right) \tag{130}$$

where $U' > 0, U'' < 0, \psi > 0, \beta^W \in [0,1)$. Workers are hand-to-mouth and they supply labor L_t at wage w_t, meaning that they consume:

$$C_t^W = w_t L_t \tag{131}$$

B.1.3 Equilibrium

An equilibrium is simply a set of all endogenous variables:

$$\{L_t^M, M_t, p_t^M, p_t^*, w_t, \{x_{it}, L_{it}, m_{it}, \pi_{it}, d_{it}, P_{it}, S_{it}\}_{i\in[0,1]}, X_t, L_t, C_t, A_t\} \tag{132}$$

such that all agents optimize as described above and markets clear given the exogenous process $\{\theta_t, \theta_t^M, \{\theta_{it}\}_{i\in[0,1]}\}_{t\in\mathbb{N}}$.

We will primarily be interested, as in the main text, in linear-quadratic equilibria where Π is approximated around its optimal level and the CES aggregator is approximated as described in Section 3.4.

B.2 Characterizing Equilibrium

We now reduce the description of equilibrium to a scalar fixed-point equation that can equivalently be formulated in terms of total production or capitalist consumption. This simplifies the analysis of the model and allows us to establish some equilibrium properties.

B.2.1 Production by Intermediate Goods Firms

Owing to CES aggregation, intermediate goods firms face the following iso-elastic demand curve:

$$q_{it} = X_t^{\frac{1}{\epsilon}} x_{it}^{-\frac{1}{\epsilon}} \tag{133}$$

They moreover perfectly cost-minimize. As a result, given production level x_{it} their unit input choices solve the following program:

$$\min_{L_{it}, m_{it}} w_t L_{it} + p_t^M m_{it} \quad \text{s.t.} \quad x_{it} = \theta_{it} L_{it}^{\alpha} m_{it}^{1-\alpha} \tag{134}$$

Taking the ratio of the two FOCs and rearranging:

$$m_{it} = \frac{1-\alpha}{\alpha} \frac{w_t}{p_t^M} L_{it} \tag{135}$$

Thus, given x_{it}, the optimal labor and materials choices are given by:

$$\begin{aligned} L_{it} &= \frac{1}{\theta_{it}} \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} \left(\frac{p_t^M}{w_t} \right)^{1-\alpha} x_{it} \\ m_{it} &= \frac{1}{\theta_{it}} \left(\frac{\alpha}{1-\alpha} \right)^{-\alpha} \left(\frac{p_t^M}{w_t} \right)^{-\alpha} x_{it} \end{aligned} \tag{136}$$

It follows that the cost of producing x_{it} is given by:

$$w_t L_{it} + p_t^M m_{it} = \frac{q_t}{\theta_{it}} x_{it} \tag{137}$$

where we define the unit marginal cost up to constant $c_\alpha > 0$:

$$q_t := \left[\left(\frac{\alpha}{1-\alpha} \right)^{-\alpha} \frac{1}{1-\alpha} \right] w_t^{\alpha} (p_t^M)^{1-\alpha} = c_\alpha w_t^{\alpha} (p_t^M)^{1-\alpha} \tag{138}$$

We now turn to solving the firm's stochastic choice problem. From the above, firm dollar profits are given by:

$$\pi_{it} = X_t^{\frac{1}{\epsilon}} x_{it}^{1-\frac{1}{\epsilon}} - \frac{q_t}{\theta_{it}} x_{it} \tag{139}$$

Recall that this is paid out as a dividend at period $t+1$, $d_{it+1} = \pi_{it}$. Note moreover in equilibrium by market clearing that $A_t = 0$ and $S_{it} = 1$ for all $i \in [0,1]$ and $t \in \mathbb{N}$. Thus, $C_{t+1} = \int_{[0,1]} d_{it+1}\,\mathrm{d}i = \int_{[0,1]} \pi_{it}\,\mathrm{d}i$. The firm's risk-adjusted profit is then given by:

$$\Pi_{it} = C_{t+1}^{-\gamma} \pi_{it} \tag{140}$$

where the firm takes C_{t+1} as given.

As in the main text, we define the optimal production level $x^*(\Lambda, \theta)$ which solves:

$$x^*(\Lambda, \theta_{it}) := \arg\max_{x \in \mathcal{X}} \; \Pi(x; \Lambda, \theta_{it}) \tag{141}$$

and $\bar{\Pi}(\Lambda, \theta_{it})$ as the maximized objective. Now let $\Pi_{xx}(\Lambda, \theta_{it})$ denote the second derivative of the profits function in x, evaluated at x^*:

$$\Pi_{xx}(\Lambda, \theta_{it}) := \left.\frac{\partial^2 \Pi}{\partial x^2}\right|_{x^*(\Lambda, \theta_{it}); \Lambda, \theta} \tag{142}$$

The approximate objective of the intermediate goods firm is:

$$\tilde{\Pi}(x; \Lambda, \theta_{it}) := \bar{\Pi}(\Lambda, \theta_{it}) + \frac{1}{2} \Pi_{xx}(\Lambda, \theta_{it})(x - x^*(\Lambda, \theta_{it}))^2 \tag{143}$$

Under this approximate objective, it follows by a slight algebraic variation of the arguments in Proposition 1 that optimal choices follow:

$$x_{it} \sim N\left(x^*_{it}, \frac{\lambda}{|\Pi_{xx,it}|}\right) \tag{144}$$

where:

$$\begin{aligned} x^*_{it} &= \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X_t \theta_{it}^{\epsilon} q_t^{-\epsilon} \\ |\Pi_{xx,it}| &= (\epsilon - 1)^{-\epsilon} \epsilon^{\epsilon - 1} C_{t+1}^{-\gamma} X_t^{-1} \theta_{it}^{-1-\epsilon} q_t^{1+\epsilon} \end{aligned} \tag{145}$$

These expressions mirror those in the main model, with $C_{t+1}^{-\gamma}$ replacing $X_t^{-\gamma}$ as the marginal utility and q_t replacing w_t as the marginal cost.

B.2.2 Finding Materials Prices and Wages

Materials producers maximize profits:

$$p_t^M \theta_t^M L_t^M - w_t L_t^M \tag{146}$$

Thus, in equilibrium, it follows that:

$$p_t^M = \frac{1}{\theta_t^M} w_t \tag{147}$$

The workers' labor supply condition is given by their Euler equation:

$$w_t = L_t^\psi \tag{148}$$

Moreover, we know that aggregate labor is equal to the sum of labor used to produce intermediates and materials:

$$L_t = \int_{[0,1]} L_{it} \mathrm{d}i + L_t^M = \int_{[0,1]} L_{it} \mathrm{d}i + \frac{1}{\theta_t^M} \int_{[0,1]} m_{it} \mathrm{d}i \tag{149}$$

where the second equality follows by market clearing for intermediates as $\int_{[0,1]} m_{it} \mathrm{d}i = M_t = \theta_t^M L_t^M$. We next substitute our expression of materials demand as a function of labor demand for intermediate goods firms to simplify the labor supply condition further:

$$\begin{aligned} L_t &= \int_{[0,1]} L_{it} di + \frac{1}{\theta_t^M} \int_{[0,1]} m_{it} di \\ &= \int_{[0,1]} L_{it}\, \mathrm{d}i + \frac{1}{\theta_t^M} \int_{[0,1]} \frac{1-\alpha}{\alpha} \frac{w_t}{p_t^M} L_{it}\, \mathrm{d}i \\ &= \left(1 + \frac{1}{\theta_t^M} \frac{1-\alpha}{\alpha} \frac{w_t}{p_t^M}\right) \int_{[0,1]} L_{it}\, \mathrm{d}i \end{aligned} \tag{150}$$

We finally substitute in the fact that the material input is priced at marginal cost to simplify further

$$\begin{aligned} L_t &= \left(1 + \frac{1-\alpha}{\alpha}\right) \int_{[0,1]} L_{it} \mathrm{d}i \\ &= \frac{1}{\alpha} \int_{[0,1]} L_{it} \mathrm{d}i \end{aligned} \tag{151}$$

We now write this in terms of prices and output choices by substituting in, from the intermediate goods firm's cost-minimization, $L_{it} = \frac{1}{\theta_{it}} \left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha} \left(\frac{p_t^M}{w_t}\right)^{1-\alpha} x_{it}$. Thus:

$$L_t = \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha} \left(\frac{p_t^M}{w_t}\right)^{1-\alpha} \int_{[0,1]} \frac{x_{it}}{\theta_{it}} \mathrm{d}i \tag{152}$$

We can then use our earlier characterization of the solution to the intermediate goods producers' stochastic choice problem to compute:

$$\begin{aligned}
\int_{[0,1]} \frac{x_{it}}{\theta_{it}} \mathrm{d}i &= \mathbb{E}\left[\frac{x_{it}}{\theta_{it}}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{x_{it}}{\theta_{it}} | \theta_{it}\right]\right] \\
&= \mathbb{E}\left[\frac{1}{\theta_{it}} x_{it}^*\right] \\
&= \mathbb{E}\left[\frac{1}{\theta_{it}} \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X_t q_t^{-\epsilon} \theta_{it}^{\epsilon}\right] \\
&= \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X_t q_t^{-\epsilon} \theta^{\epsilon-1}
\end{aligned} \tag{153}$$

where we use the definition $\theta = \left(\mathbb{E}_{\theta_i}[\theta_i^{\epsilon-1} \mid \theta]\right)^{\frac{1}{\epsilon-1}}$. By combining the previous two equations, we derive that total labor demand is given by:

$$L_t = \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha} \left(\frac{p_t^M}{w_t}\right)^{1-\alpha} \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X_t q_t^{-\epsilon} \theta^{\epsilon-1} \tag{154}$$

Substituting this into the workers' intratemporal Euler equation, and using Equation 138 to write the marginal cost in terms of materials prices and wages, we obtain:

$$\begin{aligned}
w_t^{\frac{1}{\psi}} &= \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha} \left(\frac{1}{\theta_t^M}\right)^{1-\alpha} \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} \theta_t^{\epsilon-1} c_\alpha^{-\epsilon} X_t \left(w_t \left(\frac{p_t^M}{w_t}\right)^{1-\alpha}\right)^{-\epsilon} \\
&= \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha} \left(\theta_t^M\right)^{(\epsilon-1)(1-\alpha)} \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} \theta^{\epsilon-1} c_\alpha^{-\epsilon} X_t w_t^{-\epsilon}
\end{aligned} \tag{155}$$

Moreover, we can write the marginal cost for the firm as

$$q_t = c_\alpha w_t^\alpha (p_t^M)^{1-\alpha} = \bar{q}_t X_t^\chi \tag{156}$$

where we define coefficient $\chi = \frac{\psi}{1+\epsilon\psi}$ and intercept

$$\bar{q}_t = \left[\left(\frac{\alpha}{1-\alpha}\right)^{-\alpha}\frac{1}{1-\alpha}\right]^{1-\alpha}\left(\frac{1}{\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}\left(\theta_t^M\right)^{(\epsilon-1)(1-\alpha)}\left(1-\frac{1}{\epsilon}\right)^{\epsilon}\theta^{\epsilon-1}c_\alpha^{-\epsilon}\right)^{\chi}(\theta_t^M)^{-1+\alpha} \tag{157}$$

Marginal costs, holding fixed productivity, increase in output due to upward-sloping labor supply or convex disutilty of effort. The intercept of this "cost rule" varies as a function of productivity in the intermediate-goods and materials sectors. Observe that Equation 156 is the "fully Neoclassical" analogue to our wage rule Equation 7; indeed, when $\alpha = 1$ or there is no materials factor, it reduces to a wage rule

$$w_t = \bar{w}_t X_t^{\alpha} \tag{158}$$

where $\bar{w}_t = \bar{q}_t|_{\alpha=1}$. This verifies our claim in the main text that the wage rule can be (essentially) micro-founded in the simple model. Indeed, in a model with materials or $\alpha < 1$, we obtain exactly the wage rule studied in the main text if $\theta_t^M = \theta_t^{\beta}$ where $\beta = \frac{\chi(\epsilon-1)}{(1-\chi(\epsilon-1))(1-\alpha)} > 0$, thereby canceling out the direct effect of productivity on the intercept of the wage rule.

B.2.3 Finding Equilibrium Output

We have characterized all endogenous objects in period t in terms of output X_t and capitalists' consumption C_{t+1}. It remains only to characterize these variables.

To this end, we approximate as we have throughout:

$$X = \left(\int_0^1 x^*(z_i)^{1-\frac{1}{\epsilon}}\,\mathrm{d}i\right)^{\frac{\epsilon}{\epsilon-1}} - \frac{1}{2\epsilon}\int_0^1 \frac{(x_i - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}}(x^*(z_i))^{1+\frac{1}{\epsilon}}}\,\mathrm{d}i \tag{159}$$

where the mean and variance are taken over the realizations of θ_{it}, conditional on the aggregate state θ_t. Substituting in Equation 145, this provides one equation in terms of (X_t, C_{t+1}). Consider first the computation of X_t^*. Observe that we can write:

$$x_{it}^* = \delta_t X_t^{1-\chi\epsilon}\theta_{it}^{\epsilon} \tag{160}$$

where:

$$\delta_t = \left(1-\frac{1}{\epsilon}\right)^{\epsilon}\left(c_\alpha \bar{w}_t\left(\frac{1}{\theta_t^M}\right)^{1-\alpha}\right)^{-\epsilon} \tag{161}$$

Substituting this into the expression for X_t^*, we obtain:

$$X_t^* = \left(\delta_t^{\frac{\epsilon-1}{\epsilon}} X_t^{\frac{\epsilon-1}{\epsilon}(1-\chi\epsilon)} \theta_t^{\epsilon-1}\right)^{\frac{\epsilon}{\epsilon-1}} = \delta_t X_t^{1-\chi\epsilon} \theta_t^{\epsilon} \tag{162}$$

Now consider the computation of the dispersion term. See that we can write:

$$\begin{aligned}
\mathbb{E}\left[\frac{(x_{it} - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}}(x^*(z_i))^{1+\frac{1}{\epsilon}}}\right] &= (X_t^*)^{\frac{1}{\epsilon}} \mathbb{E}\left[x^*(z_{it})^{-1-\frac{1}{\epsilon}} \mathbb{E}[(x_{it} - x^*(z_{it}))^2 | z_{it}]\right] \\
&= (X_t^*)^{\frac{1}{\epsilon}} \mathbb{E}\left[x^*(z_{it})^{-1-\frac{1}{\epsilon}} \frac{\lambda}{|\Pi_{xx,it}|}\right]
\end{aligned} \tag{163}$$

To simplify this, observe that we can write:

$$\frac{1}{|\Pi_{xx,it}|} = \zeta_t^{-1} C_{t+1}^{\gamma} X_t^{1-\chi(\epsilon+1)} \theta_{it}^{1+\epsilon} \tag{164}$$

where:

$$\zeta_t = (\epsilon - 1)^{-\epsilon} \epsilon^{\epsilon-1} \left(c_\alpha \bar{w}_t \left(\frac{1}{\theta_t^M}\right)^{1-\alpha}\right)^{1+\epsilon} \tag{165}$$

So we may express:

$$\begin{aligned}
\mathbb{E}\left[\frac{(x_{it} - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}}(x^*(z_i))^{1+\frac{1}{\epsilon}}}\right] &= (X_t^*)^{\frac{1}{\epsilon}} \mathbb{E}\left[\delta_t^{-1-\frac{1}{\epsilon}} X_t^{-(1+\frac{1}{\epsilon})(1-\chi\epsilon)} \theta_{it}^{-1-\epsilon} \lambda \zeta_t^{-1} C_{t+1}^{\gamma} X_t^{1-\chi(\epsilon+1)} \theta_{it}^{1+\epsilon}\right] \\
&= \lambda \zeta_t^{-1} \delta_t^{-1} C_{t+1}^{\gamma} X_t^{-\chi}
\end{aligned} \tag{166}$$

Putting all of the above together we have that:

$$X_t = \delta_t X_t^{1-\chi\epsilon} \theta_t^{\epsilon} - \frac{\lambda}{2\epsilon} \zeta_t^{-1} \delta_t^{-1} C_{t+1}^{\gamma} X_t^{-\chi} \theta_t \tag{167}$$

The final equation we require comes from equating capitalists' consumption with the previous period's dividends, which is implied by market clearing in the securities market and the fact that workers are hand-to-mouth. Thus:

$$C_{t+1} = \int_{[0,1]} \pi_{it} \, \mathrm{d}i \tag{168}$$

Using our running approximation:

$$\pi_{it} = \pi_{it}(x_{it}^*) + \frac{1}{2} \pi_{xx,it} (x_{it} - x_{it}^*)^2 \tag{169}$$

we obtain:

$$
\begin{aligned}
C_{t+1} &= \int_{[0,1]} \pi_{it} di \\
&= \mathbb{E}\left[\mathbb{E}\left[\pi_{it}(x_{it}^*) + \frac{1}{2}\pi_{xx,it}(x_{it}^*)(x_{it} - x_{it}^*)^2 \mid x_{it}^*\right] \mid \delta_t, \theta_t\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\pi_{it}(x_{it}^*) + \frac{1}{2}\pi_{xx,it}(x_{it}^*)\frac{\lambda}{|\Pi_{xx,it}|} \mid x_{it}^*\right] \mid \delta_t, \theta_t\right] \\
&= \mathbb{E}\left[\pi_{it}(x_{it}^*) \mid \delta_t, \theta_t\right] - \frac{\lambda}{2}C_{t+1}^{\gamma} \\
&= (\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}X_t^{1-\chi(\epsilon-1)} - \frac{\lambda}{2}C_{t+1}^{\gamma}
\end{aligned}
\tag{170}
$$

We can therefore solve for X_t as a function of C_{t+1}:

$$
X_t = \left(\frac{C_{t+1} + \frac{\lambda}{2}C_{t+1}^{\gamma}}{(\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}}\right)^{\frac{1}{1-\chi(\epsilon-1)}} \tag{171}
$$

Plugging this into the scalar fixed point equation for output then boils down the equilibrium of the model to a scalar fixed-point equation for the consumption of capitalists:

$$
\begin{aligned}
\left(\frac{C_{t+1} + \frac{\lambda}{2}C_{t+1}^{\gamma}}{(\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}}\right)^{\frac{1}{1-\chi(\epsilon-1)}} &= \delta_t \left(\frac{C_{t+1} + \frac{\lambda}{2}C_{t+1}^{\gamma}}{(\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}}\right)^{\frac{1-\chi\epsilon}{1-\chi(\epsilon-1)}} \theta_t^{\epsilon} \\
&\quad - \frac{\lambda}{2\epsilon}\zeta_t^{-1}\delta_t^{-1}C_{t+1}^{\gamma}\left(\frac{C_{t+1} + \frac{\lambda}{2}C_{t+1}^{\gamma}}{(\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}}\right)^{\frac{-\chi}{1-\chi(\epsilon-1)}} \theta_t
\end{aligned}
\tag{172}
$$

The above can be summarized in the following result:

Proposition 5. *Equilibria of the model are characterized by the solutions to Equation 172.*

B.3 Existence, Uniqueness, and Monotonicity of Equilibrium

To establish existence of equilibrium, all we require is that the above equation has a solution. As there is always a trivial equilibrium with $C_{t+1} = 0$, we will focus on when there exists an equilibrium with positive output, when it is unique, and when it is monotone. In this more general setting, we show that so long as cognitive frictions are not too large, these properties apply.

Proposition 6. *Suppose $\chi(\epsilon - 1) < 1$. There exists $\bar{\lambda} > 0$ such that there exists a unique equilibrium with positive output whenever $\lambda < \bar{\lambda}$. Moreover, equilibrium output is monotone increasing in aggregate productivity θ.*

Proof. Following Equation 172, define:

$$
\begin{aligned}
g_\lambda(C) = \left(\frac{C + \frac{\lambda}{2}C^\gamma}{(\epsilon-1)^{-1}\delta\theta^{\epsilon-1}}\right)^{\frac{1}{1-\chi(\epsilon-1)}} &- \left[\delta\left(\frac{C + \frac{\lambda}{2}C^\gamma}{(\epsilon-1)^{-1}\delta\theta^{\epsilon-1}}\right)^{\frac{1-\chi\epsilon}{1-\chi(\epsilon-1)}}\theta^\epsilon\right. \\
&\left. - \frac{\lambda}{2\epsilon}\zeta^{-1}\delta^{-1}C^\gamma\left(\frac{C + \frac{\lambda}{2}C^\gamma}{(\epsilon-1)^{-1}\delta\theta^{\epsilon-1}}\right)^{\frac{-\chi}{1-\chi(\epsilon-1)}}\theta\right]
\end{aligned} \tag{173}
$$

Observe that g_λ is continuous in λ in the sup-norm. Thus, if we can show that there is a unique value of $C \in \mathbb{R}_{++}$ such that $g_0(C) = 0$ and $g_0'(C) \neq 0$, then there exists λ such that for all $\lambda < \bar{\lambda}$ there will be a unique $C' \in \mathbb{R}_{++}$ such that $g_\lambda(C') = 0$.

To prove the result, it remains to show that there is a unique value of $C \in \mathbb{R}_{++}$ such that $g_0(C) = 0$ and $g_0'(C) \neq 0$ when $\chi(\epsilon-1) < 1$. To this end, define $\tilde{C} = \frac{C}{(\epsilon-1)^{-1}\delta\theta^{\epsilon-1}}$ and see that:

$$
g_0(\tilde{C}) = \tilde{C}^{\frac{1}{1-\chi(\epsilon-1)}} - b_{\tilde{C}}\tilde{C}^{\frac{1-\chi\epsilon}{1-\chi(\epsilon-1)}} \tag{174}
$$

where $b_{\tilde{C}} = \delta\theta^\epsilon$. We can then compute:

$$
\begin{aligned}
g_0'(\tilde{C}) &= \frac{1}{1-\chi(\epsilon-1)}\tilde{C}^{\frac{\chi(\epsilon-1)}{1-\chi(\epsilon-1)}} - b_{\tilde{C}}\frac{1-\chi\epsilon}{1-\chi(\epsilon-1)}\tilde{C}^{\frac{-\chi}{1-\chi(\epsilon-1)}} \\
g_0''(\tilde{C}) &= \frac{\chi(\epsilon-1)}{(1-\chi(\epsilon-1))^2}\tilde{C}^{\frac{\chi(\epsilon-1)}{1-\chi(\epsilon-1)}-1} + b_{\tilde{C}}\frac{\chi(1-\chi\epsilon)}{(1-\chi(\epsilon-1))^2}\tilde{C}^{\frac{-\chi}{1-\chi(\epsilon-1)}-1}
\end{aligned} \tag{175}
$$

From which we observe the following when $\chi(\epsilon-1) < 1$:

$$
\lim_{\tilde{C}\to 0} g_0'(\tilde{C}) = -\infty \qquad \lim_{\tilde{C}\to\infty} g_0'(\tilde{C}) = \infty \qquad g_0''(\tilde{C}) > 0 \quad \text{for all}\, \tilde{C} \in \mathbb{R}_{++} \tag{176}
$$

We now establish monotonicity. If we can show that the unique value of $C \in \mathbb{R}_{++}$ such that $g_0(C) = 0$ and $g_0'(C) \neq 0$ is monotone increasing in θ, then there exists $\bar{\lambda}$ such that for all $\lambda < \bar{\lambda}$ the same will be true of the unique $C' \in \mathbb{R}_{++}$ such that $g_\lambda(C') = 0$.

To this end, see that the solution when $\lambda = 0$ is given by:

$$
\ln\tilde{C} = \frac{(1-\chi(\epsilon-1))}{\chi\epsilon}\ln b_{\tilde{C}} \tag{177}
$$

We also know that $\ln\tilde{C} = \ln C + \ln(\epsilon-1) - \ln\delta - (\epsilon-1)\ln\theta$. Thus, we have that:

$$
\ln C = \frac{(1-\chi(\epsilon-1))}{\chi\epsilon}(\ln\delta + \epsilon\ln\theta) - \ln(\epsilon-1) + \ln\delta + (\epsilon-1)\ln\theta \tag{178}
$$

As $\epsilon > 1$ and $1 > \chi(\epsilon-1)$ by hypothesis, and δ is increasing in θ, the result follows. $\square$

B.4 Attention and Misoptimization Cycles in the Extended Model

Having shown that equilibrium output is monotone and increasing in the extended model, we now provide conditions under which the analogue of Proposition 3 that establishes monotonicity of attention and mistakes holds in this setting:

Proposition 7. *Assume $\chi\epsilon < 1$ and $1 > \chi(\epsilon + 1)$. There exists $\bar{\lambda}$ such that when $\lambda < \bar{\lambda}$, intermediate goods firms pay more attention and misoptimize less in lower-productivity, lower-output states.*

Proof. Recall that:

$$m(z_{it}) = \frac{\lambda_i}{|\Pi_{xx,it}|} = \lambda_i \zeta_t^{-1} C_{t+1}^{\gamma} X_t^{1-\chi(\epsilon+1)} \theta_{it}^{1+\epsilon} \tag{179}$$

Thus, the average extent of misoptimization in aggregate state θ is:

$$m(\theta) = \lambda \zeta(\theta)^{-1} C(\theta)^{\gamma} X(\theta)^{1-\chi(\epsilon+1)} \mathbb{E}[\theta_{it}^{1+\epsilon} \mid \theta] \tag{180}$$

See that $1 > \chi(\epsilon + 1)$ implies $1 > \chi(\epsilon - 1)$. As we assumed $\chi\epsilon < 1$, Proposition 6 implies that C and X are both increasing in θ. By the assumed FOSD ordering on θ, we have that $\mathbb{E}[\theta_{it}^{1+\epsilon} \mid \theta]$ is monotone increasing in θ. We moreover have that $\xi \propto \delta^{-\frac{1+\epsilon}{\epsilon}}$. Thus, as δ is increasing in θ, we have that ξ^{-1} is increasing in θ. This establishes that $m(\theta)$ is increasing in θ, and therefore that intermediate goods firms misoptimize less in lower productivity and lower output states. By the same arguments as in Proposition 3, it is immediate that the opposite pattern holds for attention. □

B.5 Macroeconomic Dynamics in the Extended Model

We can moreover derive an analogous representation of the impact of inattention on macroeconomic dynamics through an attention wedge that depresses output relative to the fully-attentive benchmark. Formally:

Proposition 8. *Output can be written in the following way:*

$$\log X(\log \theta, \lambda) = \frac{1}{\chi} \log \tilde{\theta} + \log W(\log \theta, \lambda) \tag{181}$$

where $\tilde{\theta} = \theta \delta^{\frac{1}{\epsilon}}$ and $\log W(\log \theta, 0) = 0$ for all $\theta \in \Theta$.

Proof. The representation follows immediately by combining Equations 177 and 171. That the wedge is 0 when $\lambda = 0$ follows immediately from the same equations. □

This formula differs from Proposition 4 only in so far as θ is replaced by $\tilde{\theta}$ which captures the effect of the inclusion of other factors of production and the endogenous labor supply of agents. Note that this result does not establish any properties of the wedge in this case, as the fixed point equation is challenging to manipulate. The nature of the wedge is then a quantitative question. Similarly to the main text, a concave attention wedge implies higher shock responsiveness in low states, greater responsiveness to negative than positive shocks, and volatility of output that is greater in low states.

B.6 Micro-foundation and Interpretation of the Stock Return Regressions

In Section 5.3, we showed that mistakes of the same size by firms lead to more adverse impacts on stock returns when the aggregate stock market return is low. We interpreted this as direct evidence in favor of our mechanism that risk-pricing is a key determinant of attention cycles. The simple model of Section 3 is too stylized to formally map to this regression. However, in the extended model developed in this section, we can derive exactly the regression we run from the theory and show how the estimated regression coefficients map to the risk-pricing channel in the theory.

First, from the Euler equation of capitalists, the equilibrium price of firm i at time t, P_{it} is given by:

$$u'(C_t)P_{it} = \mathbb{E}_t[\beta u'(C_{t+1})(P_{it+1} + d_{it+1})] \tag{182}$$

where $d_{it+1} = \pi_{it}$. Thus we may write:

$$u'(\pi_{t-1})P_{it} = \beta u'(\pi_t)\pi_{it} + \beta u'(\pi_t)\mathbb{E}_t[P_{it+1}] \tag{183}$$

where $\pi_t = \int_{[0,1]} \pi_{it}\, \mathrm{d}i$. It follows that:

$$P_{it} = \beta \frac{u'(\pi_t)}{u'(\pi_{t-1})}\pi_{it} + \beta \frac{u'(\pi_t)}{u'(\pi_{t-1})}\mathbb{E}_t[P_{it+1}] \tag{184}$$

A production mistake $m_{it} \equiv x_{it} - x_{it}^*$ leads to profits (under our running quadratic approximation) of:

$$\pi_{it} = \pi_{it}(x_{it}^*) + \pi_{xx,it}m_{it}^2 \tag{185}$$

Thus, the firm's stock price follows:

$$P_{it} = \beta \frac{u'(\pi_t)}{u'(\pi_{t-1})}\left(\pi_{it}(x_{it}^*) + \pi_{xx,it}m_{it}^2\right) + \beta \frac{u'(\pi_t)}{u'(\pi_{t-1})}\mathbb{E}_t[P_{it+1}] \tag{186}$$

Thus:

$$\frac{\partial P_{it}}{\partial m_{it}^2} = \beta \frac{u'(\pi_t)}{u'(\pi_{t-1})} \pi_{xx,it} \tag{187}$$

and:

$$\frac{\partial P_{it}}{\partial \pi_{it}} = \beta \frac{u'(\pi_t)}{u'(\pi_{t-1})} \tag{188}$$

To simplify this further, observe by the Euler equation for trading an equally weighted portfolio of all intermediate goods firms must satisfy, where P_t is the price of this portfolio (the stock market):

$$u'(C_t)P_t = \mathbb{E}_t[\beta u'(C_{t+1})(P_{t+1} + \pi_t)] \tag{189}$$

Or:

$$\beta \frac{u'(\pi_t)}{u'(\pi_{t-1})} = \frac{P_t}{\mathbb{E}_t[P_{t+1}] + D_{t+1}} = \frac{1}{R_t} \tag{190}$$

Which is the inverse aggregate return on equity between period t and period $t+1$, R_t. We therefore have that:

Proposition 9. *The equilibrium effect of mistakes on stock returns is given by:*

$$\frac{\partial P_{it}}{\partial m_{it}^2} = -\frac{1}{R_t} |\pi_{xx,it}| \tag{191}$$

If a mistake is measured in terms of its impact in profit units, then one obtains the simpler:

$$\frac{\partial P_{it}}{\partial \pi_{it}} = -\frac{1}{R_t} \tag{192}$$

Proof. Given in the text above. □

Of course, it is trivial to reformulate the above comparative statics in terms of firm level returns as P_{it-1} is invariant to innovations in m_{it}.

When equity returns are high, mistakes should (all else equal) have a lower price impact. Mapping this slightly more formally to our exact regression analysis: when we instrument for profits with mistakes, we should obtain a negative and significant coefficient on the interaction between profits and the aggregate stock market return. This is exactly what we find. The OLS regressions of returns on mistakes retain a similar structure but are intermediated by the curvature of dollar profits across firms. These regression models therefore provide a less sharp test of the risk-pricing channel, although empirically they produce entirely consistent results.

C Alternative Specifications of Stochastic Choice

In this section, we extend our basic class of cost function to allow for persistent mistakes as in the empirical analysis. In particular, we micro-found the AR(1) structure of mistakes that we uncovered in the data but abstracted from in the simple model. Further, we show how the core logic of attention cycles carries over to settings with alternative foundations for stochastic choice in terms of information acquisition of two forms: Gaussian signal extraction, and optimal signal processing with mutual information costs. Concretely, we will provide simple sufficient conditions under which an increase in the stakes of making mistakes—the curvature of intermediate goods firms' profits—leads to increased attention and smaller mistakes. In our baseline model, this corresponds to firms making smaller mistakes when market risk pricing is more severe and leads to an attention cycle.

C.1 Persistent Mistakes

In our empirical work we showed that firms' mistakes are persistent. The basic model we have developed, however, places no restrictions on the auto-correlation of mistakes across time within a firm. In this section, we introduce a more general class of cost functional that allows us to place restrictions on the within-firm correlation of mistakes across time and, in particular, to derive the AR(1) formulation of mistakes that we use in the empirical analysis.

We have so far considered likelihood-separable cost functions $c : \mathcal{P} \times \mathcal{Z} \to \mathbb{R}$ of the form:

$$c(p; \theta_{t-1}) = \int_{\Theta} \int_{\mathcal{X}} \phi(p(x|\theta)) \, \mathrm{d}x \, f(z|z_{t-1}) \, \mathrm{d}z \tag{193}$$

for some convex ϕ that we take to be $\phi(y) = y \log y$. To allow for persistent mistakes we now allow the cost functional to depend on the previous period's mistake v_{t-1} and today's optimal action $c : \mathcal{P} \times \mathcal{Z} \times \mathbb{R} \to \mathbb{R}$ of the form:

$$c(p; z_{t-1}, v_{t-1}) = \int_{\Theta} \int_{\mathcal{X}} \phi(p(x|\theta); v_{t-1}, x^*, x) \, \mathrm{d}x \, f(z|z_{t-1}) \, \mathrm{d}z \tag{194}$$

for ϕ convex in its first argument. In this formulation, the full non-parametric distribution of mistakes now depends on the previous period's mistake and today's optimal action.

To derive the Gaussian AR(1) formulation of mistakes, we now suppose that:

$$\phi(y; m, x^*, x) = \lambda y \log y + \omega y((x - x^*) - m)^2 \tag{195}$$

Concretely, this leads to the following cost functional:

$$c(p; z_{t-1}, v_{t-1}) = \int_\Theta \left[\lambda \int_\mathcal{X} p(x|\theta) \log p(x|\theta) \,\mathrm{d}x \right. \\ \left. + \omega \int_\mathcal{X} ((x - x^*(\theta)) - v_{t-1})^2 p(x|\theta) \,\mathrm{d}x \right] f(z|z_{t-1}) \,\mathrm{d}z \tag{196}$$

which penalizes sharply peaked distributions and those where average mistakes differ greatly from the previous period's mistake. If we moreover suppose that firm risk-adjusted profits are of their quadratic form:

$$\tilde{\Pi}(x, z) := \bar{\Pi}(z) + \frac{1}{2}\Pi_{xx}(z)(x - x^*(z))^2 \tag{197}$$

and we suppose that firms solve the problem:

$$\max_{p \in \mathcal{P}} \int_\Theta \int_\mathcal{X} \Pi(x, z)\, p(x \mid \theta) \,\mathrm{d}x\, f(z|z_{t-1}) \,\mathrm{d}z - c\,(p; z_{t-1}, v_{t-1}) \tag{198}$$

Solving this problem yields the AR(1) structure for mistakes, with Gaussian innovations.

Proposition 10. *The optimal stochastic choice pattern is given by:*

$$x = x^*(z) + v \tag{199}$$

where:

$$v = \rho v_{t-1} + u \tag{200}$$

and

$$\rho = \rho(z) = \frac{\omega}{\frac{1}{2}|\Pi_{xx}(z)| + \omega} \tag{201}$$

and

$$u \sim N\left(0, \frac{2\lambda}{\frac{1}{2}|\Pi_{xx}(z)| + \omega}\right) \tag{202}$$

Proof. We will denote $x^*(z)$ by γ and $\frac{1}{2}|\Pi_{xx}(z)|$ by β to simplify notation. We observe that the FOC characterizing optimal stochastic choice is given by:

$$-\beta(x-\gamma)^2 - \lambda\left[1 + \log p(x|z)\right] - \omega(x - \gamma - m)^2 + \mu(z) + \kappa(x, z) = 0 \tag{203}$$

where $\mu(z)$ is the Lagrange multiplier on the constraint that $p(x|z)$ integrates to unity and $\kappa(x, z)$ is the Lagrange multiplier on the non-negativity constraint that $p(x|z) \geq 0$. We can then observe that this has solution:

$$p(x|z) = \frac{\exp\left(-\tilde{\beta}(x - \tilde{z})^2\right)}{\int_{\mathcal{X}} \exp\left(-\tilde{\beta}(x' - \tilde{z})^2\right) \mathrm{d}x'} \tag{204}$$

where $\tilde{\beta} = \frac{\beta+\omega}{\lambda}$ and $\tilde{\gamma} = \gamma + \frac{\omega}{\beta+\omega} v_{t-1}$. It follows that:

$$x|z \sim N\left(\gamma + \frac{\omega}{\beta+\omega} v_{t-1}, \frac{2\lambda}{\beta+\omega}\right) \tag{205}$$

Putting this in more explicit terms, and substituting for γ and β, we obtain the desired representation. □

C.2 Transformed Gaussian Signal Extraction

In this section, we analyze attention cycles in a setting with Gaussian signal extraction. For notational simplicity, we describe this alternative model under the assumption that there is a uniform, scalar state variable θ, which represents each firm's productivity.

C.2.1 Set-up

When the state of the world is θ, the previous state is θ_{-1}, agents have priors $\pi_{\theta_{-1}} \in \Delta(\Theta)$, and the equilibrium level of output is $X(\theta, \theta_{-1})$, intermediates goods firms have payoffs given by:

$$\tilde{\Pi}(x, X(\theta, \theta_{-1}), \theta) = \alpha(X(\theta, \theta_{-1}), \theta) - \beta(X(\theta, \theta_{-1}), \theta)(x - \gamma(X(\theta, \theta_{-1}), \theta))^2 \tag{206}$$

where we will write $\beta(\theta, \theta_{-1}) = \beta(X(\theta, \theta_{-1}), \theta)$ and similarly for α and γ, and as we microfound via a second-order approximation of their true profit functions around the unconditionally optimal level of production in the main text.

Suppose moreover that agents receive a private Gaussian signal regarding their stakes-adjusted optimal action given by:

$$s_i = \frac{\beta(\theta, \theta_{-1})}{\int_{\Theta} \beta(\theta, \theta_{-1}) \mathrm{d}\pi(\theta|\theta_{-1})} \gamma(\theta, \theta_{-1}) + \frac{1}{\tau(\theta_{-1})} \varepsilon_i \tag{207}$$

where ε_i is a $N(0, 1)$ variable that is independent across agents and time periods; $\tau(\theta_{-1})$ is the (soon-to-be endogenized) square-root precision; and the agents' prior $\pi_{\theta_{-1}}$ is such that:

$$\frac{\beta(\theta,\theta_{-1})}{\int_\Theta \beta(\theta,\theta_{-1})\mathrm{d}\pi(\theta|\theta_{-1})}\gamma(\theta,\theta_{-1}) \sim N\left(\mu(\theta_{-1}),\sigma^2(\theta_{-1})\right) \tag{208}$$

This model incorporates the tractability of linear signal extraction into our non-quadratic tracking problem.

Conditional on such a signal s, the best reply of any firm is equal to the conditional expectation of the stakes-adjusted optimal action:

$$\begin{aligned} x(s) &= \mathbb{E}_{\pi_{\theta_{-1}}}\left[\frac{\beta(\theta,\theta_{-1})}{\int_\Theta \beta(\theta,\theta_{-1})\mathrm{d}\pi(\theta|\theta_{-1})}\gamma(\theta,\theta_{-1})|s\right] \\ &= \lambda(\theta_{-1})s + (1-\lambda(\theta_{-1}))\mu(\theta_{-1}) \end{aligned} \tag{209}$$

where $\lambda(\theta_{-1}) = \frac{\tau^2(\theta_{-1})}{\tau^2(\theta_{-1})+\frac{1}{\sigma^2(\theta_{-1})}}$ is the appropriate signal-to-noise ratio. Thus, the cross-sectional distribution of actions is given by:

$$\begin{aligned} x|\theta,\theta_{-1} \sim N\Bigg(&\lambda(\theta_{-1})\frac{\beta(\theta,\theta_{-1})}{\int_\Theta \beta(\theta,\theta_{-1})\mathrm{d}\pi(\theta|\theta_{-1})}\gamma(\theta,\theta_{-1}) \\ &+ (1-\lambda(\theta_{-1}))\mathbb{E}_{\pi_{\theta_{-1}}}\left[\frac{\beta(\theta,\theta_{-1})}{\int_\Theta \beta(\theta,\theta_{-1})\mathrm{d}\pi(\theta|\theta_{-1})}\gamma(\theta,\theta_{-1})\right], \frac{\lambda^2(\theta_{-1})}{\tau^2(\theta_{-1})}\Bigg) \end{aligned} \tag{210}$$

Say we empirically estimate an equation of the form:

$$x_{it} = \gamma_i + \chi_{j(i),t} + f_t(\theta_{it},\theta_{i,t-1}) + \varepsilon_{it} \tag{211}$$

which differs from our baseline specification in controlling flexibly for observed and lagged productivity.[57] The fitted values span $\mathbb{E}[x \mid \theta,\theta_{-1}]$ and capture state-dependent anchoring toward the prior mean. The residual ε_{it} captures the noise in the firm's action coming from the noise in the signal. The fact that the average action is no longer the unconditionally optimal action is an important departure from our baseline models in Section 3 and Online Appendix B. In the signal extraction model, the behavior of the stochastic residual captures some, but not all, of the effects of the "cognitive friction," since it does not directly speak to anchoring.

[57]In our main analysis, we consider some specifications with time-varying responsiveness to the fundamental shock. We have also considered specifications which depend more flexibly on lagged TFP and found broadly similar results to our baseline, but do not print these in the paper for brevity.

C.2.2 Interpreting Monotone Misoptimization

We now discuss the interpretation of our empirical exercise of studying stochastic volatility in ε_{it}. The variance of the residual is given in this model by:

$$\mathbb{V}_t[\varepsilon_{it}] = \frac{\lambda^2(\theta_{t-1})}{\tau^2(\theta_{t-1})} = \frac{\tau^2(\theta_{t-1})}{\left(\tau^2(\theta_{t-1}) + \frac{1}{\sigma^2(\theta_{t-1})}\right)^2} \tag{212}$$

Our empirical findings are consistent with $\theta_{t-1} \mapsto \mathbb{V}_t[\varepsilon_{it}]$ being an increasing function.[58]

$$\frac{\partial \tau^2(\theta_{-1})}{\partial \theta_{-1}} \left(\frac{1}{2\lambda(\theta_{-1})} - 1 \right) > \frac{\partial \frac{1}{\sigma^2(\theta_{-1})}}{\partial \theta_{-1}} \tag{213}$$

for all $\theta_{-1} \in \Theta$. Our own regression analysis in Section 5.5.2, as well as comprehensive studies of manufacturing establishments by Bloom et al. (2018) and Kehrig (2015), suggests that there is less fundamental dispersion in higher aggregate productivity states of the world. As a result, the RHS of this expression condition must be positive. Thus, there are two potential conditions under which the model is consistent with pro-cyclical misoptimization and counter-cyclical fundamentals dispersion:

1. Firms acquire sufficiently less precise signals in higher states $\frac{\partial \tau^2(\theta_{-1})}{\partial \theta_{-1}} < 0$ and the signal to noise ratio is always such that $\lambda(\theta_{-1}) > \frac{1}{2}$
2. Firms acquire sufficiently more precise signals in lower states $\frac{\partial \tau^2(\theta_{-1})}{\partial \theta_{-1}} > 0$ and the signal to noise ratio is always $\lambda(\theta_{-1}) < \frac{1}{2}$

In the former case, "high attention" measured by high signal precision correlates with low residual variance. In the latter case, "low attention" measured by low precision correlates with low residual variance. An important difference of the signal-extraction model from our baseline, then, is that additional information is required to separately identify patterns in attention and residual variance. In both models, "misoptimization" in payoff terms and attention are perfectly correlated by construction. But residual variance is monotone in misoptimization in our baseline model, but not in the signal extraction model due to the role of anchoring.

We interpret our finding that firms discuss macroeconomic developments more during recessions as qualitatively inconsistent with model case 2, and therefore an identifying piece of evidence for case 1. Elsewhere in the literature, Coibion et al. (2018a) find that firms

[58] Our specific empirical specification measured the correlation between contemporaneous output and contemporaneous dispersion of ϵ_{it}. If output is monotone in the state of nature and, along with the state of nature, very persistent, the translation to $\partial \mathbb{V}_t[\varepsilon_{it}]/\partial \theta_{t-1} \geq 0$ is immediate.

report higher demand for information when presented with bad macroeconomic news. Thus, our preferred interpretation of the model is one in which residual variance inherits the monotonicity of signal precision and attention.

C.2.3 Monotone Endogenous Precision

We now extend the model to include endogenous choice of signal precision and derive conditions under which firms obtain less precise signals in high productivity states in the model. To this end, suppose that after θ_{-1} is realized, but before θ is realized, that the agent can pay a cost $\tilde{\phi}(\tau^2)$ to achieve signal precision of τ^2, and where $\tilde{\phi}', \tilde{\phi}'' > 0$. Concretely, the optimal $\tau^2(\theta_{-1})$ solves:

$$\max_{\tau^2(\theta_{-1}) \in \mathbb{R}_+} \mathbb{E}_{\pi_{\theta_{-1}}} \left[-\beta(\theta, \theta_{-1}) \left(x^*(s, \tau^2(\theta_{-1})) - \gamma(\theta, \theta_{-1}) \right)^2 \right] - \tilde{\phi}(\tau^2(\theta_{-1})) \tag{214}$$

We moreover parameterize the scaling of quadratic losses by writing $\beta(\theta, \theta_{-1}) = \kappa \hat{\beta}(\theta, \theta_{-1})$ for all (θ, θ_{-1}) and some $\kappa \geq 1$. Our first goal will be to derive conditions under which the optimally chosen τ^2 in Program 214 is monotone increasing in κ. This demonstrates the natural incentives for firms to choose more precise information when the utility cost of a fixed posterior variance about the stakes-adjusted optimal action is higher.

Toward this end, we first simplify the agent's objective function. Using the distribution of optimal actions condition on τ^2 from in Equation 210, we write

$$\begin{aligned}
&\mathbb{E}_{\pi_{\theta_{-1}}} \left[-\beta(\theta, \theta_{-1}) \left(x^*(s, \tau^2(\theta_{-1})) - \gamma(\theta, \theta_{-1}) \right)^2 \right] \\
&= \mathbb{E}_{\pi_{\theta_{-1}}} \left[\mathbb{E} \left[-\beta(\theta, \theta_{-1}) \left(x^*(s, \tau^2(\theta_{-1})) - \gamma(\theta, \theta_{-1}) \right)^2 |\theta \right] \right] \\
&= \mathbb{E}_{\pi_{\theta_{-1}}} \left[\mathbb{E} \left[-\beta(\theta, \theta_{-1}) \left(x^*(s, \tau^2(\theta_{-1})) - \bar{x}(\theta, \theta_{-1}) + \bar{x}(\theta, \theta_{-1}) - \gamma(\theta, \theta_{-1}) \right)^2 |\theta \right] \right] \\
&= \mathbb{E}_{\pi_{\theta_{-1}}} \left[-\beta(\theta, \theta_{-1}) \mathbb{E} \left[\left(x^*(s, \tau^2(\theta_{-1})) - \bar{x}(\theta, \theta_{-1}) \right)^2 + \left(\bar{x}(\theta, \theta_{-1}) - \gamma(\theta, \theta_{-1}) \right)^2 |\theta \right] \right] \\
&= \mathbb{E}_{\pi_{\theta_{-1}}} \left[-\beta(\theta, \theta_{-1}) \left(\frac{\lambda^2(\theta_{-1})}{\tau^2(\theta_{-1})} + \left[\lambda(\theta_{-1}) \frac{\beta(\theta, \theta_{-1})}{\bar{\beta}(\theta_{-1})} \gamma(\theta, \theta_{-1}) + (1 - \lambda(\theta_{-1})) \mu(\theta_{-1}) - \gamma(\theta, \theta_{-1}) \right]^2 \right) \right]
\end{aligned} \tag{215}$$

where $\bar{x}(\theta, \theta_{-1})$ is the mean of the distribution in Equation 210 and $\bar{\beta}(\theta_{-1}) = \int_\Theta \beta(\theta, \theta_{-1}) \mathrm{d}\pi(\theta|\theta_{-1})$. Observe moreover that we simplify the second term as the following:

$$\mathbb{E}_{\pi_{\theta_{-1}}}\left[-\beta(\theta,\theta_{-1})\left(\lambda(\theta_{-1})\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})+(1-\lambda(\theta_{-1}))\mu(\theta_{-1})-\gamma(\theta,\theta_{-1})\right)^2\right]$$
$$=\mathbb{E}_{\pi_{\theta_{-1}}}\left[-\beta(\theta,\theta_{-1})\left((1-\lambda(\theta_{-1}))\left(\mu(\theta_{-1})-\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})\right)+\left(\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}-1\right)\gamma(\theta,\theta_{-1})\right)^2\right] \tag{216}$$

A necessary condition for an interior and optimal $\tau^2(\theta_{-1})$ is then the following first-order condition:

$$\begin{aligned}\tilde{\phi}'(\tau^2(\theta_{-1})) &= -\bar{\beta}(\theta_{-1})\frac{\partial}{\partial\tau^2(\theta_{-1})}\left[\frac{\lambda^2(\theta_{-1})}{\tau^2(\theta_{-1})}\right]\\ &+2(1-\lambda(\theta_{-1}))\frac{\partial\lambda(\theta_{-1})}{\partial\tau^2(\theta_{-1})}\mathbb{E}_{\pi_{\theta_{-1}}}\left[\beta(\theta,\theta_{-1})\left(\mu(\theta_{-1})-\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})\right)^2\right]\\ &+2\frac{\partial\lambda(\theta_{-1})}{\partial\tau^2(\theta_{-1})}\mathbb{E}_{\pi_{\theta_{-1}}}\left[\beta(\theta,\theta_{-1})\left(\mu(\theta_{-1})-\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})\right)\left(\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}-1\right)\gamma(\theta,\theta_{-1})\right]\end{aligned} \tag{217}$$

which reduces to:

$$\begin{aligned}\tilde{\phi}'(\tau^2(\theta_{-1})) &= -\bar{\beta}(\theta_{-1})\frac{\frac{1}{\sigma^2(\theta_{-1})}-\tau^2}{\left(\tau^2+\frac{1}{\sigma^2(\theta_{-1})}\right)^3}\\ &+2\frac{\frac{1}{\sigma^2(\theta_{-1})}}{\tau^2+\frac{1}{\sigma^2(\theta_{-1})}}\frac{\frac{1}{\sigma^2(\theta_{-1})}}{\left(\tau^2+\frac{1}{\sigma^2(\theta_{-1})}\right)^2}\mathbb{E}_{\pi_{\theta_{-1}}}\left[\beta(\theta,\theta_{-1})\left(\mu(\theta_{-1})-\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})\right)^2\right]\\ &+2\frac{\frac{1}{\sigma^2(\theta_{-1})}}{\left(\tau^2+\frac{1}{\sigma^2(\theta_{-1})}\right)^2}\mathbb{E}_{\pi_{\theta_{-1}}}\left[\beta(\theta,\theta_{-1})\left(\mu(\theta_{-1})-\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})\right)\left(\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}-1\right)\gamma(\theta,\theta_{-1})\right]\end{aligned} \tag{218}$$

We can now ask how $\tau^2(\theta_{-1})$ moves with κ. In particular, see that we can write:

$$\tilde{\phi}'(\tau^2(\theta_{-1}))=\xi(\tau^2(\theta_{-1}))\kappa \tag{219}$$

where:

$$
\begin{aligned}
\xi(\tau^2(\theta_{-1})) = & \frac{1}{\left(\tau^2 + \frac{1}{\sigma^2(\theta_{-1})}\right)^3} \Bigg[- \bar{\beta}(\theta_{-1}) \left(\frac{1}{\sigma^2(\theta_{-1})} - \tau^2 \right) \\
& + 2 \left(\frac{1}{\sigma^2(\theta_{-1})} \right)^2 \xi_1(\theta_{-1}) + 2 \frac{1}{\sigma^2(\theta_{-1})} \left(\tau^2(\theta_{-1}) + \frac{1}{\sigma^2(\theta_{-1})} \right) \xi_2(\theta_{-1}) \Bigg] \\
\xi_1(\theta_{-1}) = & \mathbb{E}_{\pi_{\theta_{-1}}} \left[\beta(\theta, \theta_{-1}) \left(\mu(\theta_{-1}) - \frac{\beta(\theta, \theta_{-1})}{\bar{\beta}(\theta_{-1})} \gamma(\theta, \theta_{-1}) \right)^2 \right] \\
\xi_2(\theta_{-1}) = & \mathbb{E}_{\pi_{\theta_{-1}}} \left[\beta(\theta, \theta_{-1}) \left(\mu(\theta_{-1}) - \frac{\beta(\theta, \theta_{-1})}{\bar{\beta}(\theta_{-1})} \gamma(\theta, \theta_{-1}) \right) \left(\frac{\beta(\theta, \theta_{-1})}{\bar{\beta}(\theta_{-1})} - 1 \right) \gamma(\theta, \theta_{-1}) \right]
\end{aligned}
\tag{220}
$$

Applying the implicit function theorem we then have that:

$$
\frac{\partial \tau^2(\theta_{-1})}{\partial \kappa} = \frac{\xi(\tau^2(\theta_{-1}))}{\tilde{\phi}''(\tau^2(\theta_{-1})) - \xi'(\tau^2(\theta_{-1}))} = \frac{\tilde{\phi}'(\tau^2(\theta_{-1}))}{\tilde{\phi}''(\tau^2(\theta_{-1})) - \xi'(\tau^2(\theta_{-1}))} \tag{221}
$$

where the denominator is positive as the marginal cost of precision is always positive. Thus, we have that:

$$
\frac{\partial \tau^2(\theta_{-1})}{\partial \kappa} > 0 \iff \tilde{\phi}''(\tau^2(\theta_{-1})) > \xi'(\tau^2(\theta_{-1})) \tag{222}
$$

A sufficient condition for $\frac{\partial \tau^2(\theta_{-1})}{\partial \kappa} > 0$, therefore, by convexity of the costs of precision is that $\xi'(\tau^2(\theta_{-1})) < 0$ for all $\tau^2(\theta_{-1})$. In words, if the benefit of precision is a concave function, then optimally set precision is increasing in κ.

Having shown the desired general comparative static, we now return to the context of our macroeconomic model. Recall that the curvature of firms profits is given by:

$$
\beta(\theta, \theta_{-1}) = v_\Pi X(\theta, \theta_{-1})^{-1-\gamma+\chi(1+\epsilon)} \theta^{-1-\epsilon} \tag{223}
$$

Thus:

$$
\bar{\beta}(\theta_{-1}) = \mathbb{E}_{\pi_{\theta_{-1}}} [v_\Pi X(\theta, \theta_{-1})^{-1-\gamma+\chi(1+\epsilon)} \theta^{-1-\epsilon}] \tag{224}
$$

Thus, whenever aggregate output is monotonically increasing in both θ and θ_{-1} and the prior $\pi_{\theta_{-1}}$ is monotone increasing in the FOSD order and $\gamma > \chi(1+\epsilon) - 1$, we have that $\bar{\beta}(\theta_{-1})$ is monotone decreasing in θ_{-1}. It then follows that $\tau^2(\theta_{-1})$ is monotone decreasing in θ_{-1} in equilibrium whenever $\xi' < 0$. Thus, the core logic of our baseline model translates exactly over to this setting with Gaussian signal extraction.

C.3 Rational Inattention

We now extend our results to the case of mutual information cost. As in the previous subsection, for notational simplicity, we describe this alternative model under the assumption that there is a uniform, scalar state variable θ, which represents each firm's productivity.

We first introduce the class of posterior-separable cost functionals. Denti (2020) provides this formulation as a representation theorem in stochastic choice space of the usual posterior-based definition of Caplin and Dean (2013):

Definition 3 (Posterior-Separable Cost Functionals)**.** *A cost functional c has a posterior-separable representation if and only if there exists a convex and continuous ϕ such that:*

$$c(p) = \int_{\mathcal{X}} \hat{\phi}(\{p(x|\theta)\}_{\theta\in\Theta}) \,\mathrm{d}x \tag{225}$$

where:

$$\hat{\phi}(\{p(x|\theta)\}_{\theta\in\Theta}) = p(x)\phi\left(\left\{\frac{p(x|\theta)\pi(\theta)}{p(x)}\right\}_{\theta\in\Theta}\right) \tag{226}$$

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 ϕ 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 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 θ 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 11. *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 $\overline{x} = -\underline{x} = \hat{x}$, the optimal stochastic choice rule is given by:*[59]

$$p(x|\theta) = \frac{1}{\sqrt{\frac{\pi\lambda}{\beta(\theta)}}} \exp\left\{ -\frac{1}{2} \left(\frac{x-\gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}} \right)^2 \right\} \tag{228}$$

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

$$x|\theta \sim N\left(\gamma(\theta), \frac{\lambda}{2\beta(\theta)}\right) \tag{229}$$

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

$$I(g) = \int_{\mathcal{S}} \int_{\Theta} g(s,\theta) \log \left(\frac{g(s,\theta)}{\pi(\theta) \int_{\Theta} g(s,\tilde{\theta})\, \mathrm{d}\tilde{\theta}} \right) \mathrm{d}\theta\, \mathrm{d}s \tag{230}$$

We now argue that it is without loss to consider a choice over stochastic choice rules $p : \Theta \to \Delta(\mathcal{X})$. Suppose x is an optimal action conditional on receiving any $s \in S_x$. Suppose that there exist $S_x^1, S_x^2 \in S_x$ 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 $\tilde{s} \in S_x^1 \cup S_x^2$ is 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 $g(s,\theta) \in \Delta(\mathcal{S} \times \Theta)$ is a choice over $g(\cdot|\cdot) : \Theta \to \Delta(S)$. Moreover, there is a unique posterior

[59] Formally, all actions are exchangeable in the prior if:

$$\int_{\Theta} \frac{\exp\{\beta(\theta)\lambda^{-1}(x-\gamma(\theta))^2\}}{\int_{\mathcal{X}} \exp\{\beta(\theta)\lambda^{-1}(\tilde{x}-\gamma(\theta))^2\} d\tilde{x}} \pi(\theta) d\theta = 1 \quad \forall x \in \mathcal{X} \tag{227}$$

$\mu(\theta|s)$ associated with each (non-dominated) action which is determined exactly by $g(\cdot|\cdot)$. Hence, the agent directly chooses a mapping $p(\cdot|\cdot) : \Theta \to \Delta(\mathcal{X})$. The agent's problem can then be directly re-written in the claimed stochastic choice form for some c_I:

$$\max_{P \in \mathcal{P}} \int_\Theta \int_\mathcal{X} u(x,\theta) \, \mathrm{d}P(x|\theta) \, \mathrm{d}\pi(\theta) - c_I(P) \tag{231}$$

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

$$c_I(p) = \int_\Theta \int_\mathcal{X} p(x|\theta) \log p(x|\theta) \, \mathrm{d}x \, \mathrm{d}\pi(\theta) - \int_\mathcal{X} p(x) \log p(x) \, \mathrm{d}x \tag{232}$$

where:

$$p(x) = \int_\Theta p(x|\theta) \, \mathrm{d}\pi(\theta) \tag{233}$$

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

$$\begin{aligned}\mathcal{L}(\{p(x|\theta), \kappa(x,\theta)\}_{x\in\mathcal{X},\theta\in\Theta}, \{\gamma(\tilde{\theta})\}_{\theta\in\Theta}) &= \int_\Theta \int_\mathcal{X} u(x,\theta) p(x|\theta) \, \mathrm{d}x \, \mathrm{d}\pi(\theta) \\ - \lambda \left(- \int_\mathcal{X} p(x) \log p(x) \, \mathrm{d}x + \int_\Theta \int_\mathcal{X} p(x|\theta) \log p(x|\theta) \, \mathrm{d}x \, \mathrm{d}\pi(\theta) \right) \\ + \kappa(x,\theta) p(x|\theta) + \tilde{\gamma}(\theta) \left(\int_\mathcal{X} p(x|\theta) \, \mathrm{d}x - 1 \right)\end{aligned} \tag{234}$$

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

$$p(x|\theta) = \frac{p(x) \exp\{u(x,\theta)\}}{\int_\mathcal{X} p(\tilde{x}) \exp\{u(\tilde{x},\theta)\} \, \mathrm{d}\tilde{x}} \tag{235}$$

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

$$\int_\Theta \frac{\exp\{u(x,\theta)\}}{\int_\mathcal{X} p(\tilde{x}) \exp\{u(\tilde{x},\theta)\} \, \mathrm{d}\tilde{x}} \, \mathrm{d}\pi(\theta) = 1 \tag{236}$$

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

$$u(x,\theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2 \tag{237}$$

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

$$\int_{\Theta} \frac{\exp\{-\beta(\theta)\lambda^{-1}(x-\gamma(\theta))^2\}}{\int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(\tilde{x}-\gamma(\theta))^2\}\,\mathrm{d}\tilde{x}} \pi(\theta)\,\mathrm{d}\theta = 1 \quad \forall x \in \mathcal{X} \tag{238}$$

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

$$p(x|\theta) = \frac{\exp\{-\beta(\theta)\lambda^{-1}(x-\gamma(\theta))^2\}}{\int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(\tilde{x}-\gamma(\theta))^2\}\,\mathrm{d}\tilde{x}} \tag{239}$$

The denominator of this expression can be computed:

$$\begin{aligned}
\int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(x-\gamma(\theta))^2\}\,\mathrm{d}x &= \int_{\mathcal{X}} \frac{\sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}}{\sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x-\gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\} \mathrm{d}x \\
&= \sqrt{2\pi\frac{\lambda}{2\beta(\theta)}} \int_{\mathcal{X}} \frac{1}{\sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x-\gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\} \\
&= \sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}
\end{aligned} \tag{240}$$

It follows that:

$$p(x|\theta) = \frac{1}{\sqrt{\frac{\pi\lambda}{\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x-\gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\} \tag{241}$$

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

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

$$\begin{aligned}
\gamma(\theta) &= x^*(X(\theta), \theta) \\
\beta(\theta) &= \frac{1}{2}|\Pi_{xx}(X(\theta), \theta)|
\end{aligned} \tag{242}$$

where $X(\theta)$ is the unique equilibrium level of aggregate production, then the model with mutual information is exactly equivalent to the model with entropic likelihood-separable cost that we studied. All results from Section 4 then carry directly.

Away from the exchangeability condition on the prior, one can still establish a comparative statics result whereby a small increase in curvature over a small set of states gives rise to an increase in precision in those states. Thus, even if attention is not ranked by curvature, small increases in curvature nevertheless increase attention. This is stated formally as Proposition 12.

Proposition 12. *Suppose that $u(x,\theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2$ and costs are posterior separable with differentiable kernel $\phi(\cdot) = \log(\cdot)$. Now perturb $\beta(\theta)$ in a neighborhood of width $\delta > 0$ of some $\hat{\theta}$ at which $\pi(\hat{\theta})$ exists and is finite by $\varepsilon > 0$:*

$$\hat{\beta}(\theta) = \beta(\theta) + \varepsilon \mathbb{I}\left[\theta \in [\hat{\theta} - \delta, \hat{\theta} + \delta]\right] \tag{243}$$

Moreover assume that the optimal policy is differentiable in ε at $\varepsilon = 0$. In the limit of $\delta \to 0$ the change in the optimal action density at $\varepsilon = 0$ in state $\hat{\theta}$ is such that $p(\hat{\theta})$ is becoming more precise about $\gamma(\hat{\theta})$ under ϕ, where precision is defined in the sense of Flynn and Sastry (2021).[60]

Proof. To prove the result, we first derive how the log density changes in ε. Recall from Proposition 11 that the optimal stochastic choice rule is given by:

$$p(x|\theta) = \frac{p(x)\exp\{u(x,\theta)\}}{\int_{\mathcal{X}} p(\tilde{x})\exp\{u(\tilde{x},\theta)\}\,\mathrm{d}\tilde{x}} \tag{244}$$

which under our assumption on payoffs is given by:

$$p(x|\theta) = \frac{p(x)\exp\{-\hat{\beta}(\theta)(x - \gamma(\theta))^2\}}{\int_{\mathcal{X}} p(\tilde{x})\exp\{-\hat{\beta}(\theta)(\tilde{x} - \gamma(\theta))^2\}\,\mathrm{d}\tilde{x}} \tag{245}$$

[60] We restate this definition below for completeness:

Definition 4 (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 $|x' - x^{*'}|$.*

On an asymmetric support, we call a distribution g symmetric if $g(x) = g(-x)$ whenever both $g(x)$ and $g(-x)$ are defined.

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

$$\begin{aligned}\frac{\mathrm{d}p(x|\theta)}{\mathrm{d}\varepsilon}|_{\varepsilon=0} = p(x|\theta)\left[-(x-\gamma(\theta))^2 + \int_{\mathcal{X}} (\tilde{x}-\gamma(\theta))^2 p(\tilde{x}|\theta)\,\mathrm{d}\tilde{x}\right]\mathbb{I}[|\theta-\hat{\theta}|\leq\delta] \\ + p(x|\theta)\left[\frac{\mathrm{d}\log p(x)}{\mathrm{d}x}|_{\varepsilon=0} - \int_{\mathcal{X}} \frac{\mathrm{d}\log p(x)}{\mathrm{d}x}|_{\varepsilon=0} p(\tilde{x}|\theta)\,\mathrm{d}\tilde{x}\right]\end{aligned} \tag{246}$$

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

$$\lim_{\delta\to 0}\frac{\mathrm{d}p(x|\theta)}{\mathrm{d}\varepsilon}|_{\varepsilon=0} = p(x|\theta)\left[\frac{\mathrm{d}\log p(x)}{\mathrm{d}\varepsilon}|_{\varepsilon=0} - \int_{\mathcal{X}} \frac{\mathrm{d}\log p(\tilde{x})}{\mathrm{d}\varepsilon}|_{\varepsilon=0}\ p(\tilde{x}|\theta)\,\mathrm{d}\tilde{x}\right] \quad \forall\theta\neq\hat{\theta} \tag{247}$$

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

$$\lim_{\delta\to 0}\frac{\mathrm{d}p(x)}{\mathrm{d}\varepsilon}|_{\varepsilon=0} = \int_\Theta \lim_{\delta\to 0}\frac{\mathrm{d}p(x|\theta)}{\mathrm{d}\varepsilon}|_{\varepsilon=0}\pi(\theta)d\theta \tag{248}$$

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

$$\lim_{\delta\to 0}\frac{\mathrm{d}p(x)}{\mathrm{d}\varepsilon}|_{\varepsilon=0} = 0 \tag{249}$$

It follows that:

$$\lim_{\delta\to 0}\frac{\mathrm{d}\log p(x|\theta)}{\mathrm{d}\varepsilon}|_{\varepsilon=0} = \left[-(x-\gamma(\theta))^2 + \int_{\mathcal{X}} (\tilde{x}-\gamma(\theta))^2 p(\tilde{x}|\theta)\,\mathrm{d}\tilde{x}\right]\mathbb{I}[|\theta-\hat{\theta}|\leq\delta] \tag{250}$$

where $\int_{\mathcal{X}}(\tilde{x}-\gamma(\theta))^2 p(\tilde{x}|\theta)d\tilde{x} > 0$. 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 for firms will tend to give rise to more precise attention, they key idea in the theory. Moreover, the same qualitative forces that gave rise to this curvature (the SDF and aggregate demand) externalities) still apply in this context.

D Additional Numerical Results

In this Appendix, we discuss robustness of our numerical findings as well as how the macroeconomic implications of Attention Cycles change under counterfactual scenarios.

D.1 Sensitivity of Main Results

D.1.1 Parameter Choice

To probe robustness to our choice of elasticity of substitution ϵ and the wage rule slope χ, we re-calibrate the model for alternative choices. For brevity, we summarize these experiments by considering "high" and "low" deviations for each parameter, holding fixed the others at baseline values, and present the proportional difference from the baseline in three summary statistics introduced in Section 6.3:

1. The relative output effect of negative and positive shocks, normalized in $\log \theta$ units such that the latter increases output by 3%;
2. The relative output effect of a "double dip" versus positive shock, holding fixed the size of the shock to productivity as above;
3. The ratio of output-growth volatility from the 10^{th} to 90^{th} of the output distribution.

We present our results in Figure A10. Lowering the elasticity of substitution or increasing the implied average markups can have ambiguous effects because it simultaneously increases the bite of a fixed level of misoptimization on misallocation, productivity, and output, while decreasing the bite of the profit-curvature channel toward cyclical attention. We find numerically that increasing markups or decreasing ϵ, toward the level implied by De Loecker et al. (2020), significantly increases the extent of our predicted asymmetries ($\approx$ 1.75x), while decreasing markups or increasing ϵ, toward the level implied by Edmond et al. (2018), modestly increases the extent of our predicted asymmetries.

Increasing the slope of the wage Phillips curve dampens our predictions, due to its dampening the economy's Keynesian-cross feedback. Decreasing the slope increases the bite of our predictions substantially by amplifying the same general-equilibrium effects.

D.1.2 Classical Labor Markets

For tractability in the theoretical section we equipped our model with a reduced-form *wage rule* rather than a micro-founded labor supply curve. As an alternative, we use the preferences of Greenwood et al. (1988) which replace Equation 5 with the following:

$$\mathcal{U}(\{C_{t+j}, L_{t+j}\}_{j\in\mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \frac{\left(C_{t+j} - \frac{{L_{t+j}}^{1+\phi}}{1+\phi}\right)^{1-\gamma}}{1-\gamma} \tag{251}$$

and remove the wage rule, Equation 7. These preferences generate a labor supply curve $w_t = L_t^{\phi}$ which closely resembles our reduced-form wage rule, but also takes seriously the implications for risk-pricing by making marginal utility a function of hours worked. We choose a parameterization of $\phi = \frac{\chi}{1-\chi\epsilon} = 0.153$, where χ and ϵ take our benchmark values indicated in Table 4. As indicated in the richer model of Online Appendix B, this calibration replicates an elasticity of $\chi = 0.095$ between real wages and real output.

Figure A11 is this model's analogue to Figure 3, showing output, the attention wedge, and labor productivity. We find comparable behavior of the attention wedge and losses from misoptimization and inattention. Figure A12 is this model's analogue to Figure 4, showing state-dependent shock response and stochastic volatility. Our results are quantitatively similar to our baseline calibration. These results demonstrate that classical labor markets do not undermine our main results in calibrations that are consistent with our calibrated wage rigidity.

D.2 Attention Cycles under Counterfactual Scenarios

Because the main amplification mechanism in our model, the reallocation of attention, is endogenous to economic conditions, we can use our framework to study how Attention Cycles and all associated macro phenomena would behave under counterfactual conditions. In this subsection, we explore the interaction of our findings with recent trends in product mark-ups; the decreasing pro-cyclicality of real wages; and external uncertainty shocks.

D.2.1 The Rise of Markups, via Lower Substitutability

An extensive recent literature documents a secular increase in markups charged by US public firms over the last half century (see, e.g., De Loecker et al., 2020; Edmond et al., 2018; Demirer, 2020). In our empirical calibration, we targeted "modern" average mark-ups as informed by this literature. Our framework would interpret any trends in aggregate mark-ups as arising from changes in the elasticity of substitution ϵ between products, which would need to have been higher in the previous, low-markup era than it is today. In our model, a lower elasticity of substitution or higher markup increases the output cost of a fixed amount of misoptimization dispersion, as it intuitively makes each individual product more "essential" to the consumed good; and it has *a priori* ambiguous effects on the extent of equilibrium attention cycles.

In our model, we run the following simple experiment. First, we adjust ϵ upward to simulate a 15 percentage point decrease in the aggregate markup, to match the estimate of Demirer (2020) for markups in the 1970s; and second, we adjust ϵ downward to extrapolate a 15 percentage point increase. We find that lower markups correspond to more severe effects of attention cycles on business cycles, as summarized by the asymmetry and state-dependence of dynamics (second and third panel of Figure A13). These results provide only speculative clues about the future, owing to both uncertainty about the true "cause" of rising markups and the plausibility of the trend continuing. But, regardless of their precision, they highlight a potentially important pathway linking market structure, firm decisionmaking, and the aggregate cost of misallocation.

D.2.2 More Rigid Real Wages

The inverse relationship between wage inflation and real conditions, the focus of Phillips's (1958), has proved elusive in modern data, particularly since the financial crisis (see, e.g., Galí and Gambetti, 2019). Our baseline estimate of a slope 0.095 between (detrended) real output and real wages since 1987 reflects this reality. In our model, more rigid real wages corresponded to a steeper Keynesian cross, and a steeper incentive toward high attention in low states of the world. For this reason, we may expect that the growing disconnect between factor prices and real conditions contributes toward the severity of our estimated macro effects.

In parallel to the previous experiment, we simulate both a "calibrated past" and "extrapolated future." For the former, we plug in the estimate of Galí and Gambetti (2019) that the wage Phillips curve has flattened by a factor of 1.9 over the last half century; for the latter, we extrapolate the same multiplicative trend as additional flattening.[61] We find, as shown in panels four and five of Figure A13, stronger effects of attention cycles in the regime with more rigid wages. This underscores the complementarity between attention cycles and the steepness of the Keynesian cross, and suggests a novel pathway by which factor price rigidity can influence patterns of macroeconomic volatility.

D.2.3 Elevated Uncertainty

Spikes in uncertainty around exceptional economic and political events have large documented effects on financial markets and firm decisionmaking (Bloom, 2009). Moreover, large,

[61] In particular, we use the ratio of the 1964-2007 estimate and 2007-2017 estimates in Table 3A of Galí and Gambetti (2019).

disorienting shocks are often either a natural consequence of poor economic performance (e.g., policy surprises during the 2007-2009 financial crisis) or their root cause (e.g., the Covid-19 pandemic). For this reason it is natural to study how changes in the "level of uncertainty," properly formalized in our model as variation in the attention cost λ, might interact with our main business cycle predictions. Proposition 4 showed that increases in uncertainty depress output in our model. These shocks also, according to the results of Proposition 1, increase the *sensitivity* of dispersion to macroeconomic conditions and hence, based on extrapolation of this partial-equilibrium logic, may amplify the extent of misoptimization cycles. For this reason, we might predict that elevated uncertainty is also complementary to the asymmetry and state-dependence generated at the macro level.

We explore this relationship by solving for the model equilibrium under scenarios with depressed and elevated attention costs, and numerically verify the predicted complementarity (panels six and seven of Figure A13). Thus our theory predicts that business cycles caused and/or amplified by background uncertainty-inducing events may induce sharper fluctuations in aggregate volatility due to endogenous reallocation of attention.

E Alternative Measurement of Macro Attention

In this appendix, we describe our methods and replication of main results about macroeconomic attention in language under two alternative methods. The first uses an alternative text data source from sales and earnings conference calls (E.1), and the second uses a "stemming" method on the list of words (E.2).

E.1 Sales and Earnings Conference Calls

We use sales and earnings conference calls as an alternative source of information about firms' relative attention to different risks and events. This analysis mirrors and supplements our main analysis of forms 10-K and 10-Q.

E.1.1 Data and Measurement

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 which cover 2003 to 2014. AWe 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.[62] 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.

To tabulate histograms of words within documents, we use the CountVectorizer function in the FeatureExtraction module of the standard Python package Scikit Learn.We then replicate the exact methodology of Section 2.1, generating a new list of 73 macroeconomic words. Like the main list of words, they are a combination of very interpretable choices ("government," "unemployment," "monetary") and false positives related to structure and pedagogy ("theory," "chapter").

E.1.2 Results

Figure A14 plots the conference-call derived measure, in log units and with quarterly fixed effects taken out, alongside the US unemployment rate. Conference-call derived macro attention, like our main measured derived from forms 10-Q/K, is cyclical and persistent. To benchmark these facts in the same way we did in the main text, we first run linear regressions of the form

$$\log \text{MacroAttentionCC}_t = \alpha + \beta_Z \cdot Z_t + \epsilon_t \tag{252}$$

for $Z_t \in \{\text{Unemployment}_t/100, \log \text{SPDetrend}_t\}$. The first two columns of Table A9 show the coefficients, which are slightly larger in absolute value than their equivalents with our 10K/Q measure (1.529 and -0.104, respectively). The third column gives our estimate of an AR(1) process, which is close to a unit root.

Casual comparison of Figure A14 and Figure 1 in the main text suggests that, while our two measures of attention have similar cyclical patterns, they do not closely track each other at the aggregate level. Conference-call derived attention is more sharply peaked around the onset of the Great Recession while 10-K/Q-derived attention remains elevated for several subsequent years. The correlation between the two measures on a common sample is a (statistically insignificant) 0.091. The relationship is closer, however, at the firm level. Columns 4-6 of Table A9 show the results of regressing the conference-call-derived measure 10K/Q-derived measure at the firm level, with increasingly more stringent fixed effects. The

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

correlation is consistently positive, though strongest in terms of cross-firm differences as opposed to within-firm differences. Moreover, as indicated in column 1 of Table A6, our finding linking firm-level misoptimization with firm-level attention is stable based on the conference call measure.

E.2 Word Stemming

E.2.1 Methods

Our main method for constructing Macro Attention treats individual words as the unit of measurement. For this reason, words like "unemployment" and "unemployed" are counted separately despite likely communicating the same meaning in all contexts. This method, while appealingly simple, may systematically under-count words that have a number of different forms or tenses, while allowing the multiple forms of certain ubiquitous words to crowd out other distinct concepts.

As an alternative method, which allays some of these concerns, we re-do our calculation of macroeconomic language using *word stems*. For each word w in the macroeconomics references and/or regulatory filings, we use the Porter Stemmer implemented in Python's `nltk` software to determine a stem $s(w)$. Stemming is an algorithmic and imperfect process. In examples relevant to our context, the Porter Stemmer associates "unemployment" and "unemployed" with the common stem "unemploy." But it also, employing the same logic, associates "nominal" with "nomin," a stem which may match to words less often used to describe aggregate prices (e.g., "nominate").

We adapt our tf-idf calculation to the stem level by calculating, for each stem s that appears in the regulatory filings,

$$\text{tf-idf}(s)_{it} := \text{tf}(s)_{it} \cdot \log\left(\frac{1}{\text{df}(s)}\right) \tag{253}$$

where $\text{tf}(s)$ is the total term frequency of all words mapped to stem s, and $\text{df}(s)$ is the minimum document frequency among words associated with the stem.[63] We calculate the top macro stems using the approach described in the main text (Section 2.1); construct the set of macro words $\mathcal{W}_M$ as the set of all words associated with a macro stem; and proceed in the standard way to calculate firm-level and aggregate macro attention.

[63] We use the minimum instead of the overall frequency due to a data limitation of having document frequencies at the word, not stem, level. We expect either method to produce broadly similar results.

E.2.2 Results

Table A10, in analogy to Table A9, presents a summary of the cyclical patterns of the stemmed Macro Attention measure as well as its relationship to our main measure. The two measures behave very similarly in the time series and are tightly connected at the firm level. Moreover, when we replicate our main model linking firm-level Macro Attention to firm-level misoptimization as in Table A6, we estimate a coefficient of -0.020 (SE: 0.004), which is comparable within error bars to our baseline estimate of -0.009 (SE 0.003).

F Details of Measuring Productivity and Misoptimization

This appendix describes in greater detail our data construction and empirical methodologies for our firm-level analysis of production and misoptimization. It serves in particular as a companion for Section 5.1 of the paper.

F.1 Sample Selection and Data Construction

Our main dataset is Compustat Annual Fundamentals, which compiles detailed information from public firms' financial statements. Table A8 describes the main sample variables that we use and their definitions, in brief. Production, in value terms, is defined as reported sales. Employment in Compustat is reported as the number of employees. To calculate a wage bill, we multiply this by the average industry wage calculated from the Census Bureau's County Business Patterns dataset in the same year, as the sector's total national wage bill divided by the number of employees. From 1998 onward, we use the 2- or 3-digit NAICS classification that is consistent with our main analysis. Prior to 1997, and the introduction of NAICS codes in the CBP data, we use 2-digit SIC industries. For materials expenditure, we measure the sum of reported variable costs (`cogs`) and sales and administrative expense (`xsga`) net of depreciation (`dp`) and the aforementioned wage bill. To measure the capital stock, we use a perpetual inventory method as in Ottonello and Winberry (2020) starting with the first reported observation of gross value of plant, property, and equipment and adding net investment or the differences in net value of plant, property, and equipment.[64]

We restrict the sample to firms based in the United States, reporting statistics in US Dollars, and present in the "Industrial" dataset. Within this sample, we apply the following additional filters:

1. Sales, material expenditures, and capital stock are strictly positive;

[64]Note that, because of our later usage of fixed effects and lack of direct calculations using capital "expenditures" evaluated at an imputed rental rate, that it is inessential to deflate the value of the capital stock.

2. Employees exceed 10;
3. 2-digit NAICS is not 52 (Finance and Insurance) or 22 (Utilities);
4. Acquisitions as a proportion of assets (`aqc` over `at`) does not exceed 0.05.

The first two ensure that all companies meaningfully report all variables of interest for our production function estimation; the second applies a stricter cut-off to eliminate firms that are very small, and lead to outlier estimates of productivity and choices. The third filter eliminates firms in two industries that, respectively, may have highly non-standard production technology and non-standard market structure. The fourth is a simple screening device for large acquisitions which may spuriously show up as large innovations in firm choices and/or productivity. We finally restrict attention to firms operating on a fiscal calendar that ends in December, for more straightforward calculations of aggregate time trends.

We categorize the data into 44 sectors. These are defined at the 2-digit NAICS level, but for the Manufacturing (31-33) and Information (51) sectors, which we classify at the 3-digit level to achieve better balance of sector size. Table A11 lists the sectors along with summary statistics for their relative size, in terms of sales and employment, in cross sections corresponding to 1990 and 2010, in the full (not selected) Compustat sample. Overall, the full dataset covers between 15-20% of US employment and 60-80% of US output, modulo the clarification that not all Compustat sales necessarily occur in the United States.

F.2 Production Function and Productivity Estimation

Our primary method for production for estimating production functions, and thereby recovering total factor productivity, is a cost share approach. In brief, we use cost shares for materials and labor to back out production elasticities, and treat the elasticity of capital as the implied "residual" given an assumed mark-up $\mu > 1$ (in our baseline, $\mu = 4/3$) and constant returns to scale. We validate, in subsection F.3 of this Appendix and in particular Lemma 3, that this method is consistent in sample up to an essentially negligible correction term, due to the underlying logic that input choices are "right on average" even in the presence of mistakes. The exact procedure is the following:

1. For all firms in industry j, calculate the estimated materials and labor shares:

$$\begin{aligned}
\text{Share}_{M,j'} &= \frac{\sum_{i:j(i)=j'} \sum_t \text{MaterialExpenditure}_{it}}{\sum_{i:j(i)=j'} \sum_t \text{Sales}_{it}} \\
\text{Share}_{L,j'} &= \frac{\sum_{i:j(i)=j'} \sum_t \text{WageBill}_{it}}{\sum_{i:j(i)=j'} \sum_t \text{Sales}_{it}}
\end{aligned} \tag{254}$$

2. If $\text{Share}_{M,j'} + \text{Share}_{L,j'} \leq \mu^{-1}$, then set

$$\begin{aligned}\alpha_{M,j'} &= \mu^{-1} \cdot \text{Share}_{M,j'} \\ \alpha_{L,j'} &= \mu^{-1} \cdot \text{Share}_{L,j'} \\ \alpha_{K,j'} &= 1 - \alpha_{M,j'} - \alpha_{L,j'}\end{aligned} \tag{255}$$

3. Otherwise, adjust shares to match the assumed returns to scale, or set

$$\begin{aligned}\alpha_{M,j'} &= \frac{\text{Share}_{M,j'}}{\text{Share}_{M,j'} + \text{Share}_{L,j'}} \\ \alpha_{L,j'} &= \frac{\text{Share}_{L,j'}}{\text{Share}_{M,j'} + \text{Share}_{L,j'}} \\ \alpha_{K,j'} &= 0\end{aligned} \tag{256}$$

A more flexible method for production function estimation might allow for the returns to scale and/or markups to vary across sectors. We opt not to make such a method a baseline because uniform returns to scale and markups are consistent with our subsequent empirical calibration of the model. In robustness checks, we have experimented in particular with extracting production function parameters under different assumed markups and found overall stable results for the empirical behavior of production misoptimizations.

To translated our production function estimates into productivity, we first calculate a "Sales Solow Residual" $\tilde{\theta}_{it}$ of the following form:

$$\tilde{\theta}_{it} = \log \text{Sales}_{it} - \frac{1}{\mu} \left(\alpha_{M,j(i)} \cdot \log \text{MatExp}_{it} - \alpha_{L,j(i)} \cdot \log \text{Empl}_{it} - \alpha_{K,j(i)} \cdot \log \text{CapStock}_{it} \right)$$

where we intentionally use the "literal" accounting notation to highlight the fact that not all variables are specified in economically meaningful *quantity* units. To alleviate this issue in a conservative way, we define our final estimate of TFP as the residual of the previous from industry-by-time fixed effects which, under our presumed model of industry-level variation in factor prices, identifies a rescaling of firm-level TFP in productivity units.

As a robustness check, we also calculate TFP using the method of Olley and Pakes (1996), applied separately to estimate the production function of each industry.[65] The methodology of Olley and Pakes (1996) aims, in particular, to correct the bias in standard least-squares estimates that under-states the output elasticity to capital, since that input is likely pre-

[65] In particular, we use the implementation by Yasar et al. (2008) of the `opreg` package in Stata. We use log investment as the proxy variable and year dummies as additional controls. We throw out estimates that imply individual elasticities that are negative or greater than 1, but do not otherwise enforce any returns to scale normalization.

determined and therefore mechanically less responsive to productivity. Since the underlying timing assumptions of Olley and Pakes (1996) are not directly compatible with our specifically assumed structure, we do not prefer it as a main estimate. But we are re-assured by these estimates' "up stream" and "down stream" similarity to our baseline estimates. To the first point, Table A12 shows the results from regressing the two TFP measures on one another in a common sample, including various levels of fixed effects. In each case, the slope is close to one and the within-R^2, or goodness of fit net of fixed effects, exceeds 0.6. To the second point, the relevant columns of Tables A1, A2, and A6 demonstrate how our main aggregate and firm-level results replicate under the alternative measurement scheme, with similar quantitative and qualitative take-aways.

F.3 Theory to Data: Micro-foundations

In this subsection, we outline more exactly the mapping from our model to our production function estimation via cost shares and our log-linear estimating equations.

Firms are monopolists within their unique product. We assume that this demand curve has a constant elasticity of substitution form, so prices q_{it} lie on the demand curve $\log q_{it} = \gamma_i - \frac{1}{\epsilon}(\log x_{it} - \log X_t)$ for some inverse elasticity $\epsilon > 1$ and aggregate output X_t.[66] Finally, firms face sector-specific input prices $(q_{j(i),L,t}, q_{j(i),M,t}, q_{j(i),K,t})$ for the three inputs, respectively.

We model firm (mis-) optimization in the following way that is uniform across inputs. Conditional on any chosen level of production, firms cost minimize over their input bundle conditional on observed input prices. Let $q_{j(i),T,t}$ denote the associated "Total" input cost per unit of produced output, and $x^*(\theta, q_T, X)$ denote the unconditionally profit-maximizing level of production. Firms choose a production level which differs from this level by a *misoptimization* m_{it}:

$$\log x_{it} = \log x^*(\theta_{it}, q_{j(i),T,t}, X_t) + m_{it} \tag{257}$$

And the dynamics of the misoptimization, as given in the main text's Equation 26 and re-printed here, are described by an AR(1) process:

$$m_{it} = \rho m_{i,t-1} + \left(\sqrt{1-\rho^2}\right) u_{it} \tag{258}$$

in which innovations u_{it} are mean zero with variance σ_{it}^2.

First, we characterize the firm's optimal production level x^* as a function of productivity, the total input price, and aggregate demand:

[66] It is straightforward, and consistent with our modeling approach, also to allow substitution within more narrowly defined industries.

Lemma 1 (Optimal Output Choice)**.** *The firm's optimal output choice is*

$$\log x^*(\theta, q_T, X) = \epsilon \log \left(1 - \frac{1}{\epsilon}\right) + \epsilon\gamma_i + \log X - \epsilon \log q_T \tag{259}$$

Proof. Immediate from the first-order conditions of the program

$$x^*(\theta, q_T, X) = \arg\max_x \left\{ x \left(e^{\gamma_i} x^{-\frac{1}{\epsilon}} X^{\frac{1}{\epsilon}} - q_T \right) \right\} \tag{260}$$

□

We next characterize the optimal choice of each input, based on cost-minimization and the expression for chosen output as a function of optimal output and the misoptimization (Equation 257):

Lemma 2 (Input Choice)**.** *For any input $Z \in \{L, M, K\}$, and firm i,*

$$\log Z_{it} = K_{Z,i,j(i)} + (1-\epsilon) \log \tilde{q}_{j(i),T,t} - \log q_{j(i),Z,t} + \log X_t + (\epsilon - 1) \log \theta_{it} + m_{it} \tag{261}$$

where $K_{Z,i,j(i)} = \left(\epsilon\gamma_i + \epsilon \log\left(1 - \frac{1}{\epsilon}\right) - \log \alpha_{Z,j(i)}\right)$ is a constant, $\tilde{q}_{j(i),T,t} = \sum_Z \alpha_{Z,j(i)} (\log q_{j(i),Z,t} -$ $\log \alpha_{Z,j(i)}$ is a normalized price index, and m_{it} is the production misoptimization.

Proof. In the cost minimization step, for any planned output choice X, the firm solves

$$\min_{L_{it}, M_{it}, K_{it}} \sum_{z \in \{L,M,K\}} q_{j(i),Z,t} Z_{it} \qquad \text{s.t. } \theta_{it} L_{it}^{\alpha_{L,j(i)}} M_{it}^{\alpha_{M,j(i)}} K_{it}^{\alpha_{K,j(i)}} \leq X \tag{262}$$

Standard first-order methods yield the solution, for each input,

$$\log Z_{it}^* = \log q_{i,T,t} + \log X - \log \alpha_{Z,j(i)} \tag{263}$$

where the price index $q_{j(i),T,t}$, which is also the Lagrange multiplier on the constraint, is

$$\log q_{i,T,t} = \sum_Z \alpha_{Z,j(i)} (\log q_{j(i),Z,t} - \log \alpha_{Z,j(i)}) - \log \theta_{it} \tag{264}$$

The desired expression comes from substituting in Equations 257 and 260 into the above. □

This calculation validates our log-linear regression model Equation 25, and its equivalent for any input on the left-hand-side, by showing how all input choices inherit the "optimal plus error" structure.

We finally describe and validate our method for recovering production function parameters from a cost shares approach. The following result shows how cost shares are recovered at the firm level, if all data were observed without noise:

Lemma 3 (Production Function Estimation)**.** *For any input $Z \in \{L, M, K\}$, and firm i,*

$$\alpha_{j(i),Z} = \left(1 - \frac{1}{\epsilon}\right)^{-1} \frac{q_{j(i),Z,t} Z_{it}}{q_{it} x_{it}} \exp\left(-\frac{m_{it}}{\epsilon}\right) \tag{265}$$

Proof. This can be calculated directly using the results of Lemmas 1 and 2. It also follows immediately from "standard" results with perfect optimization after noting that $Z_{it} = Z_{it} \cdot \exp(m_{it})$ based on the calculation above; $q_{it} x_{it} = \exp(\gamma_i) X_t (x_{it}^*)^{1-\frac{1}{\epsilon}} \cdot \exp\left(\left(1 - \frac{1}{\epsilon}\right) m_{it}\right)$ based on the definitions of q_{it} and x_{it}; and standard calculations. $\square$

In words, this result says that the ratio of expenditures on input Z to total sales, multiplied by the the markup and a correction factor related to the mistake, equals the production elasticity. In principle, we could simultaneously estimate the production function and the statistical properties of mistakes to correct for the fact that the term $\exp\left(-\frac{m_{it}}{\epsilon}\right)$ is not zero on average. In practice, our mistakes are zero mean by construction and have a variance of about 0.08 in sample. Using a log-linear calculation, and our standard value of $\epsilon = 4$, this implies an average correction factor of $\exp\left(\frac{1}{2 \cdot 4^2} \cdot 0.08\right) = 1.0025$ which is essentially negligible.

G State-Dependent Attention in the Coibion et al. (2018a) Survey

In this Appendix, we test our interpretation of attention and misoptimization cycles using the dataset of Coibion et al. (2018a) (henceforth, CGK), one of the most comprehensive datasets of firm-level operations and macro backcasts in an advanced economy. These data were assembled from detailed survey of the general managers of a representative panel of firms in New Zealand from 2013 to 2016. The final subsection of this appendix, G.3, contains more details about the survey questions and data.

G.1 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 A13 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 increases in the context of *bad* news about the macroeconomy and, if anything, contracts with *good* news of the macroeconomy. This is consistent with our hypothesis 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.

G.2 Reported Profit Function Curvature and Attention

A second test possible in the CGK data relates to our more specific prediction that higher curvature of the firm's objective, as a function of decision variables, should increase attentiveness to decision-relevant variables including macroeconomic aggregates. The CGK survey indirectly elicits information on this shifter via questions about purely hypothetical price changes and revenue increases to an "optimal point." In Section G.3 at the end of this appendix, 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.[67]

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.

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

$$Y_{it} = \alpha + \beta \cdot \text{ProfitCurv}_{it} + \gamma' Z_{it} + \epsilon_{it} \tag{266}$$

Our prediction that profit curvature drives stakes for attention corresponds to $\beta < 0$ for back-cast errors and $\beta > 0$ for reported attention. 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 A14 shows the results. For inflation we find strong evidence that higher-curvature firms make smaller errors, with some much of the effect being absorbed by control variables

[67] Table A15 shows that this curvature measure is higher for smaller firms with more within-industry competitors, though these patterns are immaterial for our reduced-form verification of our prediction.

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.

G.3 Details of Data Construction

In this final subsection, we describe more precisely how we use the raw survey results of Coibion et al. (2018a) in 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* and OpenICPSR (Coibion et al., 2018b). 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.

G.3.1 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 Δp_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 (14), in percentage units for the outcome and the choice variable), the following relationship holds between the measurable quantities and the profit function curvature ProfitCurv_i:

$$\Delta\Pi_i = \text{ProfitCurv}_i \cdot \Delta p_i^2 \tag{267}$$

We use this expression to calculate an empirical analogue of profit curvature. The top panel of Table A15 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 A15 shows firm and manager-level correlates for our measure in the CGK data. The table reports coefficients of the following regression:

$$\widehat{\text{ProfitCurv}}_i = \beta \cdot \hat{X}_i + e_{it} \tag{268}$$

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 β 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.

G.3.2 Outcomes: Back-cast Errors

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

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

Although the wording of the question is not entirely clear about what indicator is being referred to, we follow CGK and interpret this as the annual percent change in CPI, with realized value 1.6%. Firms are asked a similar question in wave 4, but we prefer the wave 1 version because the sample size is slightly larger. Table A16 recreates Table A14 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.

G.3.3 Outcomes: Tracking Indicators

We finally use, for the lower panel of Table A14, 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 combine GDP in this question (by implication, in levels) with quantitative forecasts of GDP Growth in Table A14.

H Supplemental Tables and Figures

List of Tables

List of Figures

Table A1: Cyclicality of Misoptimization, with Alternative Measurement Schemes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Outcome: $\text{MisoptimizationDispersion}_t$										
$\text{Unemployment}_t/100$	-0.841	-0.439	-0.822	-0.749	-0.501	-0.695	-0.802	-0.813	-0.605	-0.812	-1.034
	(0.341)	(0.196)	(0.336)	(0.292)	(0.159)	(0.280)	(0.337)	(0.330)	(0.293)	(0.335)	(0.549)
Baseline	✓										
Adj. Control		✓									
Leverage Control			✓								
t, t^2 Control				✓							
Manufacturing					✓						
Sector Policy Fn.						✓					
t-varying Policy Fn.							✓				
Quadratic Policy Fn.								✓			
Pre-Period TFP									✓		
t-varying Prod. Fn.										✓	
OP (96) TFP											✓
N	31	31	31	31	31	31	31	31	20	31	31
R^2	0.243	0.219	0.235	0.326	0.298	0.255	0.230	0.239	0.244	0.213	0.144

Notes: Standard errors are HAC-robust with a 3-year Bartlett Kernel. The "Adjustment Cost" and "Leverage" controls are described in the main text. The "Sector Policy Fn." estimates the policy function separately for each sector. The "t-varying Policy Fn." model interacts all coefficients in the policy function with time fixed effects. The "Quadratic Policy Fn." allows for quadratic dependence on TFP. The "Pre-Period TFP" model uses cost shares from before 1997 to construct the production function, and data after 1998 to estimate the policy function and misoptimizations. The "t-varying Prod. Fn." model estimates the Solow residual using industry-by-year-specific cost shares. The "OP (96)" model estimates productivity using the proxy-variable strategy of Olley and Pakes (1996), as detailed in Online Appendix F.2.

Table A2: Pricing of Misoptimization, with Alternative Measurement Schemes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Outcome: $\Delta \log P_{it}$					
$\hat{u}_{it}^2$	-0.097	-0.239	-0.101	-0.168	-0.090	-0.099	-0.109	-0.062	-0.057	-0.096
	(0.034)	(0.941)	(0.035)	(0.045)	(0.036)	(0.035)	(0.034)	(0.028)	(0.021)	(0.034)
$\hat{u}_{it}^2 \times \Delta \log P_t$	0.443	0.941	0.415	0.680	0.420	0.330	0.447	0.227	0.231	0.417
	(0.171)	(0.370)	(0.169)	(0.182)	(0.156)	(0.163)	(0.163)	(0.130)	(0.098)	(0.168)
Sector x Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline	✓									
Adj. Control		✓								
Leverage Control			✓							
Manufacturing				✓						
Sector Policy Fn.					✓					
t-varying Policy Fn.						✓				
Quadratic Policy Fn.							✓			
Pre-Period TFP								✓		
t-varying Prod. Fn.									✓	
OP (96) TFP										✓
N	41,206	35,388	41,016	22,902	41,197	41,203	41,203	26,206	40.078	41,166
R^2	0.385	0.387	0.385	0.367	0.385	0.384	0.385	0.429	0.382	0.385

Notes: Standard errors are double-clustered at the year and firm level. The scenarios are described in the main text and the notes of Table A1.

Table A3: The Effects of Misoptimization in Levels

	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome:					
	$\Delta \log P_{it}$		π_{it}		$\hat{m}_{it}^2$	
$\hat{m}_{it}^2$	-0.042	-0.076	-0.021	-0.025		
	(0.028)	(0.033)	(0.010)	(0.011)		
$\hat{m}_{it}^2 \times \Delta \log P_t$		0.177		0.038		
		(0.078)		(0.053)		
$\log \text{MacroAttention}_{it}$					-0.010	-0.020
					(0.006)	(0.007)
Firm FE	✓	✓	✓	✓		✓
Sector x Time FE	✓	✓	✓	✓	✓	✓
N	41,247	41,247	57,646	57,646	34,421	33,841
R^2	0.385	0.385	0.656	0.656	0.053	0.488

Notes: Standard errors are double-clustered at the year and firm level. These specifications replicate results in Tables 1, 2, 3, and A4 with with $\hat{m}_{it}$ in place of $\hat{u}_{it}$.

Table A4: Misoptimization and Firm Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Outcome: $\Delta \log P_{it}$				Outcome: π_{it}			
$\hat{u}_{it}^2$	-0.236	-0.230	-0.060	-0.051	-0.316	-0.316	-0.106	-0.105
	(0.026)	(0.026)	(0.032)	(0.032)	(0.024)	(0.024)	(0.018)	(0.017)
Sector x Time FE	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE			✓	✓			✓	✓
TFP Control		✓		✓		✓		✓
N	41,578	41,578	41,206	41,206	51,015	51,015	50,966	50,996
R^2	0.238	0.261	0.384	0.403	0.117	0.131	0.663	0.681

Notes: Standard errors are double-clustered at the year and firm level.

Table A5: Alternative Timing for Relationship Between Attention and Misoptimization

	(1)	(2)	(3)	(4)	(5)	(6)
			Outcome: $\hat{u}_{it}^2$			
$\log \text{MacroAttention}_{i,t-1}$	-0.0031	-0.0010			0.0060	0.0079
	(0.0034)	(0.0045)			(0.0048)	(0.0045)
$\log \text{MacroAttention}_{it}$					-0.0006	-0.0058
					(0.0048)	(0.0042)
$\log \text{MacroAttention}_{i,t+1}$			-0.0064	-0.0052	-0.012	-0.0067
			(0.0030)	(0.0039)	(0.004)	(0.0048)
Joint F-statistic					4.17	2.41
Joint p-value					0.019	0.097
Sector x Time FE	✓	✓	✓	✓	✓	✓
Firm FE		✓		✓		✓
N	25,657	25,122	24,094	23,312	19,330	18,649
R^2	0.054	0.375	0.062	0.382	0.065	0.383

Notes: Standard errors are double-clustered at the year and firm level.

Table A6: Attention and Misoptimization, with Alternative Measurement Schemes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
						Outcome: $\hat{u}_{it}^2$					
$\log \text{MacroAttention}_{it}$	-0.0081	-0.0163	-0.0035	-0.0076	-0.0127	-0.0107	-0.0084	-0.0062	-0.0140	-0.0080	-0.0140
	(0.0028)	(0.0066)	(0.0015)	(0.0028)	(0.0037)	(0.0028)	(0.0028)	(0.0026)	(0.0042)	(0.0028)	(0.0042)
Sector x Time FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Baseline	✓										
Conf. Call		✓									
Adj. Control			✓								
Leverage Control				✓							
Manufacturing					✓						
Sector Policy Fn.						✓					
t-varying Policy Fn.							✓				
Quadratic Policy Fn.								✓			
Pre-Period TFP									✓		
t-varying Prod. Fn.										✓	
OP (96) TFP											✓
N	28,279	5,997	24,024	28,133	14,891	28,283	28,275	28,275	24,785	28,266	24,785
R^2	0.053	0.072	0.060	0.053	0.041	0.054	0.051	0.056	0.046	0.053	0.046

Notes: Standard errors are double-clustered at the year and firm level. The scenarios are described in the main text and the notes of Table A1.

Table A7: Policy Function Estimation

	Baseline	Adj. Cost	Leverage	Quadratic
	Panel A: Persistence of Misoptimization			
	Outcome: $\hat{m}^0_{it}$			
$\hat{m}^0_{i,t-1}$ (ρ)	0.696	0.016	0.696	0.683
	(0.021)	(0.005)	(0.003)	(0.003)
	Panel B: Quasi-Differenced Policy Function			
	Outcome: $L_{it} - \hat{\rho} L_{i,t-1}$			
$\hat{\theta}_{it}$	0.418	0.381	0.419	0.463
	(0.024)	(0.026)	(0.025)	(0.029)
$\hat{\theta}_{i,t-1}$	-0.031	-0.090	-0.026	0.006
	(0.018)	(0.015)	(0.019)	(0.020)
$\hat{\theta}_{it} \times \text{Lev}_{it}$			-0.008	
			(0.003)	
$\hat{\theta}_{i,t-1} \times \text{Lev}_{i,t-1}$			-0.025	
			0.006	
Lev_{it}			-0.020	
			(0.005)	
$\text{Lev}_{i,t-1}$			-0.050	
			(0.011)	
$\hat{\theta}^2_{it}$				0.045
				(0.008)
$\hat{\theta}^2_{i,t-1}$				0.031
				(0.006)
$L_{i,t-1}$		0.811		
		(0.012)		
$L_{i,t-2}$		-0.041		
		(0.010)		
N	51,891	44,051	51,664	51,891
R^2	0.896	0.990	0.896	0.904

Notes: Standard errors are double clustered by firm and year. The four columns correspond to four of our policy-function estimation methods, as described in the main text. To limit the effect of outliers, we drop the top and bottom 1% of the $\hat{\theta}_{it}$ distribution in each regression.

Table A8: Data Definitions in Compustat

	Quantity	Expenditure
Production, x_{it}	—	`sale`
Employment, L_{it}	`emp`	`emp` $\times$ industry wage
Materials, M_{it}	—	`cogs` + `xsga` − `dp` − wage bill
Capital, K_{it}	`ppegt` plus net investment	—

Table A9: Time-Series and Cross-Sectional Properties of Conference-Call Attention

	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome:					
	$\log \text{MacroAttnCC}_t$			$\log \text{MacroAttnCC}_{it}$		
$\frac{\text{Unemployment}_t}{100}$	2.481					
	(0.596)					
$\log \text{SPDetrend}_t$		-0.270				
		(0.056)				
$\log \text{MacroAttnCC}_{t-1}$			0.949			
			(0.068)			
$\log \text{MacroAttn10K}_{it}$				0.463	0.372	0.121
				(0.034)	(0.036)	(0.028)
Firm FE						✓
Sector x Time FE					✓	✓
N	46	46	45	8,023	7,994	7,670
R^2	0.376	0.593	0.873	0.123	0.308	0.804

Notes: In the first three columns, standard errors are HAC robust with a bandwidth (Bartlett kernel) of four quarters. In the second three columns, standard errors are double-clustered by year and firm ID.

Table A10: Time-Series and Cross-Sectional Properties of Word-Stemmed Attention

	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome:					
	$\log \text{MacroAttnStem}_t$			$\log \text{MacroAttnStem}_{it}$		
$\frac{\text{Unemployment}_t}{100}$	0.994					
	(0.330)					
$\log \text{SPDetrend}_t$		-0.062				
		(0.031)				
$\log \text{MacroAttnStem}_{t-1}$			0.811			
			(0.057)			
$\log \text{MacroAttn10K}_{it}$				0.553	0.542	0.518
				(0.010)	(0.010)	(0.008)
Firm FE						✓
Sector x Time FE					✓	✓
N	92	92	92	46,612	46,590	45,458
R^2	0.118	0.140	0.675	0.561	0.639	0.867

Notes: In the first three columns, standard errors are HAC robust with a bandwidth (Bartlett kernel) of four quarters. In the second three columns, standard errors are double-clustered by year and firm ID.

Table A11: Selected Summary Statistics of Firm Micro-Data

Code	Name	1990 Sales millions	1990 Sales share	1990 Employees thousands	1990 Employees share	2010 Sales millions	2010 Sales share	2010 Employees thousands	2010 Employees share
11	Agriculture, Forestry, Fishing and Hunting	7935.69	0.22	117.03	0.54	16028.05	0.13	157.12	0.49
21	Mining, Quarrying, and Oil and Gas Extraction	143716.64	3.95	557.43	2.59	728283.11	5.88	1020.34	3.18
23	Construction	20221.05	0.56	79.43	0.37	76357.58	0.62	249.05	0.78
42	Wholesale Trade	92141.14	2.53	316.32	1.47	220566.67	1.78	286.52	0.89
44	Retail Trade (I)	90746.30	2.49	697.08	3.24	755083.01	6.10	1802.16	5.61
45	Retail Trade (II)	71844.36	1.98	606.25	2.82	103823.53	0.84	393.61	1.23
48	Transportation and Warehousing (I)	202207.84	5.56	1278.65	5.94	490525.14	3.96	1484.93	4.62
49	Transportation and Warehousing (II)	22033.22	0.61	337.61	1.57	60473.08	0.49	505.72	1.57
53	Real Estate and Rental and Leasing	29186.20	0.80	335.31	1.56	97644.47	0.79	414.75	1.29
54	Professional, Scientific, and Technical Services	39096.89	1.07	367.54	1.71	142888.17	1.15	878.65	2.74
56	Administrative and Support and Waste Management and Remediation Services	30757.47	0.85	1055.62	4.91	97139.76	0.78	1432.94	4.46
61	Educational Services	2830.78	0.08	21.45	0.10	12350.02	0.10	93.87	0.29
62	Health Care and Social Assistance	14676.48	0.40	278.49	1.29	113475.92	0.92	874.35	2.72
71	Arts, Entertainment, and Recreation	4299.00	0.12	77.61	0.36	14592.82	0.12	140.96	0.44
72	Accommodation and Food Services	28474.41	0.78	741.09	3.45	126452.99	1.02	1949.83	6.07
81	Other Services (except Public Administration)	818.90	0.02	13.61	0.06	6028.48	0.05	71.80	0.22
99	Nonclassifiable Establishments	72587.03	2.00	402.04	1.87	337554.38	2.73	957.49	2.98
311	Food Manufacturing	89888.93	2.47	393.90	1.83	280503.94	2.26	773.89	2.41
312	Beverage and Tobacco Product Manufacturing	81514.55	2.24	527.06	2.45	323117.33	2.61	1128.77	3.51
313	Textile Mills	8281.43	0.23	109.85	0.51	1698.84	0.01	14.17	0.04
314	Textile Product Mills	1642.40	0.05	15.52	0.07	6512.22	0.05	31.62	0.10
315	Apparel Manufacturing	18195.52	0.50	179.87	0.84	50193.90	0.41	204.02	0.64
316	Leather and Allied Product Manufacturing	3220.64	0.09	18.56	0.09	8007.68	0.06	26.47	0.08
321	Wood Product Manufacturing	17079.99	0.47	94.34	0.44	32329.60	0.26	79.89	0.25
322	Paper Manufacturing	80736.02	2.22	485.96	2.26	137924.31	1.11	366.40	1.14
323	Printing and Related Support Activities	6959.89	0.19	64.35	0.30	17960.41	0.14	98.55	0.31
324	Petroleum and Coal Products Manufacturing	707106.33	19.44	1054.61	4.90	2666606.41	21.53	1718.55	5.35
325	Chemical Manufacturing	376182.22	10.34	2146.13	9.98	1234732.55	9.97	2362.52	7.36
326	Plastics and Rubber Products Manufacturing	23886.12	0.66	206.51	0.96	35081.57	0.28	139.19	0.43
327	Nonmetallic Mineral Product Manufacturing	22485.57	0.62	190.56	0.89	72492.53	0.59	242.54	0.76
331	Primary Metal Manufacturing	81200.98	2.23	438.98	2.04	308524.32	2.49	883.98	2.75
332	Fabricated Metal Product Manufacturing	34872.77	0.96	296.50	1.38	51755.70	0.42	172.89	0.54
333	Machinery Manufacturing	126838.80	3.49	823.90	3.83	303757.26	2.45	1060.24	3.30
334	Computer and Electronic Product Manufacturing	169692.71	4.66	1377.51	6.40	575557.43	4.65	1976.28	6.15
335	Electrical Equipment, Appliance, and Component Manufacturing	48369.82	1.33	447.71	2.08	96056.38	0.78	394.72	1.23
336	Transportation Equipment Manufacturing	553821.60	15.22	3359.30	15.62	1210603.25	9.77	3201.16	9.97
337	Furniture and Related Product Manufacturing	2516.33	0.07	23.81	0.11	15259.16	0.12	68.65	0.21
339	Miscellaneous Manufacturing	16722.13	0.46	143.46	0.67	63225.24	0.51	259.74	0.81
511	Publishing Industries (except Internet)	18829.43	0.52	149.15	0.69	42552.96	0.34	170.81	0.53
512	Motion Picture and Sound Recording Industries	13821.56	0.38	53.44	0.25	36968.97	0.30	112.21	0.35
515	Broadcasting (except Internet)	23400.47	0.64	187.74	0.87	189150.10	1.53	435.67	1.36
517	Telecommunications	158686.32	4.36	1045.88	4.86	1107220.16	8.94	3041.71	9.47
518	Data Processing, Hosting, and Related Services	1032.31	0.03	9.77	0.05	74483.99	0.60	273.40	0.85
519	Other Information Services	77053.51	2.12	384.84	1.79	45026.66	0.36	168.24	0.52
Total		3637612	100	21512	100	12386570	100	32120	100
USA		5963000		125840		14990000		153890	
Compustat share of USA			61.00		17.09		82.63		20.87

Table A12: Comparison of TFP Measures

	(1)	(2)	(3)
	Outcome: Cost-Share TFP		
OP TFP	1.117	1.141	1.025
	(0.009)	(0.008)	(0.006)
Firm FE		✓	✓
Sector x Time FE			✓
N	68,825	68,821	67,395
R^2	0.649	0.721	0.977

Notes: Standard errors are double-clustered by year and firm ID.

Table A13: Changing Macro Attention in Response to News

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

Notes: Data are from the Coibion et al. (2018a) survey, as described in Online Appendix G.

Table A14: Profit-Function Curvature and Attention to Macro Variables

Panel 1: Back-cast Error (Absolute Value)

Variable	Inflation		GDP Growth		Unemployment	
ProfitCurv$_{it}$	-1.172	-0.328	-0.072	-0.042	0.075	0.121
	(0.195)	(0.091)	(0.041)	(0.041)	(0.072)	(0.077)
Controls		✓		✓		✓
R^2	0.024	0.457	0.001	0.006	0.001	0.032
N	3,153	3,145	1,257	1,237	1,257	1,256

Panel 2: Keeping Track

Variable	Inflation		GDP Growth		Unemployment	
ProfitCurv$_{it}$	0.170	0.050	0.015	0.019	-0.005	-0.022
	(0.039)	(0.029)	(0.022)	(0.028)	(0.035)	(0.081)
Controls		✓		✓		✓
R^2	0.032	0.332	0.000	0.074	0.000	0.065
N	1,255	1,235	1,255	1,235	1,255	1,235

Notes: Standard errors are clustered by three-digit industry. Data are from the Coibion et al. (2018a) survey, as described in Online Appendix G.

Table A15: Profit Curvature in the Data

Summary Statistics

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

Correlates

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

Notes: The top panel gives summary statistics. The bottom panel gives normalized regression coefficients for a number of possible correlates. Standard errors, used to calculate the t-statistics, are clustered by three-digit industry.

Table A16: Curvature and Inflation BCE in Waves 1 versus 4

	Outcome: absolute Inflation BCE			
Wave	1		4	
ProfitCurv$_{it}$	-1.172	-0.328	-0.884	-0.330
	(0.195)	(0.091)	(0.181)	(0.126)
Controls		✓		✓
R^2	0.024	0.457	0.033	0.268
N	3,153	3,145	1,257	1,256

Notes: Standard errors are clustered by three-digit industry.

Figure A1: Frequency over Time of Each Word in MacroAttention (Part I)

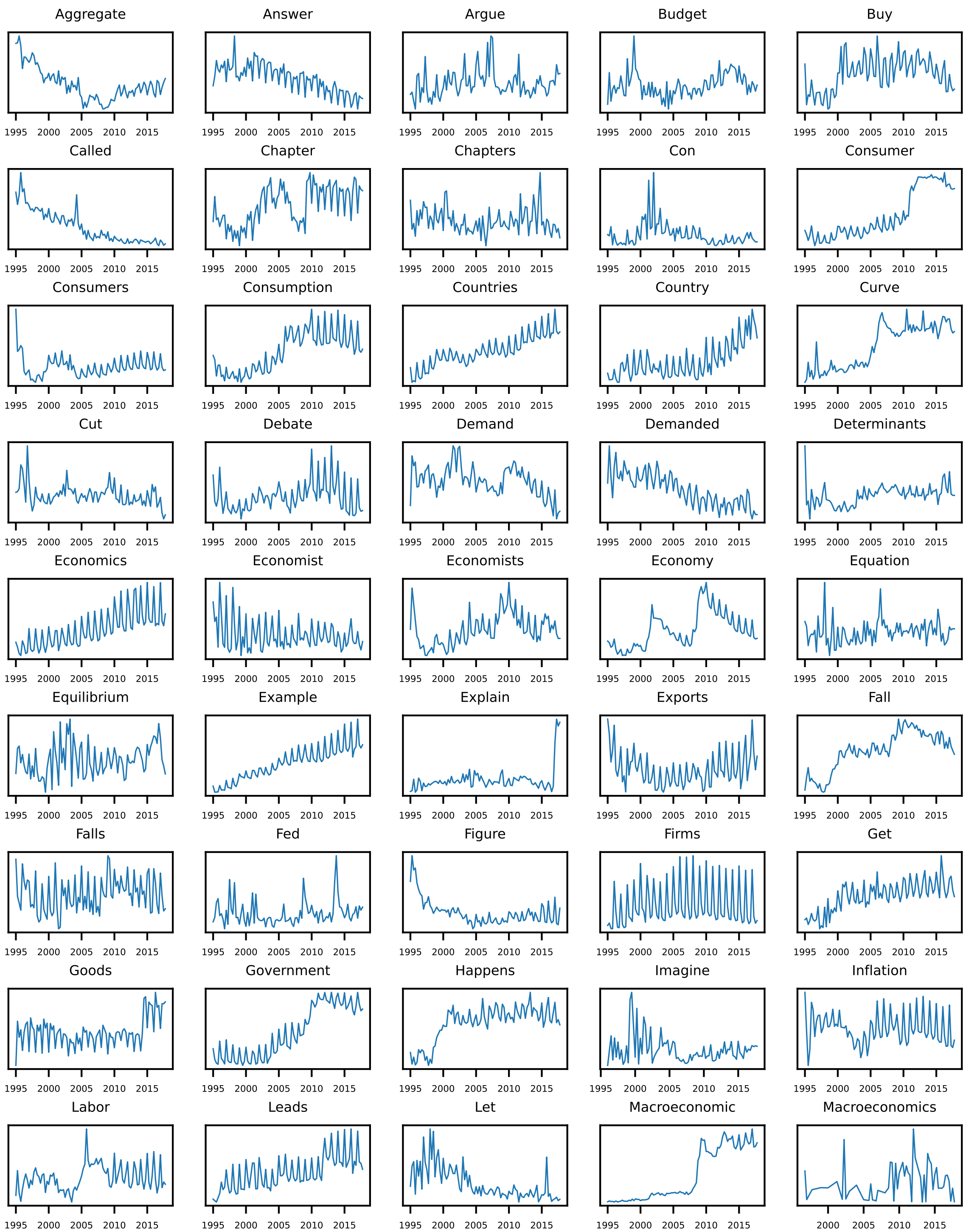


Figure A1: Frequency over Time of Each Word in MacroAttention (Part II)

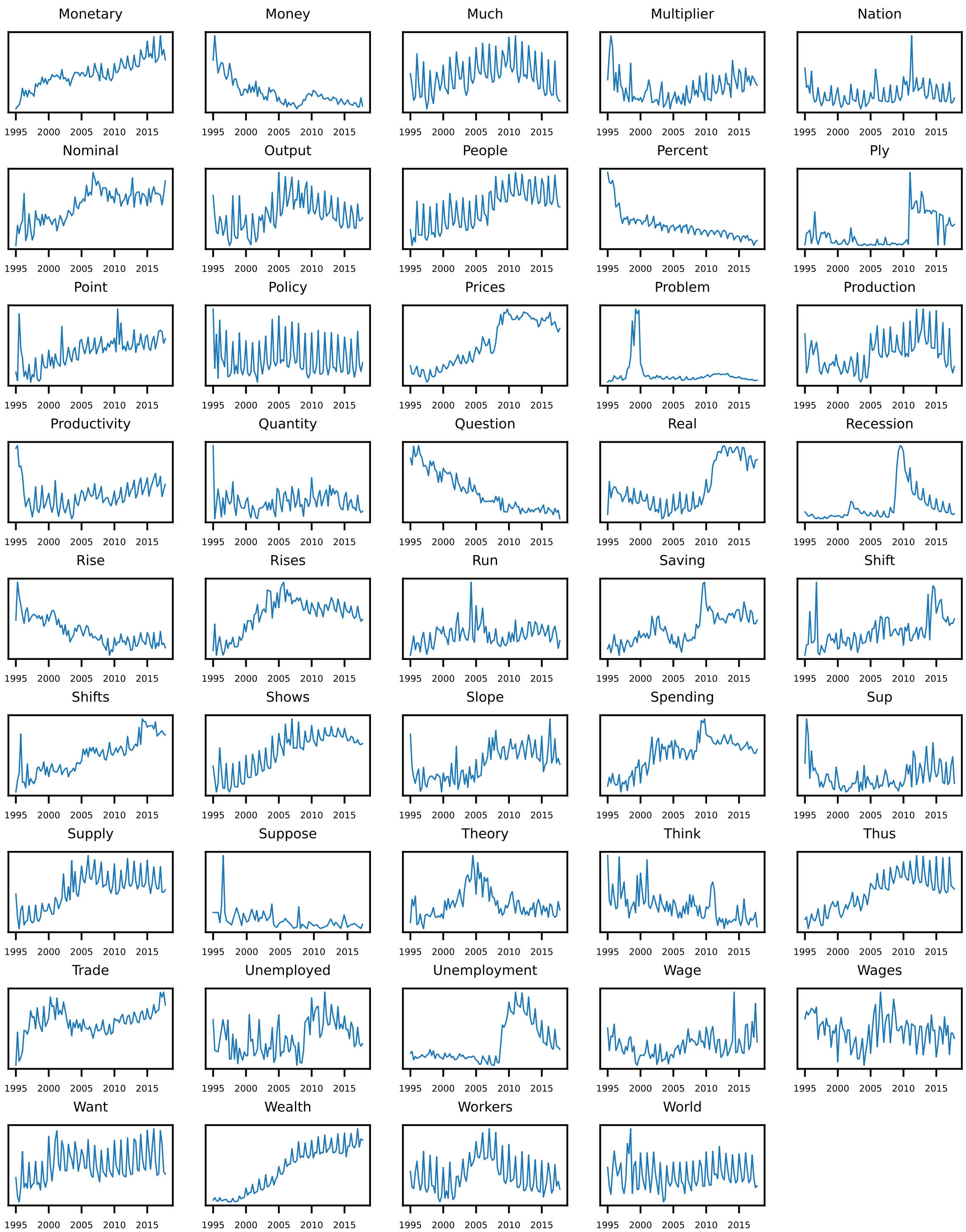
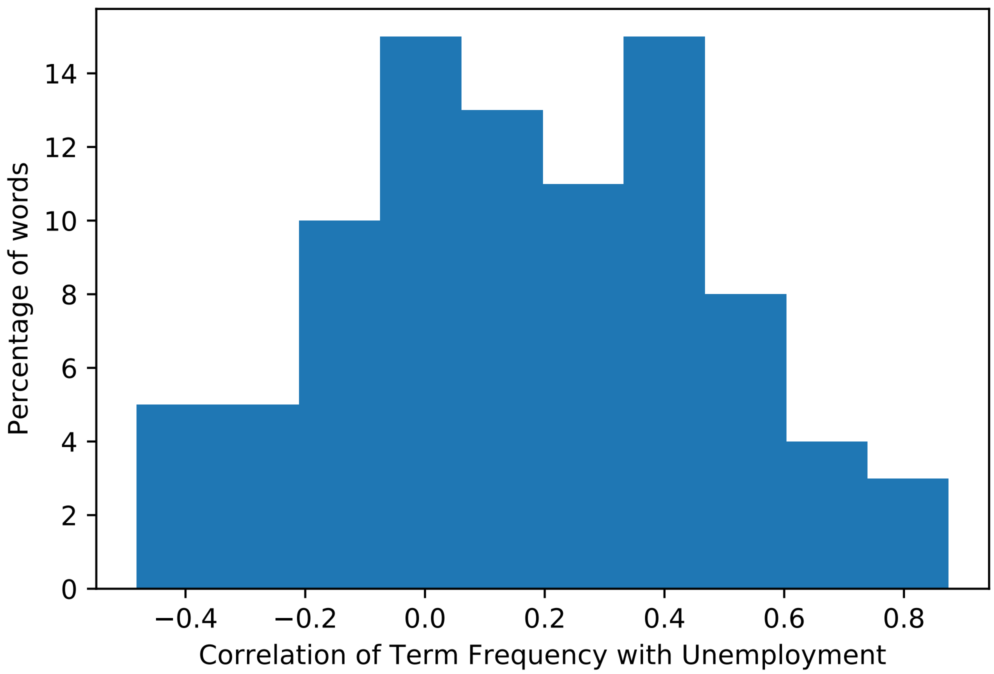
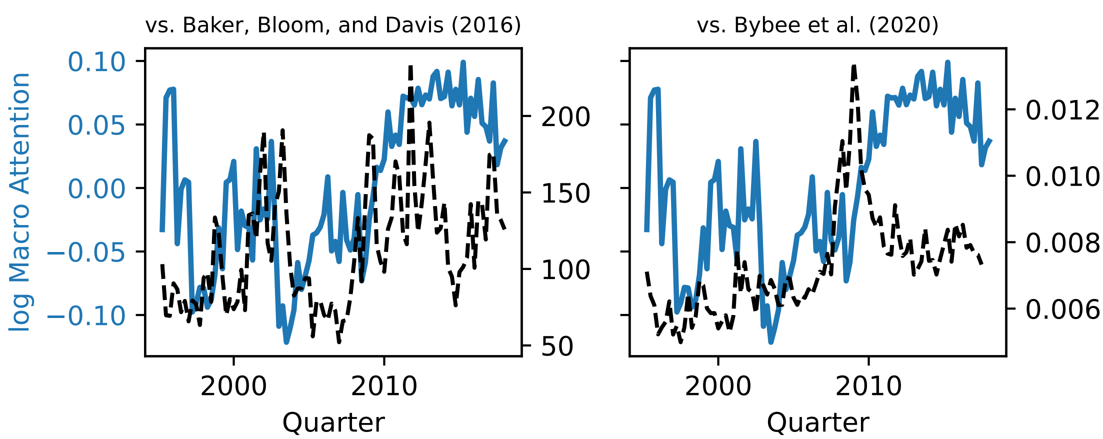


Figure A2: Correlations with Unemployment by Word



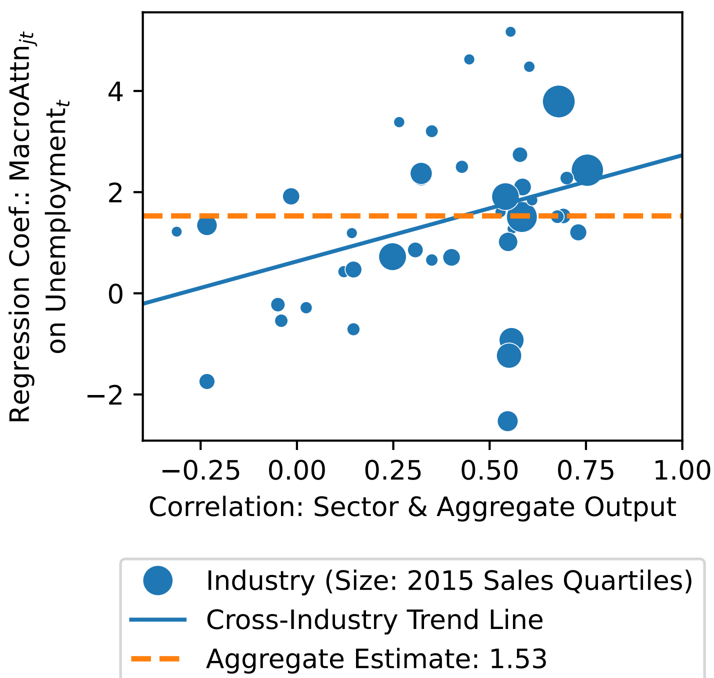
Notes: Correlations are calculated at the quarterly frequency.

Figure A3: Relationship of Macro Attention to News Indices



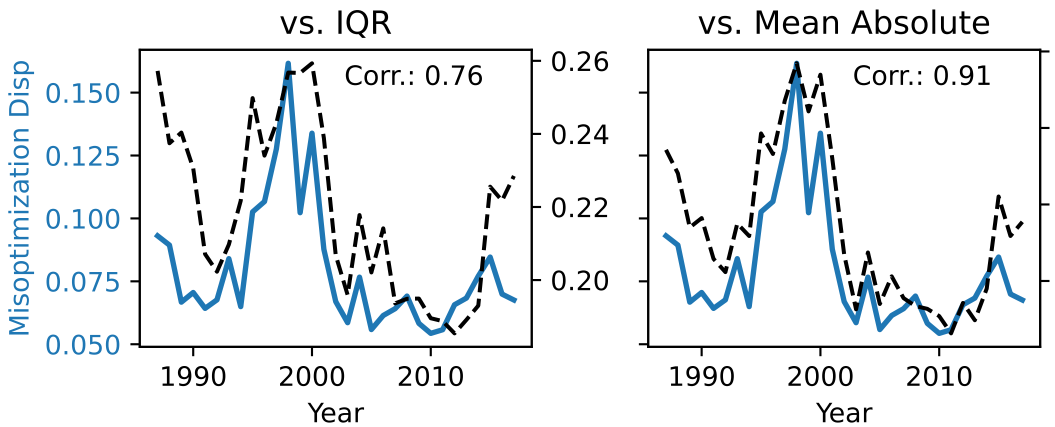
Notes: The blue line, measured by the left axis of each plot, is the log of Macro Attention, adjusted for seasonality. The left panel is the news index of Baker et al. (2016). The right panel is the sum of five macroeconomic topics from the Bybee et al. (2021) Structure of News dataset, described in Section 2.3 of the main text.

Figure A4: Industry-Specific Cyclicality of Macro Attention



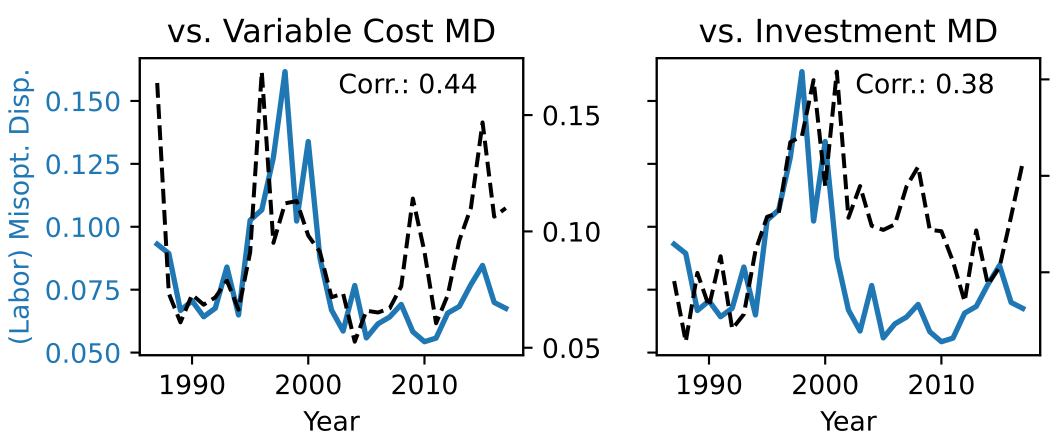
Notes: The horizontal axis is the correlation of sectoral and aggregate nominal GDP, calculated as described in Section 2.3 of the main text. The vertical axis is the regression coefficient of log sectoral macro attention, net of quarterly fixed effects, on the US unemployment rate. The dashed orange line is the estimate of the same using aggregate Macro Attention. The dots are sized based on quartiles of total sales in Compustat in 2015. The blue solid line is a cross-industry linear regression line.

Figure A5: Relationship of Misoptimization Dispersion with Other Statistics of "Spread"



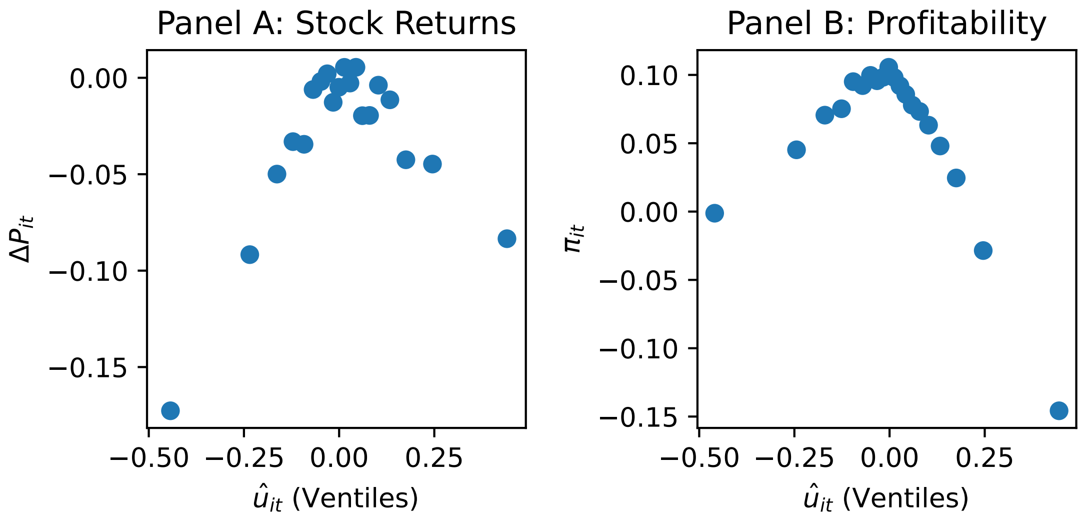
Notes: The blue line, measured by the left axis of each plot, is Misoptimization Dispersion as defined in Section 5. The black dashed line on the left is the (optimal-sale-weighted) interquartile range of the distribution of $\hat{u}_{it}$. The black dashed line on the right is the (optimal-sale-weighted) average of $|\hat{u}_{it}|$.

Figure A6: Relationship of Misoptimization Dispersion Across Inputs



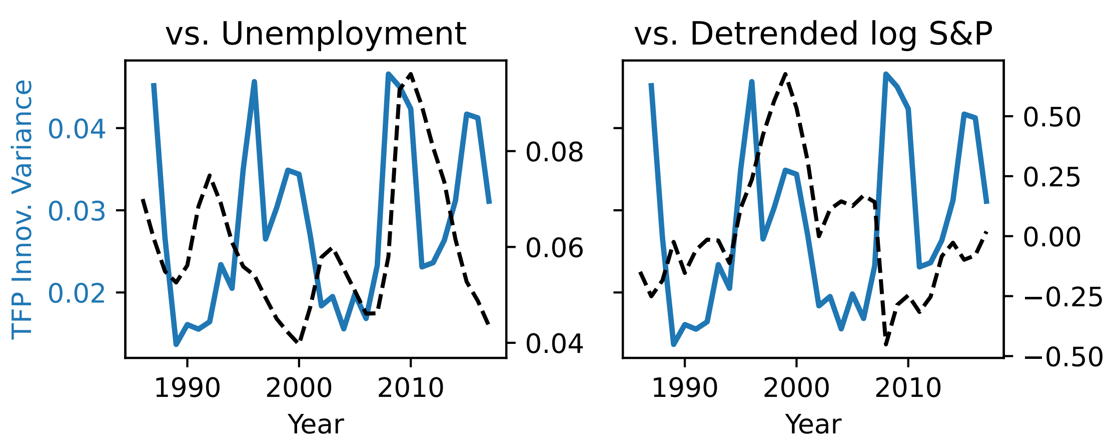
Notes: The blue line, measured by the left axis of each plot, is Misoptimization Dispersion as defined in Section 5. The black dashed line on the left is Misoptimization Dispersion for total variable cost expenditures. The black dashed line on the right is the same for investment rates.

Figure A7: Binned Scatterplots of Misoptimization and Firm Performance



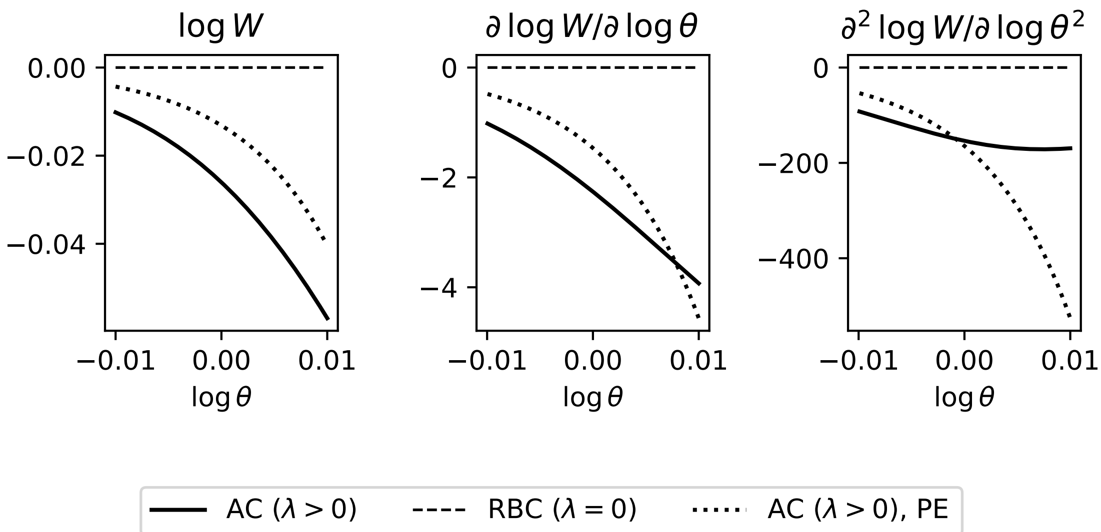
Notes: Each plot is a binned scatter plot, where dots represent means in ventiles of the x-axis variable, $\hat{u}_{it}$, and corresponding means of the outcome variable, stock returns (left) or profitability (right), after removing industry-by-time fixed effects.

Figure A8: Relationship of TFP Innovation Variance with Macro Variables



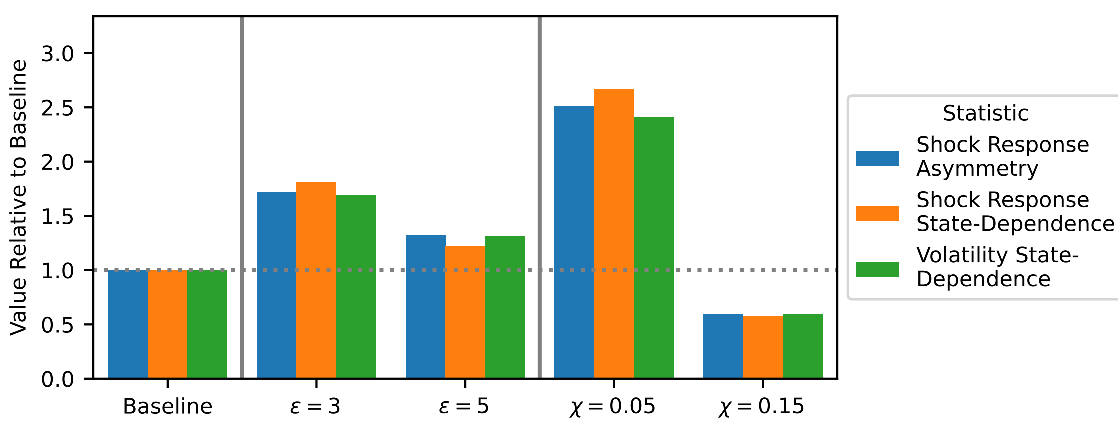
Notes: "TFP Innovations" are the residuals from an AR(1) model for θ_{it} with firm and sector-by-time fixed effects, as described in Section 5.5.2 of the main text. We calculate their average, consistent with our main calculations of Misoptimization Dispersion, using optimal-sales-weights.

Figure A9: The Attention Wedge and Its Derivatives



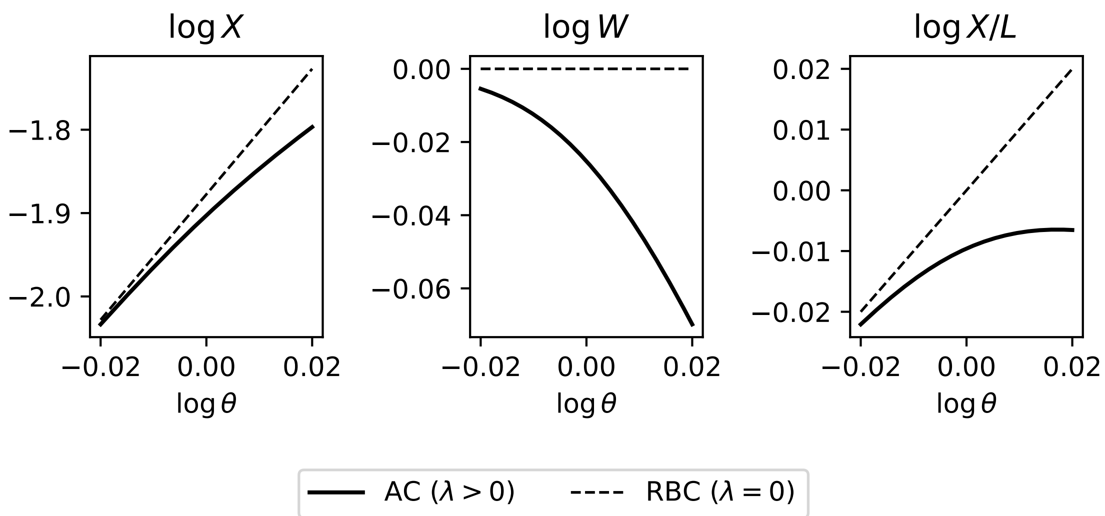
Notes: The panels respectively show the level, first derivative, and second derivative of the log attention wedge in the log state. The "Partial Equilibrium" thought experiment, plotted in each figure as a dotted line, is for firms to best-reply to the output and wages of the counterfactual RBC equilibrium.

Figure A10: Robustness of Numerical Results to Parameter Choices



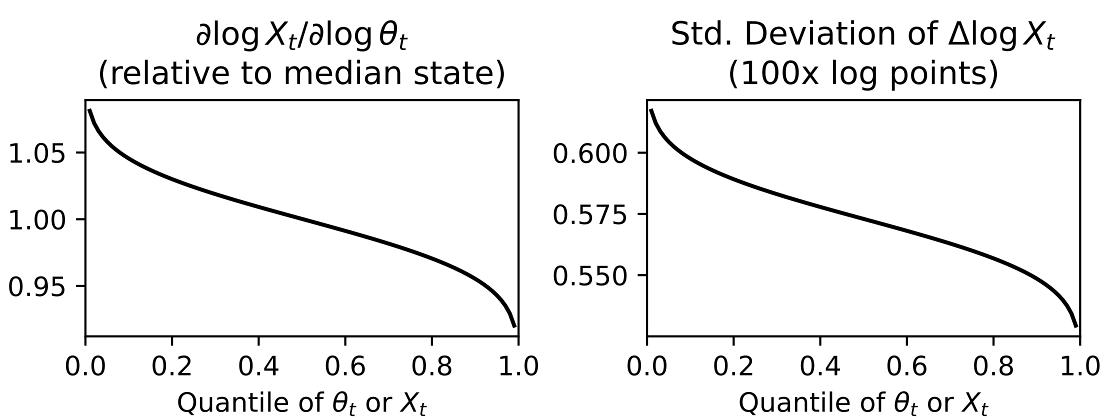
Notes: We re-calibrate the model under each indicated parameter choice. The outcomes and their interpretation are described in the text.

Figure A11: Output, Wedge, and Labor Productivity with GHH Preferences



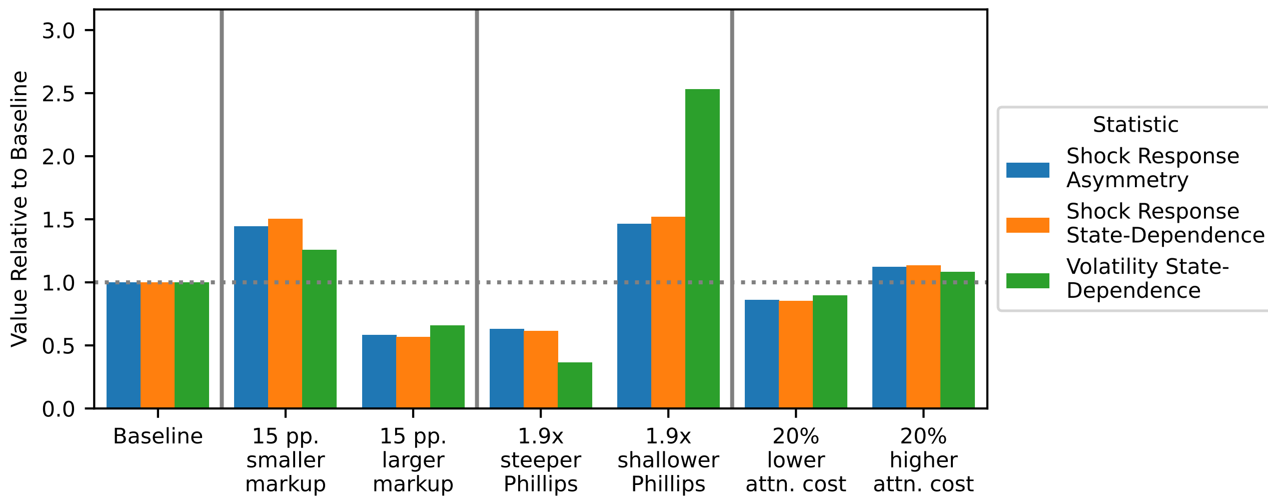
Notes: This recreates Figure 3 from the main analysis in the variant model with Greenwood et al. (1988) preferences, described in Online Appendix D.

Figure A12: Asymmetric Shock Response and Stochastic Volatility with GHH Preferences



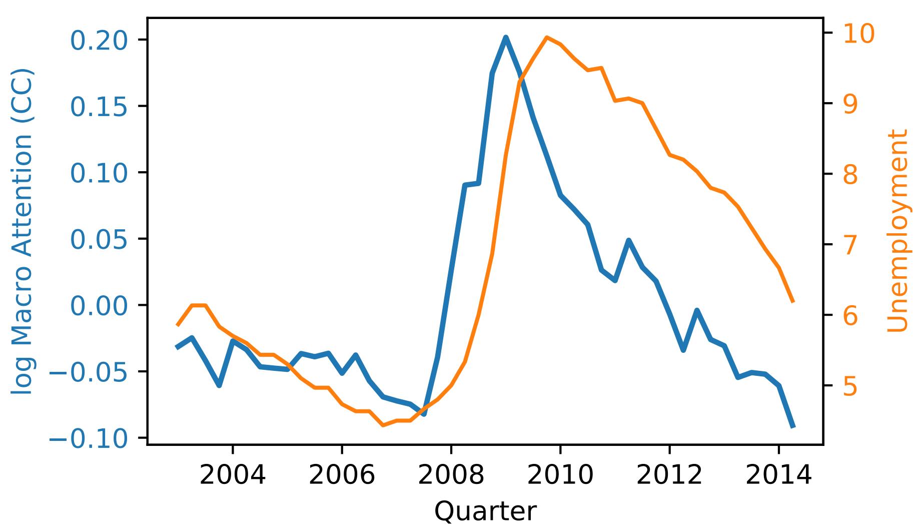
Notes: This recreates Figure 4 from the main analysis in the variant model with Greenwood et al. (1988) preferences, described in Online Appendix D.

Figure A13: Predictions in Counterfactual Scenarios



Notes: The scenarios are described in the main text. The outcomes are the same as in Figure A10, and are described in that Figure's discussion.

Figure A14: Conference-Call Macro Attention and Unemployment



Notes: The left axis and blue line show our estimate of Macro Attention based on conference-call data, in log units net of seasonal trends. The right axis and orange line show the US unemployment rate.