# Attention Cycles

Joel P. Flynn[*] Karthik A. Sastry[†]
Yale Princeton

February 7, 2024

**Abstract**

We study how the extent of bounded rationality evolves with the state of the economy. We introduce a business-cycle model in which firms face a cognitive cost of making precise decisions and decide how much to "pay attention" to specific states. We analytically characterize equilibrium with non-parametric, state-dependent stochastic choice and nonlinear aggregation and dynamics. Firms have greater incentives to direct attention toward states of lower aggregate consumption because they are owned by risk-averse households. This mechanism generates counter-cyclical attention, pro-cyclical mistakes, and an endogenous attention wedge that depresses aggregate productivity when attention is low. We test and validate these macroeconomic predictions and our proposed microeconomic mechanism using novel measures of US public firms' input-choice mistakes and a textual proxy for their attention toward macroeconomic conditions. When calibrated to match our measurements, attention cycles generate a significant fraction of the observed stochastic volatility of output growth and shock propagation asymmetries.

[*]Yale University, joel.flynn@yale.edu.

[†]Princeton University, ksastry@princeton.edu.

We thank Daron Acemoglu, Marios Angeletos, Gadi Barlevy, Martin Beraja, Ricardo Caballero, Larry Christiano, Ian Dew-Becker, Martin Eichenbaum, Xavier Gabaix, Cosmin Ilut, Alexandre Kohlhas, Chen Lian, Omar Licandro, Francesco Lippi, Jeremy Majerovitz, Stephen Morris, Giuseppe Moscarini, George Nikolakoudis, Kristoffer Nimark, Christina Patterson, Giorgio Primiceri, Richard Rogerson, Pari Sastry, Larry Schmidt, Niklas Schmitz, Chris Sims, Alp Simsek, Stephen Terry, Adrien Verdelhan, Gianluca Violante, Iván Werning, Christian Wolf, and seminar participants at the 2021 NBER Summer Institute, Northwestern, the Chicago Fed, the Minneapolis Fed, the ECB, Princeton, Stanford, Harvard, Wharton, Chicago Booth, Yale, Boston U., the 2022 Review of Economic Studies Tour (U. of Manchester, Central European U., PSE), National U. of Singapore, Zhejiang U., the IIES at Stockholm U., U.C. Berkeley, LSE, Cornell, the 2022 Econometric Society NASM, the Rome Workshop in Macroeconomics, the 2022 SED Conference, the China Star Tour, the MIT Macroeconomics Lunch, the MIT Theory Lunch, and the MIT Sloan Finance Lunch. First posted version: May 4, 2020.

Firms often make decisions which, *ex post*, are inconsistent with profit maximization. An influential explanation is that firms' managers face constraints on their attention and decisionmaking capacity that introduce "mistakes" (Simon, 1947). Although this view relaxes perfect optimization, it still preserves incentives at its core—even if plans are imperfect, an optimizing manager may design them to minimize mistakes in the economic circumstances in which mistakes are most costly. By implication, the state of the economy may be a central determinant of apparent bounded rationality. Moreover, the macroeconomic consequences of bounded rationality may depend on the state of the economy.

This paper studies the co-determination of business cycles and *attention cycles*, aggregate fluctuations in cognitive effort and mistakes. Theoretically, we develop a general-equilibrium model in which heterogeneous firms choose non-parametric, state-dependent production plans and face a cognitive control cost of choosing more precise plans. Our modeling approach allows us to analytically study state-dependent attention, a phenomenon away from which much of the existing literature on macroeconomics with decision frictions abstracts (e.g., Woodford, 2003; Nimark, 2008; Maćkowiak and Wiederholt, 2009; Lorenzoni, 2009; Angeletos and La'O, 2010; Gabaix, 2019; Lian, 2021; Kohlhas and Walther, 2021; Pavan, 2023). In our model, firms want to allocate costly attention toward the states in which decision mistakes are most costly. In particular, because firms are owned by risk-averse investors, incentives for precise, attentive behavior are high when aggregate consumption is low. Therefore, in equilibrium, the model predicts higher cognitive effort and smaller decision mistakes in downturns versus booms. Endogenous attention, in turn, induces asymmetric propagation of positive and negative macroeconomic shocks.

To test the model empirically, we introduce new strategies to measure precision in "what firms do," an exact measure of mistakes as defined by our model, and attention to the macroeconomy in "what firms say," a suggestive proxy for cognitive effort. At the macro level, firms make smaller input-choice mistakes and speak more about the macroeconomy in downturns. At the micro level, input-choice mistakes reduce firms' stock returns, this market punishment is significantly harsher during downturns, and firms that make smaller mistakes use more attentive language. When the model is calibrated to match the data, we find that endogenous fluctuations in attention contribute significantly to the observed asymmetric and state-dependent effects of macroeconomic shocks. That is, realistic attention cycles help explain fast crashes and slow recoveries.

**Model.** We build on a standard Neoclassical real business cycle model with heterogeneous firms. A representative household owns, supplies labor to, and consumes the output of monopolistically competitive firms with heterogeneous productivities. An aggregate shock shifts the productivity distribution and generates a business cycle.

We model firms' state-dependent attention by subjecting firms to a cognitive *control cost* of choosing any non-parametric stochastic choice rule, *i.e.*, a mapping from all payoff-relevant microeconomic and macroeconomic states to a distribution of production. These control costs are greater in a given state when decisionmaking is more precise in that state. Firms therefore choose costly "attention," or cognitive effort, to optimally trade off these costs with the benefit of increasing expected, stochastically discounted profits in every state of the world. Our approach of directly modelling stochastic choice allows for equilibrium analysis of rich decision frictions in a way that is robust to the specific and difficult-to-test microfoundation for why precise decisionmaking is hard, like costly information acquisition, behavioral biases, or trembling hands in implementation (see, e.g., Caplin and Martin, 2015; Morris and Yang, 2022; Flynn and Sastry, 2023). Despite this generality, we show that equilibrium analysis in our setting is well-posed and tractable.

**Attention Cycles and Their Consequences.** In partial equilibrium, bounded rationality is lower when stakes are higher. That is, firms pay more attention, measured by the extent of their cognitive effort, and make smaller misoptimizations, measured by the variance of their actions around the *ex post* ideal, when the benefit of paying attention is high. These benefits are driven by two forces: the sensitivity of firms' dollar profits to mistakes and the stochastic discount factor. The former, which we call the *profit-curvature channel*, depends on the price elasticity of demand that firms face and the elasticity of wages to real output. While this channel has an ambiguous sign in general, we argue that in realistic parameterizations it pushes toward high attention when aggregate output and firm productivity are low. The latter, which we call the *risk-pricing channel*, depends on the firm owner's risk aversion. This channel always pushes toward high attention when aggregate consumption is low.

We next describe the *general-equilibrium* properties of "attention cycles." We show that, if household relative risk aversion exceeds an empirically modest lower bound of one plus the elasticity of real wages to output and wages are not too procyclical, then equilibrium attention is counter-cyclical: cognitive effort decreases in output and the average size of agents' mistakes increases in output. Intuitively, market incentives push firms toward paying more attention to decisions when the economy is doing poorly.

We finally study the macroeconomic consequences of attention cycles. We show that output is the product of the counterfactual output under no cognitive friction with an *attention wedge* that is less than one. The wedge arises because inattentive, stochastic choices translate into dispersion of value marginal products across firms ("misallocation") and reduced aggregate total factor productivity (TFP). Due to counter-cyclical attention, the wedge widens when the economy is booming and firms optimize less precisely. Owing to this mechanism, misallocation is endogenously higher in booms than recessions. Starkly, equilibrium aggregate TFP can even be non-monotone in the distribution of underlying microeconomic productivity. Moreover, the translation of negative shocks into increased attention leads to asymmetric, state-dependent shock propagation and endogenous stochastic volatility.

**Measurement.** The model makes micro and macro predictions for both choice mistakes ("misoptimization") and cognitive effort ("attention"). Toward testing these predictions, we introduce new methods to measure each of these variables for firms.

We measure misoptimization, as defined in the model, using data on firm choices and fundamentals in Compustat from 1986-2018. We do so by combining our theoretical results with conventional econometric techniques for estimating firms' productivity and policy functions (e.g., Ilut et al., 2018).

Attention, unlike misoptimization, has no direct analog in observable choices. Our premise is that, when management uses more language to describe the relationship of their decisions with a specific topic, they are revealing greater cognitive effort toward adapting decisions to those factors. Specifically, we quantify firms' attention toward *macroeconomic conditions*, a factor faced to some extent by all firms. Using textual data from firms' regulatory filings (forms 10-Q/K) and conference calls, we measure firms' attention using a natural-language-processing technique that builds on Hassan et al. (2019). The attention measure complements the "model-based" misoptimization measure by being constructed in a "model-free" fashion.

**Testing the Theory.** As nothing in our estimation imposes that measured misoptimizations are bad for firms' performance, we first test for this in the data. As predicted by the theory, we find that both positive and negative misoptimizations (i.e., "over"- and "under"-hiring) have negative effects on firms' stock returns and profitability. Moreover, misoptimizations do not predict higher future productivity, stock returns, or profitability. Together, these results rule out stories in which misoptimizations purely reflect misspecification, wedges, or forward-looking adjustments with short-run costs and long-run benefits.

We next find that the size of firms' misoptimizations is pro-cyclical, consistent with our theoretical prediction. This result holds under a number of alternative constructions of misoptimizations, including those that account for time-varying financial frictions and adjustment costs; under alternative aggregation that adjusts for compositional bias; and on the entire Compustat sample back to 1950. This finding enriches the literature's observation, replicable in our data, that the variance of firm productivity spikes in recessions (e.g., Bloom et al., 2018). Our new fact implies that choice dispersion *conditional on* productivity decreases exactly when productivity itself becomes more volatile.

We next test the theoretical mechanism underlying pro-cyclical mistakes—that incentives to avoid mistakes vary with the state of the aggregate economy. We find that the negative effect of misoptimizations on stock returns significantly strengthens when the market is doing poorly. For example, in the stock market trough of 2008, a misoptimization was more than 6 times more costly for returns than it would have been in a year with 10% aggregate returns. By contrast, the effect of misoptimizations on *profitability* is close to constant. This supports a primary role for the risk-pricing channel and a limited role for the profit-curvature channel in driving pro-cyclical misoptimization.

We finally test the model's predictions for attention. At the aggregate level, we find that macroeconomic attention is counter-cyclical. At the firm level, we find that macroeconomic attention is associated with smaller misoptimizations. These findings, together, validate the consistency of our two approaches of measuring attention at both the microeconomic and macroeconomic levels and provide further support for the theory's prediction of counter-cyclical attention.

**Quantification.** To assess the effects of attention cycles on business cycles, we estimate our model to match our empirical findings. In the model, negative productivity shocks have a 7% larger effect on aggregate output than positive shocks of the same size. This is 25% of the asymmetry estimated by Ilut et al. (2018) using industry-level data. Similarly, in the model, a shock that replicates a "Great-Recession-sized" 5% peak-to-trough reduction in output generates also an 11% increase in the conditional volatility of output.

This is 19% of the increase in uncertainty about output growth between the trough of the Great Recession and the preceding peak as measured by Jurado et al. (2015). Our model therefore explains a significant fraction of observed non-linearities and state-dependence of macroeconomic dynamics, solely as consequences of state-dependent attention.

**Related Literature.** Our study's primary contribution is to theoretically justify and empirically validate the hypothesis that agents' inattention systematically varies with the business cycle because of changing incentives. The idea that incentives drive attention is obtained in classic single-agent models (e.g., Sims, 2003; Gabaix, 2014). By contrast, most macroeconomic models with cognitive and/or informational frictions (e.g., Woodford, 2003; Van Nieuwerburgh and Veldkamp, 2006; Maćkowiak and Wiederholt, 2009; Lorenzoni, 2009; Angeletos and La'O, 2010; Gabaix, 2020; Kohlhas and Walther, 2021) do not study how optimal state-dependent attention varies over the business cycle. Three exceptions are Mäkinen and Ohl (2015), Benhabib et al. (2016), and subsequent work by Chiang (2023).[1] In contrast to all three studies, we emphasize a distinct risk-pricing mechanism, develop novel measures of firms' attention and misoptimization, and use these measures to directly test and quantify both the macroeconomic and microeconomic predictions of the theory.

Methodologically, we show how to use a control-cost model to tractably and analytically study state-dependent attention in general equilibrium while allowing for rich choice patterns. This relates to a literature that numerically studies state-dependent attention under mutual-information costs (Afrouzi, 2023; Afrouzi and Yang, 2021; Matějka, 2016; Stevens, 2020; Turen, 2023). In contrast to these studies, our model allows for an exact analytical characterization of firms' choices. This is crucial in the theory to obtain an analytical characterization of equilibrium and in our empirical work to obtain a direct and realistic mapping from the model's predictions to observed choices.[2] Our approach is similar in spirit to that of Ilut and Valchev (2023), who model cognitive constraints as costly learning of policy rules. These authors focus on different applications and use numerical analysis of aggregative equilibrium.

[1] In Mäkinen and Ohl (2015), decreasing returns to scale lead firms to demand more information when aggregate productivity is low. Chiang (2023) considers a complementary mechanism in the context of a model where firms acquire Gaussian signals about the state. Benhabib et al. (2016) predict lower information demand in recessions, a prediction at odds with our empirical findings.

[2] Costain and Nakov (2015, 2019) also apply the control-cost model but study different questions, regarding price dynamics, and do not provide comparable analytical results.

Our findings regarding attention to the macroeconomy contribute to a literature measuring this object in firm and household surveys (Coibion et al., 2018; Link et al., 2024). Kuang et al. (2024) estimate that the conditional precision of professional forecasters' signals increases when perceived recession risk and volatility are high, consistent with our attention cycles hypothesis. Our textual measure of attention also relates to subsequent work by Song and Stern (2023), who develop a text-based measure of firms' attention to macroeconomic news and use it to study the importance of inattention for the propagation of monetary shocks.

An additional contribution is to show how endogenous attention contributes to microeconomic and macroeconomic volatility. A key lesson from our analysis is that endogenous volatility in choices conditional on fundamentals may have an opposite cyclicality to exogenous volatility in fundamentals, helping reconcile the finding of counter-cyclical TFP variance (Bloom et al., 2018) with more ambiguous evidence for the volatility of other variables (e.g., Bachmann and Bayer, 2014; Dew-Becker and Giglio, 2020).

**Outline.** In Section 1, we introduce our model. In Section 2, we present our theoretical results. In Section 3, we describe the data and measurement. In Section 4, we test the model's six main predictions. In Section 5, we calibrate our model and analyze the impact of attention cycles on macroeconomic dynamics. Section 6 concludes.

# 1 Model

We first describe our model, a Neoclassical real business cycle model with a stochastic choice friction for intermediate goods firms that captures inattention and mistake making.

## 1.1 Consumers and Final Goods

Time periods are indexed by $t \in \mathbb{N}$. A representative household has constant relative risk-aversion preferences over final-good consumption $C_t$ and labor $L_t$. Their payoffs are:

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

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 aggregate final good is produced from a continuum of intermediate goods, indexed by $i \in [0, 1]$, by a representative, perfectly competitive firm using a constant-elasticity-of-substitution production function

$$X_t = X(\{x_{it}\}_{i\in[0,1]}) = \left(\int_{[0,1]} x_{it}^{\frac{\epsilon-1}{\epsilon}}\,\mathrm{d}i\right)^{\frac{\epsilon}{\epsilon-1}} \tag{2}$$

where $\epsilon > 1$ is the elasticity of substitution. This firm buys inputs at prices $\{q_{it}\}_{i\in[0,1]}$ and sells its output at a normalized price of one. The household owns equity in firms that produce intermediate goods, receiving profits $\{\pi_{it}\}_{i\in[0,1]}$.

Wages are determined by the following wage rule:

$$w_t = \bar{w}\cdot\left(\frac{X_t}{\bar{X}}\right)^{\chi} \tag{3}$$

where $\bar{w} > 0$ and $\bar{X} > 0$ are constants, and $\chi \geq 0$ (inversely) measures the extent of real wage rigidity. Households supply labor to meet firms' labor demand at the wages from Equation 3. Describing the labor market via a wage rule is for technical simplicity as it allows us to study the economy via a scalar fixed-point equation.[3] Moreover, it allows our model to parsimoniously match the empirical acyclicality of real wages (Solon et al., 1994; Grigsby et al., 2021). In Appendix D, we micro-found this wage rule when households have Greenwood et al. (1988) preferences and markets clear in the standard fashion. We show in Appendix E.1 that our results are quantitatively robust to considering wages set in this manner.

### 1.2 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:

$$x_{it} = \theta_{it} L_{it} \tag{4}$$

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

We parameterize the stochastic process for productivity in the following way that allows for both rich cross-firm heterogeneity and "aggregate productivity" shocks. We let $\theta_t \in \Theta$ be an aggregate productivity variable. For each firm $i$, $(\theta_{it}, \theta_t)$ follows a first-order Markov process of the following form: the transition density can be factored as $h(\theta_{it}, \theta_t \mid \theta_{i,t-1}, \theta_{t-1}) = h_{agg}(\theta_t \mid \theta_{t-1}) h_{idio}(\theta_{it} \mid \theta_{i,t-1}, \theta_t, \theta_{t-1})$. In words, aggregate productivity follows a first-order Markov process by itself and firm-level productivity depends flexibly on its own lag and the current and lagged aggregate state. This model allows

[3] Blanchard and Galí (2010) and Alves et al. (2020) use similar wage-rule formulations.

for aggregate productivity shocks to affect all moments of the idiosyncratic productivity distribution. For example, it is consistent with the model empirically estimated by Bloom et al. (2018) wherein aggregate shocks affect both the mean of firm level productivity and its idiosyncratic variance.

Throughout our analysis, we will reference two aggregate summaries of the productivity distribution. The first is the cross-sectional distribution of $\theta_{it}$ at time $t$, $G_t \in \Delta(\Theta)$. The second is the statistic,

$$\hat{\theta}(G) = \left( \int_\Theta \theta_i^{\epsilon-1} \, \mathrm{d}G(\theta_i) \right)^{\frac{1}{\epsilon-1}} \tag{5}$$

which is a quasi-arithmetic mean of the productivity distribution. Because the integrand inside is an increasing function, $\hat{\theta}(G)$ is monotone in distributions $G$ and $G'$ ranked by first-order stochastic dominance.

### 1.3 Modelling Attention via Costly Control

Due to cognitive constraints, these agents struggle to perfectly adapt their production choice to the microeconomic and macroeconomic state without making mistakes. We model this by having them choose stochastic choice rules, or mappings from states to a distribution of production outcomes, subject to a cognitive cost. Both the choice set and the cost are flexible enough to allow for unrestricted, state-dependent attention, as we formalize below.

**Decision State and Choice Variable.** We first 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.[4] Firms believe that $z_{it}$ follows a first-order Markov process with transition density $f(z_{it}|z_{i,t-1})$. As we will soon clarify, this conjecture combines the first-order Markov process for $(\theta_{it}, \theta_t)$ with a conjecture for how aggregates depend on the productivity distribution. We denote the corresponding set of possible transition densities by $\mathcal{F}$. At time $t$, each firm $i$ observes the history of states, $\{z_{is}\}_{s<t}$, but not the contemporaneous value $z_{it}$.

The firm's choice variable is 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 function $p(\cdot \mid z_{it})$ when the decision state is $z_{it}$. A firm using rule $p$ commits

[4] 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 (3).

to delivering the realized quantity $x_{it} \in \mathcal{X}$ to the market and hiring sufficient labor in production. It is fully equivalent to interpret firms' choices as committing to hire $L_{it}$ workers and producing $x_{it} = \theta_{it} L_{it}$.

**Costly Control.** We model the cognitive cost of attention 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. 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 $z_{it}$.

The basic idea that we wish to embody is that playing actions that are more *precise* is more costly. To make this tension clear, and also to make the analysis tractable, we specialize to the following cost functional that equals the negative expected entropy of the action distribution multiplied by $\lambda_i > 0$:

$$c(p, \lambda_i, z_{i,t-1}, f) = \lambda_i \int_{\mathcal{Z}} \int_{\mathcal{X}} p(x \mid z_{it}) \log(p(x \mid z_{it}))\, \mathrm{d}x\, f(z_{it} \mid z_{i,t-1})\, \mathrm{d}z_{it} \tag{6}$$

These costs embody the idea that it is difficult for firms to avoid making mistakes (see Morris and Yang, 2022, and Flynn and Sastry, 2023, for a discussion). In particular, they are consistent with the experimentally verified fact that the size and character of choice mistakes respond to incentives (Woodford, 2020), the central premise of our analysis.

**Comparison to Other Models of Inattention.** This notwithstanding, there are other costs that could embody the notion of incentives-driven attention. Two prominent examples are unrestricted costly information acquisition (Caplin and Dean, 2015), including the mutual information model of Sims (2003), and behavioral sparse maximization (Gabaix, 2014).

In contrast to these alternative models, our model jointly has three properties that are important for our analysis. First, our model admits an analytical solution for the globally optimal policy, despite the model's state-dependence. In particular, no comparable results exist in the literature on rational inattention with mutual information costs.[5] Second, our model allows for tractable and well-posed analysis of equilibrium and its comparative statics. The literature on equilibrium results with general information acquisition and/or

[5]Even when approximated as a *state-dependent* quadratic function of the action, our objective is not "linear-quadratic" in the sense studied by Sims (2010), Maćkowiak et al. (2018), or Afrouzi and Yang (2021).

sparsity-based decision frictions does not describe comparable comparative statics (e.g., Angeletos and Sastry, 2023; Hébert and La'O, 2023; Gabaix, 2019). Third, our model allows for rich, state-dependent, and stochastic mistakes. This will prove a more realistic prediction in our setting than, for example, the mutual information model's prediction of sparse actions (Matějka, 2016; Jung et al., 2019; Stevens, 2020) or the sparsity model's prediction of deterministic mistakes.

A key cost relative to these alternative models, however, is that our predicted optimal policy functions will not include a force of "anchoring" toward commonly played actions and therefore under-reacting to shocks. To gauge the importance of missing this force, we discuss and analyze extensions of the model with unrestricted and parametric information acquisition in Section 2.5 and Appendices F.2 and F.3.

### 1.4 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 $M(z_{it})$. We define "risk-adjusted profits" as the product of these terms:

$$\Pi(x_{it}, z_{it}) = \pi(x_{it}, z_{it}) \cdot M(z_{it}) \tag{7}$$

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

$$\pi(x_{it}, z_{it}) = x_{it}\left(x_{it}^{-\frac{1}{\epsilon}} X_t^{\frac{1}{\epsilon}} - \frac{w_t}{\theta_{it}}\right) \qquad\qquad M(z_{it}) = X_t^{-\gamma} \tag{8}$$

The firm's dynamic choice problem reduces to a series of one-shot problems. This follows because $z_{i,t-1}$ is an observed sufficient statistic for the history of states. Therefore the firm's beliefs in future periods do not depend on its decisions in the current period.

The firm's choice at time $t$ therefore solves the following problem of maximizing expected risk-adjusted profits net of cognitive costs:

$$\max_{p\in\mathcal{P}} \left\{ \int_{\mathcal{Z}} \int_{\mathcal{X}} \Pi(x, z_{it})\, p(x \mid z_{it})\, \mathrm{d}x \; f(z_{it} \mid z_{i,t-1})\, \mathrm{d}z_{it} - c\left(p, \lambda_i, z_{i,t-1}, f\right) \right\} \tag{9}$$

Firms flexibly choose a production plan for each decision state realization $z_{it}$. For example, firms can choose how the mean and variance of their production vary with idiosyncratic productivity, aggregate output, and wages. Planning precisely, however, has a cognitive cost, which is proportional to entropy. The firm's optimal choice trades off higher expected utility from precise planning with the increased costs of so doing in each state.

**Units for Costs.** Consistent with our premise that decision frictions arise because of *cognitive* limitations, the cost $c$ is in utility units. An alternative model, motivated for instance by a story of physical costs within the firm (e.g., hiring planners or consultants), might denominate the costs in units of the numeraire good. In Section 2.5, we show that this dollar-cost model (under any stochastic discount factor) is outcome-equivalent to our utility-cost model with risk-neutral owners ($\gamma = 0$), and is therefore nested in the main analysis. We further discuss how to interpret our main results in this special case.

### 1.5 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 Appendix A.6.

Toward simplifying the intermediate goods firms' objective, we first define an intermediate firm's *ex post* optimal production level, $x^*(z_i) := \arg\max_{x\in\mathcal{X}} \Pi(x, z_i)$. We next define $\bar{\Pi}(z_i)$ as risk-adjusted profits evaluated at $(x^*(z_i), z_i)$ 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:

$$\tilde{\Pi}(x, z_i) := \bar{\Pi}(z_i) + \frac{1}{2}\Pi_{xx}(z_i)(x - x^*(z_i))^2 \tag{10}$$

The first-order term of this approximation drops out due to the envelope theorem: there are no first-order costs of deviating from $x^*(z_i)$. So that this approximate payoff 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 (2) 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 $X^* = \left(\int_0^1 x^*(z_i)^{1-\frac{1}{\epsilon}}\,\mathrm{d}i\right)^{\frac{\epsilon}{\epsilon-1}}$. We then write the approximate production function as

$$X = X^* - \frac{1}{2\epsilon}\int_0^1 \frac{(x_i - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}}(x^*(z_i))^{1+\frac{1}{\epsilon}}}\,\mathrm{d}i \tag{11}$$

We now define a rational expectations equilibrium, up to these approximations, in which all agents optimize, all markets clear, and expectations are consistent. It is convenient for us to state this definition in the following, compact way in terms of the stochastic choices of intermediate goods firms $p$ and the transition density $f$:

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

1. *Intermediate goods firms' stochastic choice rules $p$ solve program (9) given $f$, with $\tilde{\Pi}$, defined in (10), replacing $\Pi$.*
2. *The transition density $f$ is consistent with $p$ in the sense that: the marginal distribution of firm-level productivity is given by $G$; aggregate output is given by the aggregator (11) evaluated in the cross-sectional distribution of production implied by $p$ and $G$; and the wage is derived from the wage rule (3) evaluated at aggregate output.*

# 2 Attention Cycles and Their Consequences

We now present our main theoretical results, in four parts. First, we characterize firms' production and attention choices in partial equilibrium. Second, we derive primitive conditions under which attention is counter-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, isolate the role of cyclical inattention, and show how it can drive asymmetric and state-dependent shock responses. We conclude the section by summarizing six key predictions of the theory at the micro and macro levels, which will form the basis of our empirical tests and quantitative calibration.

## 2.1 Attention and Misoptimization in Partial Equilibrium

We begin by describing the firms' choices in partial equilibrium:

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

$$x_i = x^*(z_i) + \sqrt{\frac{\lambda_i}{|\pi_{xx}(z_i)|\, M(z_i)}} \cdot v_i \tag{12}$$

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

*Proof.* See Appendix A.1. □

Mathematically, this result follows from three observations. First, entropy costs make the firm's problem *linearly separable* across state realizations $z_{it}$. This implies that firms' optimal policies are independent of the prior distribution $f$. Second, the firms' payoff functions are (up to approximation) quadratic in the action conditional on each state. Third, under entropy costs and quadratic payoffs, the optimal non-parametric action distribution is Gaussian with the properties identified in Proposition 1. While superficially similar to existing results in quadratic-Gaussian rational inattention models (Sims, 2003), this characterization has two important differences from all existing results with unrestricted information acquisition of which we are aware. First, the cross-sectional action distribution is a normal-mixture distribution within each aggregate state. Second, the variance of the cross-sectional action distribution depends on the exogenous and endogenous aggregate states.

**The Profit-Curvature and Risk-Pricing Channels.** 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 $\lambda_i$), and decreases if either of two components of the marginal benefits of precision increases. In this way, two channels of incentives determine agents' apparent bounded rationality, or failure to perfectly optimize from an *ex post* perspective.

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 $|\pi_{xx}(z_i)|$ moves as a function of the aggregate business cycle depends on the demand and cost curves that firms face. We will refer to the effect of these incentives on misoptimization as the *profit-curvature channel*.

The second component of this marginal benefit is the firm owner's marginal utility, $M(z_i)$. This translates dollar losses into utility losses, which can be directly compared to the utility cost of cognition. As every firm is owned by the representative household in our model, the relevant marginal utility can be written only as a function of output $X$ (i.e., aggregate consumption). When this household is risk-averse, this marginal utility is a decreasing function of output. Thus, the representative household is "hungrier" for any firm's dollar profits, in utility terms, when the aggregate economy is doing poorly. Because of this, they are less tolerant of misoptimizations when output is low. We will refer to the effect of these incentives on misoptimization as the *risk-pricing channel*.

**What Drives Attention and Misoptimization?** 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 structure of our model to re-state the comparative statics in the discussion above in terms of firms' productivity $\theta_i$ and aggregate output $X$, after imposing the stochastic discount factor, demand curves, and cost curves that our model implies. In particular, the curvature terms of interest can be written as:[6]

$$|\pi_{xx}(z_i)| := v_\pi(\epsilon, \chi, \bar{X}, \bar{w}) \cdot \theta_i^{-1-\epsilon} X^{\chi(1+\epsilon)-1} \qquad M(z_i) = X^{-\gamma} \tag{13}$$

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 below:

**Corollary 1** (Comparative Statics for Mistakes)**.** *Consider a type-$\lambda_i$ firm in state $z_i = (\theta_i, X, w(X))$. The extent of misoptimization, $\mathbb{E}[(x_i - x^*(z_i))^2 \mid z_i]$, increases in $\theta_i$. 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 13. □

The (absolute) curvature of the profit function always increases in marginal costs, and therefore decreases in $\theta_i$. 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.

[6] The first expression and the associated constant are derived in Appendix A.6.

## 2.2 Attention and Misoptimization Cycles in Equilibrium

We now translate the partial equilibrium behavior of the firm into general equilibrium. We first provide 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 depends on the productivity distribution $G$ only through the sufficient statistic $\hat{\theta}(G) \in \Theta$ (Equation 5) and can be expressed via a function $X : \Theta \to \mathbb{R}$ that is strictly positive and strictly increasing in that sufficient statistic.*

*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 sufficient statistic $\hat{\theta}(G)$. To establish uniqueness and monotonicity, we derive conditions under which the fixed-point equation is a contraction map that depends positively on $\hat{\theta}$. 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 $\lambda_i$, and highly plausible in practice. The elasticity of detrended real wages to GDP in US data since 1987 is 0.095, and micro-level studies find similarly severe wage rigidity (Solon et al., 1994; Grigsby et al., 2021).[7] Moreover, to be consistent with the restrictions $\chi\epsilon < 1$ (upward-sloping best responses) and $\epsilon > 1$ (substitutable goods), $\chi$ cannot exceed one. The corresponding conditions $\gamma > 1.095$ or $\gamma > 2$ are likely both slack by an order of magnitude, given evidence in financial economics about the high cyclicality of the stochastic discount factor (e.g., Hansen and Jagannathan, 1991). As $\gamma$ governs *only* the properties of the stochastic discount factor in our model, this is the relevant evidence for determining its appropriate value.

[7] 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.

We now give conditions under which the economy exhibits aggregate misoptimization and attention cycles. We say that firms "misoptimize less" in a state if, averaging over idiosyncratic states $(\theta_i, \lambda_i)$, they have a lower expected mean-squared error around the *ex post* optimal action $x^*(z_i)$. We say that all firms "pay more attention" in a state if, averaging over idiosyncratic states $(\theta_i, \lambda_i)$, they pay a greater attention cost conditional on that state being realized. Formally:

**Definition 2** (Aggregate Misoptimization and Attention)**.** *Fix an equilibrium law of motion $X : \Theta \to \mathbb{R}$. Firms' aggregate misoptimization is their average mean-squared-error around the ex post optimal action, or*

$$m(G) := \mathbb{E}_{\theta_i,\lambda_i,x_i}\left[(x_i - x^*(z_i(\hat{\theta}(G))))^2\right] \tag{14}$$

*where $z_i(\theta) = (\theta_i, X(\theta), w(X(\theta)))$, $p^*(\cdot \mid z_i(\theta); \lambda_i)$ is the uniquely optimal plan of a type-$\lambda_i$ firm contingent on realized state $z_i$, and the expectation is taken over $\theta_i \sim G$, $\lambda_i \sim L$, and $x_i \sim p^*(\cdot \mid z_i(\theta); \lambda_i)$. Firms' aggregate attention is their average realized cognitive cost*

$$a(G) := \mathbb{E}_{\theta_i,\lambda_i,x_i}\left[\lambda_i\, p^*(x_i \mid z_i(\theta); \lambda_i)\, \log p^*(x_i \mid z_i(\theta); \lambda_i)\right] \tag{15}$$

We now show that, under the same assumptions as Proposition 2, the model features counter-cyclical attention and pro-cyclical misoptimization:

**Proposition 3** (Attention Cycles)**.** *Assume $\chi\epsilon < 1$ and $\gamma > \chi + 1$ and consider productivity distributions $G$ and $G'$. If $\hat{\theta}(G) \geq \hat{\theta}(G')$, then aggregate output $X$ and aggregate misoptimization $m$ are larger under $G$. If additionally $G \succsim_{FOSD} G'$, then aggregate attention $a$ is lower under $G$.*

*Proof.* See Appendix A.3. □

Economically, Proposition 3 says that any calibration of the model that is consistent with existing evidence about wage rigidity and the stochastic discount factor predicts that firms should pay more attention to their decisions and make smaller misoptimizations, conditional on fundamentals, in downturns. We note that these conditions are *sufficient* but not *necessary*. For example, misoptimization would also be pro-cyclical (and attention counter-cyclical) if the (looser) condition in Corollary 1 held and if, by assumption, output were monotone increasing in productivity.

## 2.3 Misoptimization, Output, and Productivity

Having provided conditions for attention and misoptimization cycles, we now study the effects of these phenomena on output and labor productivity. Despite unrestricted heterogeneity in the cross-sectional distributions of microeconomic productivity and attention costs, there are scalar sufficient statistics in equilibrium for each distribution. As described earlier, the cross-sectional distribution of productivity is summarized by the sufficient statistic $\hat{\theta}(G)$. 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 the mean $\lambda := \mathbb{E}_L[\lambda_i]$.

We define $\log X(\log\theta)$ as a mapping from the log productivity sufficient statistic to log output in the economy, holding fixed all other parameters. The following result describes output in log units as the sum of that which would be obtained absent inattention and an *attention wedge* $\log W(\log\theta)$:

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

$$\log X(\log\theta) = X_0 + \chi^{-1}\log\theta + \log W(\log\theta) \tag{16}$$

*where $\theta = \hat{\theta}(G)$ is a sufficient statistic for the productivity distribution, $X_0$ 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.*

*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 mistakes. The magnitude of this force increases in the extent of cognitive costs $\lambda$ and in aggregate productivity $\theta$. Both results have a partial-equilibrium and general-equilibrium component. In partial equilibrium, both increasing $\lambda$ and increasing $\theta$ make firms play more dispersed actions, as shown in Proposition 1, 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 better understand this wedge, we can re-cast it in terms of labor productivity:

**Corollary 2** (The Productivity Wedge)**.** *Equilibrium labor productivity* $A := \frac{X}{L}$ *follows:*

$$\log A(\log\theta) = \log\theta + \chi\epsilon \log W(\log\theta) \tag{17}$$

*Proof.* See Appendix A.5. □

The productivity wedge representation allows for three useful parallels between our paper's mechanism and classic arguments in the literature. First, our mechanism is like an attentional, *intensive margin* version of the "cleansing" effect studied by Caballero and Hammour (1994): conditional on a given firm operating, it is more focused on making precise choices in recessions, and this on average reduces the wedge and raises aggregate labor productivity. Second, 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. This is consistent with the unstable and often negative cyclicality found in US data (e.g., Galí and Van Rens, 2021). Third, the mechanism whereby dispersion in firm-level value marginal products depresses aggregate productivity is shared with the literature on misallocation pioneered by Hsieh and Klenow (2009) and Restuccia and Rogerson (2008).[8] What is new is our prediction about the cyclicality of this force, driven by changing incentives for (in)attention, and the implications of that cyclicality for business cycles.

### 2.4 Shock Propagation and Volatility

The fact that agents are differentially attentive to shocks across states of the world makes the economy differentially sensitive to shocks. To see this, note that the elasticity of output to a small shock can be written as

$$\frac{\partial \log X(\log\theta)}{\partial \log\theta} = \chi^{-1} + \frac{\partial \log W(\log\theta)}{\partial \log\theta} \tag{18}$$

that is, the sum of the frictionless economy's response to shocks ($\chi^{-1}$) and the response of the attention wedge to shocks. The latter term is negative in the case studied by Proposition 4, so the economy is 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). Here, the result arises not from individual-level underreaction but instead from the interaction of individual mistakes with concave

[8] David et al. (2016), Ma et al. (2020), and Barrero (2022) link misallocation to cognitive and informational frictions in steady state and study the implications for productivity.

aggregation. Moreover, this dampening from the attention wedge can depend on the state. This is a direct consequence of modeling state-dependent attention and accommodating attention cycles.

If the slope of the attention wedge varies with the productivity state, then the economy has different responses to productivity shocks depending on the initial level of productivity. An immediate implication is that, even in an environment with constant volatility in $\log \theta_t$ (e.g., it follows a standard AR(1) process), there is endogenous stochastic volatility in output growth. Finally, if the attention wedge has a non-zero second derivative, positive and negative shocks have asymmetric effects on log output.

The ultimate 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 versus positive shocks from any initial state. 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 5 will feature a concave wedge. In that section we will review the implications of this finding.

### 2.5 Extensions

**Decision Costs in Dollar Units.** Motivated by the nature of cognitive costs, we denominated the cost of precise planning in utility units. An alternative choice is to specify these costs in units of final output or "dollars." This is natural if overcoming decision frictions requires investing in inputs, like employees tasked with planning. Since risk aversion enters our analysis only in translating dollar costs to utility costs, the assumption of dollar-denominated cognitive costs is nested by setting $\gamma = 0$. Proposition 1 and Corollary 1 hold as written. If output is monotone in productivity, the model can feature equilibrium attention cycles if $1 > \chi(1 + \epsilon)$. Thus, the profit-curvature channel by itself can induce attention cycles if wages are sufficiently rigid, a condition likely to hold based on our earlier discussion. However, our empirical analysis will support a strong role for the risk-pricing channel. First, we will find weak evidence of differential sensitivity of dollar profits to misoptimizations over the business cycle (Section 4.3). Second, we will find strong procyclicality of misoptimization variance. Thus, to rationalize these facts, our quantitative model requires a strong risk-pricing channel ($\gamma = 11.5$), rejecting the dollar-cost model nested with $\gamma = 0$ (Section 5.1).

**Labor Wedge Shocks.** Fluctuations can also arise in our model from shocks to $\bar{w}$, which can be interpreted as a labor wedge shock. Inspection of Proposition 1 reveals that $\bar{w}$ is equivalent 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 and pro-cyclical misoptimization in an economy with stochastic productivitity and labor wedges. Proposition 4 holds as written, with $\theta\bar{w}^{-1}$ replacing $\theta$. Thus, our core conclusions about attention cycles and their effects on business cycles hold in a model with both supply and demand shocks.

**Costly Information Acquisition.** Our model of costly control and non-parametric choice differs from models of unrestricted costly information acquisition in several respects. Costly control allows us to tractably study several partial and general equilibrium phenomena. Nonetheless, we sacrifice the model's ability to capture individual-level underreaction (while we do generate aggregate underreaction). Motivated by this, we study the robustness of our results to information-acquisition costs by examining our model with Gaussian signal extraction with costly precision (Appendix F.2) and mutual-information costs as studied by Sims (2003) (Appendix F.3). In both cases we provide conditions under which a greater cost of making mistakes from either the risk-pricing or profit-curvature channel leads to firms making smaller mistakes. We do, however, note that application of unrestricted rational inattention to the model we study is not analytically possible using any state-of-the-art techniques (e.g., Afrouzi and Yang, 2021; Miao et al., 2022) as our firms' have non-Gaussian priors and non-quadratic payoff functions, and aggregation is non-linear.

**Multiple Inputs and Classical Labor Supply.** In Appendix D, 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 4, 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 and 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.

### 2.6 Microeconomic and Macroeconomic Predictions

We now distill our findings into six testable microeconomic and macroeconomic predictions. These will motivate our measurement strategy and empirical analysis.

The first is that firms make costly misoptimizations:

**Prediction 1.** *Firms make misoptimizations (under- and over-production relative to the ex post optimal level) that reduce profits.*

This prediction is a consequence of costly cognition (Proposition 1). It is inconsistent with models in which deviations from optimal choices arise from wedges in the sense of Chari et al. (2007) or Restuccia and Rogerson (2008), as under-production would be associated with lost profits but over-production would be associated with increased profits. Thus, this prediction directly tests our misoptimization-based interpretation of the data.

The second prediction concerns how misoptimizations move over the cycle:

**Prediction 2.** *Firms make larger misoptimizations in booms and smaller misoptimizations in downturns.*

This arises from firms' counter-cyclical incentives to rein in mistakes, given assumptions on risk aversion, wage rigidity, and substitutability that we argued are natural (Proposition 3). A similar prediction does not obtain under alternative, non-cognitive theories of stochastic wedges—for example, the natural prediction with financial frictions is that their intensity and cross-firm heterogeneity increases, rather than decreases, in downturns. We will revisit this comparison to other models in the empirical analysis.

The third prediction identifies the mechanism underlying attention cycles:

**Prediction 3.** *Firm valuations are more sensitive to misoptimizations in downturns than in booms.*

In our costly attention model, firms rein in misoptimizations exactly when it is most valuable to do so. This microeconomic prediction lies at the core of our model's mechanism for endogenous, pro-cyclical misoptimization.

The fourth prediction zeros in on the mechanisms driving Prediction 3:

**Prediction 4.** *Firm profits, in dollars, may be more or less sensitive to misoptimizations in downturns versus booms.*

Prediction 3 signs the *the total effect of the risk-pricing and profit-curvature channels* while Prediction 4 says that the *profit-curvature channel in isolation* has an ambiguous sign. Therefore, when we confront the model with data on both financial valuation and profitability, we can distinguish the relative strength of these mechanisms.

The fifth prediction concerns the cyclicality of aggregate attention:

**Prediction 5.** *Firms pay more attention to decisions in downturns and less in booms.*

This prediction, like Prediction 2, derives from Proposition 3. But, since cognitive effort is by definition unobserved, it is more challenging to directly test than its equivalent for misoptimization. In the next section, we will introduce a method to proxy for firms' attention using textual analysis.

The final prediction regards the relationship between misoptimization and attention:

**Prediction 6.** *Firms that pay more attention make smaller misoptimizations.*

This prediction allows us to microeconomically validate the decision-relevance of measured attention and to further validate the attention interpretation of measured misoptimization.

# 3 Measuring Misoptimization and Attention

Toward testing the model's microeconomic and macroeconomic predictions, we introduce new empirical measures of firms' "misoptimizations" and "attention." Our model provides a structural method by which we can recover firms' misoptimizations. Econometrically, this requires only a standard estimation of firms' productivity and policy functions. We proxy for attention, which is unobservable, by constructing a text-based measure: firms' discussion of macroeconomic topics in regulatory filings and earnings conference calls.

## 3.1 Measuring Misoptimizations

**Data.** Our dataset for public firms' production and input choices is Compustat Annual Fundamentals. We use information on sales, employment, variable input expenses, and capital measured via net and gross values of plants, property and equipment (PPE). 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. We drop financial firms and utilities due to their markedly different production functions.

In our main analysis, we use a sample period from 1986 to 2018.[9] We apply standard filters to remove firms that are based outside the US, are insufficiently large, or are likely to have been involved in a merger or acquisition. This yields a final sample of 68,198 firm-year observations, or about 2,200 per year. Appendix B.1 describes the full procedure.

[9]To probe robustness, we also show results on the entire Compustat sample since 1950.

**From Theory to Estimation.** We start by translating our characterization of firms' optimal stochastic choice (Proposition 1) into an estimable structural relationship. Letting $\sigma(\theta_{it}, X_t, w_t) = \sqrt{\frac{\lambda_i}{|\pi_{xx}(\theta_{it}, X_t, w_t)| M(X_t)}}$ denote the firms' optimally chosen volatility of mistakes and $v_{it}$ denote idiosyncratic, unit variance, and zero-mean firm trembles, we observe that Proposition 1 implies:

$$\begin{aligned} \log L_{it} &= \log L^*(\theta_{it}, X_t, w_t) + \log\left(1 + \frac{\sigma(\theta_{it}, X_t, w_t)}{L^*(\theta_{it}, X_t, w_t)} v_{it}\right) \\ &= (\epsilon - 1)\log\theta_{it} - \epsilon \log w_t + \log X_t + \log\left(1 + \frac{\sigma(\theta_{it}, X_t, w_t)}{L^*(\theta_{it}, X_t, w_t)} v_{it}\right) \\ &\approx (\epsilon - 1)\log\theta_{it} - \epsilon \log w_t + \log X_t + \underbrace{\frac{\sigma(\theta_{it}, X_t, w_t)}{L^*(\theta_{it}, X_t, w_t)} v_{it}}_{m_{it}} \end{aligned} \tag{19}$$

where the first line uses Proposition 1 (expressed in terms of labor choice, and in logs); the second line uses the expression for $L^*(\theta_{it}, X_t, w_t) = \frac{x^*(\theta_{it}, X_t, w_t)}{\theta_{it}}$ in our model; and the third approximates $\log(1 + x) \approx x$ for small $x$ and defines the residual $m_{it}$. In Appendix B.3, we show that this policy function is also obtained in a generalization of our baseline model that features multiple flexible inputs. In words, Equation 19 says that misoptimizations can be measured as residual variation in firms' choices conditional on productivity, factor prices, and demand. This is the basic blueprint for our measurement strategy.

In practice, to guard against the potential for misspecification of this relationship, we estimate a more flexible model of the following form

$$\log L_{it} = \eta_i + \chi_{j(i),t} + \beta \log\theta_{it} + m_{it} \tag{20}$$

where $j(i)$ is the industry of firm $i$, $\beta$ is an unrestricted coefficient, $\eta_i$ and $\chi_{j(i),t}$ are respectively fixed effects at the firm- and industry-by-time levels, and $m_{it}$ follows an AR(1) process $m_{it} = \rho m_{i,t-1} + \left(\sqrt{1 - \rho^2}\right) u_{it}$, where $\rho \in (0, 1)$ is the persistence and $u_{it}$ is a zero-mean innovation with variance $\tilde{\sigma}^2(\theta_{it}, X_t, w_t)$. If $\tilde{\sigma}^2(\theta_{it}, X_t, w_t)$ is sufficiently persistent, then it is approximately the variance of $m_{it}$. In the Appendix, we formally describe how the more general model nests two economically important features from which we abstracted in the theoretical analysis for simplicity. The first is time-varying differences in demand and factor prices across industries and time-invariant differences across firms, respectively captured in the industry-by-time and firm fixed effects of Equation 20 (Appendix B.3). The second is persistent mistakes, which can arise in a variant model which accommodates cognitive inertia (Appendix F.1).

**Estimation Procedure.** To empirically estimate our model, we proceed to estimate productivity and firm policy functions using standard techniques. First, we estimate productivity using a cost-shares approach, as in Foster et al. (2001), Bloom et al. (2018), and Ilut et al. (2018). We describe the full details in Appendix B and summarize the main points below. We define productivity $\log \theta_{it}$ as the residual of an industry-specific, constant-returns-to-scale, Cobb-Douglas production function over labor $L_{it}$, materials $M_{it}$, and capital $K_{it}$:

$$\log x_{it} = \log \theta_{it} + \alpha_{L,j(i)} \log L_{it} + \alpha_{M,j(i)} \log M_{it} + \alpha_{K,j(i)} \log K_{it} \tag{21}$$

where $\alpha_{L,j(i)} + \alpha_{M,j(i)} + \alpha_{K,j(i)} = 1$.[10] We measure 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 net of depreciation and labor expenditures; and the capital stock as the initial gross level of plant, property, and equipment plus net investment. We use industry-specific ratios of labor and materials expenditures to total sales to estimate the revenue elasticities of labor and materials, and translate these into output elasticities given an assumption for the elasticity of demand of $\epsilon = 4$. We measure the output elasticity of capital as $1-\alpha_{L,j(i)}-\alpha_{M,j(i)}$, using constant returns to scale. We combine these estimates with our production functions and demand curves, and partial out industry-by-time fixed effects to capture factor price and demand variation, to recover an estimate for firm-level productivity $\log \theta_{it}$. Note that $\log \theta_{it}$ could blend both firm-level productivity and demand shocks, as observed by Foster et al. (2008) for any context without access to separate data on both prices and quantities. But since both types of shocks behave identically in our mapping from theory to data, this poses no issue for the interpretation of our estimated policy functions or misoptimizations.

Next, we estimate firm policy functions as in Ilut et al. (2018) or Decker et al. (2020). We first estimate Equation 20 via ordinary least squares (OLS) and obtain a preliminary estimate $\hat{m}^0_{it}$ of the residual. We next estimate the AR(1) process for misoptimizations using $\hat{m}^0_{it}$ to obtain an estimate $\hat{\rho}$ of the residual persistence (in our main procedure, 0.70).

[10]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) and Foster et al. (2008) argue that constant returns to scale in physical terms is a reasonable approximation for large, US-based firms.

We next estimate via OLS the "quasi-differenced" equation for labor choice:[11]

$$\log L_{it} - \hat{\rho}\log L_{i,t-1} = \eta_i + \chi_{j(i),t} + \beta_0 \log\hat{\theta}_{it} + \beta_1 \log\hat{\theta}_{i,t-1} + \nu_{it} \tag{22}$$

with residual $\nu_{it} = m_{it} - \hat{\rho} m_{i,t-1}$. When $\hat{\rho} = \rho$, $\nu_{it} = \left(\sqrt{1-\rho^2}\right) u_{it}$. We therefore obtain an estimate of $u_{it}$ as $\hat{u}_{it} = \frac{\hat{\nu}_{it}}{\sqrt{1-\hat{\rho}^2}}$.

We finally define aggregate "Misoptimization Dispersion" as an estimate of the cross-sectional variance of $m_{it}$ with weights $s^*_{it}$ proportional to firms' *predicted* sales based on fundamentals:[12]

$$\text{MisoptimizationDispersion}_t = \frac{\sum_{i\in\mathcal{I}_t} s^*_{it}\cdot \hat{u}^2_{it}}{\sum_{i\in\mathcal{I}_t} s^*_{it}} \tag{23}$$

The weights are the appropriate ones for mapping average misoptimization to misallocation 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.6.3, we show that pro-cyclical MisoptimizationDispersion by our empirical definition is sufficient for pro-cyclical misoptimization as defined in Definition 2, and thus constitutes a more stringent test of that model prediction.

**Robustness to Alternative Methods.** Of course, this procedure can only be interpreted as recovering mistakes given the model that we have written. Any additional unmodeled, idiosyncratic firm-level wedges à la Restuccia and Rogerson (2008) would also be captured as a mistake.[13] Later, Prediction 1 will allow us to distinguish idiosyncratic wedges from misoptimizations: residuals from wedges should have a uniformally positive effect on firms' profitability and stock returns. By contrast, residuals from misoptimization should have a hump-shaped effect on firm performance. This notwithstanding, to explicitly account for other misspecifications, we employ several alternative strategies.

First, we construct measures that are robust to the presence of misspecification arising from adjustment costs (e.g., as in Hopenhayn and Rogerson, 1993) and financial frictions

[11] Appendix Table A8 contains our estimates of Equation 22 and the AR(1) process for $m_{it}$ 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 similar without this trimming.

[12] In particular, the weights are the exponentiated fitted values $\exp(\hat{\beta}\log\hat{\theta}_{it})$ from the following regression equation: $\log \text{Sales}_{it} = \beta \log\hat{\theta}_{it} + \eta_i + \chi_{j(i),t} + \epsilon_{it}$.

[13] Time-varying wedges at the industry or aggregate level would be absorbed in our model's fixed effect, and firm-level correlated distortions that are driven by productivity (Restuccia and Rogerson, 2008) would be included in our productivity estimates. Thus, neither would affect our estimated misoptimizations.

(e.g., as in Ottonello and Winberry, 2020). To do this, we estimate a variant policy function that adds the first lag of labor choice. In this model, there is no significant persistence of mistakes (see Table A8). To capture financial frictions, we consider an augmented policy function with a direct control for leverage (total debt/total assets), as constructed by Ottonello and Winberry (2020), and its interaction with TFP.

Second, for robustness to the cost-shares-based productivity estimation strategy, we also estimate productivity using the method of Olley and Pakes (1996). We find that the two measures are very similar to one another, with a regression coefficient close to one and a $R^2$ of 0.98 after including sector-by-time and firm fixed effects (Table A12).

Finally, we perform our analysis under a host of further policy functions that (i) are industry-specific, to combat against different market conditions and/or measurement error in TFP; (ii) have time-varying coefficients on TFP (Decker et al., 2020); or (iii) allow for TFP to affect the policy function non-linearly, to capture decreasing returns-to-scale and asymmetric hiring and firing rules (Ilut et al., 2018). We also consider time-varying production functions, to capture technological trends and, to first order, cyclical changes in input shares. In another check, we estimate production functions and policy functions in a pre-sample, while estimating misoptimization in a post-sample, to guard against over-fitting.

### 3.2 Measuring (Macroeconomic) Attention

We now describe our strategy for measuring a proxy for firms' attention. There are two key challenges to this. First, attention, unlike misoptimization, has no direct theoretical analog in observable firm choices. We instead use data on how firms describe their decisions in words and treat these as a window into the contingencies for which firms plan. Second, our model predicts that attention toward *all* factors increases in downturns. But, for the purpose of measurement, it is necessary to focus on factors that are commonly faced to some extent by all firms. We will therefore focus on measuring attention to macroeconomic conditions.

**Data.** Our main 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.[14] Our total sample consists

[14]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

of 479,403 individual documents, or about 5,000 per quarter, which we index by their date of filing.

We also apply our measurement technique to a supplemental data source of 150,000 quarterly sales and earnings conference calls from 2003 to 2014. We replicate all subsequent results on these data as well. We relegate some additional details about our data construction for the conference calls to Appendix C.1.

**Methodology.** A difficulty for identifying macroeconomic attention is to differentiate characteristic language of macroeconomics from firms' standard financial vocabulary (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 references. Following the method introduced by Hassan et al. (2019) to study firm discussion of political risks, we use introductory college-level textbooks as those references: *Macroeconomics* and *Principles of Macroeconomics* by N. Gregory Mankiw and *Macroeconomics: Principles and Policy* by William J. Baumol and Alan S. Blinder.[15] 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 filing 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 regulatory filings, measured as a proportion of total documents that use the word at least once. We define the "term frequency inverse document frequency," or tf-idf, as:

$$\text{tf-idf}(w)_{it} := \text{tf}(w)_{it} \cdot \log\left(\frac{1}{\text{df}(w)}\right) \tag{24}$$

The log functional form is a heuristic in natural language processing for scaling the relative importance of each term. It is bounded below by 0 when a word appears in all documents and smoothly penalizes words that appear in more documents and are therefore less indicative of the topic of interest. For each word that appears in the 10-Q and 10-K corpus, we calculate the tf-idf using term frequencies in each of the three textbooks and (inverse) document frequencies among regulatory filings. We rank the top 200 words

as a starting point at which a nearly comprehensive sample of firms' reports are available in the system.

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

by this metric in each textbook and take the intersection among the three books to obtain a final set of 89 words.[16] Appendix Figure A5 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, and consider an *ex post* "cleaned" word list only in a robustness check. 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:

$$\text{MacroAttention}_{it} = \sum_{w \in \mathcal{W}_M} \text{tf-idf}(w)_{it} \tag{25}$$

We generate an aggregate measure $\text{MacroAttention}_t$ by averaging $\text{MacroAttention}_{it}$ across firms. In our aggregate results, we remove seasonal trends as quarter-of-the-year means.

# 4 Testing the Model's Predictions

We now test the model's six main predictions (Predictions 1-6) and report six corresponding findings. Together, these findings are consistent with our macroeconomic predictions and proposed microeconomic mechanism of incentives-driven attention.

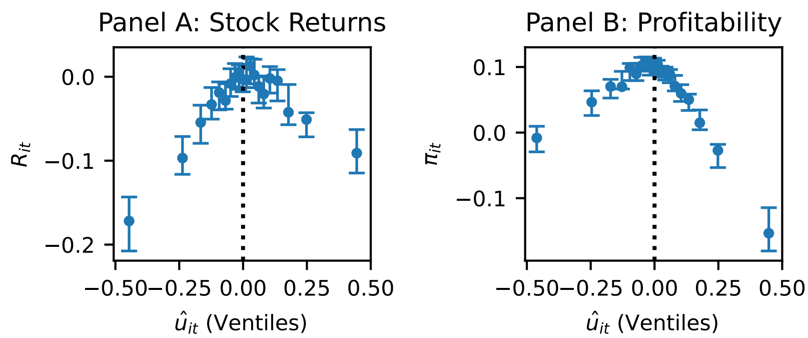
## 4.1 Fact 1: Mistakes Reduce Returns and Profitability

We first test that our measured "misoptimizations" are, as predicted by the theory, bad for firm performance (Prediction 1). We proxy for firm performance with two measures. The first is firms' log stock returns, $R_{it}$. The second is firms' "profitability" $\pi_{it}$, or this year's earnings before interest and taxes (EBIT) divided by the last year's total variable costs.[17] We examine the non-parametric binned scatter relationship of each of these measures with measured misoptimization (residuals) $\hat{u}_{it}$, net of industry-by-time fixed effects $\chi_{j(i),t}$.

[16] 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.

[17] Variable costs, by our definition, are cost of goods sold (COGS) plus administrative expenses (XSGA) net of depreciation (DP). 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.

**Figure 1:** The Negative Relationship Between Misoptimization and Firm Performance




*Notes*: Both panels are binned scatterplots. The outcome variables are log stock returns (Panel A) and profitability (Panel B). Dots represent means of the corresponding outcome conditional on ventiles of the $x$-axis variable, $\hat{u}_{it}$, and industry-by-time fixed effects. Error bars are 95% confidence intervals based on the method of Cattaneo et al. (2019), with two-way clustering by firm and year. The construction of $\hat{u}_{it}$ is described in Section 3.1.

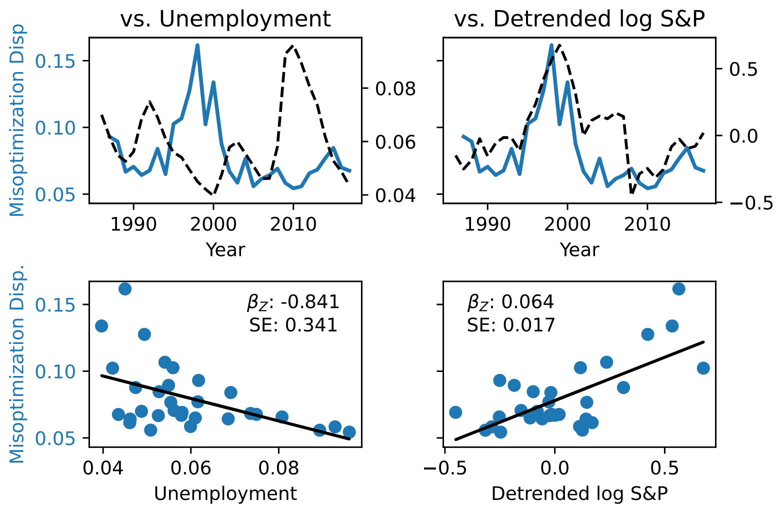
Misoptimizations in both directions (i.e., under- or over-hiring labor) hurt firm performance, measured in each way (Figure 1). This "hump-shaped" result is consistent with our interpretation of residuals as misoptimizations. It is inconsistent with the interpretation of residuals as wedges, which would predict a positive, monotone relationship, or pure noise, which would predict a flat relationship. In Appendix Table A1, we show that the result of a negative relationship between $\hat{u}_{it}^2$ and $\{R_{it}, \pi_{it}\}$ is highly statistically significant and robust to different levels of fixed effects.

The negative effects of misoptimizations on firm performance are also persistent. To measure the dynamic relationship between misoptimization and performance, we estimate projection regressions of productivity growth, profitability, and returns on $\hat{u}_{it}^2$, net of industry-by-time fixed effects:

$$X_{i,t+k} = \beta_{X,k} \cdot \hat{u}_{it}^2 + \chi_{j(i),t} + \epsilon_{it} \tag{26}$$

for $X_{i,t+k} \in \{\Delta \log \hat{\theta}_{i,t+k}, \pi_{i,t+k}, R_{i,t+k}\}$ and $k \in \{0, 1, 2\}$. We find no evidence of a quantitatively large effect on current and future TFP growth (Appendix Table A2). We also find strong evidence of persistent negative effects on profitability and stock returns. These dynamic results rule out the possibility that the observed negative effect of $\hat{u}_{it}^2$ operates through a channel related to productivity, or through dynamic trade-offs between poor performance today and improved future performance.

**Figure 2:** Misoptimization Dispersion is Pro-Cyclical




*Notes*: The top two panels plot Misoptimization Dispersion (blue line, left axis) along with, respectively, unemployment and the linearly detrended S&P 500 price (black dashed lines, right axis). The bottom two panels are scatterplots of Misoptimization Dispersion versus the corresponding macroeconomic aggregate. The black solid line is the linear regression fit. The standard errors are HAC robust based on a Bartlett kernel with a three-year bandwidth.

### 4.2 Fact 2: Misoptimizations Rise in Booms, Fall in Downturns

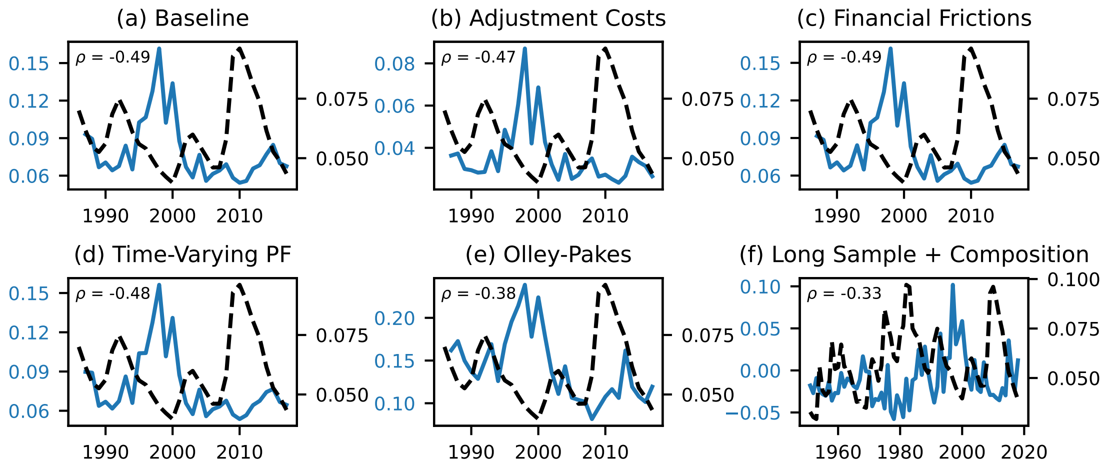
We now study the behavior of aggregate misoptimizations in the data (Prediction 2). Figure 2 plots aggregate Misoptimization Dispersion against the unemployment rate and detrended end-of-year S&P 500 price. Misoptimization Dispersion rises when the real economy and financial markets are doing well (e.g., 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. Appendix Figure A1 shows that the same pro-cyclical pattern is apparent in two other measures of dispersion, the mean of $|\hat{u}_{it}|$ and inter-quartile range of $\hat{u}_{it}$.

We benchmark the strength of this relationship with the business cycle by estimating a linear regression of Misoptimization Dispersion on each macroeconomic variable:

$$\text{MisoptimizationDispersion}_t = \alpha + \beta_Z \cdot Z_t + \epsilon_t \tag{27}$$

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. The respective correlations

**Figure 3:** Robustness of Pro-Cyclical Misoptimization to Alternative Strategies




*Notes*: In each panel, the blue line is the corresponding version of Misoptimization Dispersion, the black dashed line is the unemployment rate, and $\rho$ is their correlation. Panel (a) is the baseline measure. Panel (b) is the measure that accounts for adjustment costs via lagged labor in the policy function. Panel (c) is the measure that accounts for financial frictions via leverage and its interaction with productivity in the policy function. Panel (d) allows the elasticities of the production function to vary year-by-year. Panel (e) uses productivity estimates from the method of Olley and Pakes (1996). Panel (f) replicates our main procedure for the sample back to 1950 and applies compositional adjustment.

are -0.493 and 0.689. The regression on unemployment implies that a five percentage point swing in unemployment is associated with an increase of Misoptimization Dispersion by 0.042 log points, or 53% of its sample mean value.

To investigate the robustness of this finding to plausible misspecifications of firms' policy functions, we re-visit Fact 2 under the different constructions for "mistakes" introduced in Section 3.1. Our findings are summarized in Figure 3.

**Adjustment Dynamics.** We first account for frictional adjustment dynamics by estimating a policy function that controls for lagged input choice (labor). The residuals from this equation have negligible autocorrelation (AR(1) coefficient 0.016, as reported in Table A8). Therefore, in this variant, it is inessential to model mistakes as persistent. The resulting measure of Misoptimization Dispersion is strongly pro-cyclical (Panel (b) of Figure 3). The inability of adjustment dynamics to explain Fact 2 is consistent with our finding that misoptimizations had *persistent* negative effects on firm performance, instead of the mean-reverting effects which would obtain if firms endured poor performance today to reduce costly adjustments in the future (e.g., by hoarding labor).

**Financial Frictions.** We next consider financial frictions, corresponding to the policy function augmented with leverage and its interaction with TFP. This measure behaves almost identically to our baseline (Panel (c) of Figure 3). This is consistent with the notion that financial frictions and their heterogeneous incidence escalate in *downturns*, which would if anything lead to bias that pushes against our prediction and finding.

**Time-Varying Production Functions and Olley and Pakes (1996) Estimation.** To guard against mis-specification in estimating the production function and productivity, we also implement a method using time-varying cost shares and the method of Olley and Pakes (1996). Both measures give similar results (Panels (d) and (e) of Figure 3).

**Longer Time Period and Compositional Adjustment.** We also extend our time series back to 1950 to gauge its cyclicality properties over a longer period.[18] In so doing, we also account for firm fixed effects in $\hat{u}_{it}^2$ to adjust for the changing composition of the sample, which Davis et al. (2006) and Brown and Kapadia (2007) argue is important when studying patterns of volatility over long horizons in Compustat data. Notwithstanding significant changes in the macroeconomy over this long time period, misoptimization is similarly pro-cyclical over this longer horizon (Panel (f) of Figure 3 and Figure A2). Replicating the regressions against the unemployment rate and detrended S&P 500 price, we obtain coefficients of -0.571 (SE: 0.197) and 0.052 (SE: 0.014). Both estimates are within one standard error of our baseline estimates.[19]

**Additional Robustness Exercises.** In Appendix Table A3, we more systematically show the stability of our pro-cyclical misoptimization finding under the aforementioned variants, four other alternative constructions of policy and production functions described in Section 3.1, and variant strategies that control for linear and quadratic time trends and restrict attention to the manufacturing sector. We also recalculate misoptimizations and Misoptimization Dispersion using total variable cost expenditures and investment rates as the choice variable (Appendix Figure A3). We find broadly similar patterns, particularly in the spike of the mid-1990s and falls in the 2002 and 2009-10 downturns.

[18] One minor difference in the calculation is that, due to lack of sectoral wage data, we calculate TFP with two factors, materials (inclusive of labor) and capital.

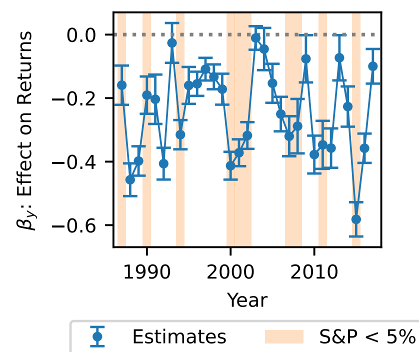
[19] When we adjust for composition in our baseline sample, the result measure has correlation 0.91 with our original version. This measure's regression slopes against unemployment and the detrended level of the S&P 500 are -0.520 (SE: 0.206) and 0.040 (SE: 0.009), comparable to our baseline estimates.

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

(a) Parametric Regression

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | | Outcome: $R_{it}$ | | |
| $\hat{u}_{it}^2$ | -0.268 | -0.262 | -0.097 | -0.087 |
| | (0.025) | (0.023) | (0.034) | (0.033) |
| $\hat{u}_{it}^2 \times R_t$ | 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 |

(b) Non-parametric Regression




*Notes*: In panel (a), $R_{it}$ is the firm-level log stock return. $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 3.1. $R_t$ is the log return of the S&P 500. Standard errors are double-clustered at the firm and year level. Panel (b) shows the estimates of $\beta_y$ from Equation 29. The blue dots are the point estimates and the blue error bars are 95% confidence intervals, based on standard errors clustered by firm and year. Years in which the S&P 500 return was less than 5% are shaded orange.

### 4.3 Fact 3: Returns Respond More to Misoptimizations During Downturns

We have shown that firms make smaller misoptimizations in downturns. We now investigate the extent to which this is explained by the incentives implied by our model: that firms have higher financial costs of misoptimization in downturns (Prediction 3).

To do this, we test whether misoptimizations have a state-dependent effect on stock returns. Specifically, we regress a firm's log stock return $R_{it}$ on the firm's squared misoptimization innovation over interacted with the log aggregate (S&P 500) stock return $R_t$:[20]

$$R_{it} = \beta \cdot \hat{u}_{it}^2 + \phi \cdot \left(\hat{u}_{it}^2 \times R_t\right) + \chi_{j(i),t} + \Gamma' X_{it} + \epsilon_{it} \tag{28}$$

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 hypothesis that the market punishes misoptimization more severely in times of distress, or low $R_t$, is captured by $\phi > 0$. In the expanded model of Appendix D, we directly derive the regression equation and the prediction $\phi \geq 0$,

[20] To be consistent with our earlier analysis, we run this regression using re-scaled mistake innovations $\hat{u}_{it}$. In Appendix Table A6, we replicate the analysis using our (first-stage) estimates of the mistake "level" $\hat{m}_{it}$ and find qualitatively similar results.

with equality only if investors are risk-neutral (there is no risk-pricing channel) and profit sensitivity to mistakes is state-independent (there is no profit-curvature channel).

We find that $\phi > 0$: misoptimization is priced more severely in states of low aggregate returns (panel (a) of Table 1).[21] 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 ($R_t = -0.52$), the model implies that pricing is 6.2 times more severe than in the "usual" states of $R_t \approx 0.10$.

In panel (b), we show estimates from the following regression which allows for a year-specific coefficient on $\hat{u}_{it}^2$ instead of imposing a parametric interaction with aggregate returns:

$$R_{it} = \sum_y \beta_y \cdot \hat{u}_{it}^2 \cdot \mathbb{I}[t = y] + \chi_{j(i),t} + \epsilon_{it} \tag{29}$$

We shade years in which the S&P 500 return is relatively low ($< 5\%$). Our estimate of $\phi > 0$ in the parametric model corresponds to the fact that, in this more non-parametric model, the plotted coefficients are more negative exactly when the S&P does poorly.

**Robustness.** Appendix Table A4 shows the stability of our main finding to all of the alternative data-construction approaches highlighted in the previous sections. In Appendix Table A5, we show that our finding of $\phi > 0$ is robust to controlling for other plausible heterogeneities in the effects of misoptimizations on stock returns. In particular, we control for the level and $\hat{u}_{it}^2$-interaction of TFP, lagged stock returns, and financial leverage to control for the observed tendencies for negative-return firms to have higher volatility (the leverage effect) and binding financial constraints; and we control for interactions of $\hat{u}_{it}^2$ with industry and firm fixed effects to model heterogeneous exposure to aggregates.

### 4.4 Fact 4: The State-Dependent Effects of Misoptimization are Driven by the Pricing of Profits and Not Profits Themselves

In our model, Fact 3 implies that at least one of the risk-pricing or profit-curvature channels drives changes in the cost of mistakes. To differentiate these explanations, we now explore the extent to which profitability is more sensitive to mistakes during downturns.

To this end, we define profitability $\pi_{it}$, as before, as this year's EBIT divided by the last year's total variable costs. We study the state-dependent effects of misoptimizations on profitability in the following regression that mirrors our previous analysis of stock

[21] Similar results are obtained using mistake "levels" (columns 3-6 of Table A6)

**Table 2:** Misoptimization, Profits, and Pricing

| | (1) Outcome: $\pi_{it}$ | (2) | (3) Outcome: $R_{it}$ | (4) |
|---|---|---|---|---|
| $\hat{u}_{it}^2$ | -0.114 | -0.021 | | |
| | (0.020) | (0.032) | | |
| $\hat{u}_{it}^2 \times R_t$ | 0.112 | | | |
| | (0.089) | | | |
| $\pi_{it}$ | | 0.400 | 0.421 | 0.690 |
| | | (0.028) | (0.034) | (0.305) |
| $\pi_{it} \times R_t$ | | | -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*: $\pi_{it}$ is firm-level profitability and $R_{it}$ is the firm-level log stock return. $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 3.1. $R_t$ is the log return of the S&P 500. Standard errors are double-clustered at the year and firm level.

returns:

$$\pi_{it} = \beta_\pi \cdot \hat{u}_{it}^2 + \phi_\pi \cdot \left(\hat{u}_{it}^2 \times R_t\right) + \chi_{j(i),t} + \gamma_i + \epsilon_{it} \tag{30}$$

We predict that $\phi_\pi = 0$ if and only if the profit function has state-independent curvature.

We find a positive, small, and statistically insignificant $\phi_\pi$ (column 1 of Table 2). Thus, misoptimizations have an almost constant effect on firms' dollar profits. This suggests that the mechanism for our earlier finding of state-dependent market punishment (Fact 3) relates primarily 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. We first regress firm stock returns on $\hat{u}_{it}^2$ and $\pi_{it}$ conditional on firm and sector by time fixed effects. 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, and not through any other channel.

We next explore whether profitability, more generally, has a larger effect on returns during downturns. This allows us to test whether the market values firm performance more strongly in low-return environments. Crucially, this tests our microeconomic mechanism without relying on our structural estimation of misoptimization. Our estimating equation is the mirror of Equation 28 with profitability in place of misoptimizations. We find that returns respond more to profitability when aggregate returns are low (column 3 of Table 2) consistent with our earlier findings and interpretation.

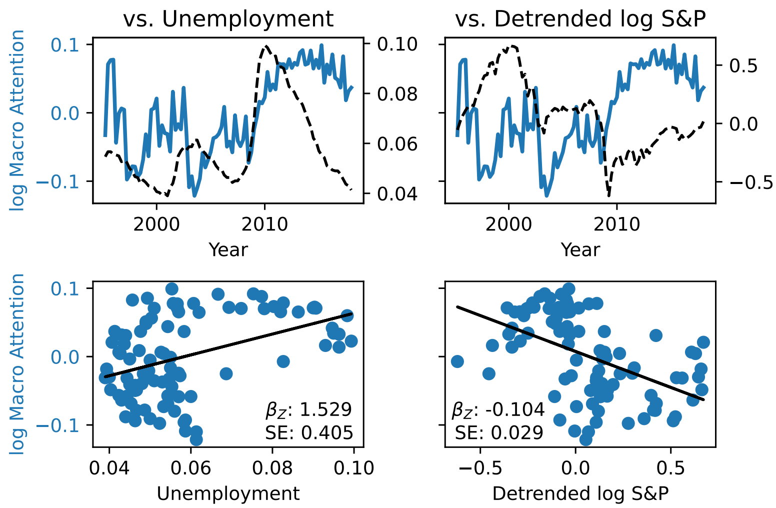
To quantify the pathway from misoptimizations to profitability to stock returns, we estimate an instrumental variables (IV) model of cyclical market responses. Specifically, we use $\hat{u}_{it}^2$ and its interaction with the market return as instruments for profitability and its interaction with the market return, to isolate the state-dependent pricing of profitability fluctuations that arise from mistakes. Our IV estimates suggest a greater state-dependence of the market response to misoptimization compared to the market response to other determinants of profits (column 4 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 misoptimization.

### 4.5 Fact 5: Macro Attention Rises in Downturns

We now test our predictions for measured attention. We find that macroeconomic attention, by our textual measure, persistently rises when the macroeconomy and financial market are distressed (Figure 4). To quantify the variable's cyclicality, we estimate the linear regression of Macro Attention on unemployment and the detrended S&P 500, akin to Equation 27. Our coefficient estimates are 1.529 (SE: 0.405) for unemployment and -0.104 (SE: 0.029) for the S&P, with $R^2$ values of 0.180 and 0.237. Our interpretation of this fact is that firms' cognitive effort to adapt their decisions to the state of the economy rises in downturns. This is consistent with Prediction 5.

**Robustness.** Appendix Figure A5 plots the time-series behavior of each word-level component of MacroAttention. In our sample, 61 of the 89 words have a positive correlation with unemployment (Appendix Figure A6). As mentioned before, some words appear *ex post* as "false positives" associated with pedagogy. Reassuringly, an index re-calculated without those words is, if anything, slightly more counter-cyclical (Appendix Figure A7). In Appendix C.1, 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-

**Figure 4:** Macro Attention is Counter-Cyclical




*Notes*: The top two panels plot log Macro Attention (blue line, left axis) along with, respectively, unemployment and the linearly detrended S&P 500 price (black dashed lines, right axis). The bottom two panels are scatterplots of log Macro Attention versus the corresponding macroeconomic aggregate. The black solid line is the linear regression fit. The standard errors are HAC robust based on a Bartlett kernel with a four-quarter bandwidth.

cyclical pattern over a smaller time period (2004-2013). In Appendix C.2, we employ an alternative procedure which uses the frequency of algorithmically determined word stems rather than full words. This yields essentially identical results.

An additional prediction is that attention is more cyclical in industries with more cyclical productivity due to an intensified profit-curvature channel, but is still cyclical in acyclical industries due to the risk-pricing channel. This prediction contrasts with that of an alternative model in which the cyclicality in macroeconomic attention is purely driven by firms' exposure to macroeconomic shocks. To study this, we compute "output cyclicality" in each of our industries as the correlation between sectoral GDP growth, calculated using quarterly BEA data since 2005 linked to our sectors, with aggregate nominal GDP growth. In Appendix Figure A8, 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. We find that the extent of counter-cyclicality increases with the industry's output cyclicality, but acyclical industries still have counter-cyclical attention. Thus, counter-cyclical attention does not arise merely as the result of increased exposure to macroeconomic conditions during downturns.

**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, Return Controls | | ✓ | | ✓ |
| $N$ | 28,279 | 24,392 | 27,875 | 23,930 |
| $R^2$ | 0.053 | 0.067 | 0.383 | 0.384 |

*Notes*: $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 3.1. $\log \text{MacroAttention}_{it}$ is the measure of firm-level macroeconomic attention. Standard errors are double-clustered at the firm and year level.

### 4.6 Fact 6: Macro Attention Predicts Smaller Misoptimizations

We next investigate the firm-level relationship between misoptimization and attention (Prediction 6). We estimate the following regression of $\hat{u}_{it}^2$, the squared innovation of the firm's model-implied misoptimization, on the log of firm-level macro attention:

$$\hat{u}_{it}^2 = \beta_a \cdot \log \text{MacroAttention}_{it} + \chi_{j(i),t} + \Gamma' X_{it} + \epsilon_{it} \tag{31}$$

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 log stock returns and TFP growth, to help further isolate variation in attention unrelated to firm-level fundamentals.

We find that $\beta_a < 0$: higher macroeconomic attention corresponds with smaller misoptimization (Table 3). The main specification in column (1) finds a strongly statistically significant effect ($p < 0.01$). The more controlled specifications in columns (2) to (4) estimate negative, similarly sized effects, but with less precision.

Fact 6 also helps rule out an important alternative interpretation of our time-series finding: that managers use the macroeconomy as a scapegoat for poor performance. This is inconsistent with the fact that macro-attentive firms make *fewer* unprofitable mistakes.

**Robustness.** Appendix Table A7 shows that our cross-sectional result is also similar using our conference-call measure of attention as well as all considered alternative models and measurement strategies for the misoptimizations.

### 4.7 Discussion and Relationship with the Literature

**Dispersion in Fundamentals Versus Misoptimizations.** Bloom et al. (2018), using micro-data on the manufacturing sector, estimate that the variance in total factor productivity rises in recessions or periods of negative growth.[22] Our analysis, by contrast, studies variance in input choices *conditional* on productivity.

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:

$$\log \theta_{it} = \gamma_i + \chi_{j(i),t} + \rho_\theta \log \theta_{i,t-1} + \epsilon_{it} \tag{32}$$

We estimate "TFP Innovation Variance" as the weighted average, $\mathbb{E}[s^*_{it}\hat{\epsilon}^2_{it}]/\mathbb{E}[s^*_{it}]$. This measure has a correlation of 0.39 with the equivalent from Bloom et al. (2018) over a common sample period, which increases to 0.47 if we restrict our data to the manufacturing sector.[23] Like Bloom et al. (2018), we find that TFP variance is significantly higher in recessions (coefficient: 0.098, SE: 0.022). The measure spikes markedly in the 2002 and 2007-09 recessions, as well as in the 1990s boom (Appendix Figure A4).

To summarize, in our data, misoptimizations are larger in booms, while TFP is less volatile. Thus, counter-cyclical TFP volatility is fully consistent with our hypothesis and finding that firms optimize more precisely conditional on productivity in downturns.

**Forecast Errors and Backcast Errors.** Our model makes no specific prediction about the accuracy of firm-level forecasts over the business cycle. To illustrate this, consider a firm's expectation in period $t$ of its production in period $t+k$, for $k>0$, conditional on the observed history of the decision state $z_{it}$ (firm-level TFP, aggregate output, and aggregate wages). Writing $\mathbb{E}_{it}[\cdot]$ as a shorthand for the firms' expectation conditional on this history and applying Proposition 1, we can decompose the variance of firms' forecast errors as:

$$\underbrace{\mathbb{E}\left[\left(x_{i,t+k} - \mathbb{E}_{it}\left[x_{i,t+k}\right]\right)^2\right]}_{\text{Forecast Error Variance}} = \underbrace{\mathbb{E}\left[\left(x^*_{i,t+k}(z_{i,t+k}) - \mathbb{E}_{it}\left[x^*_{i,t+k}\right]\right)^2\right]}_{\text{Fundamental Variance}} + \underbrace{\left(\sigma_i(z_{i,t+k})\right)^2}_{\text{Misoptimization Variance}} \tag{33}$$

[22] Kehrig (2015) reports that dispersion in levels of TFPR, or the marginal value product of all inputs under a Cobb-Douglas assumption, is counter-cyclical. In Appendix C.3, we study how this object, as well as the related calculation for the value marginal product of labor, behaves in our data.

[23] The common sample for comparing our measure with the one in Bloom et al. (2018) is 1987-2010. The measure from Bloom et al. (2018) that we study is 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.

The first term is the forecast error variance pertaining to the optimal level of production, $x^*_{i,t+k}$, which itself depends on unknown firm-level productivity, aggregate output, and aggregate wages. The second term is the forecast error variance arising from firms' future misoptimization, which has state-dependent volatility $\sigma_i(z_{i,t+k})$.

Our Fact 2 showed that Misoptimization Variance is high in booms and low in downturns. The finding of both prior work in the literature (Bloom et al., 2018), and our own reconstruction of these findings above, is that Fundamental Variance is high in downturns and low in booms. Because of these countervailing forces, our main hypotheses about misoptimizations cannot be tested by evaluating the cyclicality of forecast errors.[24]

By contrast, firm-level *backcasts* may have a more direct interpretation as "attentiveness." In Appendix G, we show two pieces of evidence consistent with our analysis in the survey of firms in New Zealand by Coibion et al. (2018). First, firms report being significantly more likely to seek out news about the macroeconomy if there were a negative aggregate shock. This is consistent with the cyclicality of the macroeconomic attention cycle. Second, firms that report a higher sensitivity of firm profits to their own choices demonstrate higher awareness of macroeconomic aggregates. This is consistent with the profit curvature channel.

# 5 Quantifying the Consequences of Attention Cycles

The previous section verified the model's microeconomic and macroeconomic predictions for misoptimization and attention, which govern the model's implications for output and productivity dynamics. In this section, we quantify the macroeconomic consequences of attention cycles by calibrating the model. We find that attention cycles generate: asymmetrically large amplification of negative shocks; greater amplification of shocks when output is low; and endogenously higher volatility when output is low. We show that these findings account for a significant portion of observed business-cycle asymmetries.

## 5.1 Calibration

In our calibration, as in our theoretical results of Section 2, the aggregate state variable $\theta_t = \left(\mathbb{E}_{G_t}[\theta_{it}^{\epsilon-1}]\right)^{\frac{1}{\epsilon-1}}$ is a one-dimensional sufficient statistic for the productivity distribution. We assume that $\log\theta_t$ follows a zero-mean, Gaussian AR(1) process, or

[24] See Charoenwong et al. (2021) and Chiang (2023) for an analysis of cyclical forecast error variance.

$\log \theta_t = \rho_\theta \log \theta_{t-1} + \nu_t$ where $\nu_t \sim^{IID} N(0, \sigma_\theta^2)$. Note that, in this formulation, the shocks $\nu_t$ may reflect changes to *any* moment of the productivity distribution that induce changes in the aggregate $\theta_t$, such as a standard shock to average productivity or an "uncertainty shock" to productivity dispersion as studied by Bloom et al. (2018).

We calibrate four parameters to standard values (see Table 4).[25] The first is the elasticity of substitution between products. We set $\epsilon = 4$, which implies an optimal average markup of $\frac{\epsilon}{\epsilon-1} = \frac{4}{3}$. This is conservative relative to estimates by De Loecker et al. (2020) (1.60) and slightly larger than the estimate by Edmond et al. (2018) (1.25). The elasticity of substitution controls the translation of misoptimization into output and productivity, with a lower value translating a fixed misoptimization dispersion 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 $\rho = 0.95$.[26] To match the elasticity of real wages to output ($\chi$), we only need information on the time-series correlation between real wages and output. 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.[27] Our estimate of 0.095 lines up with recent evidence on the weak cyclicality of wages (Galí and Gambetti, 2019). To calibrate the shock variance $\sigma_\theta^2$, we match the variance of quarterly real GDP growth over our sample period.

The key properties of inattention and misallocation are controlled by the remaining two moments: 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 (governed by $\lambda$) and the extent of its cyclicality (governed by $\gamma$). Our estimates of $\gamma$ are thus based entirely on fitting a stochastic discount factor that fits misoptimization in our model, rather than incorporating an informed prior from the asset pricing literature (see also Section 2.2). Our finding of $\gamma = 11.5$ is slightly conservative relative to the modern asset-pricing literature that estimates, in variations of

[25] The constants $\bar{w}$ and $\bar{X}$ in the wage rule scale overall production in equilibrium, but are otherwise irrelevant (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 $\log \theta = 0$.

[26] We study quarterly dynamics to compare our predictions to standard results about business-cycle asymmetries, although our measurement was annual. As our measurement is based on the long-run, cross-sectional variance of misoptimizations, the frequency of calibration is immaterial.

[27] 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.

**Table 4:** Parameters for Calibration

| | | | |
|---|---|---|---|
| Fixed Parameters | $\epsilon$ | Elasticity of substitution | 4 |
| | $\rho_\theta$ | Persistence of productivity | 0.95 |
| Free Parameters | $\chi$ | Elasticity of real wages to output | 0.095 |
| | $\lambda$ | Average weight on entropy penalty | 0.406 |
| | $\gamma$ | Coefficient of relative risk aversion | 11.5 |
| | $\sigma_\theta^2$ | Variance of the productivity innovation | $4.82 \times 10^{-7}$ |
| Matched Moments | $\beta_U$ | Slope of Misopt. Dispersion on -$\frac{\text{Unemployment}}{100}$ | 0.841 |
| | $\bar{\sigma}_M^2$ | Average level of Misopt. Dispersion | 0.080 |
| | $\chi$ | Regression of real wages on output | 0.095 |
| | $\sigma_Y^2$ | Variance of quarterly output growth | 0.337 |

*Notes*: "Fixed Parameters" are externally set. "Free Parameters" are chosen to fit the "Matched Moments," which are calculated from the data and matched exactly by the model.

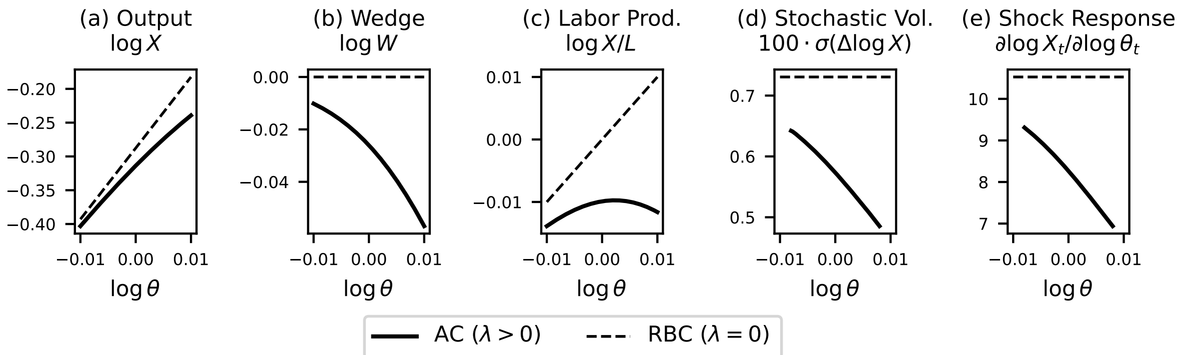
the consumption capital asset pricing model (CCAPM), $\gamma$ of about 15-20 with statistically "unfiltered" measures of consumption (Savov, 2011) or long-run variation in consumption growth (Parker and Julliard, 2005). Our finding of $\gamma > 0$ suggests that the risk-pricing channel is necessary to explain misoptimization dynamics and rules out the nested model in which control costs are denominated in dollar (rather than utility) units.

### 5.2 Quantitative Results

**Output, Productivity, and the Attention Wedge.** Panels (a), (b), and (c) of Figure 5 respectively show log output, the log attention wedge (as defined in Proposition 4), and log labor productivity in the calibrated model. 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 implied by Proposition 4 and Corollary 2, inattention reduces output and productivity relative to the fully attentive counterfactual, and this effect grows in higher productivity 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 $(1 - \chi\epsilon) \times 2.6\% = 1.6\%$. In the same state, a firm with the mean productivity loses 3.0% of its profits, on average, due to inattention.

To highlight the importance of studying misoptimization incentives in general equilibrium, we calculate also a "partial equilibrium" attention wedge based on firms' inattentive best-responses to the counterfactual RBC dynamics (Appendix Figure A9). In the mean

**Figure 5:** The Macro Effects of Attention Cycles




*Notes*: The left three panels plot log output, the log attention wedge (Proposition 4), and log labor productivity as a function of the state. The right two panels plot the shock-sensitivity of log output and the conditional volatility of output growth. In each graph, the solid line ("AC") is the calibrated model and the dotted line is a counterfactual model with $\lambda = 0$ ("RBC") and all other parameters held fixed.

state, the "partial equilibrium" attention wedge is 1.3% in terms of output, implying that general-equilibrium 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 depend on the numerical calibration. First, labor productivity is non-monotone in microeconomic productivity $\theta$ (rightmost panel) and hence also in aggregate output. This property results from the dueling forces of increased microeconomic productivity and increased misallocation from reduced attention.

Second, the attention wedge is concave. In Section 2.4, we discussed how the concavity or convexity of the attention wedge leads to business cycle asymmetries. The finding of a concave attention wedge implies that, fixing shock sizes, negative shocks have a larger effect on output than positive shocks, and that overall shock responses and volatility are higher in low-output states. We explore these predictions quantitatively in the next subsection.

**State-Dependence, Asymmetry, and Stochastic Volatility.** In our calibration, both the sensitivity of log output to productivity shocks and the conditional standard deviation of output growth are higher in low-productivity, low-output states (Panels (d) and (e) of Figure 5). This is despite the lack of asymmetry, heteroskedasticity, or stochastic volatility in the driving shocks.

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 the fundamental shock that induces a 3% change in output from steady state, or solves $\log X(\log \hat{\theta}) - \log X(0) = 0.03$. We compare the 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 $\log \theta_1 = -2 \cdot \log \hat{\theta}$. The negative shock has a 7% larger effect on log output than the positive shock, and the double dip shock has a 14% larger effect on log output than the positive shock. The same results for the response of log employment are 5% and 8%, respectively. Empirically, Ilut et al. (2018) estimate that US industries have on average a 20% larger response to negative aggregate productivity shocks than to positive shocks.[28] In these units, our model explains 25% of empirically realistic asymmetry in shock response.

To benchmark the extent of stochastic volatility, we observe that a transition from the 90th-percentile to the 10th-percentile productivity state reduces output by 4.7% and increases the conditional standard deviation of output growth by 10.6%. The peak-to-trough fall of output during the Great Recession (e.g., from early 2007 to early 2009) is comparable to this level change. Empirically, Jurado et al. (2015) estimate that the forward-looking volatility of industrial production growth at the three-month horizon increased by 57% during this episode.[29] Our model can explain about 19% of this movement.

**Parameter Robustness and Counterfactual Scenarios.** In Appendix E, we provide additional results from our numerical exercise. We first explore the robustness of our main findings to different external calibrations of wage rigidity $\chi$ and substitutability $\epsilon$ and to introducing classical labor markets using the preferences of Greenwood et al. (1988). In particular, the last has no quantitatively significant effect on our results provided that we calibrate to realistically acyclical wages.

We also study the effects of attention cycles under counterfactual scenarios. In our theory, the nature of business cycle asymmetries is endogenous to structural forces that control firms' state-dependent incentives for attention. We find that attention cycles generate greater asymmetries in regimes with larger markups, greater wage rigidity, and higher attention costs. Thus, our mechanism may interact with rising market power, the flattening wage Phillips curve, and fluctuations in macroeconomic uncertainty.

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

[29] We compare the 3-month "macro uncertainty index" from April 2007 to October 2008.

# 6 Conclusion

This paper studies how attention cycles arise as the consequence of business cycles through their effects on incentives and how attention cycles, in turn, affect business cycle dynamics. Theoretically, we introduce a Neoclassical business cycle model with flexible, state-dependent stochastic choice. We show how firms' choice of state-specific attention is shaped by the incentives embodied in the state-dependent cost of making mistakes. Due to firms' ownership by risk-averse households, the incentives to pay attention and to avoid mistakes are highest when aggregate consumption is low (a *risk-pricing channel*). Attention rises in downturns and falls in booms, leading to asymmetric propagation of macroeconomic shocks. Empirically, we introduce strategies to measure choice misoptimization, which has an exact analog in the model, and attention toward the macroeconomy, which is a more suggestive proxy for attention. Using these measures, we verify the model's macroeconomic hypotheses that attention rises and misoptimization falls in downturns. We also verify the model's microeconomic predictions that misoptimization harms firm valuations more severely in downturns and that attention coincides with lower misoptimization in the cross-section. Calibrating the model to match our evidence on cyclical misoptimization, we uncover a quantitatively important role for attention cycles in driving business-cycle asymmetries.

# References

Afrouzi, H. (2023). Strategic inattention, inflation dynamics, and the non-neutrality of money. Technical report, National Bureau of Economic Research.

Afrouzi, H. and Yang, C. (2021). Dynamic rational inattention and the Phillips curve. Working paper, Columbia University. SSRN Manuscript 3770462.

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.

Angeletos, G.-M. and Sastry, K. A. (2023). Inattentive economies. Mimeo, Northwestern University.

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

Barrero, J. M. (2022). The micro and macro of managerial beliefs. *Journal of Financial Economics*, 143(2):640–667.

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.

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.

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

Brown, G. and Kapadia, N. (2007). Firm-specific risk and equity market development. *Journal of Financial Economics*, 84(2):358–388.

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

Caplin, A. and Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203.

Caplin, A. and Martin, D. (2015). A testable theory of imperfect perception. *The Economic Journal*, 125(582):184–202.

Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2019). On binscatter. Pre-print 1902.09608, arXiv.

Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2007). Business cycle accounting. *Econometrica*, 75(3):781–836.

Charoenwong, B., Kimura, Y., Kwan, A., and Tan, E. (2021). Investment plans, uncertainty, and misallocation. Working Paper 3662347, SSRN.

Chiang, Y.-T. (2023). Attention and fluctuations in macroeconomic uncertainty. Mimeo, Federal Reserve Bank of St. Louis.

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

Costain, J. and Nakov, A. (2015). Precautionary price stickiness. *Journal of Economic Dynamics and Control*, 58:218–234.

Costain, J. and Nakov, A. (2019). Logit price dynamics. *Journal of Money, Credit and Banking*, 51(1):43–78.

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

Davis, S. J., Haltiwanger, J., Jarmin, R., Miranda, J., Foote, C., and Nagypal, E. (2006). Volatility and dispersion in business growth rates: Publicly traded versus privately held firms. *NBER Macroeconomics Annual*, 21:107–179.

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.

Dew-Becker, I. and Giglio, S. (2020). Cross-sectional uncertainty and the business cycle: evidence from 40 years of options data. Working Paper w27864, NBER.

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

Flynn, J. P. and Sastry, K. A. (2023). Strategic mistakes. *Journal of Economic Theory*, 212:105704.

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. In *New Developments in Productivity Analysis*, pages 303–372. U. of Chicago Press.

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

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 261–343. Elsevier.

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

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.

Greenwood, J., Hercowitz, Z., and Huffman, G. W. (1988). Investment, capacity utilization, and the real business cycle. *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.

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.

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.

Hébert, B. and La'O, J. (2023). Information acquisition, efficiency, and nonfundamental volatility. *Journal of Political Economy*, 131(10):2666–2723.

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. (2023). Economic agents as imperfect problem solvers. *The Quarterly Journal of Economics*, 138(1):313–362.

Jung, J., Kim, J. H. J., Matějka, F., and Sims, C. A. (2019). Discrete Actions in Information-Constrained Decision Problems. *The Review of Economic Studies*, 86(6):2643–2667.

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

Kehrig, M. (2015). The cyclical nature of the productivity distribution. Working paper.

Kohlhas, A. N. and Walther, A. (2021). Asymmetric attention. *American Economic Review*, 111(9):2879–2925.

Kuang, Y., Mislang, N., and Nimark, K. (2024). Individual and common information: Model-free evidence from probability forecasts. Working paper, Cornell University.

Lian, C. (2021). A theory of narrow thinking. *The Review of Economic Studies*, 88(5):2344–2374.

Link, S., Peichl, A., Roth, C., and Wohlfart, J. (2024). Attention to the macroeconomy. Working paper, CEBI. SSRN Manuscript 4697814.

Lorenzoni, G. (2009). A theory of demand shocks. *American Economic Review*, 99(5):2050–84.

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.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2018). Dynamic rational inattention: Analytical results. *Journal of Economic Theory*, 176:650–692.

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. (2016). Rationally inattentive seller: Sales and discrete pricing. *The Review of Economic Studies*, 83(3):1125–1155.

Miao, J., Wu, J., and Young, E. R. (2022). Multivariate rational inattention. *Econometrica*, 90(2):907–945.

Morris, S. and Yang, M. (2022). Coordination and continuous stochastic choice. *The Review of Economic Studies*, 89(5):2687–2722.

Nimark, K. (2008). Dynamic pricing and imperfect common knowledge. *Journal of Monetary Economics*, 55(2):365–382.

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.

Pavan, A. (2023). Attention, coordination, and bounded recall. Mimeo, Northwestern University.

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.

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. *J. of Monetary Econ.*, 50(3):665–690.

Sims, C. A. (2010). Rational inattention and monetary economics. In *Handbook of Monetary Economics*, volume 3, pages 155–181. Elsevier.

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. (2023). Firm inattention and the transmission of monetary policy: A text-based approach. Mimeo, University of Michigan.

Stevens, L. (2020). Coarse pricing policies. *The Review of Economic Studies*, 87(1):420–453.

Turen, J. (2023). State-dependent attention and pricing decisions. *American Economic Journal: Macroeconomics*, 15(2):161–189.

Van Nieuwerburgh, S. and Veldkamp, L. (2006). Learning asymmetries in real business cycles. *Journal of Monetary Economics*, 53(4):753–772.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. *Knowledge, Information, and Expectations in Modern Macroeconomics.*

Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, 12:579–601.

# Online Appendix

for "Attention Cycles" by Flynn and Sastry

## A Omitted Proofs

### A.1 Proof of Proposition 1

*Proof.* Consider a firm of type $\lambda_i$, 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

$$\max_{p \in \mathcal{P}} \int_{\mathcal{X}} \int_{\mathcal{Z}} u(x, z) p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z - \lambda_i \int_{\mathcal{X}} \int_{\mathcal{Z}} p(x|z) \log p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z \tag{34}$$

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 $\gamma(z)$, respectively:

$$\begin{aligned}
\mathcal{L}(\{p(x|z), \kappa(x, z)\}, \{\gamma(z)\}) = & \int_{\mathcal{Z}} \int_{\mathcal{X}} u(x, z) p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z \\
& - \lambda_i \int_{\mathcal{Z}} \int_{\mathcal{X}} p(x|z) \log p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z \\
& + \int_{\mathcal{Z}} \int_{\mathcal{X}} \kappa(x, z) p(x|z) \, \mathrm{d}x \, \pi(z) \, \mathrm{d}z \\
& + \int_{\mathcal{Z}} \gamma(z) \left( \int_{\mathcal{X}} p(x|z) \, \mathrm{d}x - 1 \right) \pi(z) \, \mathrm{d}z
\end{aligned} \tag{35}$$

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

$$u(x, z) - \lambda_i (\log p(x|z) + 1) + \kappa(x, z) + \gamma(z) = 0 \tag{36}$$

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

$$p(x|z) = \frac{\exp\left(\lambda_i^{-1} u(x, z)\right)}{\int_{\mathcal{X}} \exp\left(\lambda_i^{-1} u(x', z)\right) \mathrm{d}x'} \tag{37}$$

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

$$p(x|z_i) \propto \exp\left(-\frac{(x - x^*(z_i))^2}{2\lambda_i |\Pi_{zz}(z_i)|^{-1}}\right) \tag{38}$$

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 $|\Pi_{xx}(z_i)| = |\pi_{xx}(z_i)| M(z_i)$, we can rewrite the variance as $\lambda_i (|\pi_{xx}(z_i)| M(z_i))^{-1}$. 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. Thus:

$$x_i = x^*(z_i) + \sqrt{\frac{\lambda_i}{|\pi_{xx}(z_i)| M(z_i)}} \cdot v_i, \qquad v_i \sim \text{Normal}(0,1) \tag{39}$$

□

### A.2 Proof of Proposition 2

*Proof.* Throughout this proof, in some abuse of notation, we write $\theta = \hat{\theta}(G)$ (Equation 5) as the scalar summary of the productivity distribution. Our argument below makes, and then verifies, the conjecture that this is a scalar sufficient statistic for the productivity distribution in the definition of equilibrium output.

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 $\theta$.

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 $z_{it}$:

$$x_{it} = x^*(z_{it}) + \sqrt{\frac{\lambda_i}{|\Pi_{xx}(z_{it})|}} \cdot v_{it}, \qquad v_{it} \sim \text{Normal}(0,1) \tag{40}$$

As derived in Appendix A.6.1, the mean and variance scalings are the following, after substituting in the equilibrium conjecture $X_t = X(\theta_t)$:

$$\begin{aligned} x^*(z_{it}) &= v_x(\epsilon, \chi, \bar{w}, \bar{X}) \cdot X(\theta_t)^{1-\chi\epsilon} \theta_{it}^{\epsilon} \\ |\Pi_{xx}(z_{it})| &= v_\Pi(\epsilon, \chi, \bar{w}, \bar{X}) \cdot X(\theta_t)^{-1-\gamma+\chi(1+\epsilon)} \theta_{it}^{-1-\epsilon} \end{aligned} \tag{41}$$

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

$$\begin{aligned} v_x &:= \epsilon^{-\epsilon} (\epsilon - 1)^{\epsilon} \bar{w}^{-\epsilon} \bar{X}^{\chi\epsilon} \\ v_\Pi &:= (\epsilon - 1)^{-\epsilon} \epsilon^{\epsilon - 1} \bar{w}^{1+\epsilon} \bar{X}^{-\chi(1+\epsilon)} \end{aligned} \tag{42}$$

Conditional on the realization of any state $\theta$, aggregate output must solve the fixed point equation defined by combining the aggregate-good production function (11) and the firms' best responses. Applying the law of iterated expectations, we can re-write the aggregate-good production function as

$$X = X^* - \frac{1}{2\epsilon(X^*)^{-\frac{1}{\epsilon}}}\mathbb{E}_{\theta_i}\left[(x^*(z_i))^{-1-\frac{1}{\epsilon}}\mathbb{E}_{\lambda_i,v_i}\left[(x_i - x^*(z_i))^2 \mid \theta_i, \theta\right] \mid \theta\right] \tag{43}$$

where

$$X^* = \left(\mathbb{E}_{\theta_i}[x^*(z_i)^{1-\frac{1}{\epsilon}} \mid \theta]\right)^{\frac{\epsilon}{\epsilon-1}} \tag{44}$$

We now specialize the expressions above using the structure of the best response in Equations 40, 41, and 42. We first compute $X^*$ as

$$X^* = v_x X^{1-\chi\epsilon}\theta^{\epsilon} \tag{45}$$

where $\theta$ is the transformation defined in Equation 5. We next calculate the the second, "variance" term. We start with the "misoptimization variance"

$$\mathbb{E}_{\lambda_i,v_i}\left[(x_i - x^*(z_i))^2\right] = \frac{\lambda}{v_\Pi}X^{1+\gamma-\chi(1+\epsilon)}\theta_i^{1+\epsilon} \tag{46}$$

and then calculate the full term

$$(X^*)^{\frac{1}{\epsilon}}\mathbb{E}_{\theta_i}\left[(x^*(z_i))^{-1-\frac{1}{\epsilon}}\mathbb{E}_{\lambda_i,v_i}\left[(x_i - x^*(z_i))^2\right]\right] = \frac{\lambda}{v_\Pi v_x}X^{\gamma-\chi}\theta \tag{47}$$

where we simplify and apply the definition of $\hat{\theta}(G)$ from Equation 5.

Substituting in Equations 45 and 47, we derive that equilibrium output solves:

$$X(\theta) = v_x X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} - \frac{\lambda}{2\epsilon v_x v_\Pi}X(\theta)^{\gamma-\chi}\theta \tag{48}$$

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

$$g(X,\theta) = a_0 X^{1-\chi\epsilon}\theta^{\epsilon} - a_1 X^{\gamma-\chi}\theta \tag{49}$$

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 $X$:

$$\begin{aligned} g_X(X,\theta) &= a_0(1-\chi\epsilon)X^{-\chi\epsilon}\theta^{\epsilon} - a_1(\gamma-\chi)X^{\gamma-\chi-1}\theta \\ g_{XX}(X,\theta) &= -a_0(1-\chi\epsilon)\chi\epsilon X^{-\chi\epsilon-1}\theta^{\epsilon} - a_1(\gamma-\chi)(\gamma-\chi-1)X^{\gamma-\chi-2}\theta \end{aligned} \tag{50}$$

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

$$\lim_{X\to 0} g_X(X,\theta) = +\infty \qquad \lim_{X\to\infty} g_X(X,\theta) = -\infty \tag{51}$$

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 $\theta$. 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:

$$\begin{aligned}\frac{\mathrm{d}X(\theta)}{\mathrm{d}\theta} &= \left[a_0(1-\chi\epsilon)X(\theta)^{-\chi\epsilon}\theta^{\epsilon} - a_1(\gamma-\chi)X(\theta)^{\gamma-\chi-1}\theta\right]\frac{\mathrm{d}X(\theta)}{\mathrm{d}\theta} \\ &\quad + \left[a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^{\epsilon-1} - a_1 X(\theta)^{\gamma-\chi}\right]\end{aligned} \tag{52}$$

Yielding:

$$\frac{\mathrm{d}X(\theta)}{\mathrm{d}\theta} = \frac{a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^{\epsilon-1} - a_1 X(\theta)^{\gamma-\chi}}{1-\left[a_0(1-\chi\epsilon)X(\theta)^{-\chi\epsilon}\theta^{\epsilon} - a_1(\gamma-\chi)X(\theta)^{\gamma-\chi-1}\theta\right]} \tag{53}$$

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

$$\frac{\mathrm{d}\log X(\theta)}{\mathrm{d}\log\theta} = \frac{a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} - a_1 X(\theta)^{\gamma-\chi}\theta}{X(\theta)-\left[a_0(1-\chi\epsilon)X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} - a_1(\gamma-\chi)X(\theta)^{\gamma-\chi}\theta\right]} \tag{54}$$

We first show that the numerator is positive

$$\begin{aligned}a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} - a_1 X(\theta)^{\gamma-\chi}\theta &= a_0(\epsilon-1)X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} + X(\theta) \\ &> 0\end{aligned} \tag{55}$$

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 $\gamma > \chi + 1 > \chi$:

$$\begin{aligned}X(\theta) &= a_0 X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} - a_1 X(\theta)^{\gamma-\chi}\theta \\ &\geq a_0\underbrace{(1-\chi\epsilon)}_{\chi\epsilon<1}X(\theta)^{1-\chi\epsilon}\theta^{\epsilon} - a_1\underbrace{(\gamma-\chi)}_{\gamma-\chi>1}X(\theta)^{\gamma-\chi}\theta\end{aligned} \tag{56}$$

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

### A.3 Proof of Proposition 3

*Proof.* We start by proving 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 $(z_i, \lambda_i)$ is

$$m(z_i, \lambda_i, \theta) := \mathbb{E}_{v_i}\left[(x_i - x^*(z_i(\theta)))^2 \mid z_i, \lambda_i, \theta\right] = \frac{\lambda_i X(\theta)^{1+\gamma-\chi(1+\epsilon)}\theta_i^{1+\epsilon}}{v_\Pi(\epsilon, \chi, \bar{w}, \bar{X})} \tag{57}$$

where $v_\Pi > 0$ is defined in Equation 42. Using the law of iterated expectations, we can write $m(G) = \mathbb{E}_{\theta_i,\lambda_i}[m(z_i, \lambda_i, \theta)]$. Assessing this outer expectation, we derive

$$\begin{aligned} m(G) &= \frac{\lambda}{v_\Pi(\epsilon, \chi, \bar{w}, \bar{X})} \cdot X(\theta)^{1+\gamma-\chi(1+\epsilon)} \cdot \mathbb{E}_{\theta_i}[\theta_i^{1+\epsilon}] \\ &= \frac{\lambda}{v_\Pi(\epsilon, \chi, \bar{w}, \bar{X})} \cdot X(\theta)^{1+\gamma-\chi(1+\epsilon)} \cdot \theta^{1+\epsilon} \end{aligned} \tag{58}$$

where, in the second line, we use the definition of $\theta = \hat{\theta}(G)$. This expression clearly increases in $\theta$, conditional on $X$. 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. Thus, $m(G)$ can be written as $m(\theta)$ and $m(\theta)$ increases in $\theta$. This proves the first statement.

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

$$a(G) = (-1 - \gamma + \chi(1+\epsilon)) \log X(\theta) - (1+\epsilon)\mathbb{E}_{\theta_i}[\log \theta_i] \tag{59}$$

This is monotone decreasing in $X$ if $\gamma > \chi(1+\epsilon) - 1$, which is implied by our conditions $\chi\epsilon < 1$ and $\gamma > \chi + 1$. It is monotone decreasing in $G$ if $\mathbb{E}_{\theta_i}[\log \theta_i]$ increases in $G$. The latter claim is true because $y \mapsto \log y$ is an increasing function. Finally, $X$ is monotone in the FOSD ordering as $G \succsim_{FOSD} G' \implies \hat{\theta}(G) \geq \hat{\theta}(G')$. Therefore, $a(G)$ is monotone decreasing in $G$ in the sense of FOSD. This proves the second part of the statement.

□

### 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:

$$X(\theta) = a_0 X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1 X(\theta)^{\gamma-\chi}\theta \tag{60}$$

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

$$X_0(\theta) = a_0 X_0(\theta)^{1-\chi\epsilon}\theta^\epsilon \tag{61}$$

Or simply:

$$X_0(\theta) = a_0^{\frac{1}{\chi\epsilon}} \theta^{\frac{1}{\chi}} \tag{62}$$

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

$$W(\theta;\lambda) := \frac{X(\theta)}{X_0(\theta)} = \frac{X(\theta)}{a_0^{\frac{1}{\chi\epsilon}}\theta^{\frac{1}{\chi}}} \tag{63}$$

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

$$\log X(\log\theta) = X_0 + \chi^{-1}\log\theta + \log W(\log\theta) \tag{64}$$

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 60, we obtain

$$W(\theta) = W(\theta)^{1-\chi\epsilon} - a_1(\lambda)a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}} \tag{65}$$

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 $\theta$. First, we implicitly differentiate the fixed point condition:

$$\begin{aligned}\frac{\mathrm{d}W}{\mathrm{d}\theta} &= \left[(1-\chi\epsilon)W(\theta)^{-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}(\gamma-\chi)W(\theta)^{\gamma-\chi-1}\theta^{\frac{\gamma-1}{\chi}}\right]\frac{\mathrm{d}W}{\mathrm{d}\theta} \\ &\quad - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}\frac{\gamma-1}{\chi}W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-\chi-1}{\chi}}\end{aligned} \tag{66}$$

or:

$$\frac{\mathrm{d}W}{\mathrm{d}\theta} = \frac{-a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}\frac{\gamma-1}{\chi}W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-\chi-1}{\chi}}}{1-\left[(1-\chi\epsilon)W(\theta)^{-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}(\gamma-\chi)W(\theta)^{\gamma-\chi-1}\theta^{\frac{\gamma-1}{\chi}}\right]} \tag{67}$$

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

$$\frac{\mathrm{d}\log W}{\mathrm{d}\log\theta} = \frac{-a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}\frac{\gamma-1}{\chi}W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}}}{W(\theta)-\left[(1-\chi\epsilon)W(\theta)^{1-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}}(\gamma-\chi)W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}}\right]} \tag{68}$$

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:

$$
\begin{aligned}
W(\theta) &= W(\theta)^{1-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}} W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}} \\
&\geq \underbrace{(1-\chi\epsilon)}_{\chi\epsilon<1} W(\theta)^{1-\chi\epsilon} - a_1 a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}} \underbrace{(\gamma-\chi)}_{\gamma>\chi+1} W(\theta)^{\gamma-\chi}\theta^{\frac{\gamma-1}{\chi}}
\end{aligned} \tag{69}
$$

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:

$$L = \int_{[0,1]} L_i \,\mathrm{d}i \tag{70}$$

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

$$x_i = x^*(z_i) + \tilde{v}_i = v_x X(\theta)^{1-\chi\epsilon}\theta_i^{\epsilon} + \tilde{v}_i \tag{71}$$

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

$$L = v_x X(\theta)^{1-\chi\epsilon} \int_0^1 \theta_i^{\epsilon-1} \,\mathrm{d}i \tag{72}$$

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

$$L_t = L(\theta) = v_x X(\theta)^{1-\chi\epsilon}\theta^{\epsilon-1} \tag{73}$$

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

Combining the definition $\log A(\theta) = \log X(\theta) - \log L(\theta)$ with Equation 73, we derive

$$\log A(\theta) = -\log v_x + \chi\epsilon \log X(\theta) - (\epsilon - 1)\log\theta \tag{74}$$

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

$$\log A(\theta) = (\chi\epsilon X_0 - \log v_x) + \log\theta + \chi\epsilon \log W(\theta) \tag{75}$$

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. □

### A.6 Additional Calculations

#### A.6.1 Quadratic Approximation of Risk-Adjusted Profits

Using the expressions for dollar profits and marginal utility in Equation 8, we can write firms' risk-adjusted profits as the following:

$$\Pi(x, z_i) := X^{-\gamma}\left(x^{1-\frac{1}{\epsilon}} X^{\frac{1}{\epsilon}} - x\frac{w}{\theta_i}\right) \tag{76}$$

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

$$\left(1 - \frac{1}{\epsilon}\right) x^{*^{-\frac{1}{\epsilon}}} X^{\frac{1}{\epsilon}} = \frac{w}{\theta_i} \tag{77}$$

which can be re-arranged to define

$$x^*(z_i) = \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X \left(\frac{w}{\theta_i}\right)^{-\epsilon} \tag{78}$$

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

$$\begin{aligned}\Pi(x, z_i) &= \Pi(x^*(z_i), z_i) + \Pi_x(z_i)(x - x^*(z_i)) + \frac{1}{2}\Pi_{xx}(z_i)(x - x^*(z_i))^2 + O^3(x) \\ &=: \tilde{\Pi}(x, z_i) + O^3(x)\end{aligned} \tag{79}$$

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:

$$\tilde{\Pi}(x, z_i) = \Pi(x^*(z_i), z_i) + \frac{1}{2}\Pi_{xx}(z_i)(x - x^*(z_i))^2 \tag{80}$$

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

$$\begin{aligned}\Pi(x^*(z_i), z_i) &= X^{-\gamma}\left(X\left(\frac{w}{\theta_i}\right)^{1-\epsilon}\right)\left(\left[1 - \frac{1}{\epsilon}\right]^{\epsilon\left(1-\frac{1}{\epsilon}\right)} - \left[1 - \frac{1}{\epsilon}\right]^{\epsilon}\right) \\ &= X^{-\gamma}\left(X\left(\frac{w}{\theta_i}\right)^{1-\epsilon}\right)\epsilon^{-\epsilon}\left(\epsilon - 1\right)^{\epsilon-1}\end{aligned} \tag{81}$$

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

$$\Pi_{xx}(z_i) = X^{-\gamma} \cdot \pi_{xx}(z_i) \tag{82}$$

We calculate, using the form of the profit function from Equation 8, the dollar profit function's second derivative:[30]

$$\begin{aligned}\pi_{xx}(x^*(z_i),X) &= -\frac{1}{\epsilon}\left(1-\frac{1}{\epsilon}\right)(x^*(z_i))^{-1-\frac{1}{\epsilon}}X^{\frac{1}{\epsilon}} \\ &= -\frac{1}{\epsilon}\left(1-\frac{1}{\epsilon}\right)\left(1-\frac{1}{\epsilon}\right)^{-\left(1+\frac{1}{\epsilon}\right)\epsilon}X^{-1-\frac{1}{\epsilon}}X^{\frac{1}{\epsilon}}\left(\frac{w}{\theta_i}\right)^{\epsilon\left(1+\frac{1}{\epsilon}\right)} \\ &= -\epsilon^{\epsilon-1}(\epsilon-1)^{-\epsilon}X^{-1}\left(\frac{w}{\theta_i}\right)^{1+\epsilon}\end{aligned} \tag{83}$$

We substitute in the wage rule, Equation 3, to derive

$$\pi_{xx}(z_i) = -\upsilon_\pi(\epsilon,\chi,\bar{X},\bar{w})\cdot\theta_i^{-1-\epsilon}X^{\chi(1+\epsilon)-1} \tag{84}$$

as in Equation 13, where the constant is

$$\upsilon_\pi(\epsilon,\chi,\bar{X},\bar{w}) := (\epsilon-1)^{-\epsilon}\epsilon^{\epsilon-1}\bar{w}^{1+\epsilon}\bar{X}^{-\chi(1+\epsilon)} > 0 \tag{85}$$

### A.6.2 Quadratic Approximation of Final-Goods Technology

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

$$X(\{x_i\}_{i\in[0,1]}) = \left(\int_0^1 x_i^{\frac{\epsilon-1}{\epsilon}}\,\mathrm{d}i\right)^{\frac{\epsilon}{\epsilon-1}} \tag{86}$$

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:

$$X_{KK'K''}(\{x_{kk'k''}\}) = \left(\frac{1}{K}\sum_{k=1}^{K}\frac{1}{K'}\sum_{k'=1}^{K'}\frac{1}{K''}\sum_{k''=1}^{K''}x(\omega_{kk'k''})^{1-\frac{1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}} \tag{87}$$

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

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

$$X_{KK'K''} = X_K^* + \frac{1}{K}\sum_{k=1}^{K}\frac{1}{K'}\sum_{k'=1}^{K'}\frac{1}{K''}\sum_{k''=1}^{K''} D_k(x(\omega_{kk'k''}) - x^*(\theta_k))$$
$$+ \frac{1}{K}\sum_{k=1}^{K}\frac{1}{K'}\sum_{k'=1}^{K'}\frac{1}{K''}\sum_{k''=1}^{K''}\frac{1}{K}\sum_{\tilde{k}=1}^{K}\frac{1}{K'}\sum_{\tilde{k}'=1}^{K'}\frac{1}{K''}\sum_{\tilde{k}''=1}^{K''}\frac{1}{2}D^2_{k\tilde{k}}(x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{\tilde{k}\tilde{k}'\tilde{k}''}) - x^*(\theta_{\tilde{k}})) \tag{88}$$

where:

$$X_K^* = \left(\frac{1}{K}\sum_{k=1}^{K} x^*(\theta_k)^{1-\frac{1}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}} \tag{89}$$

and:

$$D_k = (X_K^*)^{\frac{1}{\epsilon}} x^*(\theta_k)^{-\frac{1}{\epsilon}} \tag{90}$$

and:

$$D^2_{k\tilde{k}} = \begin{cases} -KK'K''\frac{1}{\epsilon}x^*(\theta_k)^{-\frac{1}{\epsilon}-1}(X_K^*)^{\frac{1}{\epsilon}} + \frac{1}{\epsilon}\frac{\partial X_K^*}{\partial x_{kk'k''}}(X_K^*)^{\frac{1}{\epsilon}-1}x^*(\theta_k)^{-\frac{1}{\epsilon}} & \text{if } kk'k'' = \tilde{k}\tilde{k}'\tilde{k}'' \\ \frac{1}{\epsilon}(X_K^*)^{\frac{1}{\epsilon}-1}(x^*(\theta_k))^{-\frac{1}{\epsilon}}(x^*(\theta_{\tilde{k}}))^{-\frac{1}{\epsilon}} & \text{if } kk'k'' \neq \tilde{k}\tilde{k}'\tilde{k}''. \end{cases} \tag{91}$$

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:

$$\frac{1}{K''}\sum_{k''=1}^{K''} D_k(x(\omega_{kk'k''}) - x^*(\theta_k)) \to^{a.s} 0 \tag{92}$$

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:

$$Q = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{K'}\sum_{k'=1}^{K'}\frac{1}{K}\sum_{\tilde{k}=1}^{K}\frac{1}{K'}\sum_{\tilde{k}'=1}^{K'}\frac{1}{2}D^2_{k\tilde{k}}\left(\frac{1}{K''^2}\sum_{k''=1}^{K''}\sum_{\tilde{k}''=1}^{K''}(x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{\tilde{k}\tilde{k}'\tilde{k}''}) - x^*(\theta_{\tilde{k}}))\right) \tag{93}$$

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:

$$\frac{1}{K''^2} \sum_{k''=1}^{K''} \sum_{\tilde{k}'' \neq k''}^{K''} (x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{kk'\tilde{k}''}) - x^*(\theta_k)) \to^{a.s} 0 \tag{94}$$

$$\frac{1}{K''} \sum_{k''=1}^{K''} (x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{\tilde{k}\tilde{k}'k''}) - x^*(\theta_{\tilde{k}})) \to^{a.s} \sigma^2_{kk'} \tag{95}$$

Thus, using the observation that $\lim_{K'' \to \infty} \frac{\partial X^*_K}{\partial x_{kk'k''}} = 0$, we substitute in $D^2_{kk}$ to obtain:

$$Q = -\frac{1}{2\epsilon} \frac{1}{K} \sum_{k=1}^{K} \frac{1}{K'} \sum_{k'=1}^{K'} \frac{\sigma^2_{kk'}}{x^*(\theta_k)^{1+\frac{1}{\epsilon}} (X^*_K)^{-\frac{1}{\epsilon}}} \tag{96}$$

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

$$Q \to^{a.s.} -\frac{1}{2\epsilon} \frac{1}{K} \sum_{k=1}^{K} \frac{\lambda \sigma^2_k}{x^*(\theta_k)^{1+\frac{1}{\epsilon}} (X^*_K)^{-\frac{1}{\epsilon}}} \tag{97}$$

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

$$Q \to^{a.s} -\frac{1}{2\epsilon} \mathbb{E}\left[ \frac{\lambda \sigma^2_k}{x^*(\theta_k)^{1+\frac{1}{\epsilon}} (X^*_K)^{-\frac{1}{\epsilon}}} \right] \tag{98}$$

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

$$X^*_K \to^{a.s.} \left( \mathbb{E}[x^*(\theta_k)^{1-\frac{1}{\epsilon}}] \right)^{\frac{\epsilon}{\epsilon-1}} \tag{99}$$

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

$$X \approx \left( \mathbb{E}[(x^*(\theta_k))^{1-\frac{1}{\epsilon}}] \right)^{\frac{\epsilon}{\epsilon-1}} - \frac{1}{2\epsilon} \mathbb{E}\left[ \frac{\lambda \sigma^2_k}{x^*(\theta_k)^{1+\frac{1}{\epsilon}} (X^*_K)^{-\frac{1}{\epsilon}}} \right] \tag{100}$$

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

$$X = \left( \int_0^1 x^*(z_i)^{1-\frac{1}{\epsilon}} \, \mathrm{d}i \right)^{\frac{\epsilon}{\epsilon-1}} - \frac{1}{2\epsilon} \int_0^1 \frac{(x_i - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}} (x^*(z_i))^{1+\frac{1}{\epsilon}}} \, \mathrm{d}i \tag{101}$$

### A.6.3 Mapping Misoptimization Dispersion to the Model

Here, we explicitly calculate the within-model analog to Misoptimization Dispersion. We show that monotone misoptimization dispersion implies our within-model measure of misoptimization is monotone and therefore confirms the misoptimization cycles prediction.

Recall from Definition 2 the definition of aggregate misoptimization,

$$m(\theta) := \mathbb{E}_{\theta_i,\lambda_i,v_i}\left[(x_i - x^*(z_i(\theta)))^2 \mid \theta\right] \tag{102}$$

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

$$m(\theta) = \frac{\lambda}{v_\Pi(\epsilon,\chi,\bar{w},\bar{X})} \cdot X(\theta)^{1+\gamma-\chi(1+\epsilon)} \cdot \mathbb{E}_{\theta_i}[\theta_i^{1+\epsilon} \mid \theta] \tag{103}$$

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

$$\tilde{m}(\theta) = \mathbb{E}_{\theta_i,\lambda_i,v_i}\left[\hat{s}^*(\theta_i)\left(\frac{(x_i - x^*(z_i(\theta)))}{x^*(z_i)}\right)^2 \mid \theta\right] \tag{104}$$

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

$$\hat{s}^*(\theta_i) := \frac{q^*(z_i)x^*(z_i)}{\mathbb{E}_{\theta_i}[q^*(z_i)x^*(z_i)]} = \frac{X^{\frac{1}{\epsilon}}(v_x X^{1-\chi\epsilon}\theta_i^\epsilon)^{1-\frac{1}{\epsilon}}}{\mathbb{E}_{\theta_i}[X^{\frac{1}{\epsilon}}(v_x X^{1-\chi\epsilon}\theta_i^\epsilon)^{1-\frac{1}{\epsilon}}]} = \frac{\theta_i^{\epsilon-1}}{\theta^{\epsilon-1}} \tag{105}$$

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

$$\tilde{m}(z_i,\lambda_i,\theta) := \mathbb{E}_{v_i}\left[s^*(\theta_i)\left(\frac{x_i - x^*(z_i(\theta))}{x^*(z_i)}\right)^2 \mid z_i,\lambda_i,\theta\right] = \frac{\lambda_i X(\theta)^{\gamma+\chi(\epsilon-1)-1}\theta^{1-\epsilon}}{v_\Pi v_x^2} \tag{106}$$

where $v_\Pi, v_X > 0$ are defined in Equation 42. It is trivial to integrate over $(\theta_i,\lambda_i)$ to derive

$$\tilde{m}(\theta) = \frac{\lambda X(\theta)^{\gamma+\chi(\epsilon-1)-1}\theta^{1-\epsilon}}{v_\Pi v_x^2} \tag{107}$$

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

$$\frac{m(\theta)}{\tilde{m}(\theta)} = v_x^2\theta^{\epsilon-1}\mathbb{E}_{\theta_i}[\theta_i^{1+\epsilon} \mid \theta]X^{2(1-\chi\epsilon)} \tag{108}$$

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

# B Measuring Productivity and Misoptimization

This appendix describes in full detail our data construction and empirical methodologies for our firm-level analysis of misoptimization.

### B.1 Sample Selection and Data Construction

We use data from Compustat Annual Fundamentals. We define production, in value terms, 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.[31]

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;
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.

[31] Because of our later usage of fixed effects and lack of direct calculations using capital "expenditures" evaluated at an imputed rental rate, it is inessential to deflate the value of the capital stock.

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.

### B.2 Production Function and Productivity Estimation

Our primary method 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 markup $\mu > 1$ (in our baseline, $\mu = 4/3$) and constant returns to scale. We validate, in subsection B.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:
$$\begin{aligned} \text{Share}_{M,j'} &= \frac{\sum_{i:j(i)=j'} \sum_t \text{MaterialExpenditure}_{it}}{\sum_{i:j(i)=j'} \sum_t \text{Sales}_{it}} \\ \text{Share}_{L,j'} &= \frac{\sum_{i:j(i)=j'} \sum_t \text{WageBill}_{it}}{\sum_{i:j(i)=j'} \sum_t \text{Sales}_{it}} \end{aligned} \tag{109}$$
2. If $\text{Share}_{M,j'} + \text{Share}_{L,j'} \leq \mu^{-1}$, then set
$$\begin{aligned} \alpha_{M,j'} &= \mu \cdot \text{Share}_{M,j'} \\ \alpha_{L,j'} &= \mu \cdot \text{Share}_{L,j'} \\ \alpha_{K,j'} &= 1 - \alpha_{M,j'} - \alpha_{L,j'} \end{aligned} \tag{110}$$
3. Otherwise, adjust shares to match the assumed returns to scale, or set
$$\begin{aligned} \alpha_{M,j'} &= \frac{\text{Share}_{M,j'}}{\text{Share}_{M,j'} + \text{Share}_{L,j'}} \\ \alpha_{L,j'} &= \frac{\text{Share}_{L,j'}}{\text{Share}_{M,j'} + \text{Share}_{L,j'}} \\ \alpha_{K,j'} &= 0 \end{aligned} \tag{111}$$

We have experimented with extracting production function parameters under different assumed markups and found overall stable results for the behavior of misoptimizations.

To translate our production function estimates into productivity, we first calculate a "Sales Solow Residual" $\tilde{\theta}_{it}$ of the following form:

$$\log \tilde{\theta}_{it} = \log \mathrm{Sales}_{it} - \frac{1}{\mu}\left(\alpha_{M,j(i)} \cdot \log \mathrm{MatExp}_{it} - \alpha_{L,j(i)} \cdot \log \mathrm{Empl}_{it} - \alpha_{K,j(i)} \cdot \log \mathrm{CapStock}_{it}\right) \tag{112}$$

Because these variables are not in quantity units, we define our final estimate of TFP as the residual of the previous from industry-by-time fixed effects. This procedure, under our presumed model of industry-level variation in factor prices, identifies (log) TFP rescaled by $\mu^{-1}$. The rescaling is immaterial for our analysis.

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.[32] 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. We are re-assured by these estimates' "upstream" and "downstream" 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 A3, A4, and A7 demonstrate how our main aggregate and firm-level results replicate under the alternative measurement scheme, with similar quantitative and qualitative take-aways.

### B.3 Theory to Data: Micro-foundations

In this subsection, we outline the mapping from our model to our production function estimation via cost shares and our log-linear estimating equations.

Firms face a CES demand curve, or $\log q_{it} = \gamma_i - \frac{1}{\epsilon}(\log x_{it} - \log X_t)$ for some inverse elasticity $\epsilon > 1$ and aggregate output $X_t$.[33] 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.

[32] 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.

[33] It is straightforward, and consistent with our modeling approach, also to allow substitution within more narrowly defined industries.

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 $q_{i,T,t}$ denote the associated "Total" input cost per unit of produced output, and $x^*(q_T, X)$ denote the unconditionally profit-maximizing level of production. Firms choose a production level which differs from this level by a *misoptimization* $m_{it}$:

$$\log x_{it} = \log x^*(q_{i,T,t}, X_t) + m_{it} \tag{113}$$

And the dynamics of the misoptimization are described by an AR(1) process in which innovations $u_{it}$ are mean zero with variance $\sigma^2_{it}$: $m_{it} = \rho m_{i,t-1} + \left(\sqrt{1-\rho^2}\right) u_{it}$.

First, we characterize the firm's optimal production level $x^*$ as the total input price (which itself will depend on productivity) and aggregate demand:

**Lemma 1** (Optimal Output Choice)**.** *The firm's optimal output choice is*

$$\log x^*(\theta, q_T, X) = \epsilon \log\left(1 - \frac{1}{\epsilon}\right) + \epsilon\gamma_i + \log X - \epsilon \log q_T \tag{114}$$

*Proof.* Immediate from the first-order conditions of the program

$$x^*(\theta, q_T, X) = \arg\max_x \left\{ x \left( e^{\gamma_i} x^{-\frac{1}{\epsilon}} X^{\frac{1}{\epsilon}} - q_T \right) \right\} \tag{115}$$

□

We next characterize the optimal choice of each input:

**Lemma 2** (Input Choice)**.** *For any input $Z \in \{L, M, K\}$,*

$$\log Z_{it} = \eta_i + \chi_{j(i),Z,t} + (\epsilon - 1)\log\theta_{it} + m_{it} \tag{116}$$

*where*

$$\eta_i = \epsilon\gamma_i + \epsilon\log\left(1 - \frac{1}{\epsilon}\right) \tag{117}$$

*and*

$$\chi_{j(i),Z,t} = \log\alpha_{Z.j(i)} - \log q_{j(i),Z,t} + \log X_t + (1-\epsilon)\sum_Z \alpha_{Z,j(i)}(\log q_{j(i),Z,t} - \log\alpha_{Z,j(i)}) \tag{118}$$

*Proof.* In the cost minimization step, for any planned output choice $Q$, the firm solves

$$\min_{L_{it}, M_{it}, K_{it}} \sum_{z\in\{L,M,K\}} q_{j(i),Z,t} Z_{it} \qquad \text{s.t. } \theta_{it} L_{it}^{\alpha_{L,j(i)}} M_{it}^{\alpha_{M,j(i)}} K_{it}^{\alpha_{K,j(i)}} \geq Q \tag{119}$$

Standard first-order methods yield the solution, for each input,

$$\log Z_{it} = \log q_{i,T,t} + \log Q + \log\alpha_{Z,j(i)} - \log q_{j(i),Z,t} \tag{120}$$

where the price index $q_{i,T,t}$, which is also the Lagrange multiplier on the constraint, is

$$\log q_{i,T,t} = \sum_{Z} \alpha_{Z,j(i)}(\log q_{j(i),Z,t} - \log \alpha_{Z,j(i)}) - \log \theta_{it} \tag{121}$$

The desired expression comes from substituting in Equations 113 and 115 into the above.

□

This calculation validates our log-linear regression model Equation 20. It also has the same loading on the misoptimization $m_{it}$ and log productivity $\log \theta_{it}$ for all inputs $Z$. Thus, all separate inputs, as well as total production in physical units, inherit the "optimal choice 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 $i$,*

$$\alpha_{j(i),Z} = \left(1 - \frac{1}{\epsilon}\right)^{-1} \frac{q_{j(i),Z,t} Z_{it}}{q_{it} x_{it}} \exp\left(-\frac{m_{it}}{\epsilon}\right) \tag{122}$$

*Proof.* This can be calculated directly using the results of Lemmas 1 and 2. We work backwards starting from the result and substitute input demand from Equation 120 and calculate.

$$\begin{aligned} \alpha_{j(i),Z} &= \left(1 - \frac{1}{\epsilon}\right)^{-1} \frac{q_{j(i),Z,t} Z_{it}}{q_{it} x_{it}} \exp\left(-\frac{m_{it}}{\epsilon}\right) \\ &= \left(1 - \frac{1}{\epsilon}\right)^{-1} \frac{\exp\left(\log q_{i,T,t} + \log x_{it} + \log \alpha_{j(i),Z,t}\right)}{\exp\left(\log q_{it} + \log x_{it}\right)} \exp\left(-\frac{m_{it}}{\epsilon}\right) \\ &= \left(1 - \frac{1}{\epsilon}\right)^{-1} \exp\left(\log q_{i,T,t} - \log q_{it} + \log \alpha_{j(i),Z,t} - \frac{m_{it}}{\epsilon}\right) \end{aligned} \tag{123}$$

where the third line cancels out $x_{it}$ in the fraction. We take the demand curve $\log q_{it} = \gamma_i - \frac{1}{\epsilon}(\log x_{it} - \log X_t)$ and observe that, using the expression for the unconstrained optimal output in Equation 114 and the fact that $\log x_{it} = \log x_{it}^* + m_{it}$:

$$\begin{aligned} \log q_{it} &= \gamma_i + \frac{1}{\epsilon} \log X_t - \frac{1}{\epsilon}\left(\epsilon \log\left(1 - \frac{1}{\epsilon}\right) + \epsilon\gamma_i + \log X_t - \epsilon \log q_{i,T,t} + \log m_{it}\right) \\ &= -\log\left(1 - \frac{1}{\epsilon}\right) + \log q_{i,T,t} - \frac{m_{it}}{\epsilon} \end{aligned} \tag{124}$$

We then substitute the above back into Equation 123 to get

$$\begin{aligned}
\alpha_{j(i),Z} &= \left(1-\frac{1}{\epsilon}\right)^{-1} \exp\left(\log q_{i,T,t} - \left(-\log\left(1-\frac{1}{\epsilon}\right) + \log q_{i,T,t} - \frac{m_{it}}{\epsilon}\right) + \log \alpha_{j(i),Z} - \frac{m_{it}}{\epsilon}\right) \\
&= \left(1-\frac{1}{\epsilon}\right)^{-1} \exp\left((\log q_{i,T,t} - \log q_{i,T,t}) + \log\left(1-\frac{1}{\epsilon}\right) + \log \alpha_{j(i),Z} + \left(\frac{m_{it}}{\epsilon} - \frac{m_{it}}{\epsilon}\right)\right) \\
&= \alpha_{j(i),Z}
\end{aligned} \tag{125}$$

as desired. □

In words, this result says that the ratio of expenditures on input $Z$ to total sales, multiplied by 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.

We finally show, in the theory, how the calculation of Equation 112, net of fixed effects, recovers a re-scaling of TFP. In this subsection's language, that calculation is

$$\log \tilde{\theta}_{it} = \log x_{it} + \log q_{it} - \frac{1}{\mu}\left(\sum_Z \alpha_{Z,j(i)}(\log q_{j(i),Z,t} + \log Z_{it})\right) \tag{126}$$

Substituting in the demand curve $\log q_{it} = \gamma_i - \frac{1}{\epsilon}(\log x_{it} - \log X_t)$

$$\log \tilde{\theta}_{it} = \left(1-\frac{1}{\epsilon}\right)\log x_{it} + \gamma_i + \frac{1}{\epsilon}\log X_t - \frac{1}{\mu}\left(\sum_Z \alpha_{Z,j(i)}(\log q_{j(i),Z,t} + \log Z_{it})\right) \tag{127}$$

We next observe that $\mu = 1 - \frac{1}{\epsilon}$ and that $\log \theta_{it} = \log x_{it} - \left(\sum_Z \alpha_{Z,j(i)} \log Z_{it}\right)$. Thus,

$$\log \tilde{\theta}_{it} = \frac{1}{\mu}\log \theta_{it} + \gamma_i + \frac{1}{\epsilon}\log X_t - \frac{1}{\mu}\left(\sum_Z \alpha_{Z,j(i)} \log q_{j(i),Z,t}\right) \tag{128}$$

Grouping terms into fixed effects, this is

$$\log \tilde{\theta}_{it} = \frac{1}{\mu}\log \theta_{it} + \tau_{j(i),t} + \gamma_i \tag{129}$$

where $\tau_{j(i),t} = \frac{1}{\epsilon}\log X_t - \frac{1}{\mu}\left(\sum_Z \alpha_{Z,j(i)} \log q_{j(i),Z,t}\right)$ is the industry-by-time fixed effect capturing aggregate demand and factor prices. and $\gamma_i$ is the firm fixed effect from the demand curve. Thus, net of fixed effects, we recover $\frac{1}{\mu}\log \theta_{it}$.

# C Additional Empirical Results

## C.1 Macro Attention With Conference Calls

In this Appendix, we describe an alternative construction of Macro Attention measure using textual information from sales and earnings conference.

**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. We subset to documents that have reported firm names and stock tickers, which are automatically associated with documents by Lexis Nexis. When matches are probabilistic, we use the first (highest probability) match. We finally restrict to firms that are listed on one of three US stock exchanges: the NYSE, the NASDAQ, or the NYSE-MKT (Small Cap). We finally connect tickers to firm identifiers (GVKEY) using the master cross-walk available on Wharton Research Data Services (WRDS). These operations together reduce the sample size to 164,805 calls. We finally restrict to conference calls that are sales or earnings reports. This further reduces the sample to 158,810 total observations, by removing conference calls related to other activities (e.g., mergers). All in all, this sample is about 3,600 firm observations per quarter, or about 60% of the per-quarter observations we obtained via the SEC filings. Using these data, we replicate the exact methodology of Section 3.2 to measure macroeconomic attention. The procedure yields a new list of 73 macroeconomic words.

**Results.** Figure A10 plots the conference-call-derived measure alongside the US unemployment rate. Conference-call-derived macro attention, like our main measure 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 on unemployment and the detrended S&P 500. 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).

Casual comparison of Figure A10 and Figure 4 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 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 A7, our finding linking firm-level misoptimization with firm-level attention is robust to using the conference call measure.

### C.2 Macro Attention With Word Stemming

**Measurement.** 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,

$$\text{tf-idf}(s)_{it} := \text{tf}(s)_{it} \cdot \log\left(\frac{1}{\text{df}(s)}\right) \tag{130}$$

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.[34] We calculate the top macro stems using the approach described in the main text (Section 3.2); construct

[34] 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.

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.

**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 A7, 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).

### C.3 Dispersion in TFPR and Value Marginal Products

In this Appendix, we theoretically and empirically study the behavior of revenue-TFP (TFPR) dispersion in our analysis. This analysis builds a bridge between our findings and those of Kehrig (2015), who shows that cross-firm variance in revenue total-factor-productivity, or the product of prices and physical productivity, is counter-cyclical for the US durable manufacturing sector between 1972 and 2007.

**Definitions and Theoretical Context.** In our model, with a three-input production function, log physical TFP is defined as

$$\log \theta_{it} = \log x_{it} - (\alpha_{j(i),M} \log M_{it} + \alpha_{j(i),K} \log K_{it} + \alpha_{j(i),L} \log L_{it}) \tag{131}$$

where $x_{it}$ is physical sales, in quantity units; $(M_{it}, K_{it}, L_{it})$ are materials, capital, and labor; and $(\alpha_{j(i),M}, \alpha_{j(i),K}, \alpha_{j(i),L})$ are (industry-specific) weights on these inputs, which sum to one. Prices are defined by the demand curves $\log p_{it} = \frac{1}{\epsilon}(\log X_t - \log x_{it})$ and revenue-based TFP is therefore

$$\log \theta^R_{it} = \log \theta_{it} + p_{it} = \log \text{Sales}_{it} - (\alpha_{j(i),M} \log M_{it} + \alpha_{j(i),K} \log K_{it} + \alpha_{j(i),L} \log L_{it}) \tag{132}$$

where $\text{Sales}_{it} = p_{it} x_{it}$. We can calculate exactly what TFPR is in our empirical model with inattentive firms, introduced in Appendix B.3, by combining this definition with the input-choice policy functions derived in Lemma 2:

**Lemma 4.** *TFPR in our model is given by*

$$\log \theta^R_{it} = \tilde{\gamma}_i + \Xi_{j(i),t} - \frac{1}{\epsilon} m_{it} \tag{133}$$

*where $\tilde{\gamma}_i$ is a constant at the firm level and $\Xi_{j(i),t}$ is a constant that varies at the industry-by-time level.*

Thus, the only sources of within-industry variation in revenue-based TFP in our model are the firm fixed effects and misoptimizations. This is another way of stating our theoretical result that misoptimizations matter for aggregate output and productivity via their effects on "misallocation," as TFPR measures the value marginal product of the (minimal cost) input bundle. A simple corollary, under our assumption of cost minimization, is that the value marginal product $\log \theta_{it}^{Z} = \log \frac{\text{Sales}_{it}}{Z_{it}}$, for any of the three inputs, can also be written as $\log \theta_{it}^{Z} = \tilde{\gamma}_{Z,i} + \Xi_{j(i),Z,t} - \frac{1}{\epsilon} m_{it}$ with now input-specific fixed effects (firm-level and industry-by-time level).

In our model, therefore, the presence of TFPR dispersion or value-marginal-product dispersion indicates that there is non-zero misoptimization. The monotonicity of TFPR or value-marginal-product dispersion over the business cycle, once we project out firm and industry-by-time fixed effects, therefore provides an alternative test of the model's prediction of procyclical misoptimization dispersion.

**TFPR Dispersion and VMPL Dispersion.** We calculate log TFPR in our data by first calculating

$$\log \tilde{\theta}_{it}^{R} = \log \text{Sales}_{it} - \left(\alpha_{M,j(i)} \cdot \log \text{MatExp}_{it} - \alpha_{L,j(i)} \cdot \log \text{Empl}_{it} - \alpha_{K,j(i)} \cdot \log \text{CapStock}_{it}\right) \tag{134}$$

where variable definitions follow the convention of Appendix B. We then remove industry-by-time fixed effects and firm fixed effects to remove factor prices, as suggested by Lemma 4, to generate the variable $\log \hat{\theta}_{it}^{R}$.[35] We also calculate the log value marginal product of labor (VMPL) by first calculating

$$\log \tilde{\theta}_{it}^{L} = \log \text{Sales}_{it} - \log L_{it} \tag{135}$$

and removing industry-by-time and firm fixed effects to generate the final measure $\log \hat{\theta}_{it}^{L}$.

Appendix Figure A11 shows the cyclical behavior of TFPR dispersion, the variance of $\log \hat{\theta}_{it}^{R}$, and VMPL dispersion, the variance of $\log \hat{\theta}_{it}^{L}$. In line with our earlier results, VMPL dispersion is markedly pro-cyclical with respect to both the unemployment rate and the return on the S&P500. TFPR dispersion has no significant relationship with either unemployment or the S&P500. Taken together, these exercises suggest that value-marginal-product-based measures of misallocation give results consistent with our main finding of pro-cyclical misoptimization.

[35] Compare with Equation 112, which resembles Equation 134 but for deflating the input shares by the markup. This is exactly the within-model adjustment for prices $p_{it}$.

# Supplementary Materials

for "Attention Cycles" by Flynn and Sastry

(not for publication)

## D Extended Model

In this appendix, we formally develop the extension of the baseline model from Section 1 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 4.3.

### D.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 and 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 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 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.

**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:

$$M_t = \theta_t^M L_t^M \tag{136}$$

where $\theta_t^M$ 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 $\theta_{it}$ and use materials $m_{it}$ and labor $L_{it}$ to produce output $x_{it}$ with Cobb-Douglas production technology:

$$x_{it} = \theta_{it} L_{it}^{\alpha} m_{it}^{1-\alpha} \tag{137}$$

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 1.2. 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 $\Theta$ by the mapping $G : \Theta \to \Delta(\Theta)$, where we denote the productivity distribution in any state $\theta_t$ by $G_t = G(\theta_t)$ with corresponding density $g_t$. We assume that the total order on $\theta_t$ ranks distributions $G_t$ by first-order stochastic dominance, or $\theta \geq \theta'$ implies $G(\theta) \succsim_{FOSD} G(\theta')$. Finally, materials productivity $\theta_t^M$ is determined as an increasing function of the overall productivity state $\theta_t$.

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 9 in the main model, with a different definition of the decision state and profits function:

$$\max_{p \in \mathcal{P}} \int_{\mathcal{Z}} \int_{\mathcal{X}} \Pi(x, z_{it})\, p(x \mid z_{it})\, \mathrm{d}x \; f(z_{it} \mid z_{i,t-1})\, \mathrm{d}z - c\left(p, \lambda_i, z_{i,t-1}, f\right) \tag{138}$$

Intermediate goods firms generate profits in units of consumption goods given by $\pi_{it}$. The firms store these consumption goods and pay them out as dividends $d_{it}$ to their owners in the following period, or $d_{it+1} = \pi_{it}$. A unit supply of stock in the firm, which confers the right to the dividend stream, is available at price $P_{it}$.

The outputs of intermediate goods firms are 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:

$$X_t = X(\{x_{it}\}_{i \in [0,1]}) = \left( \int_{[0,1]} x_{it}^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{139}$$

**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:

$$\mathcal{U}^C(\{C_{t+j}\}_{j\in\mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta_C^j \frac{C_{t+j}^{1-\gamma}}{1-\gamma} \tag{140}$$

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

$$C_t + A_{t+1} + \int_{[0,1]} P_{it} S_{it+1} \, \mathrm{d}i \leq \int_{[0,1]} d_{it} S_{it} \, \mathrm{d}i + (1+r_t) A_t + \int_{[0,1]} P_{it} S_{it} \, \mathrm{d}i \tag{141}$$

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

Workers have preferences over streams of consumption and labor given by:

$$\mathcal{U}^W(\{C_{t+j}^W, L_{t+j}\}_{j\in\mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta_W^j U\left(C_t^W - \frac{L_{t+j}^{1+\psi}}{1+\psi}\right) \tag{142}$$

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:

$$C_t^W = w_t L_t \tag{143}$$

**Equilibrium.** An equilibrium is a set of all endogenous variables:

$$\{L_t^M, M_t, p_t^M, p_t^*, w_t, \{x_{it}, L_{it}, m_{it}, \pi_{it}, d_{it}, P_{it}, S_{it}\}_{i\in[0,1]}, X_t, L_t, C_t, A_t\} \tag{144}$$

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 be interested, as in the main text, in linear-quadratic equilibria where $\Pi$ is approximated around its optimal level and the CES aggregator is approximated as described in Section 1.5.

## D.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.

**Production by Intermediate Goods Firms.** Owing to CES aggregation, intermediate goods firms face an following iso-elastic demand curve, $q_{it} = X_t^{\frac{1}{\epsilon}} x_{it}^{-\frac{1}{\epsilon}}$. They moreover perfectly cost-minimize. As a result, given production level $x_{it}$ their unit input choices solve the following program:

$$\min_{L_{it}, m_{it}} w_t L_{it} + p_t^M m_{it} \quad \text{s.t.} \quad x_{it} = \theta_{it} L_{it}^{\alpha} m_{it}^{1-\alpha} \tag{145}$$

Taking the ratio of the two FOCs and rearranging, we obtain $m_{it} = \frac{1-\alpha}{\alpha}\frac{w_t}{p_t^M}L_{it}$. Thus, given $x_{it}$, the optimal labor and materials choices are given by:

$$L_{it} = \frac{1}{\theta_{it}}\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}\left(\frac{p_t^M}{w_t}\right)^{1-\alpha} x_{it}, \qquad m_{it} = \frac{1}{\theta_{it}}\left(\frac{\alpha}{1-\alpha}\right)^{-\alpha}\left(\frac{p_t^M}{w_t}\right)^{-\alpha} x_{it} \tag{146}$$

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

$$w_t L_{it} + p_t^M m_{it} = \frac{q_t}{\theta_{it}} x_{it} \tag{147}$$

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

$$q_t := \left[\left(\frac{\alpha}{1-\alpha}\right)^{-\alpha}\frac{1}{1-\alpha}\right] w_t^\alpha (p_t^M)^{1-\alpha} = c_\alpha w_t^\alpha (p_t^M)^{1-\alpha} \tag{148}$$

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

$$\pi_{it} = X_t^{\frac{1}{\epsilon}} x_{it}^{1-\frac{1}{\epsilon}} - \frac{q_t}{\theta_{it}} x_{it} \tag{149}$$

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:

$$\Pi_{it} = C_{t+1}^{-\gamma}\pi_{it} \tag{150}$$

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:

$$x^*(\Lambda, \theta_{it}) := \arg\max_{x \in \mathcal{X}} \; \Pi(x; \Lambda, \theta_{it}) \tag{151}$$

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 $x^*$:

$$\Pi_{xx}(\Lambda, \theta_{it}) := \left.\frac{\partial^2 \Pi}{\partial x^2}\right|_{x^*(\Lambda,\theta_{it});\Lambda,\theta} \tag{152}$$

The approximate objective of the intermediate goods firm is:

$$\tilde{\Pi}(x; \Lambda, \theta_{it}) := \bar{\Pi}(\Lambda, \theta_{it}) + \frac{1}{2}\Pi_{xx}(\Lambda, \theta_{it})(x - x^*(\Lambda, \theta_{it}))^2 \tag{153}$$

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

$$\begin{aligned} x_{it} &\sim N\left(x_{it}^*, \frac{\lambda}{|\Pi_{xx,it}|}\right) \\ x_{it}^* &= \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X_t \theta_{it}^{\epsilon} q_t^{-\epsilon} \\ |\Pi_{xx,it}| &= (\epsilon - 1)^{-\epsilon}\epsilon^{\epsilon-1} C_{t+1}^{-\gamma} X_t^{-1} \theta_{it}^{-1-\epsilon} q_t^{1+\epsilon} \end{aligned} \tag{154}$$

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.

**Finding Materials Prices and Wages.** Materials producers maximize their profits, $p_t^M \theta_t^M L_t^M - w_t L_t^M$. Thus, in equilibrium, it follows that $p_t^M = \frac{1}{\theta_t^M} w_t$. The worker's labor supply condition is $w_t = L_t^{\psi}$. Moreover, we know that aggregate labor is equal to the sum of labor used to produce intermediates and materials:

$$L_t = \int_{[0,1]} L_{it} \mathrm{d}i + L_t^M = \int_{[0,1]} L_{it} \mathrm{d}i + \frac{1}{\theta_t^M} \int_{[0,1]} m_{it} \mathrm{d}i \tag{155}$$

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:

$$\begin{aligned} L_t &= \int_{[0,1]} L_{it} di + \frac{1}{\theta_t^M} \int_{[0,1]} m_{it} di = \int_{[0,1]} L_{it} \, \mathrm{d}i + \frac{1}{\theta_t^M} \int_{[0,1]} \frac{1-\alpha}{\alpha} \frac{w_t}{p_t^M} L_{it} \, \mathrm{d}i \\ &= \left(1 + \frac{1}{\theta_t^M} \frac{1-\alpha}{\alpha} \frac{w_t}{p_t^M}\right) \int_{[0,1]} L_{it} \, \mathrm{d}i = \frac{1}{\alpha} \int_{[0,1]} L_{it} \mathrm{d}i \end{aligned} \tag{156}$$

Where the final equality follows from the fact that the material input is priced at marginal cost. 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:

$$L_t = \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha} \left(\frac{p_t^M}{w_t}\right)^{1-\alpha} \int_{[0,1]} \frac{x_{it}}{\theta_{it}} \mathrm{d}i \tag{157}$$

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

$$\begin{aligned} \int_{[0,1]} \frac{x_{it}}{\theta_{it}} \mathrm{d}i &= \mathbb{E}\left[\frac{x_{it}}{\theta_{it}}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{x_{it}}{\theta_{it}} | \theta_{it}\right]\right] = \mathbb{E}\left[\frac{1}{\theta_{it}} x_{it}^*\right] \\ &= \mathbb{E}\left[\frac{1}{\theta_{it}} \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X_t q_t^{-\epsilon} \theta_{it}^{\epsilon}\right] = \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X_t q_t^{-\epsilon} \theta^{\epsilon - 1} \end{aligned} \tag{158}$$

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:

$$L_t = \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha} \left(\frac{p_t^M}{w_t}\right)^{1-\alpha} \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} X_t q_t^{-\epsilon} \theta^{\epsilon-1} \tag{159}$$

Substituting this into the workers' intratemporal Euler equation, and using Equation 148 to write the marginal cost in terms of materials prices and wages, we obtain:

$$\begin{aligned} w_t^{\frac{1}{\psi}} &= \frac{1}{\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}\left(\frac{1}{\theta_t^M}\right)^{1-\alpha}\left(1-\frac{1}{\epsilon}\right)^{\epsilon}\theta_t^{\epsilon-1}c_\alpha^{-\epsilon}X_t\left(w_t\left(\frac{p_t^M}{w_t}\right)^{1-\alpha}\right)^{-\epsilon} \\ &= \frac{1}{\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}\left(\theta_t^M\right)^{(\epsilon-1)(1-\alpha)}\left(1-\frac{1}{\epsilon}\right)^{\epsilon}\theta^{\epsilon-1}c_\alpha^{-\epsilon}X_t w_t^{-\epsilon} \end{aligned} \tag{160}$$

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

$$q_t = c_\alpha w_t^\alpha (p_t^M)^{1-\alpha} = \bar{q}_t X_t^\chi \tag{161}$$

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

$$\bar{q}_t = \left[\left(\frac{\alpha}{1-\alpha}\right)^{-\alpha}\frac{1}{1-\alpha}\right]^{1-\alpha}\left(\frac{1}{\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}\left(\theta_t^M\right)^{(\epsilon-1)(1-\alpha)}\left(1-\frac{1}{\epsilon}\right)^{\epsilon}\theta^{\epsilon-1}c_\alpha^{-\epsilon}\right)^{\chi}(\theta_t^M)^{-1+\alpha} \tag{162}$$

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 161 is the "fully Neoclassical" analog to our wage rule Equation 3; indeed, when $\alpha = 1$ or there is no materials factor, it reduces to a wage rule

$$w_t = \bar{w}_t X_t^\alpha \tag{163}$$

where $\bar{w}_t = \bar{q}_t|_{\alpha=1}$. This verifies our claim in the main text that the wage rule can be 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.

**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:

$$X = \left(\int_0^1 x^*(z_i)^{1-\frac{1}{\epsilon}}\,\mathrm{d}i\right)^{\frac{\epsilon}{\epsilon-1}} - \frac{1}{2\epsilon}\int_0^1 \frac{(x_i - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}}(x^*(z_i))^{1+\frac{1}{\epsilon}}}\,\mathrm{d}i \tag{164}$$

where the mean and variance are taken over the realizations of $\theta_{it}$, conditional on the aggregate state $\theta_t$. Substituting in Equation 154, this provides one equation in terms of $(X_t, C_{t+1})$. Consider first the computation of $X_t^*$. Observe that we can write:

$$x_{it}^* = \delta_t X_t^{1-\chi\epsilon}\theta_{it}^\epsilon \tag{165}$$

where:

$$\delta_t = \left(1 - \frac{1}{\epsilon}\right)^{\epsilon} \left(c_\alpha \bar{w}_t \left(\frac{1}{\theta_t^M}\right)^{1-\alpha}\right)^{-\epsilon} \tag{166}$$

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

$$X_t^* = \left(\delta_t^{\frac{\epsilon-1}{\epsilon}} X_t^{\frac{\epsilon-1}{\epsilon}(1-\chi\epsilon)} \theta_t^{\epsilon-1}\right)^{\frac{\epsilon}{\epsilon-1}} = \delta_t X_t^{1-\chi\epsilon} \theta_t^{\epsilon} \tag{167}$$

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

$$\begin{aligned} \mathbb{E}\left[\frac{(x_{it} - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}}(x^*(z_i))^{1+\frac{1}{\epsilon}}}\right] &= (X_t^*)^{\frac{1}{\epsilon}} \mathbb{E}\left[x^*(z_{it})^{-1-\frac{1}{\epsilon}} \mathbb{E}[(x_{it} - x^*(z_{it}))^2 | z_{it}]\right] \\ &= (X_t^*)^{\frac{1}{\epsilon}} \mathbb{E}\left[x^*(z_{it})^{-1-\frac{1}{\epsilon}} \frac{\lambda}{|\Pi_{xx,it}|}\right] \end{aligned} \tag{168}$$

To simplify this, observe that we can write:

$$\frac{1}{|\Pi_{xx,it}|} = \zeta_t^{-1} C_{t+1}^{\gamma} X_t^{1-\chi(\epsilon+1)} \theta_{it}^{1+\epsilon} \tag{169}$$

where:

$$\zeta_t = (\epsilon - 1)^{-\epsilon} \epsilon^{\epsilon-1} \left(c_\alpha \bar{w}_t \left(\frac{1}{\theta_t^M}\right)^{1-\alpha}\right)^{1+\epsilon} \tag{170}$$

So we may express:

$$\begin{aligned} \mathbb{E}\left[\frac{(x_{it} - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}}(x^*(z_i))^{1+\frac{1}{\epsilon}}}\right] &= (X_t^*)^{\frac{1}{\epsilon}} \mathbb{E}\left[\delta_t^{-1-\frac{1}{\epsilon}} X_t^{-(1+\frac{1}{\epsilon})(1-\chi\epsilon)} \theta_{it}^{-1-\epsilon} \lambda \zeta_t^{-1} C_{t+1}^{\gamma} X_t^{1-\chi(\epsilon+1)} \theta_{it}^{1+\epsilon}\right] \\ &= \lambda \zeta_t^{-1} \delta_t^{-1} C_{t+1}^{\gamma} X_t^{-\chi} \end{aligned} \tag{171}$$

Putting all of the above together we have that:

$$X_t = \delta_t X_t^{1-\chi\epsilon} \theta_t^{\epsilon} - \frac{\lambda}{2\epsilon} \zeta_t^{-1} \delta_t^{-1} C_{t+1}^{\gamma} X_t^{-\chi} \theta_t \tag{172}$$

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:

$$C_{t+1} = \int_{[0,1]} \pi_{it} \, \mathrm{d}i \tag{173}$$

Using our running approximation:

$$\pi_{it} = \pi_{it}(x_{it}^*) + \frac{1}{2} \pi_{xx,it} (x_{it} - x_{it}^*)^2 \tag{174}$$

we obtain:

$$\begin{aligned} C_{t+1} &= \int_{[0,1]} \pi_{it} di = \mathbb{E}\left[\mathbb{E}\left[\pi_{it}(x_{it}^*) + \frac{1}{2}\pi_{xx,it}(x_{it}^*)(x_{it} - x_{it}^*)^2 \mid x_{it}^*\right] \mid \delta_t, \theta_t\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\pi_{it}(x_{it}^*) + \frac{1}{2}\pi_{xx,it}(x_{it}^*)\frac{\lambda}{|\Pi_{xx,it}|} \mid x_{it}^*\right] \mid \delta_t, \theta_t\right] = \mathbb{E}\left[\pi_{it}(x_{it}^*) \mid \delta_t, \theta_t\right] - \frac{\lambda}{2}C_{t+1}^{\gamma} \\ &= (\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}X_t^{1-\chi(\epsilon-1)} - \frac{\lambda}{2}C_{t+1}^{\gamma} \end{aligned} \tag{175}$$

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

$$X_t = \left(\frac{C_{t+1} + \frac{\lambda}{2}C_{t+1}^{\gamma}}{(\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}}\right)^{\frac{1}{1-\chi(\epsilon-1)}} \tag{176}$$

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:

$$\begin{aligned} \left(\frac{C_{t+1} + \frac{\lambda}{2}C_{t+1}^{\gamma}}{(\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}}\right)^{\frac{1}{1-\chi(\epsilon-1)}} &= \delta_t\left(\frac{C_{t+1} + \frac{\lambda}{2}C_{t+1}^{\gamma}}{(\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}}\right)^{\frac{1-\chi\epsilon}{1-\chi(\epsilon-1)}}\theta_t^{\epsilon} \\ &\quad - \frac{\lambda}{2\epsilon}\zeta_t^{-1}\delta_t^{-1}C_{t+1}^{\gamma}\left(\frac{C_{t+1} + \frac{\lambda}{2}C_{t+1}^{\gamma}}{(\epsilon - 1)^{-1}\delta_t\theta_t^{\epsilon-1}}\right)^{\frac{-\chi}{1-\chi(\epsilon-1)}}\theta_t \end{aligned} \tag{177}$$

The above can be summarized in the following result:

**Proposition 5.** *Equilibria are characterized by the solutions to Equation 177.*

### D.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 $\theta$.*

*Proof.* Following Equation 177, define:

$$
\begin{aligned}
g_\lambda(C) = \left(\frac{C + \frac{\lambda}{2}C^\gamma}{(\epsilon-1)^{-1}\delta\theta^{\epsilon-1}}\right)^{\frac{1}{1-\chi(\epsilon-1)}} - \Bigg[\delta\left(\frac{C + \frac{\lambda}{2}C^\gamma}{(\epsilon-1)^{-1}\delta\theta^{\epsilon-1}}\right)^{\frac{1-\chi\epsilon}{1-\chi(\epsilon-1)}}\theta^\epsilon \\
- \frac{\lambda}{2\epsilon}\zeta^{-1}\delta^{-1}C^\gamma\left(\frac{C + \frac{\lambda}{2}C^\gamma}{(\epsilon-1)^{-1}\delta\theta^{\epsilon-1}}\right)^{\frac{-\chi}{1-\chi(\epsilon-1)}}\theta\Bigg]
\end{aligned}
\tag{178}
$$

Observe that $g_\lambda$ is continuous in $\lambda$ 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 $\lambda$ 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:

$$
g_0(\tilde{C}) = \tilde{C}^{\frac{1}{1-\chi(\epsilon-1)}} - b_{\tilde{C}}\tilde{C}^{\frac{1-\chi\epsilon}{1-\chi(\epsilon-1)}}
\tag{179}
$$

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

$$
\begin{aligned}
g_0'(\tilde{C}) &= \frac{1}{1-\chi(\epsilon-1)}\tilde{C}^{\frac{\chi(\epsilon-1)}{1-\chi(\epsilon-1)}} - b_{\tilde{C}}\frac{1-\chi\epsilon}{1-\chi(\epsilon-1)}\tilde{C}^{\frac{-\chi}{1-\chi(\epsilon-1)}} \\
g_0''(\tilde{C}) &= \frac{\chi(\epsilon-1)}{(1-\chi(\epsilon-1))^2}\tilde{C}^{\frac{\chi(\epsilon-1)}{1-\chi(\epsilon-1)}-1} + b_{\tilde{C}}\frac{\chi(1-\chi\epsilon)}{(1-\chi(\epsilon-1))^2}\tilde{C}^{\frac{-\chi}{1-\chi(\epsilon-1)}-1}
\end{aligned}
\tag{180}
$$

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

$$
\lim_{\tilde{C}\to 0} g_0'(\tilde{C}) = -\infty \qquad \lim_{\tilde{C}\to\infty} g_0'(\tilde{C}) = \infty \qquad g_0''(\tilde{C}) > 0 \quad \text{for all}\, \tilde{C} \in \mathbb{R}_{++}
\tag{181}
$$

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 $\theta$, 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:

$$
\ln\tilde{C} = \frac{(1-\chi(\epsilon-1))}{\chi\epsilon}\ln b_{\tilde{C}}
\tag{182}
$$

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

$$
\ln C = \frac{(1-\chi(\epsilon-1))}{\chi\epsilon}(\ln\delta + \epsilon\ln\theta) - \ln(\epsilon-1) + \ln\delta + (\epsilon-1)\ln\theta
\tag{183}
$$

As $\epsilon > 1$ and $1 > \chi(\epsilon - 1)$ by hypothesis, and $\delta$ is increasing in $\theta$, the result follows. □

### D.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 analog 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:

$$m(z_{it}) = \frac{\lambda_i}{|\Pi_{xx,it}|} = \lambda_i \zeta_t^{-1} C_{t+1}^{\gamma} X_t^{1-\chi(\epsilon+1)} \theta_{it}^{1+\epsilon} \tag{184}$$

Thus, the average extent of misoptimization in aggregate state $\theta$ is:

$$m(\theta) = \lambda \zeta(\theta)^{-1} C(\theta)^{\gamma} X(\theta)^{1-\chi(\epsilon+1)} \mathbb{E}[\theta_{it}^{1+\epsilon} \mid \theta] \tag{185}$$

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 $\theta$. By the assumed FOSD ordering on $\theta$, we have that $\mathbb{E}[\theta_{it}^{1+\epsilon} \mid \theta]$ is monotone increasing in $\theta$. We moreover have that $\xi \propto \delta^{-\frac{1+\epsilon}{\epsilon}}$. Thus, as $\delta$ is increasing in $\theta$, we have that $\xi^{-1}$ is increasing in $\theta$. This establishes that $m(\theta)$ is increasing in $\theta$, 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. □

### D.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:*

$$\log X(\log\theta, \lambda) = \frac{1}{\chi} \log\tilde{\theta} + \log W(\log\theta, \lambda) \tag{186}$$

*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 182 and 176. 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 $\theta$ 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.

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

In Section 4.3, 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 1 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 $t$, $P_{it}$ solves:

$$u'(C_t)P_{it} = \mathbb{E}_t[\beta u'(C_{t+1})(P_{it+1} + d_{it+1})] \tag{187}$$

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

$$u'(\pi_{t-1})P_{it} = \beta u'(\pi_t)\pi_{it} + \beta u'(\pi_t)\mathbb{E}_t[P_{it+1}] \tag{188}$$

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

$$P_{it} = \beta\frac{u'(\pi_t)}{u'(\pi_{t-1})}\pi_{it} + \beta\frac{u'(\pi_t)}{u'(\pi_{t-1})}\mathbb{E}_t[P_{it+1}] \tag{189}$$

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

$$\pi_{it} = \pi_{it}(x_{it}^*) + \pi_{xx,it}m_{it}^2 \tag{190}$$

Thus, the firm's stock price follows:

$$P_{it} = \beta\frac{u'(\pi_t)}{u'(\pi_{t-1})}\left(\pi_{it}(x_{it}^*) + \pi_{xx,it}m_{it}^2\right) + \beta\frac{u'(\pi_t)}{u'(\pi_{t-1})}\mathbb{E}_t[P_{it+1}] \tag{191}$$

Thus:

$$\frac{\partial P_{it}}{\partial m_{it}^2} = \beta\frac{u'(\pi_t)}{u'(\pi_{t-1})}\pi_{xx,it} \qquad\qquad \frac{\partial P_{it}}{\partial \pi_{it}} = \beta\frac{u'(\pi_t)}{u'(\pi_{t-1})} \tag{192}$$

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):

$$u'(C_t)P_t = \mathbb{E}_t[\beta u'(C_{t+1})(P_{t+1} + \pi_t)] \tag{193}$$

Or:

$$\beta \frac{u'(\pi_t)}{u'(\pi_{t-1})} = \frac{P_t}{\mathbb{E}_t[P_{t+1}] + D_{t+1}} = \frac{1}{R_t} \tag{194}$$

Which is the inverse aggregate return on equity between period $t$ and period $t+1$, $R_t$. Therefore:

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

$$\frac{\partial P_{it}}{\partial m_{it}^2} = -\frac{1}{R_t} |\pi_{xx,it}| \tag{195}$$

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

$$\frac{\partial P_{it}}{\partial \pi_{it}} = -\frac{1}{R_t} \tag{196}$$

*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.

# E 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.

## E.1 Sensitivity of Main Results

**Parameter Choice.** To probe robustness to our choice of elasticity of substitution $\epsilon$ and the wage rule slope $\chi$, we re-calibrate the model for alternative choices. 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 5.2: (i) the relative output effect of negative and positive shocks, normalized in $\log \theta$ units such that the latter increases output by 3%; (ii) the relative output effect of a "double dip" versus positive shock, holding fixed the size of the shock to productivity as above; and (iii) the ratio of output-growth volatility from the 10th to the 90th percentile of the output distribution.

We present our results in Figure A12. 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 $\epsilon$, toward the level implied by De Loecker et al. (2020), significantly increases the extent of our predicted asymmetries ($\approx 1.75$x), while decreasing markups or increasing $\epsilon$, toward the level implied by Edmond et al. (2018), modestly increases the extent of our predicted asymmetries. Increasing the slope of the wage rule 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.

**Classical Labor Markets.** For tractability, we assumed a 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 1 with the following:

$$\mathcal{U}(\{C_{t+j}, L_{t+j}\}_{j\in\mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \frac{\left(C_{t+j} - \frac{L_{t+j}{}^{1+\phi}}{1+\phi}\right)^{1-\gamma}}{1-\gamma} \tag{197}$$

and remove the wage rule, Equation 3. 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 $\chi$ and $\epsilon$ take our benchmark values indicated in Table 4. As indicated in the richer model of Appendix D, this calibration replicates an elasticity of $\chi = 0.095$ between real wages and real output.

Figure A13 is this model's analog to panels (a)-(c) of Figure 5, showing output, the attention wedge, and labor productivity. We find comparable behavior of the attention wedge and losses from misoptimization and inattention. Figure A14 is this model's analog to Figure panels (d) and (e) of Figure 5, 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.

### E.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.

**The Rise of Markups, via Lower Substitutability.** A large 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 markups as informed by this literature. Our framework would interpret any trends in aggregate markups as arising from changes in the elasticity of substitution $\epsilon$ 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 $\epsilon$ upward to simulate a 15 percentage point decrease in the aggregate markup, to match the estimate of Demirer (2020) for markups since the 1960s; and second, we adjust $\epsilon$ downward to match 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 A15).

**More Rigid Real Wages.** The relationship between wage inflation and real growth has proved elusive in modern data, particularly since the financial crisis (see, e.g., Galí and Gambetti, 2019). 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.[36] We find, as shown in panels four and five of Figure A15, 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.

**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, 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," formalized in our model as variation in the attention cost $\lambda$, 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 A15). 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.

# F 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

[36] In particular, we use the ratio of the 1964-2007 estimate and 2007-2017 estimates in Table 3A of Galí and Gambetti (2019).

extraction, and optimal signal processing with mutual information costs.

### F.1 Persistent Mistakes

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 state-separable cost functions $c : \mathcal{P} \times \mathcal{Z} \to \mathbb{R}$ of the form:

$$c(p; \theta_{t-1}) = \int_{\Theta} \int_{\mathcal{X}} \phi(p(x|\theta)) \,\mathrm{d}x \, f(z|z_{t-1}) \,\mathrm{d}z \tag{198}$$

for some convex $\phi$ 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:

$$c(p; z_{t-1}, v_{t-1}) = \int_{\Theta} \int_{\mathcal{X}} \phi(p(x|\theta); v_{t-1}, x^*, x) \,\mathrm{d}x \, f(z|z_{t-1}) \,\mathrm{d}z \tag{199}$$

for $\phi$ 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:

$$\phi(y; m, x^*, x) = \lambda y \log y + \omega y((x - x^*) - m)^2 \tag{200}$$

Concretely, this leads to the following cost functional:

$$\begin{aligned} c(p; z_{t-1}, v_{t-1}) = \int_{\Theta} \Bigg[ &\lambda \int_{\mathcal{X}} p(x|\theta) \log p(x|\theta) \,\mathrm{d}x \\ &+ \omega \int_{\mathcal{X}} ((x - x^*(\theta)) - v_{t-1})^2 p(x|\theta) \,\mathrm{d}x \Bigg] f(z|z_{t-1}) \,\mathrm{d}z \end{aligned} \tag{201}$$

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:

$$\tilde{\Pi}(x, z) := \bar{\Pi}(z) + \frac{1}{2} \Pi_{xx}(z)(x - x^*(z))^2 \tag{202}$$

and we suppose that firms solve the problem:

$$\max_{p \in \mathcal{P}} \int_{\Theta} \int_{\mathcal{X}} \Pi(x, z) \, p(x \mid \theta) \,\mathrm{d}x \, f(z|z_{t-1}) \,\mathrm{d}z - c\,(p; z_{t-1}, v_{t-1}) \tag{203}$$

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

**Proposition 10.** *The optimal stochastic choice pattern is given by* $x = x^*(z) + v$*, where* $v = \rho v_{t-1} + u$ *and* $\rho = \rho(z) = \frac{\omega}{\frac{1}{2}|\Pi_{xx}(z)|+\omega}$ *and* $u \sim N\left(0, \frac{2\lambda}{\frac{1}{2}|\Pi_{xx}(z)|+\omega}\right)$.

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

$$-\beta(x-\gamma)^2 - \lambda\left[1 + \log p(x|z)\right] - \omega(x-\gamma-m)^2 + \mu(z) + \kappa(x,z) = 0 \tag{204}$$

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:

$$p(x|z) = \frac{\exp\left(-\tilde{\beta}(x-\tilde{z})^2\right)}{\int_{\mathcal{X}} \exp\left(-\tilde{\beta}(x'-\tilde{z})^2\right) \mathrm{d}x'} \tag{205}$$

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

$$x|z \sim N\left(\gamma + \frac{\omega}{\beta+\omega} v_{t-1}, \frac{2\lambda}{\beta+\omega}\right) \tag{206}$$

Putting this in more explicit terms, and substituting for $\gamma$ and $\beta$, we obtain the desired representation. □

### F.2 Transformed Gaussian Signal Extraction

In this section, we analyze attention cycles in a setting with Gaussian signal extraction and show how the basic logic of our main model carries over to this setting. For notational simplicity, we describe this alternative model under the assumption that there is a common, scalar state variable $\theta$, which represents each firm's productivity.

**Set-up.** When the state of the world is $\theta$, the previous state is $\theta_{-1}$, 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:

$$\tilde{\Pi}(x, X(\theta,\theta_{-1}), \theta) = \alpha(X(\theta,\theta_{-1}),\theta) - \beta(X(\theta,\theta_{-1}),\theta)(x - \gamma(X(\theta,\theta_{-1}),\theta))^2 \tag{207}$$

where we will write $\beta(\theta,\theta_{-1}) = \beta(X(\theta,\theta_{-1}),\theta)$ and similarly for $\alpha$ and $\gamma$, and as we micro-found 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:

$$s_i = \frac{\beta(\theta,\theta_{-1})}{\int_\Theta \beta(\theta,\theta_{-1}) \mathrm{d}\pi(\theta|\theta_{-1})} \gamma(\theta,\theta_{-1}) + \frac{1}{\tau(\theta_{-1})} \varepsilon_i \tag{208}$$

where $\varepsilon_i$ is a $N(0,1)$ variable, 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:

$$\frac{\beta(\theta,\theta_{-1})}{\int_\Theta \beta(\theta,\theta_{-1}) \mathrm{d}\pi(\theta|\theta_{-1})}\gamma(\theta,\theta_{-1}) \sim N\left(\mu(\theta_{-1}),\sigma^2(\theta_{-1})\right) \tag{209}$$

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:

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

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:

$$\begin{aligned} x|\theta,\theta_{-1} \sim N\Bigg(&\lambda(\theta_{-1})\frac{\beta(\theta,\theta_{-1})}{\int_\Theta \beta(\theta,\theta_{-1}) \mathrm{d}\pi(\theta|\theta_{-1})}\gamma(\theta,\theta_{-1}) \\ &+ (1-\lambda(\theta_{-1}))\mathbb{E}_{\pi_{\theta_{-1}}}\left[\frac{\beta(\theta,\theta_{-1})}{\int_\Theta \beta(\theta,\theta_{-1}) \mathrm{d}\pi(\theta|\theta_{-1})}\gamma(\theta,\theta_{-1})\right], \frac{\lambda^2(\theta_{-1})}{\tau^2(\theta_{-1})}\Bigg) \end{aligned} \tag{211}$$

Say we empirically estimate an equation of the form:

$$x_{it} = \gamma_i + \chi_{j(i),t} + f_t(\theta_{it},\theta_{i,t-1}) + \varepsilon_{it} \tag{212}$$

which differs from our baseline specification in controlling flexibly for observed and lagged productivity.[37] The fitted values span $\mathbb{E}[x \mid \theta,\theta_{-1}]$ and capture state-dependent anchoring toward the prior mean. The residual $\varepsilon_{it}$ 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 1 and Appendix D. 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.

**Interpreting Monotone Misoptimization.** We now discuss the interpretation of our empirical exercise of studying stochastic volatility in $\varepsilon_{it}$. The variance of the residual is:

$$\mathbb{V}_t[\varepsilon_{it}] = \frac{\lambda^2(\theta_{t-1})}{\tau^2(\theta_{t-1})} = \frac{\tau^2(\theta_{t-1})}{\left(\tau^2(\theta_{t-1}) + \frac{1}{\sigma^2(\theta_{t-1})}\right)^2} \tag{213}$$

[37] In our main analysis, we have experimented with such specifications and found similar results.

Our empirical findings are consistent with $\theta_{t-1} \mapsto \mathbb{V}_t[\varepsilon_{it}]$ being an increasing function.[38] This holds in this model exactly when:

$$\frac{\partial \tau^2(\theta_{-1})}{\partial \theta_{-1}} \left( \frac{1}{2\lambda(\theta_{-1})} - 1 \right) > \frac{\partial \frac{1}{\sigma^2(\theta_{-1})}}{\partial \theta_{-1}} \tag{214}$$

for all $\theta_{-1} \in \Theta$. Our own regression analysis in Section 4.7, 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 right-hand-side of this condition must be positive. Thus, there are two potential conditions under which this variant of our 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 between 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 of counter-cyclical attention in language (Fact 5) as qualitatively inconsistent with model case 2, and therefore an identifying piece of evidence for case 1. Elsewhere in the literature, Coibion et al. (2018) find that firms have higher demand for information when presented with bad macroeconomic news. Chiang (2023) also finds evidence of higher attention (interpreted as precision of signals) in downturns. And Kuang et al. (2024) finds a similar result, via structural recovery of signal precision in the Survey of Professional Forecasters. Thus, our preferred interpretation of the model is one in which residual variance inherits the monotonicity of signal precision and attention.

[38] Our specific empirical specification measured the correlation between contemporaneous output and contemporaneous dispersion of $\epsilon_{it}$. 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.

**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 $\theta_{-1}$ is realized, but before $\theta$ is realized, that the agent can pay a cost $\tilde{\phi}(\tau^2)$ to achieve signal precision of $\tau^2$, and where $\tilde{\phi}', \tilde{\phi}'' > 0$. Concretely, the optimal $\tau^2(\theta_{-1})$ solves:

$$\max_{\tau^2(\theta_{-1}) \in \mathbb{R}_+} \mathbb{E}_{\pi_{\theta_{-1}}} \left[ -\beta(\theta, \theta_{-1}) \left( x^*(s, \tau^2(\theta_{-1})) - \gamma(\theta, \theta_{-1}) \right)^2 \right] - \tilde{\phi}(\tau^2(\theta_{-1})) \tag{215}$$

We moreover parameterize the scaling of quadratic losses by writing $\beta(\theta, \theta_{-1}) = \kappa \hat{\beta}(\theta, \theta_{-1})$ for all $(\theta, \theta_{-1})$ and some $\kappa \geq 1$. Our first goal will be to derive conditions under which the optimally chosen $\tau^2$ in Program 215 is monotone increasing in $\kappa$. 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 $\tau^2$ from in Equation 211, we write

$$\begin{aligned}
&\mathbb{E}_{\pi_{\theta_{-1}}} \left[ -\beta(\theta, \theta_{-1}) \left( x^*(s, \tau^2(\theta_{-1})) - \gamma(\theta, \theta_{-1}) \right)^2 \right] \\
&= \mathbb{E}_{\pi_{\theta_{-1}}} \left[ \mathbb{E} \left[ -\beta(\theta, \theta_{-1}) \left( x^*(s, \tau^2(\theta_{-1})) - \gamma(\theta, \theta_{-1}) \right)^2 | \theta \right] \right] \\
&= \mathbb{E}_{\pi_{\theta_{-1}}} \left[ \mathbb{E} \left[ -\beta(\theta, \theta_{-1}) \left( x^*(s, \tau^2(\theta_{-1})) - \bar{x}(\theta, \theta_{-1}) + \bar{x}(\theta, \theta_{-1}) - \gamma(\theta, \theta_{-1}) \right)^2 | \theta \right] \right] \\
&= \mathbb{E}_{\pi_{\theta_{-1}}} \left[ -\beta(\theta, \theta_{-1}) \mathbb{E} \left[ \left( x^*(s, \tau^2(\theta_{-1})) - \bar{x}(\theta, \theta_{-1}) \right)^2 + \left( \bar{x}(\theta, \theta_{-1}) - \gamma(\theta, \theta_{-1}) \right)^2 | \theta \right] \right] \\
&= \mathbb{E}_{\pi_{\theta_{-1}}} \left[ -\beta(\theta, \theta_{-1}) \left( \frac{\lambda^2(\theta_{-1})}{\tau^2(\theta_{-1})} + \left[ \lambda(\theta_{-1}) \frac{\beta(\theta, \theta_{-1})}{\bar{\beta}(\theta_{-1})} \gamma(\theta, \theta_{-1}) + (1 - \lambda(\theta_{-1})) \mu(\theta_{-1}) - \gamma(\theta, \theta_{-1}) \right]^2 \right) \right]
\end{aligned} \tag{216}$$

where $\bar{x}(\theta, \theta_{-1})$ is the mean of the distribution in Equation 211 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:

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

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

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

which reduces to:

$$\begin{aligned}\tilde{\phi}'(\tau^2(\theta_{-1})) &= -\bar{\beta}(\theta_{-1})\frac{\frac{1}{\sigma^2(\theta_{-1})}-\tau^2}{\left(\tau^2+\frac{1}{\sigma^2(\theta_{-1})}\right)^3} \\ &+ 2\frac{\frac{1}{\sigma^2(\theta_{-1})}}{\tau^2+\frac{1}{\sigma^2(\theta_{-1})}}\frac{\frac{1}{\sigma^2(\theta_{-1})}}{\left(\tau^2+\frac{1}{\sigma^2(\theta_{-1})}\right)^2}\mathbb{E}_{\pi_{\theta_{-1}}}\left[\beta(\theta,\theta_{-1})\left(\mu(\theta_{-1})-\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})\right)^2\right] \\ &+ 2\frac{\frac{1}{\sigma^2(\theta_{-1})}}{\left(\tau^2+\frac{1}{\sigma^2(\theta_{-1})}\right)^2}\mathbb{E}_{\pi_{\theta_{-1}}}\left[\beta(\theta,\theta_{-1})\left(\mu(\theta_{-1})-\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})\right)\left(\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}-1\right)\gamma(\theta,\theta_{-1})\right]\end{aligned} \tag{219}$$

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

$$\tilde{\phi}'(\tau^2(\theta_{-1})) = \xi(\tau^2(\theta_{-1}))\kappa \tag{220}$$

where:

$$\begin{aligned}\xi(\tau^2(\theta_{-1})) &= \frac{1}{\left(\tau^2+\frac{1}{\sigma^2(\theta_{-1})}\right)^3}\Bigg[-\bar{\beta}(\theta_{-1})\left(\frac{1}{\sigma^2(\theta_{-1})}-\tau^2\right) \\ &\quad + 2\left(\frac{1}{\sigma^2(\theta_{-1})}\right)^2\xi_1(\theta_{-1}) + 2\frac{1}{\sigma^2(\theta_{-1})}\left(\tau^2(\theta_{-1})+\frac{1}{\sigma^2(\theta_{-1})}\right)\xi_2(\theta_{-1})\Bigg] \\ \xi_1(\theta_{-1}) &= \mathbb{E}_{\pi_{\theta_{-1}}}\left[\beta(\theta,\theta_{-1})\left(\mu(\theta_{-1})-\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})\right)^2\right] \\ \xi_2(\theta_{-1}) &= \mathbb{E}_{\pi_{\theta_{-1}}}\left[\beta(\theta,\theta_{-1})\left(\mu(\theta_{-1})-\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}\gamma(\theta,\theta_{-1})\right)\left(\frac{\beta(\theta,\theta_{-1})}{\bar{\beta}(\theta_{-1})}-1\right)\gamma(\theta,\theta_{-1})\right]\end{aligned} \tag{221}$$

Applying the implicit function theorem we then have that:

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

where the denominator is positive as the marginal cost of precision is always positive.

Thus, we have that:

$$\frac{\partial \tau^2(\theta_{-1})}{\partial \kappa} > 0 \iff \tilde{\phi}''(\tau^2(\theta_{-1})) > \xi'(\tau^2(\theta_{-1})) \tag{223}$$

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 $\kappa$.

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:

$$\beta(\theta, \theta_{-1}) = \upsilon_\Pi X(\theta, \theta_{-1})^{-1-\gamma+\chi(1+\epsilon)} \theta^{-1-\epsilon} \tag{224}$$

Thus:

$$\bar{\beta}(\theta_{-1}) = \mathbb{E}_{\pi_{\theta_{-1}}}[\upsilon_\Pi X(\theta, \theta_{-1})^{-1-\gamma+\chi(1+\epsilon)} \theta^{-1-\epsilon}] \tag{225}$$

Thus, whenever aggregate output is monotonically increasing in both $\theta$ and $\theta_{-1}$ 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 $\theta_{-1}$. It then follows that $\tau^2(\theta_{-1})$ is monotone decreasing in $\theta_{-1}$ in equilibrium whenever $\xi' < 0$. Thus, the core logic of our baseline model translates exactly over to this setting with Gaussian signal extraction.

### F.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 $\theta$, 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 $\phi$ such that:*

$$c(p) = \int_{\mathcal{X}} \hat{\phi}(\{p(x|\theta)\}_{\theta \in \Theta}) \, \mathrm{d}x \tag{226}$$

*where:*

$$\hat{\phi}(\{p(x|\theta)\}_{\theta \in \Theta}) = p(x) \phi\left(\left\{\frac{p(x|\theta)\pi(\theta)}{p(x)}\right\}_{\theta \in \Theta}\right) \tag{227}$$

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

Intuitively, such a cost functional considers the cost to the agent of arriving at any given posterior and adds that up over the distribution of posteriors that are realized. Important cost functionals such as the mutual information cost functional considered in the literature on rational inattention are members of this class. Indeed, mutual information is the special case of the above where $\phi$ returns the entropy of the distribution that is its argument.

The mathematical structure of posterior-separable cost functionals does not admit the same prior-independence property as state-separable cost functionals. As a result, we will not be able to carry all of our results over to this setting. Nevertheless, as we will argue, the key qualitative forces apply.

In the setting with state-separable choice in the single-agent context, we showed that greater curvature of payoffs leads to more precise actions (Proposition 1). With posterior-separable choice, the above result does not hold in general. This is because the prior also influences the states in which the agent would like to learn precisely. In particular, even if a state features high curvature, if it is unlikely to arise, the agent may not care to acquire precise information in that state. A particular case where this complication can be bypassed is when costs are given by mutual information and all actions are exchangeable in the prior in the sense that all actions are *ex ante* equally attractive (Matějka and McKay, 2015). This is a natural case to consider and yields a particularly revealing structure to the optimal policy: the agent's actions in state $\theta$ are given by a normal distribution centered on the objective optimum and with variance inversely proportional to the curvature of their objective in that state – a normal mixture model.

**Proposition 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:*[39]

$$p(x|\theta) = \frac{1}{\sqrt{\frac{\pi\lambda}{\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x - \gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\} \tag{229}$$

*Which is to say that the agent's actions follow a normal mixture model with conditional*

[39] Formally, all actions are exchangeable in the prior if:

$$\int_\Theta \frac{\exp\{\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\}}{\int_\mathcal{X} \exp\{\beta(\theta)\lambda^{-1}(\tilde{x} - \gamma(\theta))^2\}d\tilde{x}}\pi(\theta)d\theta = 1 \quad \forall x \in \mathcal{X} \tag{228}$$

*action density given by:*

$$x|\theta \sim N\left(\gamma(\theta), \frac{\lambda}{2\beta(\theta)}\right) \tag{230}$$

*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:

$$I(g) = \int_{\mathcal{S}} \int_{\Theta} g(s,\theta) \log\left(\frac{g(s,\theta)}{\pi(\theta)\int_{\Theta} g(s,\tilde{\theta})\,\mathrm{d}\tilde{\theta}}\right) \mathrm{d}\theta\,\mathrm{d}s \tag{231}$$

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 $\mu(\theta|s)$ 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 $c_I$:

$$\max_{P \in \mathcal{P}} \int_{\Theta} \int_{\mathcal{X}} u(x,\theta)\,\mathrm{d}P(x|\theta)\,\mathrm{d}\pi(\theta) - c_I(P) \tag{232}$$

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

$$c_I(p) = \int_{\Theta} \int_{\mathcal{X}} p(x|\theta) \log p(x|\theta)\,\mathrm{d}x\,\mathrm{d}\pi(\theta) - \int_{\mathcal{X}} p(x) \log p(x)\,\mathrm{d}x \tag{233}$$

where:

$$p(x) = \int_{\Theta} p(x|\theta)\,\mathrm{d}\pi(\theta) \tag{234}$$

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)

$$\mathcal{L}(\{p(x|\theta), \kappa(x,\theta)\}_{x\in\mathcal{X},\theta\in\Theta}, \{\gamma(\tilde{\theta})\}_{\theta\in\Theta}) = \int_\Theta \int_\mathcal{X} u(x,\theta)p(x|\theta)\,\mathrm{d}x\,\mathrm{d}\pi(\theta)$$
$$- \lambda \left( - \int_\mathcal{X} p(x)\log p(x)\,\mathrm{d}x + \int_\Theta \int_\mathcal{X} p(x|\theta)\log p(x|\theta)\,\mathrm{d}x\,\mathrm{d}\pi(\theta) \right) \tag{235}$$
$$+ \kappa(x,\theta)p(x|\theta) + \tilde{\gamma}(\theta) \left( \int_\mathcal{X} p(x|\theta)\,\mathrm{d}x - 1 \right)$$

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

$$p(x|\theta) = \frac{p(x)\exp\{u(x,\theta)\}}{\int_\mathcal{X} p(\tilde{x})\exp\{u(\tilde{x},\theta)\}\,\mathrm{d}\tilde{x}} \tag{236}$$

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

$$\int_\Theta \frac{\exp\{u(x,\theta)\}}{\int_\mathcal{X} p(\tilde{x})\exp\{u(\tilde{x},\theta)\}\,\mathrm{d}\tilde{x}}\,\mathrm{d}\pi(\theta) = 1 \tag{237}$$

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:

$$u(x,\theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2 \tag{238}$$

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

$$\int_\Theta \frac{\exp\{-\beta(\theta)\lambda^{-1}(x-\gamma(\theta))^2\}}{\int_\mathcal{X} \exp\{-\beta(\theta)\lambda^{-1}(\tilde{x}-\gamma(\theta))^2\}\,\mathrm{d}\tilde{x}}\pi(\theta)\,\mathrm{d}\theta = 1 \quad \forall x \in \mathcal{X} \tag{239}$$

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:

$$p(x|\theta) = \frac{\exp\{-\beta(\theta)\lambda^{-1}(x-\gamma(\theta))^2\}}{\int_\mathcal{X} \exp\{-\beta(\theta)\lambda^{-1}(\tilde{x}-\gamma(\theta))^2\}\,\mathrm{d}\tilde{x}} \tag{240}$$

The denominator of this expression can be computed:

$$\begin{aligned}
\int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(x-\gamma(\theta))^2\}\,\mathrm{d}x &= \int_{\mathcal{X}} \frac{\sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}}{\sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x-\gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\}\mathrm{d}x \\
&= \sqrt{2\pi\frac{\lambda}{2\beta(\theta)}} \int_{\mathcal{X}} \frac{1}{\sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x-\gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\} \\
&= \sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}
\end{aligned} \tag{241}$$

It follows that:

$$p(x|\theta) = \frac{1}{\sqrt{\frac{\pi\lambda}{\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x-\gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\} \tag{242}$$

Thus, $X|\theta$ 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:

$$\begin{aligned}
\gamma(\theta) &= x^*(X(\theta), \theta) \\
\beta(\theta) &= \frac{1}{2}|\Pi_{xx}(X(\theta), \theta)|
\end{aligned} \tag{243}$$

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 state-separable cost that we studied. All results from Section 2 then carry directly.

# G State-Dependent Attention in Survey Data

In this Appendix, we test our interpretation of attention and misoptimization cycles using the dataset of Coibion et al. (2018) (henceforth, CGK), one of the most comprehensive datasets of firm-level operations and macro backcasts in an advanced economy. These data were assembled from a detailed survey of the general managers of a representative panel of firms in New Zealand from 2013 to 2016.

## G.1 Reported Attention and the Business Cycle

Although the CGK survey took place during relatively tranquil times for the New Zealand economy, it asks two hypothetical questions about how firms' desire to collect information would vary with the macroeconomic state:

> 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." Self-reported demand for information is higher in the "bad news" state. 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. It is also consistent with our findings regarding macroeconomic attention in language (Section 4.5).

## G.2 Reported Profit Function Curvature and Attention

A second test possible in the CGK data relates to our 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, we show exactly how we 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.[40]

We consider two "macro attention" outcomes. The first is the absolute-value error in firms' one-year back-casts for 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 $\text{ProfitCurv}_{it}$ and a vector of controls $Z_{it}$:

$$Y_{it} = \alpha + \beta \cdot \text{ProfitCurv}_{it} + \gamma' Z_{it} + \epsilon_{it} \tag{244}$$

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.

[40] Curvature is higher for smaller firms with more within-industry competitors (Table A15).

Table A14 shows the results. For inflation we find strong evidence that higher-curvature firms make smaller errors, with some of the effect being absorbed by control variables when added. For GDP growth we find estimates that are much less precise but have the same signs; and for unemployment, results that are further imprecise and have the wrong signs. We take this as further evidence consistent with Facts 3, 4, and 5: in the firm survey, the differential stakes of making mistakes contributes to macro attention.

### G.3 Details of Data Construction

**Profit Function Curvature.** We draw our measure of profit function curvature from the answers to two questions about hypothetical price changes. These are jointly asked as Question 17 of Wave 5, Part B:

> If this firm was free to change its price (i.e. suppose there was no cost to renegotiating contracts with clients, no costs of reprinting catalogues, etc.) right now, by how much would it change its price? Please provide a percentage answer. By how much do you think profits would change as a share of revenues? Please provide a numerical answer in percent.

Denote the answer for prices as $\Delta p_i$ and the answer for profits as $\Delta \Pi_i$. Under the assumption that the following second-order approximation holds for the deviation of profits from their frictionless optimum (e.g., a version of Equation 10), the following relationship holds between the measurable quantities and the profit function curvature $\text{ProfitCurv}_i$:

$$\Delta \Pi_i = \text{ProfitCurv}_i \cdot \Delta p_i^2 \qquad \Rightarrow \qquad \text{ProfitCurv}_i = \frac{\Delta \Pi_i}{\Delta p_i^2} \tag{245}$$

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 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:

$$\widehat{\text{ProfitCurv}}_i = \beta \cdot \hat{X}_i + e_{it} \tag{246}$$

where the hat denotes that both variables have been normalized to z-score units (i.e., with means subtracted and standard deviation divided out), so the coefficient $\beta$ is the standard-deviation-to-standard-deviation effect. Firms with higher profit function curvature are smaller and have more competitors. There is weak evidence that the associated managers are more skilled and/or better rewarded. Thus, 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.

**Outcomes: Back-cast Errors.** For back-cast errors, we use the following questions. 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?

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.

**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

**Table A1:** Misoptimization and Firm Performance

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Outcome: $R_{it}$ | | | | Outcome: $\pi_{it}$ | | | |
| $\hat{u}_{it}^2$ | -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*: $R_{it}$ is the firm-level log stock return and $\pi_{it}$ is firm-level profitability. $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 3.1. Standard errors are double-clustered at the year and firm level.

**Table A2:** Dynamic Effects of Misoptimization

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Outcome: | | | | |
| | | $\Delta \log \hat{\theta}_{i,t+k}$ | | | $R_{i,t+k}$ | | | $\pi_{i,t+k}$ | |
| Horizon $k$ | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| $\hat{u}_{it}^2$ | -0.009 | 0.014 | -0.007 | -0.236 | -0.252 | -0.251 | -0.316 | -0.286 | -0.265 |
| | (0.007) | (0.008) | (0.010) | (0.026) | (0.027) | (0.038) | (0.024) | (0.018) | (0.019) |
| Sector x Time FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $N$ | 50,455 | 40,671 | 32,362 | 41,578 | 34,643 | 28,103 | 51,015 | 42,014 | 33,934 |
| $R^2$ | 0.231 | 0.245 | 0.263 | 0.238 | 0.241 | 0.248 | 0.117 | 0.123 | 0.126 |

*Notes*: Each column is the estimate of a separate projection regression. The outcomes are TFP growth $\Delta \log \hat{\theta}_{it}$ (first three columns), log stock returns $R_{it}$ (second three columns), and profitability $\pi_{it}$ (last three columns). $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 3.1. Standard errors are double-clustered at the year and firm level.

**Table A3:** Cyclicality of Misoptimization, with Alternative Measurement Schemes

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Outcome: MisoptimizationDispersion$_t$ | | | | | | | | | | |
| 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) |
| Correlation | -0.493 | -0.468 | -0.485 | -0.466 | -0.546 | -0.505 | -0.479 | -0.489 | -0.494 | -0.462 | -0.379 |
| 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*: The first row reports the coefficient from the regression of MisoptimizationDispersion$_t$ on Unemployment$_t$/100, with standard errors that are HAC-robust with a 3-year Bartlett Kernel. The following row reports the correlation of these variables (in column 4, conditional on projecting out controls). 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 Appendix B.2.

**Table A4:** Pricing of Misoptimization, with Alternative Measurement Schemes

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Outcome: $R_{it}$ | | | | | |
| $\hat{u}_{it}^2$ | -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) |
| $\hat{u}_{it}^2 \times R_t$ | 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*: $R_{it}$ is the firm-level log stock return. $\hat{u}_{it}$ is the firm-level misoptimization residual, constructed using the methods described in Section 3.1 and the indicated variants described in the main text and the notes of Table A3. $R_t$ is the log return of the S&P 500. Standard errors are double-clustered at the year and firm level. The scenarios are described in the main text and the notes of Table A3.

**Table A5:** Markets Punish Misoptimizations Harder in Low States, Additional Controls

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Outcome: $R_{it}$ | | | | | |
| $\hat{u}_{it}^2 \times R_t$ | 0.376 | 0.378 | 0.345 | 0.321 | 0.330 | 0.489 |
| | (0.123) | (0.109) | (0.118) | (0.173) | (0.094) | (0.296) |
| Sector x Time FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $\hat{u}_{it}^2$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| TFP and Interaction | | ✓ | | | | |
| Leverage and Interaction | | | ✓ | | | |
| Lag Return and Interaction | | | | ✓ | | |
| Industry FE and Interaction | | | | | ✓ | |
| Firm FE and Interaction | | | | | | ✓ |
| $N$ | 41,578 | 41,578 | 41,429 | 34,805 | 41,206 | 41,206 |
| $R^2$ | 0.239 | 0.261 | 0.246 | 0.239 | 0.379 | 0.498 |

*Notes*: Column 1 reports the baseline estimate. Columns 2-6 add additional variables and their interaction with $\hat{u}_{it}^2$ : the level of log firm TFP, $\log \hat{\theta}_{it}$; leverage, $\text{Lev}_{it}$; the previous year's stock return, $R_{i,t-1}$; an industry fixed effect $\chi_{j(i)}$; and a firm fixed effect $\gamma_i$. Standard errors are double-clustered at the year and firm level.

**Table A6:** The Effects of Misoptimization in Levels

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Outcome: | | | | | |
| | $R_{it}$ | | $\pi_{it}$ | | $\hat{m}_{it}^2$ | |
| $\hat{m}_{it}^2$ | -0.042 | -0.076 | -0.021 | -0.025 | | |
| | (0.028) | (0.033) | (0.010) | (0.011) | | |
| $\hat{m}_{it}^2 \times R_t$ | | 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*: $\hat{m}_{it}$ is the firm-level misoptimization. These specifications replicate results in Tables 1, 2, 3, and A1 with with $\hat{m}_{it}$ in place of $\hat{u}_{it}$. Standard errors are double-clustered at the year and firm level.

**Table A7:** 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*: $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 3.1 and the indicated variants described in the main text and the notes of Table A3. $\log \text{MacroAttention}_{it}$ is the measure of firm-level macroeconomic attention, constructed using the methods described in Section 3.2 and Appendix C.1. Standard errors are double-clustered at the year and firm level. The scenarios are described in the main text and the notes of Table A3.

**Table A8:** Policy Function Estimation

| | Baseline | Adj. Cost | Leverage | Quadratic |
|---|---|---|---|---|
| | Panel A: Persistence of Misoptimization | | | |
| | Outcome: $\hat{m}^0_{it}$ | | | |
| $\hat{m}^0_{i,t-1}$ | 0.696 | 0.016 | 0.696 | 0.683 |
| | (0.021) | (0.005) | (0.003) | (0.003) |
| | Panel B: Quasi-Differenced Policy Function | | | |
| | Outcome: $\log L_{it} - \hat{\rho} \log L_{i,t-1}$ | | | |
| $\log \hat{\theta}_{it}$ | 0.418 | 0.381 | 0.419 | 0.463 |
| | (0.024) | (0.026) | (0.025) | (0.029) |
| $\log \hat{\theta}_{i,t-1}$ | -0.031 | -0.090 | -0.026 | 0.006 |
| | (0.018) | (0.015) | (0.019) | (0.020) |
| $\log \hat{\theta}_{it} \times \text{Lev}_{it}$ | | | -0.008 | |
| | | | (0.003) | |
| $\log \hat{\theta}_{i,t-1} \times \text{Lev}_{i,t-1}$ | | | -0.025 | |
| | | | 0.006 | |
| $\text{Lev}_{it}$ | | | -0.020 | |
| | | | (0.005) | |
| $\text{Lev}_{i,t-1}$ | | | -0.050 | |
| | | | (0.011) | |
| $(\log \hat{\theta}_{it})^2$ | | | | 0.045 |
| | | | | (0.008) |
| $(\log \hat{\theta}_{i,t-1})^2$ | | | | 0.031 |
| | | | | (0.006) |
| $\log L_{i,t-1}$ | | 0.811 | | |
| | | (0.012) | | |
| $\log L_{i,t-2}$ | | -0.041 | | |
| | | (0.010) | | |
| $N$ | 51,891 | 44,051 | 51,664 | 51,891 |
| $R^2$ | 0.896 | 0.990 | 0.896 | 0.904 |

*Notes*: The four columns correspond to four of our policy-function estimation methods, as described in the main text. The coefficient on $\hat{m}^0_{i,t-1}$ in panel A corresponds in the model to parameter $\rho$. Standard errors are double clustered by firm and year.

**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*: The "CC" MacroAttention is constructed using the methods described in Appendix C.1. The "10K" MacroAttention is our baseline measure from Section 4.5. 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*: The "Stem" MacroAttention is constructed using the methods described in Appendix C.2. The "10K" MacroAttention is our baseline measure from Section 4.5. 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 |

| 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 |
|---|---|---|---|---|---|---|---|---|---|
| 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*: "Cost-Share TFP" is our baseline measure of $\log \hat{\theta}_{it}$ used in the main analysis. "OP TFP" is the alternative measure based on the method of Olley and Pakes (1996). 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. (2018) survey, as described in Appendix G.

**Table A14:** Profit-Function Curvature and Attention to Macro Variables

Panel 1: Back-cast Error (Absolute Value)

| Variable | Inflation | | GDP Growth | | Unemployment | |
|---|---|---|---|---|---|---|
| $\text{ProfitCurv}_{it}$ | -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 | |
|---|---|---|---|---|---|---|
| $\text{ProfitCurv}_{it}$ | 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. (2018) survey, as described in 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 |

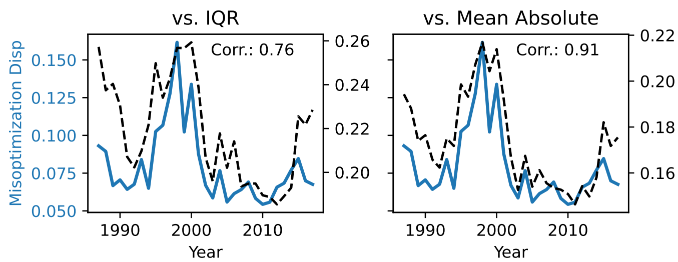
*Notes*: The top panel gives summary statistics. The bottom panel gives normalized regression coefficients for a number of possible correlates. Standard errors, used to calculate the $t$-statistics, are clustered by three-digit industry.

**Table A16:** Curvature and Inflation BCE in Waves 1 versus 4

| | Outcome: absolute Inflation BCE | | | |
|---|---|---|---|---|
| Wave | 1 | | 4 | |
| $\text{ProfitCurv}_{it}$ | -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 |

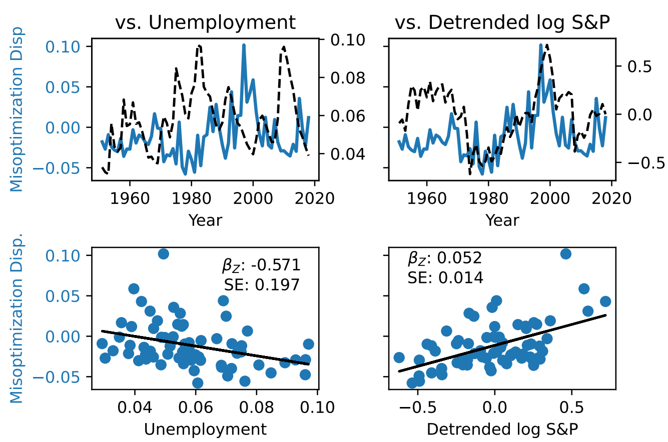
*Notes*: Standard errors are clustered by three-digit industry.

**Figure A1:** Relationship of Misoptimization Dispersion with Other Statistics



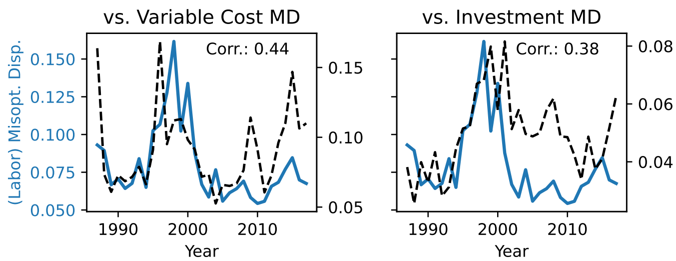

*Notes*: The blue line, measured by the left axis of each plot, is Misoptimization Dispersion as defined in Section 4. The black dashed line on the left is the (optimal-sale-weighted) interquartile range of the distribution of $\hat{u}_{it}$. The black dashed line on the right is the (optimal-sale-weighted) average of $|\hat{u}_{it}|$.

**Figure A2:** Misoptimization Dispersion is Pro-Cyclical, Long Sample



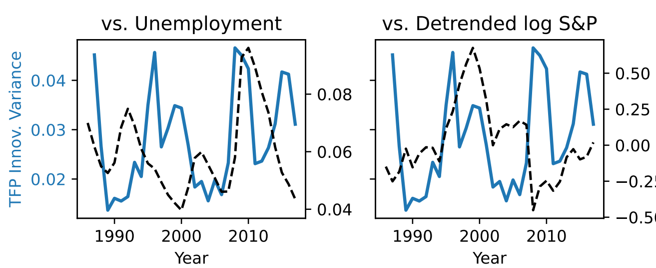

*Notes*: This Figure replicates Figure 2 using our long-sample (1950-2018) calculation of Misoptimization Dispersion, described in Section 4.2 ("Robustness to Studied Time Period."). The top two panels plot Misoptimization Dispersion (blue line, left axis) along with, respectively, unemployment and the linearly detrended S&P 500 price (black dashed lines, right axis). Because of the composition adjustment, the metric can be negative. The bottom two panels are scatterplots of Misoptimization Dispersion versus each macroeconomic aggregate. The black solid line is the linear regression fit. The standard errors are HAC robust based on a Bartlett kernel with a three-year bandwidth.

**Figure A3:** 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 4. The black dashed line on the left is Misoptimization Dispersion for total variable cost expenditures (total labor plus materials expenditure). The black dashed line on the right is the same for investment rates (log growth rates of the capital stock).

**Figure A4:** Relationship of TFP Innovation Variance with Macro Variables




*Notes*: "TFP Innovations" are the residuals from an AR(1) model for $\log \theta_{it}$ with firm and sector-by-time fixed effects, as described in Section 4.7 of the main text. We calculate their average, consistent with our main calculations of Misoptimization Dispersion, using optimal-sales-weights.

**Figure A5:** Frequency over Time of Each Word in MacroAttention (Part I)

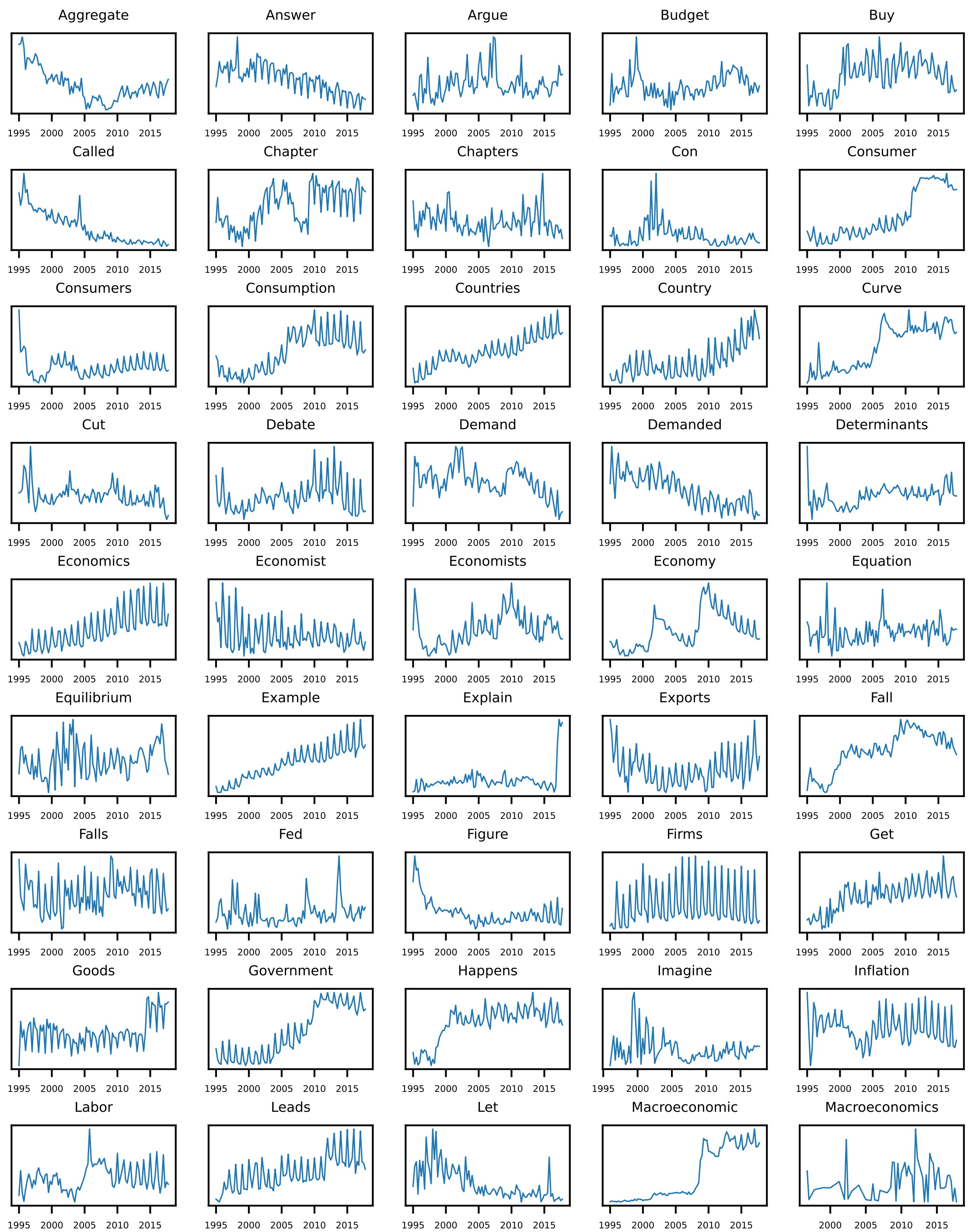


**Figure A5:** Frequency over Time of Each Word in MacroAttention (Part II)

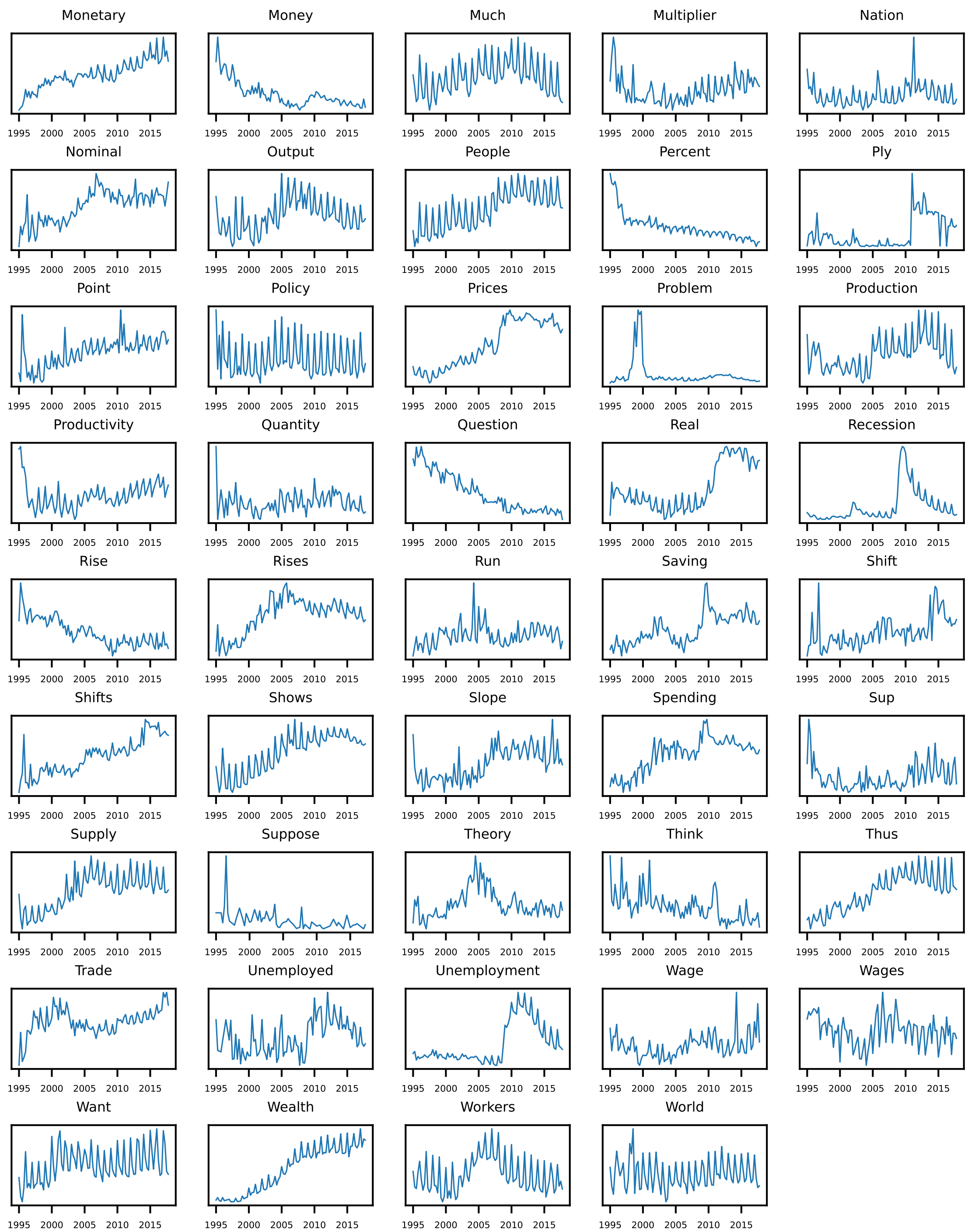
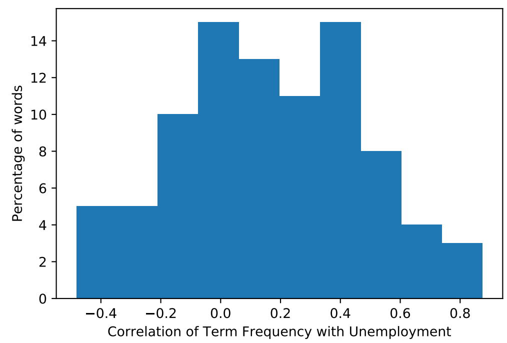
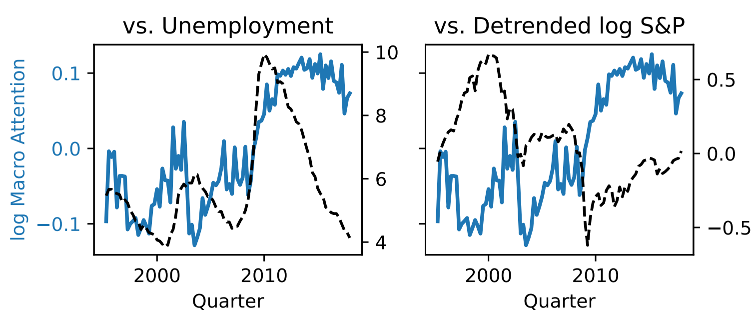


**Figure A6:** Correlations with Unemployment by Word



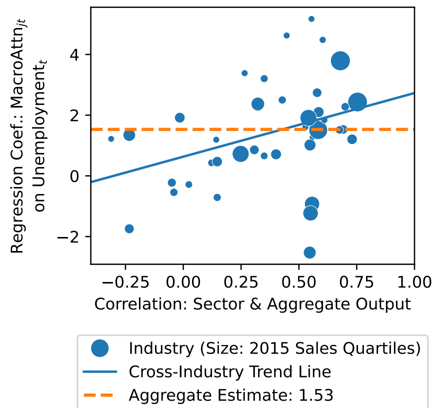

*Notes*: Correlations are calculated at the quarterly frequency.

**Figure A7:** Macroeconomic Attention with "Cleaned" Word List



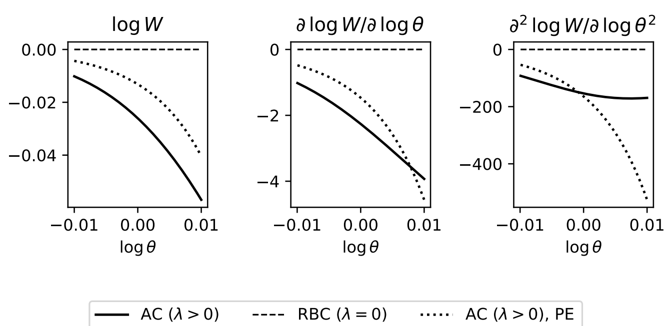

*Notes*: This replicates our main analysis of macroeconomic attention, but with *ex post* removal of the following words: answer, argue, chapter, chapters, equation, example, explain, figure, get, happens, imagine, leads, let, much, people, ply, point, problem, question, rise, rises, run, shift, shifts, shows, suppose, theory, think, thus, want.

**Figure A8:** Industry-Specific Cyclicality of Macro Attention



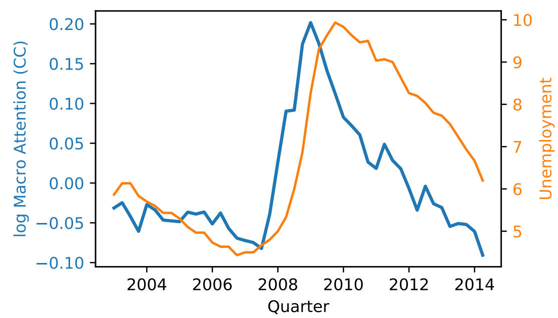

*Notes*: The horizontal axis is the correlation of sectoral and aggregate nominal GDP. The vertical axis is the regression coefficient of log sectoral macro attention, net of quarterly fixed effects, on the US unemployment rate. The dashed orange line is the estimate of the same using aggregate Macro 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 A9:** The Attention Wedge and Its Derivatives



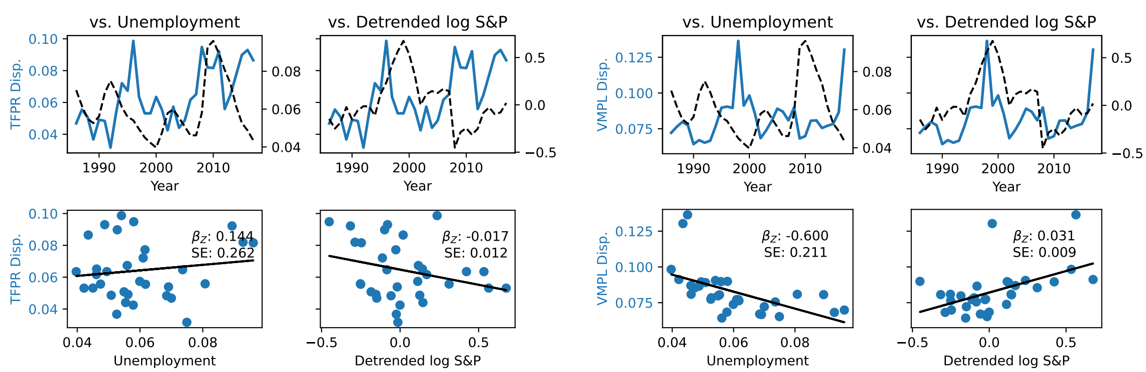

*Notes*: The panels respectively show the level, first derivative, and second derivative of the log attention wedge in the log state. The "Partial Equilibrium" thought experiment, plotted in each figure as a dotted line, is for firms to best-reply to the output and wages of the counterfactual RBC equilibrium.

**Figure A10:** 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.

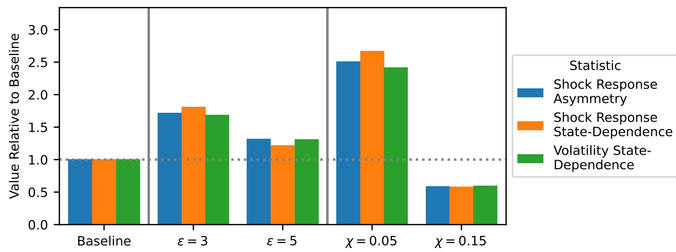
**Figure A11:** The Cyclical Behavior of TFPR and VMPL Dispersion




(a) TFPR Dispersion (b) VMPL Dispersion

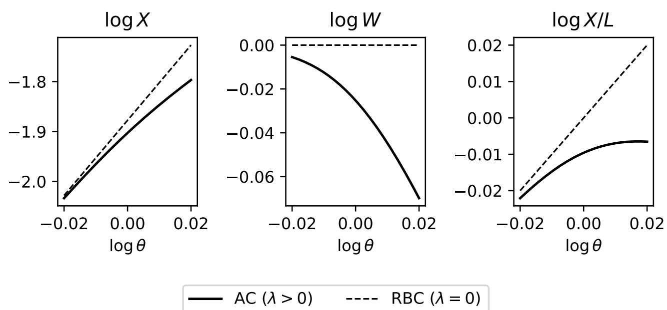
*Notes*: Each half of this Figure replicates Figure 2 using our measures of TFPR Dispersion (top) and VMPL Dispresion (bottom). Variable construction is defined in Appendix C.3. Because of the composition adjustment, the variance metrics can be negative. The standard errors for the linear fits are HAC robust baesd on a Bartlett kernel with a three-year bandwidth.

**Figure A12:** Robustness of Numerical Results to Parameter Choices



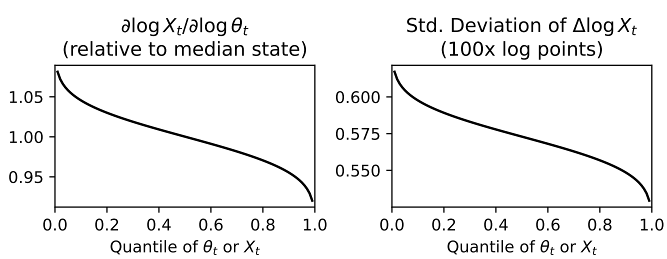

*Notes*: We re-calibrate the model under each indicated parameter choice. The outcomes and their interpretation are described in Appendix E.1.

**Figure A13:** Output, Wedge, and Labor Productivity with GHH Preferences



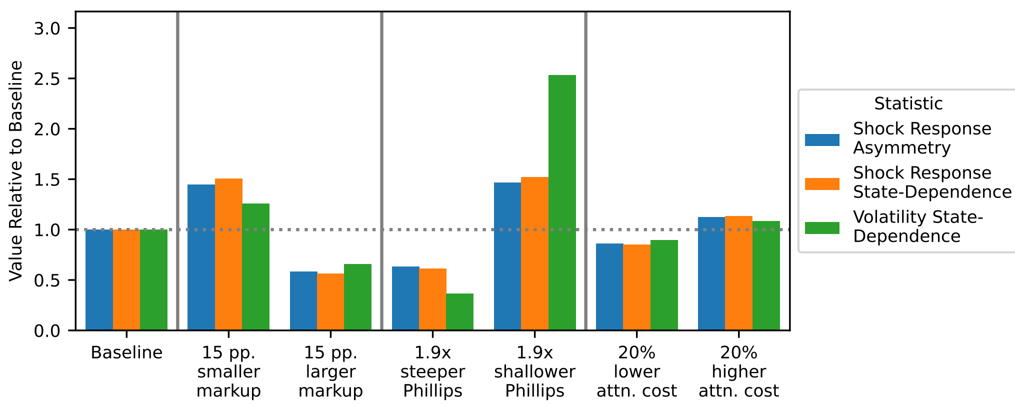

*Notes*: This recreates panels (a)-(c) of Figure 5 from the main analysis in the variant model with Greenwood et al. (1988) preferences, described in Appendix E.

**Figure A14:** Asymmetric Shock Response and Stochastic Volatility with GHH Preferences




*Notes*: This recreates panels (d) and (e) from Figure 5 from the main analysis in the variant model with Greenwood et al. (1988) preferences, described in Appendix E.

**Figure A15:** Predictions in Counterfactual Scenarios




*Notes*: The scenarios are described in the main text. The outcomes are the same as in Figure A12, and are described in Appendix E.1.

## References


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.

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

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

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.

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.

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

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.

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

Kehrig, M. (2015). The cyclical nature of the productivity distribution. Working paper.

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.

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.

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

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.