

Attention Cycles

Joel P. Flynn*
Yale

Karthik A. Sastry†
Princeton

October 12, 2026

Abstract

We study how bounded rationality coevolves with the business cycle. We introduce a business-cycle model in which firms face a cognitive cost of making precise decisions. Theoretically, we characterize equilibrium with non-parametric, state-dependent stochastic choice. Firms have greater incentives to pay attention in downturns because they are owned by risk-averse households. Correspondingly, the model generates counter-cyclical attention, pro-cyclical mistakes, and an endogenous attention wedge that depresses aggregate productivity when attention is low. Empirically, we test the model's predictions using novel measures of firms' mistakes and attention. Quantitatively, attention cycles generate significant 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.

1 Introduction

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 largely abstracted from in the existing literature on macroeconomics with decision frictions (e.g., Woodford, 2003; Nimark, 2008; Maćkowiak and Wiederholt, 2009; Lorenzoni, 2009; Angeletos and La’O, 2010; Gabaix, 2019). 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 modeling 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 pushes toward high attention when aggregate consumption is low.

We next describe the *general-equilibrium* properties of “attention cycles.” 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. Output is the product of the counterfactual output under no cognitive friction and 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, misallo-

cation 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. This asymmetry takes the form of a concave relationship of aggregate output (and the attention wedge) with productivity, in logarithmic units: that is, negative shocks have larger effects than positive shocks, and all shocks have larger effects when output and productivity are low.

Measurement. The model makes micro and macro predictions for both choice mistakes (“misoptimization”) and cognitive effort (“attention”). To test 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 1986-2017. 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 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 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, macroeconomic attention is counter-cyclical. At the firm level, macroeconomic attention is associated with smaller misoptimizations. These findings, together, validate the consistency of our two approaches to measuring attention at both the microeconomic and macroeconomic levels and further support 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 corresponds to 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 also generates 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 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](#); [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 [Chiang \(2023\)](#).¹ In contrast to these studies, we emphasize a distinct risk-pricing mechanism, develop

¹In [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

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 study state-dependent attention in general equilibrium. This relates to a literature that numerically studies state-dependent attention under mutual-information costs (Kacperczyk et al., 2016; 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.

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 cyclical 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).

Our focus on state-dependence is shared by a long literature that studies macroeconomic dynamics when firms face “menu costs” for decision variables (e.g., Bertola and Caballero, 1990; Golosov and Lucas, 2007). Apart from differences in microfoundations, an important difference in our analysis is that our mechanism operates through the endogenous *curvature* of firms’ payoff functions, as determined in equilibrium. Thus, the mechanism driving state-dependence in our model is complementary to, and distinct from, the channel by which higher shock volatility leads to more frequent adjustment in menu-cost models (Vavra, 2014).

In the same vein, our focus on business-cycle asymmetries links to studies of information production over the business cycle. One strand of this literature hypothesizes that information about fundamental shocks is a by-product of production and is therefore in pro-cyclical supply (e.g., Zeira, 1994; Veldkamp, 2005; Van Nieuwerburgh and Veldkamp, 2006; Ordoñez, 2013). Others emphasize that information grows stale during booms and is created in crises, amplifying the crises’ magnitude (e.g., Gorton and Ordoñez, 2014; Asriyan et al., 2022; Howes

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.

²Costain and Nakov (2015, 2019) also apply the control-cost model but study different questions, regarding price dynamics, and do not provide comparable analytical results. In the context of households’ attention to savings choices, Macaulay (2020) studies the special case of exchangeable choices under mutual information costs, which is isomorphic to a control cost approach.

and Weitzner, 2026). Our mechanism is similar to that proposed by the latter literature, but it arises from different foundations: the key decisionmakers with time-varying incentives are firms themselves, not lenders, and mistakes arise from flexible cognitive frictions rather than (lack of) information *per se*.

Finally, 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 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.

Outline. In Section 2, we introduce our model. In Section 3, we present our theoretical results. In Section 4, we describe the data and measurement. In Section 5, we test the model's six main predictions. In Section 6, we calibrate our model and analyze the impact of attention cycles on macroeconomic dynamics. Section 7 concludes.

2 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.

2.1 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:

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

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

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

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:

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

where $\bar{w} > 0$ and $\bar{X} > 0$ are constants, and $\chi > 0$ measures the extent of real wage flexibility. 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 this allows us to study the economy via a scalar fixed-point equation.³ Moreover, it allows our model to parsimoniously match the empirical acyclicity of real wages (Solon et al., 1994; Grigsby et al., 2021). In Appendix B, 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 F.1 that our results are quantitatively robust to considering wages set in this manner.⁴

2.2 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:

$$x_{it} = \theta_{it} L_{it} \quad (4)$$

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

We parameterize the stochastic process for 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 , (θ_{it}, θ_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 for aggregate productivity shocks to

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

⁴Under Greenwood et al. (1988) preferences, the household's marginal utility depends directly on labor hours: $\partial \mathcal{U} / \partial C_t = (C_t - v(L_t))^{-\gamma}$. If a household with GHH preferences owns the firms, this makes the analysis messier. In Appendix B, we nevertheless extend our model to a setting with workers and capitalists with GHH preferences and show that our main results carry over. Practically speaking, in our numerical example, the “correction” of marginal utility due to labor hours is negligible, and all our quantitative and qualitative results about equilibrium attention cycles go through.

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 θ_{it} at time t , $G_t \in \Delta(\Theta)$. The second is the statistic:⁵

$$\hat{\theta}(G) = \left(\int_{\Theta} \theta_i^{\epsilon-1} dG(\theta_i) \right)^{\frac{1}{\epsilon-1}} \quad (5)$$

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.

2.3 Modeling 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.⁶ 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 (θ_{it}, θ_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 $p : \mathcal{Z} \rightarrow \Delta(\mathcal{X})$ in set $\mathcal{P}$, or a mapping from states of the world to *distributions of actions* described by a probability density function $p(\cdot | z_{it})$ when the decision state is z_{it} . A firm using rule p commits 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 $c : \mathcal{P} \times \Lambda \times \mathcal{Z} \times \mathcal{F} \rightarrow \mathbb{R}$ 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

⁵Throughout, we assume that this integral and $\mathbb{E}[\log \theta_i]$ exist.

⁶As will become clear, $\mathcal{Z} = \Theta \times \mathcal{X} \times \mathcal{W}$, where $\mathcal{X}$ is the feasible set of production, and $\mathcal{W}$ is the image of $\mathcal{X}$ via the wage rule (3).

independent from the decision state and distributed in the cross-section as $L \in \Delta(\Lambda)$, and the previous value of the decision state $z_{i,t-1}$. We assume that $\mathbb{E}[\lambda_i]$ and $\mathbb{E}[\lambda_i |\log \lambda_i|]$ exist.

The basic idea that we wish to embody is that firms pay a cost to have more precise production strategies. To make this tension clear, while also keeping the analysis tractable, we specialize to the following state-separable entropy cost with a (potentially firm-specific) scaling $\lambda_i > 0$:

$$c(p, \lambda_i, z_{i,t-1}, f) = \lambda_i \int_{\mathcal{Z}} \int_{\mathcal{X}} p(x | z_{it}) \log(p(x | z_{it})) dx f(z_{it} | z_{i,t-1}) dz_{it} \quad (6)$$

These costs embody the idea that it is costly for firms to avoid making mistakes in their production choices (see [Morris and Yang, 2022](#), and [Flynn and Sastry, 2023](#), for a discussion). To see this, note that we can interpret Equation 6 as an expectation over realized states z of the negative entropy of the conditional distribution of production. We can think of the firm’s choice variable as a “plan” for each state: given a particular value of productivity, aggregate output, and the wage, how much should the firm produce? Since entropy measures *imprecision*, these costs penalize the *precision* of the firm’s plan in any given state. We adopt the flexible interpretation that these costs can represent both any *ex ante* costs of formulating a precise plan and any *ex post* costs of implementing such a plan. The choice of entropy units is primarily one of convenience, as they will yield particularly simple choice rules (e.g., as in [McFadden, 1973](#); [McKelvey and Palfrey, 1995](#)). As we will soon see, the firm’s optimal choices will intuitively embody a state-by-state tradeoff of the benefits of precision against these costs. As a result, the size and character of firms’ choice mistakes will respond to incentives, consistent with broad experimental evidence on imperfect choice ([Woodford, 2020](#)) as well as the central premise of this paper’s analysis.

In Section 3.5, we discuss how our approach compares in premise and results to other candidate models of incentives-driven attention. In particular, we contrast our model to models of information acquisition and discuss the robustness of our key conclusions to models of stochastic choice with different functional forms.

2.4 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:

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

We can also think of this as a measure of the risk-adjusted *valuation* of the firm's cash flow. Under our assumptions about the firms' cost and revenue structure and the household's utility function, the profit function and marginal utility are respectively

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

Since the firm observes the history of realized states and their actions do not affect this history, the firm's dynamic choice problem reduces to a series of one-shot problems. The firm's choice at time t therefore solves the following problem of maximizing expected risk-adjusted profits net of cognitive costs:

$$\max_{p \in \mathcal{P}} \left\{ \int_{\mathcal{Z}} \int_{\mathcal{X}} \Pi(x, z_{it}) p(x | z_{it}) dx f(z_{it} | z_{i,t-1}) dz_{it} - c(p, \lambda_i, z_{i,t-1}, f) \right\} \quad (9)$$

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 negative 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.⁷ An alternative model, motivated for instance by a story of labor costs within the firm (e.g., hiring planners or consultants), might denominate the costs in units of the numeraire good. In Section 3.5, 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. We further discuss how to interpret our main results in this case.

2.5 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.7.

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

⁷Our analysis focuses on business cycle fluctuations rather than growth. To accommodate balanced growth in this model, it would be natural for the cognitive cost shifter λ_i to scale with trend growth in marginal utility. That is, one could set $\lambda_{it} = \bar{\lambda}_i \exp((1 - \gamma)gt)$, where $\bar{\lambda}_i$ is a time-invariant shifter and g is the trend growth of consumption. This would not affect our main conclusions about business cycles, understood as stochastic deviations from this trend.

$\bar{\Pi}(z_i)$ 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:

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

The first-order term of this approximation drops out as 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 $X^* = \left(\int_0^1 x^*(z_i)^{1-\frac{1}{\epsilon}} di\right)^{\frac{\epsilon}{\epsilon-1}}$. We then write the approximate production function as

$$X = X^* - \frac{1}{2\epsilon} \int_0^1 \frac{(x_i - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}} (x^*(z_i))^{1+\frac{1}{\epsilon}}} di \quad (11)$$

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 Π .*
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.*

3 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.

3.1 Attention and Misoptimization in Partial Equilibrium

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

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

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

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. □

Proposition 1 describes firms' state-dependent strategies to control mistakes in closed form. We are able to do so tractably for three reasons. 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.

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 (i.e., the second, stochastic term of Equation 12). The variance of the misoptimization increases if the marginal cost of precision increases (higher λ_i), and decreases if either of two components of the marginal benefits of precision increases. 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 state depends on the demand curves that firms face. In particular, this curvature term is high if revenues are very concave in quantities sold, which occurs when demand curves are steep (i.e., prices are very sensitive to quantities) and/or become steeper as higher quantities are sold.⁸ 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; equivalently, the household places a higher valuation on the same cash flows. 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 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 θ_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

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

where $v_\pi(\epsilon, \chi, \bar{X}, \bar{w}) > 0$ (see derivations in Appendix A.7). Profit curvature is higher when quantities sold are lower and the firm operates in the steeper segment of its demand curve. This occurs when marginal costs are high, contributing a term $(w/\theta_i)^{1+\epsilon} \propto X^{\chi(1+\epsilon)} \theta_i^{-1-\epsilon}$, or aggregate output and its associated aggregate demand externality are low, contributing a term X^{-1} . Marginal utility is high when aggregate consumption is low, and the coefficient of relative risk aversion, γ , controls the elasticity of marginal utility to aggregate output.

We next 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 can combine this observation with our microfoundations for the profit-curvature and risk-pricing channels

⁸Because of assumed constant returns to scale, total costs are linear in quantities sold. Otherwise, more convex costs (e.g., from more decreasing returns to scale) would also contribute to higher profit curvature.

to describe how firms’ “mistakes” depend on the economic state.

Corollary 1 (Comparative Statics for Mistakes). *Consider a type- λ_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 θ_i . Moreover, $\mathbb{E}[(x_i - x^*(z_i))^2 \mid z_i]$ strictly increases in X if and only if $\gamma > \chi(1 + \epsilon) - 1$.*

Proof. Immediate from combining Proposition 1 with Equation 13. $\square$

The (absolute) curvature of the profit function always increases in marginal costs, and therefore decreases in θ_i . The monotonicity of curvature in aggregate output depends jointly on the cyclicalities 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.

3.2 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). *If $\chi\epsilon < 1$ and $\gamma > \chi + 1$, there exists 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 $X : \Theta \rightarrow \mathbb{R}$ that is strictly positive and strictly increasing in that sufficient statistic.*

Proof. See Appendix A.2. $\square$

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 $\hat{\theta}(G)$. 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 complementarity is well-behaved.

The latter condition $\gamma > \chi + 1$ is both conservative in the model, as it ensures uniqueness and monotonicity for any possible distribution of λ_i , and highly plausible in practice. The

elasticity of detrended real wages to GDP in US data since 1987 is 0.095, and micro-level studies find similarly severe wage rigidity (Solon et al., 1994; Grigsby et al., 2021).⁹ Moreover, to be consistent with the restrictions $\chi\epsilon < 1$ (upward-sloping best responses) and $\epsilon > 1$ (substitutable goods), χ cannot exceed one. The corresponding conditions $\gamma > 1.095$ or $\gamma > 2$ are likely both slack by an order of magnitude, given evidence in financial economics about the high cyclicalities of the stochastic discount factor (e.g., Hansen and Jagannathan, 1991). As γ governs *only* the properties of the stochastic discount factor in our model, this is the relevant evidence for determining its appropriate value.

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 (θ_i, λ_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 (θ_i, λ_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 $X : \Theta \rightarrow \mathbb{R}$. Firms’ aggregate misoptimization is their average sales-weighted mean-squared-error around the ex post optimal action, or*

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

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- λ_i firm contingent on realized state z_i , $q^*(z_i)$ is the optimal frictionless price, 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

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

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 $G \succeq_{FOSD} G'$, then aggregate attention a is lower under G .*

⁹This 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.

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.

3.3 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:*

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

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 λ and in aggregate productivity θ . Both results have a partial-equilibrium and general-equilibrium

component. In partial equilibrium, both increasing λ and increasing θ 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, this logic is preserved.

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:*

$$\log A(\log \theta) = \log \theta + \chi \epsilon \log W(\log \theta) \quad (17)$$

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 and increased misallocation. This is consistent with the unstable and often negative cyclicalities 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).¹⁰ What is new is our prediction about the cyclicalities of this force, driven by changing incentives for (in)attention, and the implications of that cyclicalities for business cycles.

3.4 Shock Propagation and Volatility

Firms’ endogenous inattention shapes the response of aggregate output to shocks. To see this, note that the elasticity of output to the productivity state is

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

that is, the sum of the frictionless economy’s response to shocks (χ^{-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

¹⁰David 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.

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 aggregation.

We next study how the responsiveness of the attention wedge to shocks depends on the state—that is, we characterize the *second derivative* of the attention wedge. Under our maintained assumptions, the attention wedge is *concave* in the state and, as a result, productivity shocks propagate asymmetrically.

Corollary 3 (Asymmetric Propagation). *If $\chi\epsilon < 1$ and $\gamma > \chi + 1$, then the attention wedge is globally concave, or*

$$\frac{\partial^2 \log W(\log \theta)}{\partial \log \theta^2} < 0 \quad (19)$$

Therefore:

1. *Output is more sensitive to shocks when productivity is low: $\frac{\partial \log X(\log \theta)}{\partial \log \theta}$ decreases in θ .*
2. *Propagation is asymmetrically larger for negative productivity shocks: for any θ_0 and $\Delta > 0$,*

$$\log X(\log \theta_0) - \log X(\log \theta_0 - \Delta) > \log X(\log \theta_0 + \Delta) - \log X(\log \theta_0) \quad (20)$$

Proof. See Appendix [A.6](#). □

Intuitively, when productivity is very low, firms make very few misoptimizations and the model behaves like the standard RBC model. As productivity increases, firms make more imprecise choices and the pass-through of productivity to output becomes dampened. As a direct consequence, shocks pass more fully into output when productivity is already low or when the shocks reduce productivity. A final implication, which we will explore more fully when we specialize the law of motion for productivity in our quantitative analysis in Section [6](#), is endogenous stochastic volatility. In particular, if $\log \theta_t$ follows a persistent stochastic process with constant (state-independent) conditional volatility, then log output will have higher conditional volatility in low-output, low-productivity states, because the pass-through of shocks is higher in those states.

3.5 Discussion and Extensions

Decision Costs in Dollars. 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.” To be precise, consider an alternative model in which

planning in cognitive state z_{it} requires making expenditures in that state which are proportional to the (negative) expected entropy of the action distribution. In this case, the firm's flow profits as a function of production x_{it} , in state z_{it} , are

$$x_{it} \left(x_{it}^{-\frac{1}{\epsilon}} X_t^{\frac{1}{\epsilon}} - \frac{w_t}{\theta_{it}} \right) - \lambda_i \int_{\mathcal{X}} p(x | z_{it}) \log(p(x | z_{it})) dx \quad (21)$$

where the second term captures planning expenditures. The firm's choice problem under dollar costs is therefore (cf. Equation 9, under cognitive costs):

$$\max_{p \in \mathcal{P}} \int_{\mathcal{Z}} M(z_{it}) \left(\underbrace{\int_{\mathcal{X}} \pi(x, z_{it}) p(x | z_{it}) dx}_{\text{Expected profits in } z_{it}} - \underbrace{\lambda_i \int_{\mathcal{X}} p(x | z_{it}) \log(p(x | z_{it})) dx}_{\text{Expected costs in } z_{it}} \right) f(z_{it} | z_{i,t-1}) dz_{it} \quad (22)$$

where, crucially, the stochastic discount factor modifies both the dollar profits and the planning cost paid in each state z . Therefore, the risk pricing channel is not operative: high marginal utility equally affects the benefits and costs of attention. Because the firm's stochastic choice problem is state separable, this scenario therefore generates the same firm behavior (and downstream macroeconomic outcomes) as a case of our original model in which the risk-pricing channel is eliminated entirely, or firm owners are risk neutral ($\gamma = 0$). That is, even though the firm's owners *are* risk averse, this force cancels out on the cost and benefit sides of their attention decisions.

In this case with only the profit-curvature channel, if output is increasing in productivity, then the model cannot feature pro-cyclical misoptimizations by firms (see Equation 66 with $\gamma = 0$). Moreover, our empirical analysis will support a strong role for the risk-pricing channel. First, at the microeconomic level, we will find that the sensitivity of dollar profits to mistakes increases only slightly during bad macroeconomic states, while the sensitivity of stock returns, which take into account risk pricing, increases greatly in the same states (Section 5.3). Second, at the macroeconomic level, we will find strong procyclicality of misoptimization variance and, to rationalize this fact, a strong risk-pricing channel is necessary.

Other Models of Inattention. Our model of costly control has three principal advantages for our analysis over alternative models of incentives-driven inattention, such as costly information acquisition (Caplin and Dean, 2015), the version of that model with mutual-information costs (Sims, 2003), or sparsity (Gabaix, 2014). First, our model admits an analytical solution for the optimal policy (Proposition 1), despite the model's state-dependence. In particular, no comparable results exist in the literature on rational inattention with mu-

tual information costs.¹¹ Second, our model allows for tractable and well-posed analysis of equilibrium and its comparative statics (Propositions 2, 3, and 4). 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.

One limitation of our model is that it does not generate individual-level underreaction to shocks. Moreover, it is of independent interest to know how robust our macroeconomic conclusions are to the exact form of inattention that we assume. Motivated by this, we study how incentives from the profit-curvature and risk-pricing channels affect firms’ propensity for stochastic mistakes in an alternative model with mutual-information costs (Appendix C.2). In this model, the concavity of benefits also affects firms’ propensity to underreact, due to endogenous anchoring toward prior beliefs; this, in turn, complicates the relationship between incentives and the variance of stochastic mistakes. However, we nonetheless provide sufficient conditions under which the comparative statics of firms’ mistakes are analogous to those in Corollary 1 and consistent with our subsequent empirical findings. We note also that equilibrium aggregation in this variant is more challenging, precluding a straightforward extension of our results on equilibrium attention cycles or business-cycle asymmetries even conditional on obtaining the same firm-level comparative statics.

Relaxing Functional Form Assumptions. Within the class of control-cost models, our results can be generalized beyond the specific functional forms that are considered here for tractability. In Flynn and Sastry (2023), we show generalized versions of Propositions 1, 2, and 3 under a larger class of state-separable costs of stochastic choice, payoff functions, and aggregators. Under different cost functionals, we retain the core intuition that higher objective function curvature leads to more concentration of actions around the *ex post* optimum, but lose the exact structure of a Gaussian mixture distribution and the exact mapping to the conditional variance of actions (which we will later exploit when mapping to measurement). Under more general payoff functions and aggregators, we retain the core intuition that state-dependent attention affects output by inducing misallocation, but lose the closed-form characterization of output dynamics.

Labor Wedge Shocks. Fluctuations can also arise in our model from shocks to $\bar{w}$, which can be interpreted as labor wedge shocks. 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

¹¹Even 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).

pro-cyclical misoptimization in an economy with stochastic productivity and labor wedges. Proposition 4 holds as written, with $\theta\bar{w}^{-1}$ replacing θ . Thus, our core conclusions about attention cycles and their effects on business cycles hold in a model with both supply and demand shocks.

Multiple Inputs and Classical Labor Supply. In Appendix B, we extend the model to allow for intermediate inputs, separate capital owners and laborers, and market-clearing wages rather than a wage rule. The first two features will be useful in mapping the model to the data in Section 5, while the third enables a more Neoclassical micro-foundation. We show under general conditions how the main results derived in this section, regarding the cyclicity 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.

3.6 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 increase, rather than decrease, 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. *Firms’ risk-adjusted payoffs 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 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 cyclicity 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.

4 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.

4.1 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 property, plant, and equipment (PPE). We organize firms into 44 industries, which are defined at the NAICS 2-digit level, except 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 2017.¹² 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 68,875 firm-year observations, or about 2,200 per year. Appendix D.1 describes the full procedure.

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 (up to constants):

$$\begin{aligned} \log L_{it} &= \log L^*(\theta_{it}, X_t, w_t) + \log \left(1 + \frac{\sigma(\theta_{it}, X_t, w_t)}{x^*(\theta_{it}, X_t, w_t)} v_{it} \right) \\ &= (\epsilon - 1) \log \theta_{it} - \epsilon \log w_t + \log X_t + \log \left(1 + \frac{\sigma(\theta_{it}, X_t, w_t)}{x^*(\theta_{it}, X_t, w_t)} v_{it} \right) \\ &\approx (\epsilon - 1) \log \theta_{it} - \epsilon \log w_t + \log X_t + \underbrace{\frac{\sigma(\theta_{it}, X_t, w_t)}{x^*(\theta_{it}, X_t, w_t)} v_{it}}_{m_{it}} \end{aligned} \quad (23)$$

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 D.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 23 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

$$\log L_{it} = \eta_i + \chi_{j(i),t} + \beta \log \theta_{it} + m_{it} \quad (24)$$

where $j(i)$ is the industry of firm i , β is an unrestricted coefficient, η_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

¹²To probe robustness, we also show results on the entire Compustat sample since 1950.

$$m_{it} = \rho m_{i,t-1} + \left(\sqrt{1 - \rho^2}\right) u_{it} \quad (25)$$

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 24 (Appendix D.3). The second is persistent mistakes, which can arise in a variant model which accommodates cognitive inertia (Appendix C.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 D 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} :

$$\log x_{it} = \log \theta_{it} + \alpha_{L,j(i)} \log L_{it} + \alpha_{M,j(i)} \log M_{it} + \alpha_{K,j(i)} \log K_{it} \quad (26)$$

where $\alpha_{L,j(i)} + \alpha_{M,j(i)} + \alpha_{K,j(i)} = 1$.¹³ 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

¹³The 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.

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 24 via ordinary least squares (OLS) and obtain a preliminary estimate $\hat{m}_{it}^0$ of the residual. We next estimate the AR(1) process for $\hat{m}_{it}^0$ to obtain an estimate $\hat{\rho}$ of the residual persistence (in our main procedure, 0.70). We next estimate via OLS the “quasi-differenced” equation for labor choice:¹⁴

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

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 $\hat{u}_{it} = \frac{\hat{\nu}_{it}}{\sqrt{1 - \hat{\rho}^2}}$. Below, we refer to $\hat{u}_{it}$ as our main empirical proxy for firms’ misoptimization, after correcting for persistence.

We finally define aggregate “Misoptimization Dispersion” as an estimate of the cross-sectional variance of misoptimizations with weights s_{it}^* proportional to firms’ *predicted* sales based on fundamentals:¹⁵

$$\text{MisoptimizationDispersion}_t = \frac{\sum_{i \in \mathcal{I}_t} s_{it}^* \cdot \hat{u}_{it}^2}{\sum_{i \in \mathcal{I}_t} s_{it}^*} \quad (28)$$

This is the direct empirical analog of our theoretical definition of misoptimization dispersion.

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.¹⁶ Later, Prediction 1 will allow us to distinguish idiosyncratic wedges from misoptimizations: residuals from wedges should have a uniformly 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. These strategies alter the policy function to account for adjustment costs, financial frictions, non-log-linear terms ([Ilut et al., 2018](#)), and time-varying coefficients

¹⁴Table A5 contains our estimates of Equation 27 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.

¹⁵In particular, the weights are the exponentiated fitted values $\exp(\hat{\beta} \log \hat{\theta}_{it})$ from the following regression equation: $\log \text{Sales}_{it} = \beta \log \hat{\theta}_{it} + \eta_i + \chi_{j(i),t} + \epsilon_{it}$.

¹⁶Time-varying wedges at the industry or aggregate level would be absorbed in our model’s fixed effects, 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.

(Decker et al., 2020); and alter the calculation of TFP to account for time-varying coefficients or to use the structural method of Olley and Pakes (1996). We discuss these alternative methods in more detail in Section 5.2.

4.2 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 2017 in our main analysis.¹⁷ Our total sample consists of 459,013 individual documents, or about 5,000 per quarter, which we index by their date of filing.

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.¹⁸ 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

¹⁷The relevant digitized documents are hosted by the Securities and Exchange Commission’s EDGAR (Electronic Data Gathering, Analysis, and Retrieval), which began operation in 1994. We choose 1995 as a starting point at which a nearly comprehensive sample of firms’ reports is available in the system.

¹⁸We use electronic copies of the 7th, 3rd, and 12th editions of these books, respectively.

“term frequency inverse document frequency,” or tf-idf, as:

$$\text{tf-idf}(w)_{it} := \text{tf}(w)_{it} \cdot \log \left(\frac{1}{\text{df}(w)} \right) \quad (29)$$

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 by this metric in each textbook and take the intersection among the three books to obtain a final set of 89 words.¹⁹ Figure A4 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 of economics 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 terms from the main analysis, and then verify *ex post* that these terms do not drive our main conclusions. 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 frequencies:

$$\text{MacroAttention}_{it} = \sum_{w \in \mathcal{W}_M} \text{tf-idf}(w)_{it} \quad (30)$$

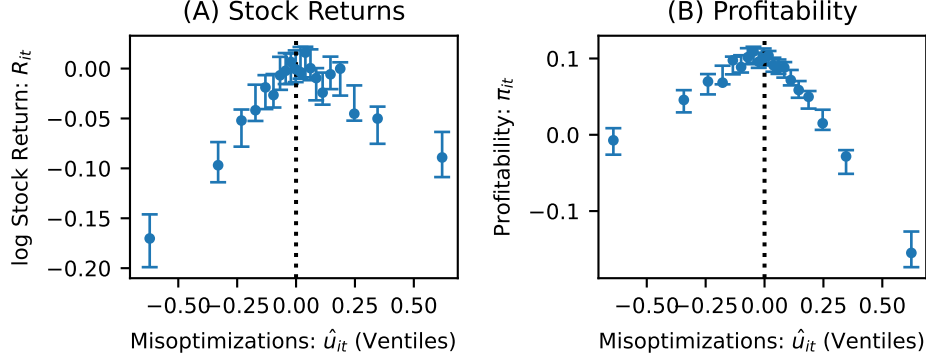
We generate an aggregate measure MacroAttention_t by averaging $\text{MacroAttention}_{it}$ across firms. In our aggregate results, we remove seasonal trends as quarter-of-the-year means.

5 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.

¹⁹Taking 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.

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 standard errors clustered by firm. The construction of $\hat{u}_{it}$ is described in Section 4.1.

5.1 Fact 1: Mistakes Reduce Returns and Profitability

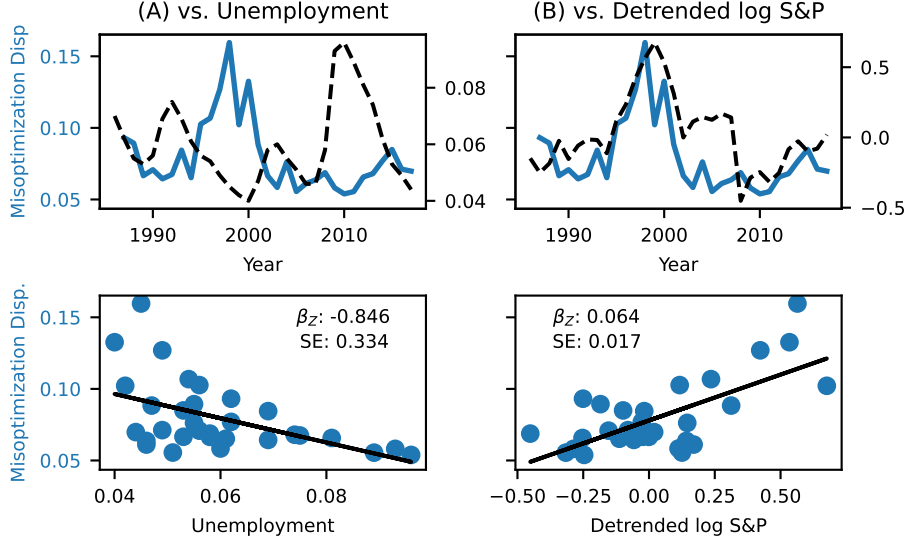
We first test that our measured “misoptimizations” are, as predicted by the theory, bad for firm performance (Prediction 1). We proxy for firm performance with two measures. The first is firms’ log stock returns, R_{it} . The second is firms’ “profitability” π_{it} , or this year’s earnings before interest and taxes (EBIT) divided by last year’s total variable costs.²⁰ 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 $\chi_{j(i),t}$.

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 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 including different levels of fixed effects and/or controlling for productivity growth. In Figure A1, we verify that the distribution of mistakes itself is symmetric around zero.

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-

²⁰Variable 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 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.

by-time fixed effects:

$$X_{i,t+k} = \beta_{X,k} \cdot \hat{u}_{it}^2 + \chi_{j(i),t} + \epsilon_{it} \quad (31)$$

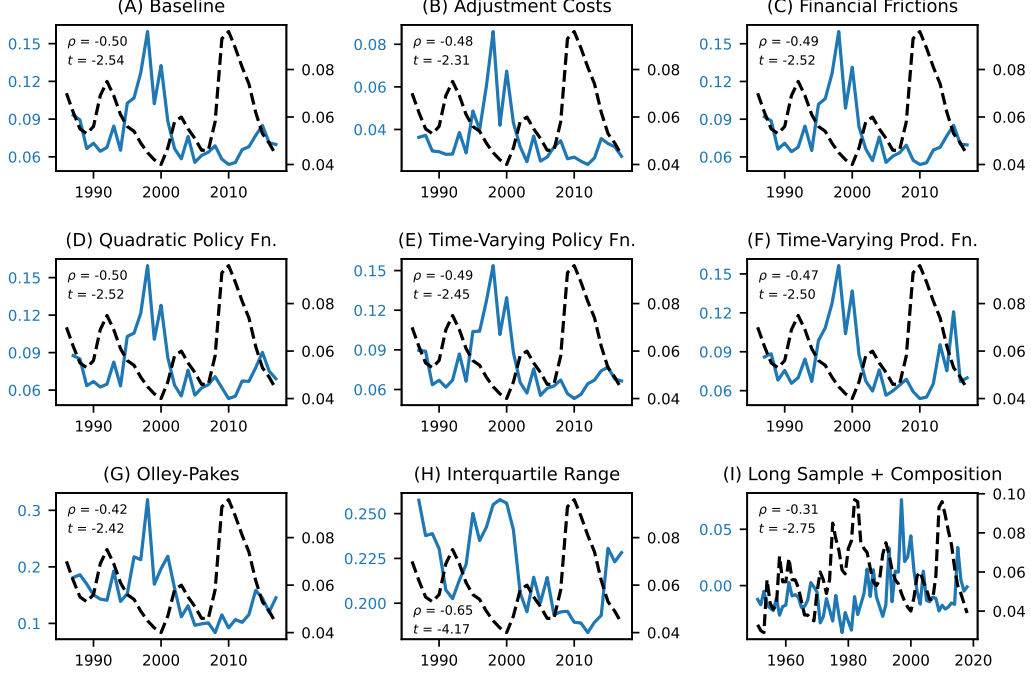
for $X_{i,t+k} \in \{\Delta \log \hat{\theta}_{i,t+k}, \pi_{i,t+k}, R_{i,t+k}\}$ and $k \in \{0, 1, 2\}$. We find no evidence of a quantitatively large effect on current and future TFP growth (Table A2). We also find evidence of persistent negative effects on profitability and stock returns. These dynamic 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 (driven, for instance, by restructuring).

5.2 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.

We benchmark the strength of this relationship with the business cycle by estimating a

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 ρ 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 for a non-linear (quadratic) policy function. Panel (E) allows for time-variation in the policy function. Panel (F) allows for time-variation in the production function. Panel (G) uses productivity estimates from the method of [Olley and Pakes \(1996\)](#). Panel (H) reports the inter-quartile range of mistakes rather than their dispersion. Panel (I) replicates our main procedure for the sample back to 1950 and applies compositional adjustment.

linear regression of Misoptimization Dispersion on each macroeconomic variable:

$$\text{MisoptimizationDispersion}_t = \alpha + \beta_Z \cdot Z_t + \epsilon_t \quad (32)$$

for $Z_t \in \{\text{Unemployment}_t/100, \log \text{SPDetrend}_t\}$, over our 31 annual observations. We estimate a slope of -0.846 (SE: 0.334, $p = 0.02$) with respect to unemployment and 0.064 (SE: 0.017, $p = 0.001$) with respect to the detrended S&P 500. The respective correlations are -0.503 and 0.688. The regression on unemployment implies that a five-percentage-point swing in unemployment is associated with an increase in Misoptimization Dispersion of 0.042 squared log units, or 53% of its sample mean value.

The most important threat to our interpretation of Fact 2 is that any unmodeled determinants of firm choices might be spuriously identified as “mistakes,” and the extent of these misspecifications may vary in magnitude over the business cycle. To investigate this pos-

sibility, we revisit Fact 2 under various alternative constructions for “mistakes” introduced in Section 4.1. Our findings, summarized in Figure 3, convey that Fact 2 emerges under a number of different empirical approaches.

We first account for frictional adjustment dynamics by estimating a policy function that controls for lagged labor choice. The residuals from this equation have negligible autocorrelation (AR(1) coefficient 0.015, as reported in Table A5). Therefore, in this variant, it is inessential to model mistakes as persistent. This implies misoptimizations that are smaller in magnitude, but have nearly the same responsiveness to the business cycle as our baseline measure (Panel B). 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). We next account for financial frictions by augmenting the policy function with leverage and its interaction with TFP. This measure behaves almost identically to our baseline (Panel C). This is consistent with the hypothesis that financial frictions and their heterogeneous incidence escalate in *downturns*, which would if anything bias against our finding of higher mistakes in booms.

We find similar time-series patterns when we estimate policy functions with a quadratic term in log productivity (D) and time-varying coefficients (E). We also find similar patterns when we estimate the production function with time-varying coefficients (F) and the method of Olley and Pakes (1996) (G). It is also inessential to focus on the variance as our measure of dispersion; our results are qualitatively similar using the inter-quartile range (H).

To test our prediction over a longer period, we extend our time series using historical Compustat data since 1950.²¹ This allows us to assess how our prediction for cyclical misoptimization generalizes beyond the key business cycle episodes in our main sample (the Dot-Com Boom, the Early 2000s Recession, the Mid 2000s Boom, the Great Recession, and the post-Great Recession Recovery). 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 I of Figure 3 and Figure A2). Replicating the regressions against the unemployment rate and detrended S&P 500 price, we obtain coefficients of -0.380 (SE: 0.138) and 0.037 (SE: 0.010).

Finally, while most of the tests above focus on defining misoptimizations under *more* stringent criteria (i.e., with more flexible policy or production functions), we can also ex-

²¹One 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.

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

	Outcome: log Stock Return, R_{it}			
	(1)	(2)	(3)	(4)
$\hat{u}_{it}^2$	-0.264 (0.0246)	-0.260 (0.0236)	-0.0961 (0.0342)	-0.0862 (0.0335)
$\hat{u}_{it}^2 \times R_t$	0.376 (0.122)	0.384 (0.127)	0.446 (0.173)	0.445 (0.171)
Sector x Time FE	Yes	Yes	Yes	Yes
Firm FE	No	No	Yes	Yes
Controls	No	Yes	No	Yes
N	42026	42026	41650	41650
R^2	0.239	0.262	0.385	0.403

Notes: This table reports estimates of Equation 33. The outcome, R_{it} , is the log stock return and $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 4.1. R_t is the log return of the S&P 500. Standard errors are double-clustered at the firm and year level.

plore what happens if we *relax* assumptions that we argued were necessary in the baseline. In Figure A3, we consider less stringent constructions that (i) use all variable input expenditures, in dollars (rather than labor inputs in quantity units); (ii) simply residualize labor choices on fixed effects, rather than constructed TFP; (iii) use lagged TFP, observable to the firm when it makes its decisions, in place of current TFP in the policy function estimation; and (iv) use the “first-stage” residuals $\hat{m}_{it}^0$, neglecting their autocorrelation. In each case, the alternative constructions of Misoptimization Dispersion correlate negatively with unemployment, although some of these results are less precise than our baseline.

5.3 Fact 3: Returns Respond More to Misoptimizations During Downturns

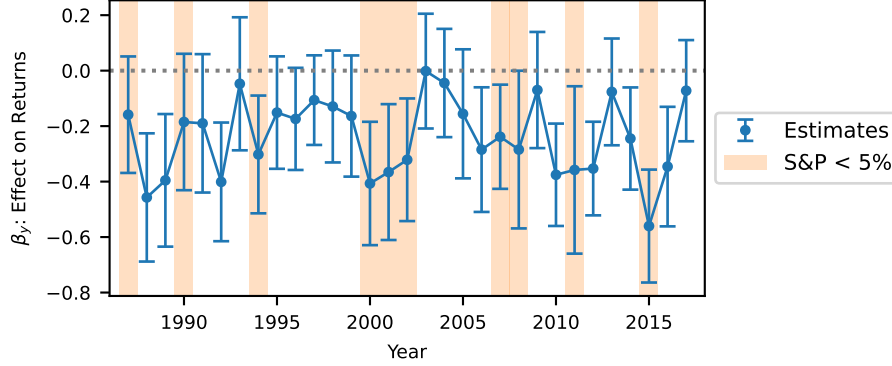
We have shown that firms make smaller misoptimizations in downturns. We now investigate the extent to which this is explained by the incentives implied by our model: that firms have higher financial costs of misoptimization in downturns (Prediction 3).

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 interacted with the log aggregate (S&P 500) stock return R_t .²²

$$R_{it} = \beta \cdot \hat{u}_{it}^2 + \phi \cdot (\hat{u}_{it}^2 \times R_t) + \chi_{j(i),t} + \Gamma' X_{it} + \epsilon_{it} \quad (33)$$

²²In Table A4, we replicate the analysis using our (first-stage) estimates of the mistake “level” $\hat{m}_{it}$ and find qualitatively similar results.

Figure 4: Time-Varying Effects of Misoptimization on Stock Returns



Notes: This figure reports estimates of Equation 34, which models time-varying effects of misoptimization on stock returns in a non-parametric way. Each dot is an estimate of β_y . Error bars are 95% confidence intervals, based on standard errors that are clustered at the firm level. The orange shaded regions are periods in which the S&P 500 return is less than 5%.

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 B, we motivate the regression equation and the prediction $\phi \geq 0$, 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 (Table 1).²³ Our estimates in column 3, in particular, suggest that mistakes have a zero price if the S&P return is 24%, close to its value in the late 1990s, the height of the dot com bubble. By contrast, in the trough of 2008 ($R_t = -0.52$), the model implies that pricing is 6.4 times more severe than in the “usual” states of $R_t \approx 0.10$.

To better convey the time-variation in marginal effects that underlies this result, we also estimate the following regression which allows for a year-specific coefficient on $\hat{u}_{it}^2$ instead of imposing a parametric interaction with aggregate returns:

$$R_{it} = \sum_y \beta_y \cdot \hat{u}_{it}^2 \cdot \mathbb{I}[t = y] + \chi_{j(i),t} + \epsilon_{it} \quad (34)$$

Our estimates, shown in Figure 4, illustrate that misoptimizations harm returns the most in years of relatively poor stock returns—for example, 2001, 2008, or 2015.

The most significant threat to our interpretation of these results is that the timing of

²³Similar results are obtained using mistake “levels” (columns 1-2 of Table A4).

the financial cycle correlates with other variables that independently affect the sensitivity of stock returns to firm performance. In Table A3, we show that our core finding is unaffected by controlling for a number of such factors and their interactions with $\hat{u}_{it}^2$. We control for time-varying firm-level conditions in the form of lagged stock returns and leverage, suggesting that our main finding does not merely pick up the known “leverage effect” that distressed firms have more volatile stock returns. We use stronger, nonparametric methods to control for *time-invariant* factors across firms, by saturating the regression with interactions between $\hat{u}_{it}^2$ and industry or firm fixed effects. The last, most conservative specification, which fully absorbs firm-specific coefficients on $\hat{u}_{it}^2$, suggests if anything *larger* variation in the pricing of mistakes over the financial cycle than our baseline model does.

5.4 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 π_{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 returns:

$$\pi_{it} = \beta_{\pi} \cdot \hat{u}_{it}^2 + \phi_{\pi} \cdot (\hat{u}_{it}^2 \times R_t) + \chi_{j(i),t} + \gamma_i + \epsilon_{it} \quad (35)$$

We treat ϕ_{π} as a gauge of the strength of the profit-curvature channel.

We find a positive, small, and statistically insignificant ϕ_{π} (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 π_{it} conditional on firm and sector-by-time fixed effects. Conditional on profitability, misoptimizations have a severely attenuated effect on stock returns that is statistically indistinguishable from zero (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

Table 2: Misoptimization, Profits, and Pricing

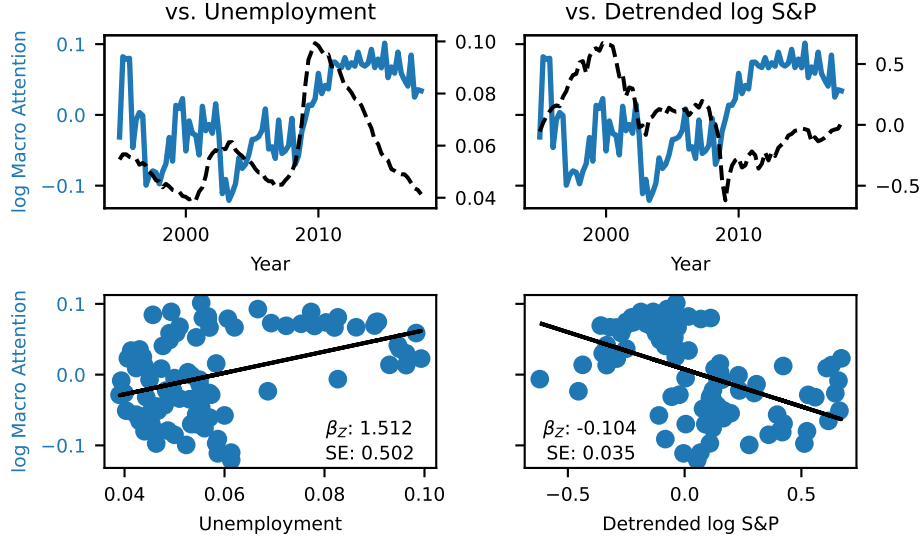
	Profitability, π_{it}	log Stock Return, R_{it}		
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	IV
$\hat{u}_{it}^2$	-0.121 (0.0217)	-0.0162 (0.0304)		
$\hat{u}_{it}^2 \times R_t$	0.123 (0.0980)			
π_{it}		0.402 (0.0283)	0.423 (0.0336)	0.627 (0.277)
$\pi_{it} \times R_t$			-0.309 (0.167)	-1.616 (0.622)
Sector x Time FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	51515	41325	41325	41325
R^2	0.663	0.402	0.402	0.010
F-stat				18.34

Notes: In column 1, the outcome π_{it} is firm-level profitability, defined as contemporaneous EBIT divided by lagged variable costs. In columns 2-4, the outcome R_{it} is the log stock return. $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 4.1. R_t is the log return of the S&P 500. Column 4 reports results from an instrumental variable regression (two-stage least squares) in which $(\hat{u}_{it}^2, \hat{u}_{it}^2 \times R_t)$ are the instruments. Standard errors are double-clustered at the year and firm level.

without relying on our structural estimation of misoptimization. Our estimating equation is the mirror of Equation 33 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.

Figure 5: 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 twelve-quarter bandwidth.

5.5 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 5). To quantify the variable’s cyclicity, we estimate the linear regression of log Macro Attention on unemployment and the detrended S&P 500, akin to Equation 32. Our coefficient estimates are 1.512 (SE: 0.502) for unemployment and -0.104 (SE: 0.035) for the S&P. The raw correlations with the unemployment rate and detrended stock price are, respectively, 0.402 and -0.495.

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, which more generally states that firms’ attention toward *all* variables that matter for their decisions—both microeconomic and macroeconomic—rises in downturns. As discussed in Section 4.2, the feasible test in our data requires measuring the cyclicity of attention toward factors that are commonly faced by all firms (i.e., macroeconomic developments), whereas an ideal, but ultimately infeasible, test would also check whether attention to microeconomic factors rises in downturns.

To build confidence that this result is not driven by specific words used in our construction of Macro Attention, Figure A4 plots the time-series behavior of each component word. 58

of the 89 words have a positive correlation with unemployment (Figure A5).

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. To study this, we compute “output cyclical” in each of our industries as the correlation between sectoral GDP growth, calculated using quarterly BEA data since 2005 linked to our sectors, and aggregate nominal GDP growth. In Figure A6, we plot in the cross-section of industries the relationship between this output cyclical and the coefficient of sector-level Macro Attention on the US unemployment rate. Consistent with our interpretation, we find that the extent of counter-cyclical increases with the industry’s output cyclical, but acyclical industries still have counter-cyclical attention. This finding contrasts with the prediction of an alternative model in which the cyclical in macroeconomic attention is purely driven by firms’ exposure to macroeconomic shocks—under this alternative model, only sufficiently cyclical sectors would exhibit cyclical Macro Attention.

5.6 Fact 6: Macro Attention Predicts Smaller Misoptimizations

We next investigate the firm-level relationship between misoptimization and attention. We estimate the following regression of $\hat{u}_{it}^2$, the squared misoptimization, on the log of firm-level macro attention:

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

Absorbed effects at the sector-by-time level partial out all trends, including the cyclical patterns studied earlier. Additional controls X_{it} can include individual fixed effects, to isolate variation at the firm level; and 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 ($p \approx 0.003$). The more controlled specifications in columns (2) to (4) estimate negative, similarly sized effects, but with less precision. We interpret this result as consistent with Prediction 6, which states that firms more attentive to economic conditions make smaller misoptimizations. This is despite the limitation that our textual approach measures attention only toward a subset of the factors relevant for firms (i.e., macroeconomic factors), whereas our outcome variable would pick up mistakes driven by either firm-level or aggregate shocks. The finding is consistent, instead, with the hypothesis that observed attention to macroeconomic variables and unobserved attention to microeconomic variables

Table 3: Macro-Attentive Firms Make Smaller Misoptimizations

	(1)	(2)	(3)	(4)
$\log \text{MacroAttention}_{it}$	-0.00960 (0.00327)	-0.00593 (0.00333)	-0.00614 (0.00403)	-0.00574 (0.00407)
Sector x Time FE	Yes	Yes	Yes	Yes
Firm FE	No	No	Yes	Yes
TFP and Return Controls	No	Yes	No	Yes
N	28709	24781	28309	24320
R^2	0.054	0.068	0.383	0.384

Notes: This table reports estimates of Equation 36. $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 4.1. $\log \text{MacroAttention}_{it}$ is the measure of firm-level macroeconomic attention. Standard errors are clustered at the firm level.

move together, as our theory would suggest.

Fact 6 also pushes against an important alternative interpretation of our time-series finding: that managers use the macroeconomy as a scapegoat for poor performance, engaging in a form of “attribution bias.” If this were the only reason that firms mentioned the macroeconomy in their filings, one might expect supposedly macro-attentive firms to be making *more* unprofitable mistakes, contrary to the evidence in Table 3. Nonetheless, we cannot entirely rule out that our measured Macro Attention at least partially reflects this phenomenon.

5.7 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.²⁴ 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:

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

We estimate “TFP Innovation Variance” as the weighted average, $\mathbb{E}[s_{it}^* \hat{\epsilon}_{it}^2] / \mathbb{E}[s_{it}^*]$. This measure has a correlation of 0.38 with the equivalent from Bloom et al. (2018) over a common

²⁴Kehrig (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 E, we study how this object, as well as the related calculation for the value marginal product of labor, behaves in our data.

sample period.²⁵ Like Bloom et al. (2018), we find that TFP variance is significantly higher in recessions. The measure spikes markedly in the Early 2000s Recession and the Great Recession, as well as in the 1990s boom (Figure A7).

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 a firm’s expectation conditional on this history and applying Proposition 1, we can decompose the variance of firms’ forecast errors as:

$$\underbrace{\mathbb{E}[(x_{i,t+k} - \mathbb{E}_{it}[x_{i,t+k}])^2]}_{\text{Forecast Error Variance}} = \underbrace{\mathbb{E}[(x_{i,t+k}^*(z_{i,t+k}) - \mathbb{E}_{it}[x_{i,t+k}^*])^2]}_{\text{Fundamental Variance}} + \underbrace{\mathbb{E}[\sigma_i(z_{i,t+k})^2]}_{\text{Misoptimization Variance}} \quad (38)$$

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 cyclicalities of forecast errors.²⁶

6 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;

²⁵The 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.

²⁶See Charoenwong et al. (2021) and Chiang (2023) for an analysis of cyclical forecast error variance.

and endogenously higher volatility when output is low. We show that these findings account for a significant portion of observed business-cycle asymmetries.

6.1 Calibration

In our calibration, as in our theoretical results of Section 3, the aggregate state variable $\theta_t = (\mathbb{E}_{G_t}[\theta_{it}^{\epsilon-1}])^{\frac{1}{\epsilon-1}}$ 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 $\log \theta_t = \rho_\theta \log \theta_{t-1} + \nu_t$ where $\nu_t \sim^{IID} N(0, \sigma_\theta^2)$. Note that, in this formulation, the shocks ν_t may reflect changes to *any* moment of the productivity distribution that induce changes in the aggregate θ_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).²⁷ 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 $\rho_\theta = 0.95$. To match the elasticity of real wages to output (χ), we only need information on the time-series relationship between real wages and output. We match directly an OLS regression of linearly detrended real wages on linearly detrended GDP, at the quarterly frequency from 1986-2017.²⁸ Our estimate of 0.094 lines up with recent evidence on the weak cyclicalities of wages (Galí and Gambetti, 2019). To calibrate the shock variance σ_θ^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 and the (negative) slope of Misoptimization Dispersion in unemployment. Intuitively, these moments describe the level of misoptimization and the extent of its cyclicalities. These forces are respectively governed by λ , the average entropy cost scaling (itself a sufficient statistic for the entire cross-sectional distribution of λ_i), and γ , the coefficient of relative risk aversion. Panels A and B of Figure A9

²⁷The 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 $\log \theta = 0$.

²⁸Our 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	ϵ	Elasticity of substitution	4
	ρ_θ	Persistence of productivity	0.95
Free Parameters	χ	Elasticity of real wages to output	0.094
	λ	Average weight on entropy penalty	0.420
	γ	Coefficient of relative risk aversion	11.6
	σ_θ^2	Variance of the productivity innovation	4.74×10^{-7}
Matched Moments	β_U	Slope of Misopt. Dispersion on $-\frac{\text{Unemployment}}{100}$	0.846
	$\bar{\sigma}_M^2$	Average level of Misopt. Dispersion	0.080
	χ	Regression of real wages on output	0.094
	σ_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.

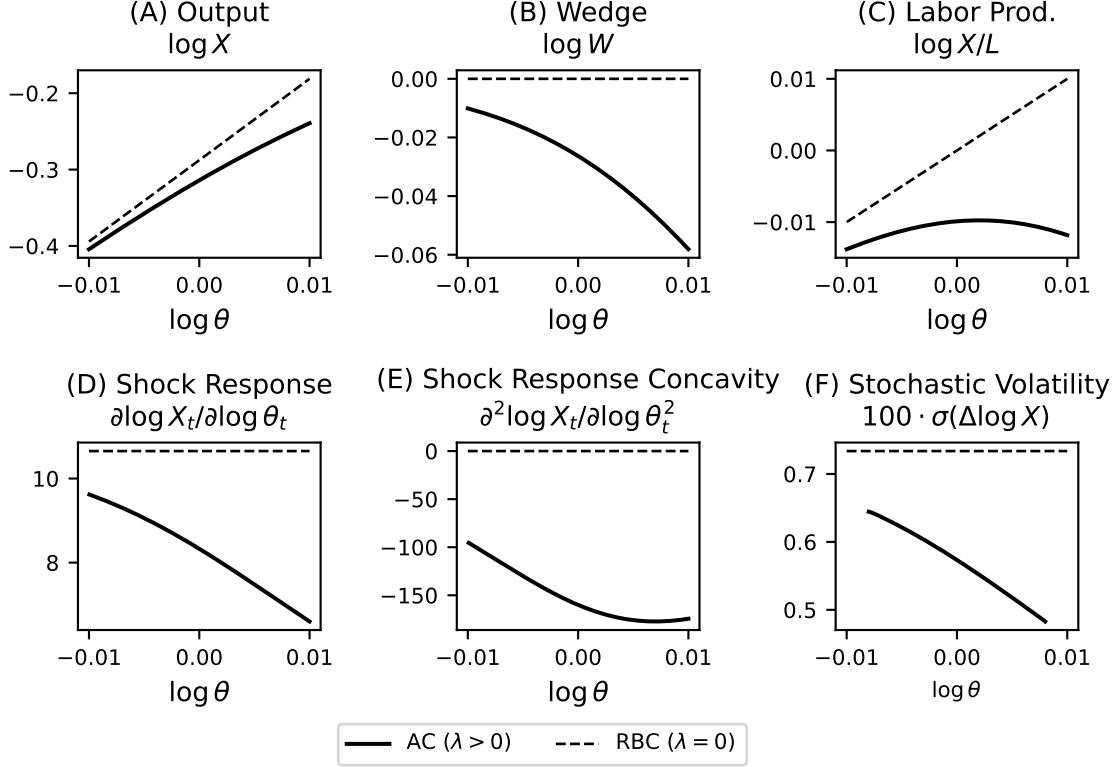
visualize this argument by showing how, as we vary γ and λ , we affect the targeted moments: higher γ makes misoptimizations smaller but more sensitive to the business cycle, and higher λ increases the size of misoptimizations. Both parameters, together, affect the scope of business cycle asymmetries, which we proxy in Panel C of Figure A9 by the $\partial^2 \log X / \partial \log \theta^2$ (recall Corollary 3): asymmetries are stronger under high risk aversion and high costs of attention. We will treat the extent of these asymmetries as a non-targeted prediction, which we compare to outside economic evidence below.

Our calibration of $\gamma = 11.6$ is conservative relative to the modern asset-pricing literature that estimates, in variations of the consumption capital asset pricing model (CCAPM), γ 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. Our calibration of λ is best interpreted by observing that the median firm loses 3% of its profits to inattention, versus a benchmark with zero costs ($\lambda = 0$).

6.2 Quantitative Results

Output, Productivity, and the Attention Wedge. Panels A, B, and C of Figure 6 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

Figure 6: The Macroeconomic Effects of Attention Cycles



Notes: This figure reports the macroeconomic predictions of the calibrated model. Panels A, B, and C respectively plot log output, the log attention wedge (Proposition 4), and log labor productivity as a function of the state. Panels D and E respectively plot the first and second derivatives of the mapping from log productivity to log output (as shown in Panel A). Panel F plots the conditional volatility (standard deviation) of output growth. In each graph, the solid line (“AC”) is the calibrated model and the dashed line is a counterfactual model with $\lambda = 0$ (“RBC”) and all other parameters held fixed.

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 $(1 - \chi\epsilon) \times 2.6\% = 1.6\%$.

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 (Figure A10). In the mean 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.

The attention wedge, as predicted by Corollary 3, is concave. This contributes toward state-dependent propagation of productivity shocks, which we will explore in due course. A quantitative result that hinges on the *extent* of this concavity is that labor productivity is non-monotone in microeconomic productivity θ (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.

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 F of Figure 6). This is despite the lack of asymmetry, heteroskedasticity, or stochastic volatility in the driving shocks. This arises because the attention wedge, and therefore aggregate output, is concave (in logarithms) in productivity (Panel E).

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 9%, 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.²⁹ 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 10.9%. 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.³⁰ Our model can explain about 19% of this movement.

Parameter Robustness and Counterfactual Scenarios. In Appendix F, we provide additional results from our numerical exercise. We first explore the robustness of our main findings to different external calibrations of wage rigidity χ and substitutability ϵ (see Figure A11) and to introducing classical labor markets using the preferences of [Greenwood et al. \(1988\)](#) (see Figure A12). 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

²⁹This calculation is based on comparing the “data” estimates in columns 6 and 7 of Table 9.

³⁰We compare the 3-month “macro uncertainty index” from April 2007 to October 2008.

asymmetries in regimes with smaller 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 (see Figure A13).

7 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.
- Asriyan, V., Laeven, L., and Martin, A. (2022). Collateral booms and information depletion. *The Review of Economic Studies*, 89(2):517–555.

- 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.
- Bertola, G. and Caballero, R. J. (1990). Kinked adjustment costs and aggregate dynamics. In *NBER Macroeconomics Annual*, volume 5, pages 237–288. MIT Press.
- 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.
- Golosov, M. and Lucas, R. E. (2007). Menu costs and phillips curves. *Journal of Political Economy*, 115(2):171–199.
- Gorton, G. and Ordoñez, G. (2014). Collateral crises. *American Economic Review*, 104(2):343–378.
- 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.

- Howes, C. and Weitzner, G. (2026). Bank information production over the business cycle. *Review of Economic Studies*, 93(5):3209–3244.
- 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.
- Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2016). A rational theory of mutual funds’ attention allocation. *Econometrica*, 84(2):571–626.
- Kehrig, M. (2015). The cyclical nature of the productivity distribution. Working paper.
- 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.
- 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.
- Macaulay, A. (2020). Cyclical attention to saving. Working Paper 140, UniCredit Foundation.
- 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.
- McFadden, D. L. (1973). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.
- McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

- 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.
- Ordoñez, G. (2013). The asymmetric effects of financial frictions. *Journal of Political Economy*, 121(5):844–895.
- Parker, J. A. and Julliard, C. (2005). Consumption risk and the cross section of expected returns. *Journal of Political Economy*, 113(1):185–222.
- 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 cyclicalities 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.
- Vavra, J. (2014). Inflation dynamics and time-varying volatility: New evidence and an ss interpretation. *The Quarterly Journal of Economics*, 129(1):215–258.
- Veldkamp, L. L. (2005). Slow boom, sudden crash. *Journal of Economic Theory*, 124(2):230–257.
- 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.
- Zeira, J. (1994). Informational cycles. *The Review of Economic Studies*, 61(1):31–44.

Online Appendix

for “Attention Cycles” by Flynn and Sastry

A Omitted Proofs

A.1 Proof of Proposition 1

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

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

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

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

The Lagrangian is concave in the collection $\{p(x | z)\}$, 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

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

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

$$p(x|z) = \frac{\exp(\lambda_i^{-1} u(x, z))}{\int_{\mathcal{X}} \exp(\lambda_i^{-1} u(x', z)) dx'} \quad (42)$$

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

To solve our firm’s problem, we replace u in the above with $\tilde{\Pi}$ and z with z_i . Performing

this substitution, and ignoring the normalizing constant, we get

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

Taking $\mathcal{X} = \mathbb{R}$, it is then immediate that $p(x|z_i)$ is a Gaussian density 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 re-write 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:

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

□

A.2 Proof of Proposition 2

Proof. Throughout this proof, with 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 $X : \Theta \rightarrow \mathbb{R}$. In particular, they believe that output is given by $X(\theta)$ in each state θ .

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

$$x_{it} = x^*(z_{it}) + \sqrt{\frac{\lambda_i}{|\Pi_{xx}(z_{it})|}} \cdot v_{it}, \quad v_{it} \sim \text{Normal}(0, 1) \quad (45)$$

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

$$\begin{aligned} x^*(z_{it}) &= v_x(\epsilon, \chi, \bar{w}, \bar{X}) \cdot X(\theta_t)^{1-\chi\epsilon} \theta_{it}^\epsilon \\ |\Pi_{xx}(z_{it})| &= v_\Pi(\epsilon, \chi, \bar{w}, \bar{X}) \cdot X(\theta_t)^{-1-\gamma+\chi(1+\epsilon)} \theta_{it}^{-1-\epsilon} \end{aligned} \quad (46)$$

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

$$\begin{aligned} v_x &:= \epsilon^{-\epsilon} (\epsilon - 1)^\epsilon \bar{w}^{-\epsilon} \bar{X}^{\chi\epsilon} \\ v_\Pi &:= (\epsilon - 1)^{-\epsilon} \epsilon^{\epsilon-1} \bar{w}^{1+\epsilon} \bar{X}^{-\chi(1+\epsilon)} \end{aligned} \quad (47)$$

Conditional on the realization of any state θ , 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

$$X = X^* - \frac{1}{2\epsilon(X^*)^{-\frac{1}{\epsilon}}} \mathbb{E}_{\theta_i} \left[(x^*(z_i))^{-1-\frac{1}{\epsilon}} \mathbb{E}_{\lambda_i, v_i} [(x_i - x^*(z_i))^2 \mid \theta_i, \theta] \mid \theta \right] \quad (48)$$

where

$$X^* = \left(\mathbb{E}_{\theta_i} [x^*(z_i)^{1-\frac{1}{\epsilon}} \mid \theta] \right)^{\frac{\epsilon}{\epsilon-1}} \quad (49)$$

We now specialize the expressions above using the structure of the best response in Equations 45, 46, and 47. We first compute X^* as

$$X^* = v_x X^{1-\chi\epsilon\theta^\epsilon} \quad (50)$$

where θ is the transformation defined in Equation 5. We next calculate the second, “variance” term. We start with the “misoptimization variance”

$$\mathbb{E}_{\lambda_i, v_i} [(x_i - x^*(z_i))^2] = \frac{\lambda}{v_\Pi} X^{1+\gamma-\chi(1+\epsilon)} \theta_i^{1+\epsilon} \quad (51)$$

and then calculate the full term

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

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

Substituting in Equations 50 and 52, we derive that equilibrium output solves:

$$X(\theta) = v_x X(\theta)^{1-\chi\epsilon\theta^\epsilon} - \frac{\lambda}{2\epsilon v_x v_\Pi} X(\theta)^{\gamma-\chi\theta} \quad (53)$$

Toward proving existence and uniqueness of an equilibrium with strictly positive output, define:

$$g(X, \theta) = a_0 X^{1-\chi\epsilon\theta^\epsilon} - a_1 X^{\gamma-\chi\theta} \quad (54)$$

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 :

$$\begin{aligned} g_X(X, \theta) &= a_0(1 - \chi\epsilon) X^{-\chi\epsilon\theta^\epsilon} - a_1(\gamma - \chi) X^{\gamma-\chi-1}\theta \\ g_{XX}(X, \theta) &= -a_0(1 - \chi\epsilon)\chi\epsilon X^{-\chi\epsilon-1}\theta^\epsilon - a_1(\gamma - \chi)(\gamma - \chi - 1) X^{\gamma-\chi-2}\theta \end{aligned} \quad (55)$$

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

$$\lim_{X \rightarrow 0} g_X(X, \theta) = +\infty \quad \lim_{X \rightarrow \infty} g_X(X, \theta) = -\infty \quad (56)$$

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

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

fixed point condition:

$$\begin{aligned} \frac{dX(\theta)}{d\theta} &= [a_0(1 - \chi\epsilon)X(\theta)^{-\chi\epsilon}\theta^\epsilon - a_1(\gamma - \chi)X(\theta)^{\gamma-\chi-1}\theta] \frac{dX(\theta)}{d\theta} \\ &\quad + [a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^{\epsilon-1} - a_1X(\theta)^{\gamma-\chi}] \end{aligned} \quad (57)$$

Yielding:

$$\frac{dX(\theta)}{d\theta} = \frac{a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^{\epsilon-1} - a_1X(\theta)^{\gamma-\chi}}{1 - [a_0(1 - \chi\epsilon)X(\theta)^{-\chi\epsilon}\theta^\epsilon - a_1(\gamma - \chi)X(\theta)^{\gamma-\chi-1}\theta]} \quad (58)$$

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

$$\frac{d \log X(\theta)}{d \log \theta} = \frac{a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1X(\theta)^{\gamma-\chi}\theta}{X(\theta) - [a_0(1 - \chi\epsilon)X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1(\gamma - \chi)X(\theta)^{\gamma-\chi}\theta]} \quad (59)$$

We first show that the numerator is positive

$$\begin{aligned} a_0\epsilon X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1X(\theta)^{\gamma-\chi}\theta &= a_0(\epsilon - 1)X(\theta)^{1-\chi\epsilon}\theta^\epsilon + X(\theta) \\ &> 0 \end{aligned} \quad (60)$$

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 $\frac{d \log X(\theta)}{d \log \theta} > 0$ it now suffices to show that the denominator is positive, which follows from $\chi\epsilon < 1$ and $\gamma > \chi + 1 > \chi$:

$$\begin{aligned} X(\theta) &= a_0X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1X(\theta)^{\gamma-\chi}\theta \\ &> a_0 \underbrace{(1 - \chi\epsilon)}_{\chi\epsilon < 1} X(\theta)^{1-\chi\epsilon}\theta^\epsilon - a_1 \underbrace{(\gamma - \chi)}_{\gamma - \chi > 1} X(\theta)^{\gamma-\chi}\theta \end{aligned} \quad (61)$$

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

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, λ_i) is

$$\begin{aligned} m(z_i, \lambda_i, \theta) &:= \mathbb{E}_{v_i} \left[\frac{q^*(z_i)x^*(z_i)}{\mathbb{E}_{\theta_i}^G[q^*(z_i)x^*(z_i)]} \left(\frac{x_i - x^*(z_i)}{x^*(z_i)} \right)^2 \mid z_i, \lambda_i, \theta \right] \\ &= \frac{\theta_i^{\epsilon-1} \frac{\lambda_i}{v_\Pi} X(\theta)^{1+\gamma-\chi(1+\epsilon)} \theta_i^{1+\epsilon}}{\theta^{\epsilon-1} \frac{v_x^2}{v_\Pi} X(\theta)^{2(1-\chi\epsilon)} \theta_i^{2\epsilon}} \\ &= \frac{\lambda_i}{v_\Pi v_x^2} X(\theta)^{\gamma-1+\chi(\epsilon-1)} \theta^{1-\epsilon} \end{aligned} \quad (62)$$

where $v_\Pi, v_x > 0$ are defined in Equation 47 and where we use the definition of $\theta = \hat{\theta}(G)$. Using the law of iterated expectations, we can write with some abuse of notation $m(\theta) =$

$\mathbb{E}_{\theta_i, \lambda_i}[m(z_i, \lambda_i, \theta)]$. Assessing this outer expectation, we derive

$$m(\theta) = \frac{\lambda}{v_{\Pi} v_x^2} X(\theta)^{\gamma-1+\chi(\epsilon-1)} \theta^{1-\epsilon} \quad (63)$$

From this, we have that:

$$\frac{d \log m(\theta)}{d \log \theta} = (\gamma - 1 + \chi(\epsilon - 1)) \frac{d \log X(\theta)}{d \log \theta} - (\epsilon - 1) \quad (64)$$

By re-arranging Equation 59, we obtain that:

$$\frac{d \log X(\theta)}{d \log \theta} = \frac{\epsilon + (\epsilon - 1)a_1 \theta X(\theta)^{\gamma-\chi-1}}{\chi\epsilon + (\gamma - 1 + \chi(\epsilon - 1))a_1 \theta X(\theta)^{\gamma-\chi-1}} \quad (65)$$

and so we obtain that:

$$\frac{d \log m(\theta)}{d \log \theta} = \frac{\epsilon(\gamma - 1)}{\chi\epsilon + (\gamma - 1 + \chi(\epsilon - 1))a_1 \theta X(\theta)^{\gamma-\chi-1}} > 0 \quad (66)$$

as required.

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

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

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. $\square$

A.4 Proof of Proposition 4

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

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

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

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

Or simply:

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

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

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

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

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

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

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

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

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

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

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

or:

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

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

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

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

that the denominator is positive. To this end, we see that:

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

This completes the proof. $\square$

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:

$$L = \int_{[0,1]} L_i \, di \quad (78)$$

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

$$x_i = x^*(z_i) + \tilde{v}_i = v_x X(\theta)^{1-\chi\epsilon} \theta_i^\epsilon + \tilde{v}_i \quad (79)$$

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

$$L = v_x X(\theta)^{1-\chi\epsilon} \int_0^1 \theta_i^{\epsilon-1} \, di \quad (80)$$

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

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

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

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

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

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

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

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

A.6 Proof of Corollary 3

As derived in the proof of Proposition 4, the attention wedge in state θ solves the fixed-point equation

$$W(\theta) = W(\theta)^{1-\chi\epsilon} - a_1(\lambda)a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}} W(\theta)^{\gamma-\chi} \theta^{\frac{\gamma-1}{\chi}} \quad (84)$$

where $a_0, a_1 > 0$. Dividing both sides by $W(\theta)^{1-\chi\epsilon}$ and re-arranging yields

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

We next re-write this function in terms of $t = \log \theta$ and $\omega : \mathbb{R} \rightarrow \mathbb{R}$, defined by $\omega(\log \theta) = \log W(\theta)$. This transformation is valid under the argument that $W(\theta) > 0$ for all $\theta \in \Theta$, as shown in the proof of Proposition 4 under the maintained conditions $\chi\epsilon < 1$ and $\gamma > \chi + 1$. We also, for convenience, define $A(\lambda) = a_1(\lambda)a_0^{\frac{\gamma-\chi-1}{\chi\epsilon}} > 0$.

$$e^{\chi\epsilon\omega(t)} + A(\lambda)e^{(\gamma-\chi(1-\epsilon)-1)\omega(t)+\frac{\gamma-1}{\chi}t} = 1 \quad (86)$$

Differentiating both sides yields

$$\chi\epsilon\omega'(t)e^{\chi\epsilon\omega(t)} + A(\lambda)e^{(\gamma-\chi(1-\epsilon)-1)\omega(t)+\frac{\gamma-1}{\chi}t} \left((\gamma - \chi(1 - \epsilon) - 1)\omega'(t) + \frac{\gamma - 1}{\chi} \right) = 0 \quad (87)$$

Differentiating again and collecting terms yields

$$\begin{aligned} \omega''(t) \left(\chi\epsilon e^{\chi\epsilon\omega(t)} + A(\lambda)e^{(\gamma-\chi(1-\epsilon)-1)\omega(t)+\frac{\gamma-1}{\chi}t} (\gamma - \chi(1 - \epsilon) - 1) \right) + (\omega'(t))^2 (\chi\epsilon)^2 e^{\chi\epsilon\omega(t)} + \\ A(\lambda)e^{(\gamma-\chi(1-\epsilon)-1)\omega(t)+\frac{\gamma-1}{\chi}t} \left((\gamma - \chi(1 - \epsilon) - 1)\omega'(t) + \frac{\gamma - 1}{\chi} \right)^2 = 0 \end{aligned} \quad (88)$$

And therefore,

$$\omega''(t) = - \frac{(\omega'(t))^2 (\chi\epsilon)^2 e^{\chi\epsilon\omega(t)} + A(\lambda)e^{(\gamma-\chi(1-\epsilon)-1)\omega(t)+\frac{\gamma-1}{\chi}t} \left((\gamma - \chi(1 - \epsilon) - 1)\omega'(t) + \frac{\gamma - 1}{\chi} \right)^2}{\chi\epsilon e^{\chi\epsilon\omega(t)} + A(\lambda)e^{(\gamma-\chi(1-\epsilon)-1)\omega(t)+\frac{\gamma-1}{\chi}t} (\gamma - \chi(1 - \epsilon) - 1)} \quad (89)$$

The numerator is strictly positive given that $A(\lambda) > 0$ and $\omega'(t) < 0$ (the latter of which is established, under the maintained conditions, by Proposition 4). A sufficient condition for the denominator to be positive is that $\gamma - \chi(1 - \epsilon) - 1 > 0$ or $\gamma > 1 + \chi - \chi\epsilon$, which follows from the maintained assumption $\gamma > 1 + \chi$ since $\epsilon > 0$. Therefore,

$$\omega''(t) = \frac{\partial^2 \log W}{\partial \log \theta^2} < 0 \quad (90)$$

We now prove the two remaining statements. The first is immediate after observing that $\frac{\partial^2 \log W}{\partial \log \theta^2} = \frac{\partial^2 \log X}{\partial \log \theta^2}$. The second is immediate by the strict log-concavity of the attention wedge in log-productivity.

A.7 Additional Calculations

A.7.1 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:

$$\Pi(x, z_i) := X^{-\gamma} \left(x^{1-\frac{1}{\epsilon}} X^{\frac{1}{\epsilon}} - x \frac{w}{\theta_i} \right) \quad (91)$$

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

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

which can be re-arranged to define

$$x^*(z_i) = \left(1 - \frac{1}{\epsilon} \right)^{\epsilon} X \left(\frac{w}{\theta_i} \right)^{-\epsilon} \quad (93)$$

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

$$\begin{aligned} \Pi(x, z_i) &= \Pi(x^*(z_i), z_i) + \Pi_x(z_i)(x - x^*(z_i)) + \frac{1}{2}\Pi_{xx}(z_i)(x - x^*(z_i))^2 + O^3(x) \\ &=: \tilde{\Pi}(x, z_i) + O^3(x) \end{aligned} \quad (94)$$

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 first-order condition, $\Pi_x(z_i) = 0$. Thus, our approximation reduces to the quadratic utility function in the Linear-Quadratic equilibrium:

$$\tilde{\Pi}(x, z_i) = \Pi(x^*(z_i), z_i) + \frac{1}{2}\Pi_{xx}(z_i)(x - x^*(z_i))^2 \quad (95)$$

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

$$\begin{aligned} \Pi(x^*(z_i), z_i) &= X^{-\gamma} \left(X \left(\frac{w}{\theta_i} \right)^{1-\epsilon} \right) \left(\left[1 - \frac{1}{\epsilon} \right]^{\epsilon(1-\frac{1}{\epsilon})} - \left[1 - \frac{1}{\epsilon} \right]^{\epsilon} \right) \\ &= X^{-\gamma} \left(X \left(\frac{w}{\theta_i} \right)^{1-\epsilon} \right) \epsilon^{-\epsilon} (\epsilon - 1)^{\epsilon-1} \end{aligned} \quad (96)$$

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

$$\Pi_{xx}(z_i) = X^{-\gamma} \cdot \pi_{xx}(z_i) \quad (97)$$

We calculate, using the form of the profit function from Equation 8, the dollar profit func-

tion's second derivative:³¹

$$\begin{aligned}
\pi_{xx}(x^*(z_i), X) &= -\frac{1}{\epsilon} \left(1 - \frac{1}{\epsilon}\right) (x^*(z_i))^{-1-\frac{1}{\epsilon}} X^{\frac{1}{\epsilon}} \\
&= -\frac{1}{\epsilon} \left(1 - \frac{1}{\epsilon}\right) \left(1 - \frac{1}{\epsilon}\right)^{-(1+\frac{1}{\epsilon})\epsilon} X^{-1-\frac{1}{\epsilon}} X^{\frac{1}{\epsilon}} \left(\frac{w}{\theta_i}\right)^{\epsilon(1+\frac{1}{\epsilon})} \\
&= -\epsilon^{\epsilon-1} (\epsilon-1)^{-\epsilon} X^{-1} \left(\frac{w}{\theta_i}\right)^{1+\epsilon}
\end{aligned} \tag{98}$$

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

$$\pi_{xx}(z_i) = -v_\pi(\epsilon, \chi, \bar{X}, \bar{w}) \cdot \theta_i^{-1-\epsilon} X^{\chi(1+\epsilon)-1} \tag{99}$$

as in Equation 13, where the constant is

$$v_\pi(\epsilon, \chi, \bar{X}, \bar{w}) := (\epsilon-1)^{-\epsilon} \epsilon^{\epsilon-1} \bar{w}^{1+\epsilon} \bar{X}^{-\chi(1+\epsilon)} > 0 \tag{100}$$

A.7.2 Quadratic Approximation of Final-Goods Technology

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

$$X(\{x_i\}_{i \in [0,1]}) = \left(\int_0^1 x_i^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} \tag{101}$$

Technically speaking, we take a quadratic approximation of a discretized version of this aggregator, and then take 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 $\omega_{kk'k''} = (\theta_k, \lambda_{k'}, v_{k''})$ with corresponding production level $x(\omega_{kk'k''})$. Define the CES aggregator in this economy as:

$$X_{KK'K''}(\{x_{kk'k''}\}) = \left(\frac{1}{K} \sum_{k=1}^K \frac{1}{K'} \sum_{k'=1}^{K'} \frac{1}{K''} \sum_{k''=1}^{K''} x(\omega_{kk'k''})^{1-\frac{1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \tag{102}$$

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

$$\begin{aligned}
X_{KK'K''} &= X_K^* + \frac{1}{K} \sum_{k=1}^K \frac{1}{K'} \sum_{k'=1}^{K'} \frac{1}{K''} \sum_{k''=1}^{K''} D_k(x(\omega_{kk'k''}) - x^*(\theta_k)) \\
&+ \frac{1}{K} \sum_{k=1}^K \frac{1}{K'} \sum_{k'=1}^{K'} \frac{1}{K''} \sum_{k''=1}^{K''} \frac{1}{K} \sum_{\tilde{k}=1}^K \frac{1}{K'} \sum_{\tilde{k}'=1}^{K'} \frac{1}{K''} \sum_{\tilde{k}''=1}^{K''} \frac{1}{2} D_{\tilde{k}\tilde{k}}^2(x(\omega_{kk'k''}) - x^*(\theta_k))(x(\omega_{\tilde{k}\tilde{k}'\tilde{k}''}) - x^*(\theta_{\tilde{k}}))
\end{aligned} \tag{103}$$

³¹Because marginal costs are constant, this curvature arises purely from the curvature of the revenue function.

where:

$$X_K^* = \left(\frac{1}{K} \sum_{k=1}^K x^*(\theta_k)^{1-\frac{1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}} \quad (104)$$

and:

$$D_k = (X_K^*)^{\frac{1}{\epsilon}} x^*(\theta_k)^{-\frac{1}{\epsilon}} \quad (105)$$

and:

$$D_{k\tilde{k}}^2 = \begin{cases} -KK'K''\frac{1}{\epsilon}x^*(\theta_k)^{-\frac{1}{\epsilon}-1}(X_K^*)^{\frac{1}{\epsilon}} + \frac{1}{\epsilon}(X_K^*)^{\frac{2}{\epsilon}-1}x^*(\theta_k)^{-\frac{2}{\epsilon}} & \text{if } kk'k'' = \tilde{k}\tilde{k}'\tilde{k}'' \\ \frac{1}{\epsilon}(X_K^*)^{\frac{2}{\epsilon}-1}(x^*(\theta_k))^{-\frac{1}{\epsilon}}(x^*(\theta_{\tilde{k}}))^{-\frac{1}{\epsilon}} & \text{if } kk'k'' \neq \tilde{k}\tilde{k}'\tilde{k}'' \end{cases} \quad (106)$$

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

$$\frac{1}{K''} \sum_{k''=1}^{K''} D_k(x(\omega_{kk'k''}) - x^*(\theta_k)) \xrightarrow{a.s.} 0 \quad (107)$$

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

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

$$\begin{aligned} Q &= \frac{X_K^{*\frac{2}{\epsilon}-1}}{2\epsilon} \left[\frac{1}{K} \sum_{k=1}^K \frac{1}{K'} \sum_{k'=1}^{K'} \frac{1}{K''} \sum_{k''=1}^{K''} x^*(\theta_k)^{-\frac{1}{\epsilon}} (x(\omega_{kk'k''}) - x^*(\theta_k)) \right]^2 \\ &\quad - \frac{X_K^{*\frac{1}{\epsilon}}}{2\epsilon} \frac{1}{K} \sum_{k=1}^K \frac{1}{K'} \sum_{k'=1}^{K'} \frac{1}{K''} \sum_{k''=1}^{K''} x^*(\theta_k)^{-1-\frac{1}{\epsilon}} (x(\omega_{kk'k''}) - x^*(\theta_k))^2 \end{aligned} \quad (108)$$

Fixing K and K' , and noting that the draws across k'' are IID conditional on k and k' , we have that (almost surely):

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{K'} \sum_{k'=1}^{K'} \frac{1}{K''} \sum_{k''=1}^{K''} x^*(\theta_k)^{-\frac{1}{\epsilon}} (x(\omega_{kk'k''}) - x^*(\theta_k)) \xrightarrow{a.s.} 0 \quad \text{as } K'' \rightarrow \infty \quad (109)$$

Hence, we have that:

$$Q \xrightarrow{a.s.} -\frac{1}{2\epsilon} \frac{1}{K} \sum_{k=1}^K \frac{1}{K'} \sum_{k'=1}^{K'} \frac{\sigma_{kk'}^2}{x^*(\theta_k)^{1+\frac{1}{\epsilon}} X_K^{*-1-\frac{1}{\epsilon}}} \quad \text{as } K'' \rightarrow \infty \quad (110)$$

Defining σ_k^2 as the conditional variance of actions when the attention cost is one, we observe

by Proposition 1 that $\sigma_{kk'}^2 = \lambda_{k'} \sigma_k^2$. Thus, taking the $K' \rightarrow \infty$ limit we have that:

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

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

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

By applying the law of large numbers and the continuous mapping theorem we have that:

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

Combining all of the above, we have shown that under the second-order approximation, in the limit, almost surely:

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

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

$$X = \left(\int_0^1 x^*(z_i)^{1-\frac{1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} - \frac{1}{2\epsilon} \int_0^1 \frac{(x_i - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}} (x^*(z_i))^{1+\frac{1}{\epsilon}}} di \quad (115)$$

B Extended Model

In this appendix, we formally develop the extension of the baseline model from Section 2 to include multiple inputs and market-clearing wages. In the process, we will provide a micro-foundation for the wage rule and motivate the stock return regression analysis in Section 5.3.

B.1 Set-up

Time is discrete, and indexed by $t \in \mathbb{N}$. There are three kinds of firms: perfectly competitive materials firms who use labor to produce materials; intermediate goods producers who differ in their productivity 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:

$$M_t = \theta_t^M L_t^M \quad (116)$$

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

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

$$x_{it} = \theta_{it} L_{it}^\alpha m_{it}^{1-\alpha} \quad (117)$$

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

Intermediate goods producers perfectly cost-minimize facing wages w_t and intermediate goods prices p_t^M . That is, for a given production level x_{it} , they always choose the cost-minimizing input bundle. We define the firm-level *decision state* $z_{it} = (\theta_{it}, X_t, w_t, p_t^M, C_{t+1}) \in \mathcal{Z}$ 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 and future capitalist consumption C_{t+1} (described shortly). All firms believe that the vector z_{it} follows a first-order Markov process with transition densities described by $f : \mathcal{Z} \rightarrow \Delta(\mathcal{Z})$, 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:

$$\max_{p \in \mathcal{P}} \int_{\mathcal{Z}} \int_{\mathcal{X}} \Pi(x, z_{it}) p(x | z_{it}) dx f(z_{it} | z_{i,t-1}) dz - c(p, \lambda_i, z_{i,t-1}, f) \quad (118)$$

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

$$X_t = X(\{x_{it}\}_{i \in [0,1]}) = \left(\int_{[0,1]} x_{it}^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} \quad (119)$$

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:

$$\mathcal{U}^C(\{C_{t+j}\}_{j \in \mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta_C^j \frac{C_{t+j}^{1-\gamma} - 1}{1-\gamma} \quad (120)$$

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

$$C_t + A_{t+1} + \int_{[0,1]} P_{it} S_{it+1} di \leq \int_{[0,1]} d_{it} S_{it} di + (1 + r_t) A_t + \int_{[0,1]} P_{it} S_{it} di \quad (121)$$

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:

$$\mathcal{U}^W(\{C_{t+j}^W, L_{t+j}\}_{j \in \mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta_W^j U \left(C_{t+j}^W - \frac{L_{t+j}^{1+\psi}}{1+\psi} \right) \quad (122)$$

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:

$$C_t^W = w_t L_t \quad (123)$$

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

$$\{L_t^M, M_t, p_t^M, r_t, C_t^W, w_t, \{x_{it}, L_{it}, m_{it}, \pi_{it}, d_{it}, P_{it}, S_{it}\}_{i \in [0,1]}, X_t, L_t, C_t, A_t\} \quad (124)$$

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 Π is approximated around its optimal level and the CES aggregator is approximated as described in Section 2.5.

B.2 Characterizing Equilibrium

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

Production by Intermediate Goods Firms. Owing to CES aggregation, intermediate goods firms face an 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:

$$\min_{L_{it}, m_{it}} w_t L_{it} + p_t^M m_{it} \quad \text{s.t.} \quad x_{it} = \theta_{it} L_{it}^\alpha m_{it}^{1-\alpha} \quad (125)$$

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:

$$L_{it} = \frac{1}{\theta_{it}} \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} \left(\frac{p_t^M}{w_t} \right)^{1-\alpha} x_{it}, \quad m_{it} = \frac{1}{\theta_{it}} \left(\frac{\alpha}{1-\alpha} \right)^{-\alpha} \left(\frac{p_t^M}{w_t} \right)^{-\alpha} x_{it} \quad (126)$$

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

$$w_t L_{it} + p_t^M m_{it} = \frac{q_t}{\theta_{it}} x_{it} \quad (127)$$

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

$$q_t := \left[\left(\frac{\alpha}{1-\alpha} \right)^{-\alpha} \frac{1}{1-\alpha} \right] w_t^\alpha (p_t^M)^{1-\alpha} = c_\alpha w_t^\alpha (p_t^M)^{1-\alpha} \quad (128)$$

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

$$\pi_{it} = X_t^{\frac{1}{\epsilon}} x_{it}^{1-\frac{1}{\epsilon}} - \frac{q_t}{\theta_{it}} x_{it} \quad (129)$$

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, $C_{t+1} = \int_{[0,1]} d_{it+1} di = \int_{[0,1]} \pi_{it} di$. The firm's risk-adjusted profit is then given by:

$$\Pi_{it} = C_{t+1}^{-\gamma} \pi_{it} \quad (130)$$

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

As in the main text, we define the optimal production level $x^*(z_{it})$ which solves:

$$x^*(z_{it}) := \arg \max_{x \in \mathcal{X}} \Pi(x; z_{it}) \quad (131)$$

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

$$\Pi_{xx}(z_{it}) := \left. \frac{\partial^2 \Pi}{\partial x^2} \right|_{x^*(z_{it}); z_{it}} \quad (132)$$

The approximate objective of the intermediate goods firm is:

$$\tilde{\Pi}(x; z_{it}) := \bar{\Pi}(z_{it}) + \frac{1}{2} \Pi_{xx}(z_{it}) (x - x^*(z_{it}))^2 \quad (133)$$

Under this approximate objective, it follows by a slight algebraic variation of the arguments

in Proposition 1 that optimal choices follow:

$$\begin{aligned} x_{it} &\sim N\left(x_{it}^*, \frac{\lambda_i}{|\Pi_{xx,it}|}\right) \\ x_{it}^* &= \left(1 - \frac{1}{\epsilon}\right)^\epsilon X_t \theta_{it}^\epsilon q_t^{-\epsilon} \\ |\Pi_{xx,it}| &= (\epsilon - 1)^{-\epsilon} \epsilon^{\epsilon-1} C_{t+1}^{-\gamma} X_t^{-1} \theta_{it}^{-1-\epsilon} q_t^{1+\epsilon} \end{aligned} \quad (134)$$

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:

$$L_t = \int_{[0,1]} L_{it} di + L_t^M = \int_{[0,1]} L_{it} di + \frac{1}{\theta_t^M} \int_{[0,1]} m_{it} di \quad (135)$$

where the second equality follows by market clearing for intermediates as $\int_{[0,1]} m_{it} di = M_t = \theta_t^M L_t^M$. 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:

$$\begin{aligned} L_t &= \int_{[0,1]} L_{it} di + \frac{1}{\theta_t^M} \int_{[0,1]} m_{it} di = \int_{[0,1]} L_{it} di + \frac{1}{\theta_t^M} \int_{[0,1]} \frac{1-\alpha}{\alpha} \frac{w_t}{p_t^M} L_{it} di \\ &= \left(1 + \frac{1}{\theta_t^M} \frac{1-\alpha}{\alpha} \frac{w_t}{p_t^M}\right) \int_{[0,1]} L_{it} di = \frac{1}{\alpha} \int_{[0,1]} L_{it} di \end{aligned} \quad (136)$$

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:

$$L_t = \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha} \left(\frac{p_t^M}{w_t}\right)^{1-\alpha} \int_{[0,1]} \frac{x_{it}}{\theta_{it}} di \quad (137)$$

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

$$\begin{aligned} \int_{[0,1]} \frac{x_{it}}{\theta_{it}} di &= \mathbb{E} \left[\frac{x_{it}}{\theta_{it}} \right] = \mathbb{E} \left[\mathbb{E} \left[\frac{x_{it}}{\theta_{it}} | \theta_{it} \right] \right] = \mathbb{E} \left[\frac{1}{\theta_{it}} x_{it}^* \right] \\ &= \mathbb{E} \left[\frac{1}{\theta_{it}} \left(1 - \frac{1}{\epsilon}\right)^\epsilon X_t q_t^{-\epsilon} \theta_{it}^\epsilon \right] = \left(1 - \frac{1}{\epsilon}\right)^\epsilon X_t q_t^{-\epsilon} \theta_t^{\epsilon-1} \end{aligned} \quad (138)$$

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

$$L_t = \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} \left(\frac{p_t^M}{w_t} \right)^{1-\alpha} \left(1 - \frac{1}{\epsilon} \right)^\epsilon X_t q_t^{-\epsilon} \theta_t^{\epsilon-1} \quad (139)$$

Substituting this into the workers' intratemporal Euler equation, and using Equation 128 to write the marginal cost in terms of materials prices and wages, we obtain:

$$\begin{aligned} w_t^{\frac{1}{\psi}} &= \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} \left(\frac{1}{\theta_t^M} \right)^{1-\alpha} \left(1 - \frac{1}{\epsilon} \right)^\epsilon \theta_t^{\epsilon-1} c_\alpha^{-\epsilon} X_t \left(w_t \left(\frac{p_t^M}{w_t} \right)^{1-\alpha} \right)^{-\epsilon} \\ &= \frac{1}{\alpha} \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} (\theta_t^M)^{(\epsilon-1)(1-\alpha)} \left(1 - \frac{1}{\epsilon} \right)^\epsilon \theta_t^{\epsilon-1} c_\alpha^{-\epsilon} X_t w_t^{-\epsilon} \end{aligned} \quad (140)$$

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

$$q_t = c_\alpha w_t^\alpha (p_t^M)^{1-\alpha} = \bar{q}_t X_t^\chi \quad (141)$$

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

$$\bar{q}_t = c_\alpha^{1-\chi(\epsilon-1)} \left(1 - \frac{1}{\epsilon} \right)^{\epsilon\chi} \theta_t^{\chi(\epsilon-1)} \theta_t^{M-(1-\alpha)(1-\chi(\epsilon-1))} \quad (142)$$

Marginal costs, holding fixed productivity, increase in output due to upward-sloping labor supply or convex disutility of effort. The intercept of this “cost rule” varies as a function of productivity in the intermediate-goods and materials sectors. Observe that Equation 141 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

$$w_t = \bar{w}_t X_t^\chi \quad (143)$$

where $\bar{w}_t = \bar{q}_t|_{\alpha=1}$. This verifies our claim in the main text that a wage rule (albeit with an exogenously productivity-dependent slope) can be micro-founded in the simple model. However, in a model with materials or $\alpha < 1$, we obtain exactly the wage rule studied in the main text (but for firms' marginal costs, which are the relevant object to which wages correspond in our baseline model) 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 marginal-cost 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:

$$X = \left(\int_0^1 x^*(z_i)^{1-\frac{1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} - \frac{1}{2\epsilon} \int_0^1 \frac{(x_i - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}} (x^*(z_i))^{1+\frac{1}{\epsilon}}} di \quad (144)$$

where the mean and variance are taken over the realizations of θ_{it} , conditional on the aggregate state θ_t . Substituting in Equation 134, this provides one equation in terms of (X_t, C_{t+1}) . Consider first the computation of X_t^* . Observe that we can write:

$$x_{it}^* = \delta_t X_t^{1-\chi\epsilon} \theta_{it}^\epsilon \quad (145)$$

where:

$$\delta_t = \left(1 - \frac{1}{\epsilon}\right)^\epsilon \bar{q}_t^{-\epsilon} \quad (146)$$

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

$$X_t^* = \left(\delta_t^{\frac{\epsilon-1}{\epsilon}} X_t^{\frac{\epsilon-1}{\epsilon}(1-\chi\epsilon)} \theta_t^{\epsilon-1}\right)^{\frac{\epsilon}{\epsilon-1}} = \delta_t X_t^{1-\chi\epsilon} \theta_t^\epsilon \quad (147)$$

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

$$\begin{aligned} \mathbb{E} \left[\frac{(x_{it} - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}} (x^*(z_i))^{1+\frac{1}{\epsilon}}} \right] &= (X_t^*)^{\frac{1}{\epsilon}} \mathbb{E} \left[x^*(z_{it})^{-1-\frac{1}{\epsilon}} \mathbb{E}[(x_{it} - x^*(z_{it}))^2 | z_{it}] \right] \\ &= (X_t^*)^{\frac{1}{\epsilon}} \mathbb{E} \left[x^*(z_{it})^{-1-\frac{1}{\epsilon}} \frac{\lambda}{|\Pi_{xx,it}|} \right] \end{aligned} \quad (148)$$

To simplify this, observe that we can write:

$$\frac{1}{|\Pi_{xx,it}|} = \zeta_t^{-1} C_{t+1}^\gamma X_t^{1-\chi(\epsilon+1)} \theta_{it}^{1+\epsilon} \quad (149)$$

where:

$$\zeta_t = (\epsilon - 1)^{-\epsilon} \epsilon^{\epsilon-1} \bar{q}_t^{1+\epsilon} \quad (150)$$

So we may express:

$$\begin{aligned} \mathbb{E} \left[\frac{(x_{it} - x^*(z_i))^2}{(X^*)^{-\frac{1}{\epsilon}} (x^*(z_i))^{1+\frac{1}{\epsilon}}} \right] &= (X_t^*)^{\frac{1}{\epsilon}} \mathbb{E} \left[\delta_t^{-1-\frac{1}{\epsilon}} X_t^{-(1+\frac{1}{\epsilon})(1-\chi\epsilon)} \theta_{it}^{-1-\epsilon} \lambda \zeta_t^{-1} C_{t+1}^\gamma X_t^{1-\chi(\epsilon+1)} \theta_{it}^{1+\epsilon} \right] \\ &= \lambda \zeta_t^{-1} \delta_t^{-1} C_{t+1}^\gamma X_t^{-\chi} \theta_t \end{aligned} \quad (151)$$

Putting all of the above together we have that:

$$X_t = \delta_t X_t^{1-\chi\epsilon} \theta_t^\epsilon - \frac{\lambda}{2\epsilon} \zeta_t^{-1} \delta_t^{-1} C_{t+1}^\gamma X_t^{-\chi} \theta_t \quad (152)$$

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

$$C_{t+1} = \int_{[0,1]} \pi_{it} \, di \quad (153)$$

Using our running approximation:

$$\pi_{it} = \pi_{it}(x_{it}^*) + \frac{1}{2}\pi_{xx,it}(x_{it} - x_{it}^*)^2 \quad (154)$$

we obtain:

$$\begin{aligned} C_{t+1} &= \int_{[0,1]} \pi_{it} di = \mathbb{E} \left[\mathbb{E} \left[\pi_{it}(x_{it}^*) + \frac{1}{2}\pi_{xx,it}(x_{it}^*)(x_{it} - x_{it}^*)^2 \mid x_{it}^* \right] \mid \delta_t, \theta_t \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\pi_{it}(x_{it}^*) + \frac{1}{2}\pi_{xx,it}(x_{it}^*) \frac{\lambda}{|\Pi_{xx,it}|} \mid x_{it}^* \right] \mid \delta_t, \theta_t \right] = \mathbb{E} [\pi_{it}(x_{it}^*) \mid \delta_t, \theta_t] - \frac{\lambda}{2} C_{t+1}^\gamma \quad (155) \\ &= (\epsilon - 1)^{-1} \bar{q}_t \delta_t \theta_t^{\epsilon-1} X_t^{1-\chi(\epsilon-1)} - \frac{\lambda}{2} C_{t+1}^\gamma \end{aligned}$$

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

$$X_t = \left(\frac{C_{t+1} + \frac{\lambda}{2} C_{t+1}^\gamma}{(\epsilon - 1)^{-1} \bar{q}_t \delta_t \theta_t^{\epsilon-1}} \right)^{\frac{1}{1-\chi(\epsilon-1)}} \quad (156)$$

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:

$$\begin{aligned} \left(\frac{C_{t+1} + \frac{\lambda}{2} C_{t+1}^\gamma}{(\epsilon - 1)^{-1} \bar{q}_t \delta_t \theta_t^{\epsilon-1}} \right)^{\frac{1}{1-\chi(\epsilon-1)}} &= \delta_t \left(\frac{C_{t+1} + \frac{\lambda}{2} C_{t+1}^\gamma}{(\epsilon - 1)^{-1} \bar{q}_t \delta_t \theta_t^{\epsilon-1}} \right)^{\frac{1-\chi\epsilon}{1-\chi(\epsilon-1)}} \theta_t^\epsilon \\ &\quad - \frac{\lambda}{2\epsilon} \zeta_t^{-1} \delta_t^{-1} C_{t+1}^\gamma \left(\frac{C_{t+1} + \frac{\lambda}{2} C_{t+1}^\gamma}{(\epsilon - 1)^{-1} \bar{q}_t \delta_t \theta_t^{\epsilon-1}} \right)^{\frac{-\chi}{1-\chi(\epsilon-1)}} \theta_t \end{aligned} \quad (157)$$

We can further re-arrange this equation to obtain:

$$\frac{\bar{q}_t}{\theta_t} \left(\frac{C_{t+1} + \frac{\lambda}{2} C_{t+1}^\gamma}{(\epsilon - 1)^{-1} \bar{q}_t \delta_t \theta_t^{\epsilon-1}} \right)^{\frac{1+\chi}{1-\chi(\epsilon-1)}} = (\epsilon - 1) C_{t+1} + (\epsilon - 2) \frac{\lambda}{2} C_{t+1}^\gamma \quad (158)$$

and then we can again manipulate it to obtain:

$$\left((\epsilon - 1) C_{t+1} + (\epsilon - 2) \frac{\lambda}{2} C_{t+1}^\gamma \right)^{\frac{\epsilon\chi}{1+\chi}} \frac{1 + \frac{\lambda}{2} C_{t+1}^{\gamma-1}}{(\epsilon - 1) + (\epsilon - 2) \frac{\lambda}{2} C_{t+1}^{\gamma-1}} = \frac{(\epsilon - 1)^{\epsilon-1}}{\epsilon^\epsilon} \left(\frac{\theta_t}{\bar{q}_t} \right)^{\frac{\epsilon}{1+\chi}} \quad (159)$$

To establish existence of equilibrium, all we require is that the above equation has a solution. We will focus on when there exists an equilibrium with positive output, when it is unique, and when it is monotone. In what follows, we assume that $\theta_t \in [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_{++}$ and $\theta_t^M = \theta^M(\theta_t)$ for some positive, continuously differentiable and increasing function. We have that:

Proposition 5. *Suppose $\gamma \geq 1$ and $\epsilon \geq 2$. There exists a unique equilibrium with positive*

output. Moreover, equilibrium output and capitalist consumption are strictly increasing in aggregate productivity θ .

Proof. We observe from the facts that $\gamma \geq 1$ and $\epsilon \geq 2$ that the first term on the left-hand side of Equation 159 is strictly positive and strictly increasing, and the second term is also strictly positive and increasing. Moreover, as $C_{t+1} \rightarrow 0$, the left-hand side converges to zero and, as $C_{t+1} \rightarrow \infty$, the left-hand side converges to infinity. Hence, as the right-hand side is strictly positive, there exists a unique solution with $C_{t+1} > 0$ to Equation 159. Moreover, the solution is monotone so long as the right-hand side is monotone, which is the case whenever $\theta_t/\bar{q}_t$ is monotone. Moreover, this ratio satisfies:

$$\frac{\theta_t}{\bar{q}_t} \propto \theta_t^{1-\chi(\epsilon-1)} \theta_t^{M(1-\alpha)(1-\chi(\epsilon-1))} \quad (160)$$

This is strictly increasing whenever $1 - \chi(\epsilon - 1) > 0$. Moreover, we have that:

$$1 - \chi(\epsilon - 1) = 1 - \frac{\psi(\epsilon - 1)}{1 + \epsilon\psi} = \frac{1 + \psi}{1 + \epsilon\psi} > 0 \quad (161)$$

It follows immediately that equilibrium output is also strictly increasing in θ . $\square$

B.3 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 6. *Suppose $\gamma > 1$ and $\epsilon \geq 2$. In the unique equilibrium, aggregate misoptimization $m(\theta)$ is strictly increasing in θ . Moreover, if $\mathbb{E}[\log \theta_i | \theta] - \log \theta$ is increasing, then aggregate attention is strictly decreasing in θ .³²*

Proof. We have that:

$$\begin{aligned} m(\theta) &= \mathbb{E} \left[\frac{q^*(z_i)x^*(z_i)}{\mathbb{E}^\theta[q^*(z_i)x^*(z_i)|\theta]} \left(\frac{x_i - x^*(z_i)}{x^*(z_i)} \right)^2 | \theta \right] \\ &= \frac{1}{\mathbb{E}^\theta[q^*(z_i)x^*(z_i)|\theta]} \mathbb{E} \left[\frac{q^*(z_i)x^*(z_i)}{x^*(z_i)^2} \mathbb{E}[(x_i - x^*(z_i))^2 | z_i, \lambda_i] | \theta \right] \\ &= \frac{1}{\mathbb{E}^\theta[q^*(z_i)x^*(z_i)|\theta]} \mathbb{E} \left[\frac{q^*(z_i)x^*(z_i)}{x^*(z_i)^2} \frac{\lambda_i C^\gamma}{|\pi_{xx}(z_i)|} | \theta \right] \end{aligned} \quad (162)$$

We now use the fact that $|\pi_{xx}(z_i)| = \frac{\epsilon-1}{\epsilon^2} \frac{q^*(z_i)}{x^*(z_i)}$, which allows us to write:

$$m(\theta) = \frac{1}{\mathbb{E}^\theta[q^*(z_i)x^*(z_i)|\theta]} \frac{\epsilon^2 \lambda C^\gamma}{(\epsilon - 1)} \quad (163)$$

³²This is true if firm productivity is proportional to aggregate productivity, *i.e.*, $\theta_i = \kappa_i \theta$ for $\kappa_i > 0$.

We also have that $\pi_i^*(z_i) = \frac{1}{\epsilon} q^*(z_i) x^*(z_i)$. Under the quadratic approximation for profits, we have that:

$$\mathbb{E}[\pi_i(z_i)|z_i, \lambda_i] = \pi_i^*(z_i) + \frac{1}{2} \pi_{xx}(z_i) \frac{\lambda_i C^\gamma}{|\pi_{xx}(z_i)|} = \frac{q^*(z_i) x^*(z_i)}{\epsilon} - \frac{\lambda_i}{2} C^\gamma \quad (164)$$

As capitalist consumption equals past profits, we have that:

$$C = \frac{\mathbb{E}^\theta[q^*(z_i) x^*(z_i) | \theta]}{\epsilon} - \frac{\lambda}{2} C^\gamma \quad (165)$$

It follows that:

$$m(\theta) = \frac{\epsilon}{\epsilon - 1} \frac{\lambda C(\theta)^{\gamma-1}}{1 + \frac{\lambda}{2} C(\theta)^{\gamma-1}} \quad (166)$$

As C is a strictly increasing function and $\gamma > 1$, it follows that aggregate misoptimization is strictly increasing in θ .

We now turn to attention and we recall that $a(\theta) = -\frac{1}{2} \mathbb{E}[\lambda_i \log(2\pi e \sigma_i^2) | \theta]$, where $\sigma_i^2 = \mathbb{E}[(x_i - x_i^*)^2 | z_i, \lambda_i]$. Moreover, up to constants, we have that $\sigma_i^2 \propto C^\gamma X \left(\frac{\theta_i}{q}\right)^{1+\epsilon}$. Hence, we have that:

$$a(\theta) = \text{cons.} - \frac{\lambda}{2} [\gamma \log C(\theta) + \log X(\theta) - (\epsilon + 1) \log q(\theta) + (\epsilon + 1) \mathbb{E}[\log \theta_i | \theta]] \quad (167)$$

As C and X are strictly increasing, it suffices to show that $\mathbb{E}[\log \theta_i | \theta] - \log q(\theta)$ is increasing. From the second-order approximation for output, we have that:

$$\frac{X}{X^*} = 1 - \frac{m(\theta)}{2\epsilon} \quad (168)$$

Moreover, by aggregating firms' optimal outputs, we have that:

$$X^* = \left(1 - \frac{1}{\epsilon}\right)^\epsilon X \left(\frac{\theta}{q}\right)^\epsilon \quad (169)$$

Combining these equations, we have that:

$$\frac{q(\theta)}{\theta} = \frac{\epsilon - 1}{\epsilon} \left(1 - \frac{m(\theta)}{2\epsilon}\right)^{\frac{1}{\epsilon}} \quad (170)$$

As $m(\theta)$ is strictly increasing and $m(\theta) < 2\frac{\epsilon}{\epsilon-1} \leq 2\epsilon$ (as $\epsilon \geq 2$), $q(\theta)/\theta$ is strictly decreasing. Now we write:

$$\mathbb{E}[\log \theta_i | \theta] - \log q(\theta) = \mathbb{E}[\log \theta_i | \theta] - \log \theta - \log \left(\frac{q(\theta)}{\theta}\right) \quad (171)$$

The conclusion then follows immediately from the assumption that $\mathbb{E}[\log \theta_i | \theta] - \log \theta$ is increasing. $\square$

B.4 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 7. *Output can be written in the following way:*

$$\log X(\log \theta, \lambda) = \frac{1}{\chi} \log \tilde{\theta} + \log W(\log \theta, \lambda) \quad (172)$$

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 Equation 156. That the wedge is 0 when $\lambda = 0$ follows immediately from the same equations. $\square$

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

B.5 Interpretation of the Stock Return Regressions

In Section 5.3, we showed that mistakes of the same size by firms lead to more adverse impacts on stock returns when the aggregate stock market return is low. We interpreted this as direct evidence in favor of our mechanism that risk-pricing is a key determinant of attention cycles. The simple model of Section 2 is too stylized to formally map to this regression. However, in the extended model developed in this section, we can motivate 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:

$$u'(C_t)P_{it} = \mathbb{E}_t[\beta_C u'(C_{t+1})(P_{it+1} + d_{it+1})] \quad (173)$$

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

$$u'(\pi_{t-1})P_{it} = \beta_C u'(\pi_t)\pi_{it} + \beta_C u'(\pi_t)\mathbb{E}_t[P_{it+1}] \quad (174)$$

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

$$P_{it} = \beta_C \frac{u'(\pi_t)}{u'(\pi_{t-1})}\pi_{it} + \beta_C \frac{u'(\pi_t)}{u'(\pi_{t-1})}\mathbb{E}_t[P_{it+1}] \quad (175)$$

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

$$\pi_{it} = \pi_{it}(x_{it}^*) + \frac{1}{2}\pi_{xx,it}m_{it}^2 \quad (176)$$

Thus, the firm's stock price follows:

$$P_{it} = \beta_C \frac{u'(\pi_t)}{u'(\pi_{t-1})} \left(\pi_{it}(x_{it}^*) + \frac{1}{2}\pi_{xx,it}m_{it}^2 \right) + \beta_C \frac{u'(\pi_t)}{u'(\pi_{t-1})} \mathbb{E}_t[P_{it+1}] \quad (177)$$

Thus:

$$\frac{\partial P_{it}}{\partial m_{it}^2} = -\frac{1}{2}\beta_C \frac{u'(\pi_t)}{u'(\pi_{t-1})} |\pi_{xx,it}| \quad \frac{\partial P_{it}}{\partial \pi_{it}} = \beta_C \frac{u'(\pi_t)}{u'(\pi_{t-1})} \quad (178)$$

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

$$u'(C_t)P_t = \mathbb{E}_t[\beta_C u'(C_{t+1})(P_{t+1} + \pi_t)] \quad (179)$$

Or:

$$\beta_C \frac{u'(\pi_t)}{u'(\pi_{t-1})} = \frac{P_t}{\mathbb{E}_t[P_{t+1}] + \pi_t} := \frac{1}{R_t} \quad (180)$$

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

Proposition 8. *The equilibrium effect of mistakes on stock prices is given by:*

$$\frac{\partial P_{it}}{\partial m_{it}^2} = -\frac{1}{2R_t} |\pi_{xx,it}| \quad (181)$$

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

$$\frac{\partial P_{it}}{\partial \pi_{it}} = \frac{1}{R_t} \quad (182)$$

Proof. Given in the text above. □

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, which we use as our empirical proxy for the expected aggregate return on equity. This is exactly what we find. The OLS regressions of returns on mistakes retain a similar structure but are intermediated by the curvature of dollar profits across firms. These regression models therefore provide a less sharp test of the risk-pricing channel, although empirically they produce entirely consistent results.

C Alternative Specifications of Stochastic Choice

In this section, we extend our basic class of cost functions to allow for persistent mistakes as in the empirical analysis. In particular, we motivate 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 optimal signal processing under mutual information costs.

C.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 functionals that allows us to place restrictions on the within-firm correlation of mistakes across time and, in particular, to motivate the AR(1) formulation of mistakes that we use in the empirical analysis.

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

$$c(p; \theta_{t-1}) = \int_{\Theta} \int_{\mathcal{X}} \phi(p(x|z)) \, dx \, f(z|z_{t-1}) \, dz \quad (183)$$

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

$$c(p; z_{t-1}, v_{t-1}) = \int_{\Theta} \int_{\mathcal{X}} \phi(p(x|z); v_{t-1}, x^*, x) \, dx \, f(z|z_{t-1}) \, dz \quad (184)$$

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

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

$$\phi(y; m, x^*, x) = \lambda y \log y + \omega y ((x - x^*) - m)^2 \quad (185)$$

Concretely, this leads to the following cost functional:

$$\begin{aligned} c(p; z_{t-1}, v_{t-1}) = \int_{\Theta} \left[\lambda \int_{\mathcal{X}} p(x|z) \log p(x|\theta) \, dx \right. \\ \left. + \omega \int_{\mathcal{X}} ((x - x^*(z)) - v_{t-1})^2 p(x|z) \, dx \right] f(z|z_{t-1}) \, dz \end{aligned} \quad (186)$$

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

$$\tilde{\Pi}(x, z) := \bar{\Pi}(z) + \frac{1}{2} \Pi_{xx}(z) (x - x^*(z))^2 \quad (187)$$

and we suppose that firms solve the one-period problem:

$$\max_{p \in \mathcal{P}} \int_{\Theta} \int_{\mathcal{X}} \tilde{\Pi}(x, z) p(x | z) dx f(z | z_{t-1}) dz - c(p; z_{t-1}, v_{t-1}) \quad (188)$$

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

Proposition 9. *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 $u \sim N\left(0, \frac{\lambda}{|\Pi_{xx}(z)| + 2\omega}\right)$.*

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

$$-\beta(x - \gamma)^2 - \lambda[1 + \log p(x|z)] - \omega(x - \gamma - m)^2 + \mu(z) + \kappa(x, z) = 0 \quad (189)$$

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 the solution:

$$p(x|z) = \frac{\exp\left(-\tilde{\beta}(x - \tilde{\gamma})^2\right)}{\int_{\mathcal{X}} \exp\left(-\tilde{\beta}(x' - \tilde{\gamma})^2\right) dx'} \quad (190)$$

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

$$x|z \sim N\left(\gamma + \frac{\omega}{\beta + \omega} v_{t-1}, \frac{\lambda}{2(\beta + \omega)}\right) \quad (191)$$

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

C.2 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 common, scalar state variable θ , which represents each firm's productivity.

We first introduce the class of posterior-separable cost functionals. Denti (2020) provides this formulation as a representation theorem in stochastic choice space of the usual posterior-based definition of Caplin and Dean (2013):

Definition 3 (Posterior-Separable Cost Functionals). *A cost functional c has a posterior-separable representation if and only if there exists a convex and continuous ϕ such that:*

$$c(p) = \int_{\mathcal{X}} \hat{\phi}(\{p(x|\theta)\}_{\theta \in \Theta}) dx \quad (192)$$

where:

$$\hat{\phi}(\{p(x|\theta)\}_{\theta \in \Theta}) = p(x) \phi\left(\left\{\frac{p(x|\theta)\pi(\theta)}{p(x)}\right\}_{\theta \in \Theta}\right) \quad (193)$$

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

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

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 may be a natural case to consider if the firm is particularly uncertain, and therefore cannot distinguish *ex ante* between actions before they begin processing information. This yields a particularly revealing structure to the optimal policy: the agent's actions in state θ are given by a normal distribution centered on the objective optimum and with variance inversely proportional to the curvature of their objective in that state – a normal mixture model.

Proposition 10. *Suppose that $u(x, \theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2$ with $\beta(\theta) > 0$ and costs are posterior separable with relative entropy kernel $\lambda\phi(\cdot)$ for some $\lambda > 0$. If all actions are exchangeable in the prior, then in the limit as the support of the action set expands to infinity, $\hat{x} \rightarrow \infty$ for $\bar{x} = -\underline{x} = \hat{x}$, an optimal stochastic choice rule converges in distribution to.*³³

$$x|\theta \sim N\left(\gamma(\theta), \frac{\lambda}{2\beta(\theta)}\right) \quad (195)$$

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:

$$I(g) = \int_{\mathcal{S}} \int_{\Theta} g(s, \theta) \log \left(\frac{g(s, \theta)}{\pi(\theta) \int_{\Theta} g(s, \tilde{\theta}) d\tilde{\theta}} \right) d\theta ds \quad (196)$$

³³Formally, all actions are exchangeable in the prior at action support end-point $\hat{x} > 0$ if:

$$\int_{\Theta} \frac{\exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\}}{\int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(\tilde{x} - \gamma(\theta))^2\} d\tilde{x}} \pi_{\hat{x}}(\theta) d\theta = \frac{1}{2\hat{x}} \quad \forall x \in [-\hat{x}, \hat{x}] \quad (194)$$

We now argue that it is without loss to consider a choice over stochastic choice rules $p : \Theta \rightarrow \Delta(\mathcal{X})$. Suppose x is an optimal action conditional on receiving any $s \in S_x$. Suppose that there exist $S_x^1, S_x^2 \subset S_x$ 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 that $\tilde{s} \in S_x^1 \cup S_x^2$ 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 $I(g') < I(g)$ for the mutual information I . 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 $g(\cdot|\cdot) : \Theta \rightarrow \Delta(\mathcal{S})$. 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 $p(\cdot|\cdot) : \Theta \rightarrow \Delta(\mathcal{X})$. The agent's problem can then be directly re-written in the claimed stochastic choice form for some c_I :

$$\max_{P \in \mathcal{P}} \int_{\Theta} \int_{\mathcal{X}} u(x, \theta) dP(x|\theta) d\pi(\theta) - \lambda c_I(P) \quad (197)$$

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

$$c_I(p) = \int_{\Theta} \int_{\mathcal{X}} p(x|\theta) \log p(x|\theta) dx d\pi(\theta) - \int_{\mathcal{X}} p(x) \log p(x) dx \quad (198)$$

where:

$$p(x) = \int_{\Theta} p(x|\theta) d\pi(\theta) \quad (199)$$

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

$$\begin{aligned} \mathcal{L}(\{p(x|\theta), \kappa(x, \theta)\}_{x \in \mathcal{X}, \theta \in \Theta}, \{\tilde{\gamma}(\theta)\}_{\theta \in \Theta}) &= \int_{\Theta} \int_{\mathcal{X}} u(x, \theta) p(x|\theta) dx d\pi(\theta) \\ &- \lambda \left(- \int_{\mathcal{X}} p(x) \log p(x) dx + \int_{\Theta} \int_{\mathcal{X}} p(x|\theta) \log p(x|\theta) dx d\pi(\theta) \right) \\ &+ \int_{\Theta} \int_{\mathcal{X}} \kappa(x, \theta) p(x|\theta) dx d\pi(\theta) + \int_{\Theta} \tilde{\gamma}(\theta) \left(\int_{\mathcal{X}} p(x|\theta) dx - 1 \right) d\pi(\theta) \end{aligned} \quad (200)$$

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

$$p(x|\theta) = \frac{p(x) \exp\{u(x, \theta)/\lambda\}}{\int_{\mathcal{X}} p(\tilde{x}) \exp\{u(\tilde{x}, \theta)/\lambda\} d\tilde{x}} \quad (201)$$

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

$$\int_{\Theta} \frac{\exp\{u(x, \theta)/\lambda\}}{\int_{\mathcal{X}} p(\tilde{x}) \exp\{u(\tilde{x}, \theta)/\lambda\} d\tilde{x}} d\pi(\theta) = 1 \quad (202)$$

Up to now we have applied standard techniques from [Matějka and McKay \(2015\)](#) for any fixed prior π and support $\mathcal{X} = [-\hat{x}, \hat{x}]$. We now use our utility function and exchangeability assumption to derive our result. In particular, we take the utility function as:

$$u(x, \theta) = \alpha(\theta) - \beta(\theta)(x - \gamma(\theta))^2 \quad (203)$$

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

$$\int_{\Theta} \frac{\exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\}}{\int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(\tilde{x} - \gamma(\theta))^2\} d\tilde{x}} \pi_{\hat{x}}(\theta) d\theta = \frac{1}{2\hat{x}} \quad \forall x \in [-\hat{x}, \hat{x}], \hat{x} \in \mathbb{R}_{++} \quad (204)$$

Under this condition, the unconditional action distribution has density $p_{\hat{x}}(x) = \frac{1}{2\hat{x}}$ on $[-\hat{x}, \hat{x}]$. The conditional action distribution then becomes:

$$p(x|\theta) = \frac{\exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\}}{\int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(\tilde{x} - \gamma(\theta))^2\} d\tilde{x}} \quad (205)$$

The denominator of this expression can be computed:

$$\begin{aligned} \int_{\mathcal{X}} \exp\{-\beta(\theta)\lambda^{-1}(x - \gamma(\theta))^2\} dx &= \int_{\mathcal{X}} \frac{\sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}}{\sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x - \gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\} dx \\ &= \sqrt{2\pi\frac{\lambda}{2\beta(\theta)}} \int_{\mathcal{X}} \frac{1}{\sqrt{2\pi\frac{\lambda}{2\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x - \gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\} dx \\ &\rightarrow \sqrt{2\pi\frac{\lambda}{2\beta(\theta)}} \quad \text{as } \hat{x} \rightarrow \infty \end{aligned} \quad (206)$$

It follows that, in that $\hat{x} \rightarrow \infty$ limit:

$$p(x|\theta) \rightarrow \frac{1}{\sqrt{\frac{\pi\lambda}{\beta(\theta)}}} \exp\left\{-\frac{1}{2}\left(\frac{x - \gamma(\theta)}{\sqrt{\frac{\lambda}{2\beta(\theta)}}}\right)^2\right\} \quad (207)$$

Thus, $x|\theta$ is a Gaussian random variable with mean $\gamma(\theta)$ and variance $\frac{\lambda}{2\beta(\theta)}$. $\square$

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:

$$\begin{aligned} \gamma(\theta) &= x^*(X(\theta), \theta) \\ \beta(\theta) &= \frac{1}{2} |\Pi_{xx}(X(\theta), \theta)| \end{aligned} \quad (208)$$

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.³⁴ All results from Section 3 then carry over directly.

D Measuring Productivity and Misoptimization

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

D.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 1998, 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.³⁵

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`) do 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

³⁴We observe that this mapping is approximate, as $x^* > 0$, while this appendix has allowed for $\gamma < 0$. In practice, this occurs with low probability.

³⁵Because 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.

trends. We categorize the data into 44 sectors. These are defined at the 2-digit NAICS level, except for the Manufacturing (31-33) and Information (51) sectors, which we classify at the 3-digit level to achieve better balance of sector size.

D.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 D.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:

$$\begin{aligned} \text{Share}_{M,j'} &= \frac{\sum_{i:j(i)=j'} \sum_t \text{MaterialExpenditure}_{it}}{\sum_{i:j(i)=j'} \sum_t \text{Sales}_{it}} \\ \text{Share}_{L,j'} &= \frac{\sum_{i:j(i)=j'} \sum_t \text{WageBill}_{it}}{\sum_{i:j(i)=j'} \sum_t \text{Sales}_{it}} \end{aligned} \quad (209)$$

2. If $\text{Share}_{M,j'} + \text{Share}_{L,j'} \leq \mu^{-1}$, then set

$$\alpha_{M,j'} = \mu \cdot \text{Share}_{M,j'}, \quad \alpha_{L,j'} = \mu \cdot \text{Share}_{L,j'}, \quad \alpha_{K,j'} = 1 - \alpha_{M,j'} - \alpha_{L,j'} \quad (210)$$

3. Otherwise, adjust shares to match the assumed returns to scale, or set

$$\alpha_{M,j'} = \frac{\text{Share}_{M,j'}}{\text{Share}_{M,j'} + \text{Share}_{L,j'}}, \quad \alpha_{L,j'} = \frac{\text{Share}_{L,j'}}{\text{Share}_{M,j'} + \text{Share}_{L,j'}}, \quad \alpha_{K,j'} = 0 \quad (211)$$

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

$$\log \tilde{\theta}_{it} = \log \text{Sales}_{it} - \frac{1}{\mu} (\alpha_{M,j(i)} \cdot \log \text{MatExp}_{it} + \alpha_{L,j(i)} \cdot \log \text{Empl}_{it} + \alpha_{K,j(i)} \cdot \log \text{CapStock}_{it}) \quad (212)$$

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 μ^{-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.³⁶ The methodology

³⁶In 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 capital elasticities that are negative or greater than 1, but do not otherwise enforce any returns to

of Olley and Pakes (1996) aims, in particular, to correct the bias in standard least-squares estimates that understates the output elasticity to capital. We are reassured by these estimates’ “upstream” and “downstream” similarity to our baseline estimates. To the first point, Table A6 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. To the second point, Panel G of Figure 3 shows that Fact 2 is quantitatively and qualitatively very similar under a measure of Misoptimization Dispersion built from the Olley and Pakes (1996) estimates, compared to our baseline with cost-share TFP (Panel A).

D.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 elasticity $\epsilon > 1$ and aggregate output X_t .³⁷ Finally, firms face sector-specific input prices $(q_{j(i),L,t}, q_{j(i),M,t}, q_{j(i),K,t})$ for the three inputs, respectively.

We model firm (mis-) optimization in the following way that is uniform across inputs. Conditional on any chosen level of production, firms cost minimize over their input bundle conditional on observed input prices. Let $q_{i,T,t}$ 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} :

$$\log x_{it} = \log x^*(q_{i,T,t}, X_t) + m_{it} \quad (213)$$

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

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

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

$$\log x^*(\theta, q_T, X) = \epsilon \log \left(1 - \frac{1}{\epsilon}\right) + \epsilon \gamma_i + \log X - \epsilon \log q_T \quad (214)$$

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

$$x^*(\theta, q_T, X) = \arg \max_x \left\{ x \left(e^{\gamma_i} x^{-\frac{1}{\epsilon}} X^{\frac{1}{\epsilon}} - q_T \right) \right\} \quad (215)$$

□

We next characterize the optimal choice of each input with positive input share:

scale normalization.

³⁷It is straightforward, and consistent with our modeling approach, also to allow substitution within more narrowly defined industries.

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

$$\log Z_{it} = \eta_i + \chi_{j(i),Z,t} + (\epsilon - 1) \log \theta_{it} + m_{it} \quad (216)$$

where

$$\eta_i = \epsilon \gamma_i + \epsilon \log \left(1 - \frac{1}{\epsilon} \right) \quad (217)$$

and

$$\chi_{j(i),Z,t} = \log \alpha_{Z,j(i)} - \log q_{j(i),Z,t} + \log X_t + (1 - \epsilon) \sum_Z \alpha_{Z,j(i)} (\log q_{j(i),Z,t} - \log \alpha_{Z,j(i)}) \quad (218)$$

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

$$\min_{L_{it}, M_{it}, K_{it}} \sum_{Z \in \{L, M, K\}} q_{j(i),Z,t} Z_{it} \quad \text{s.t.} \quad \theta_{it} L_{it}^{\alpha_{L,j(i)}} M_{it}^{\alpha_{M,j(i)}} K_{it}^{\alpha_{K,j(i)}} \geq Q \quad (219)$$

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

$$\log Z_{it} = \log q_{i,T,t} + \log Q + \log \alpha_{Z,j(i)} - \log q_{j(i),Z,t} \quad (220)$$

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

$$\log q_{i,T,t} = \sum_Z \alpha_{Z,j(i)} (\log q_{j(i),Z,t} - \log \alpha_{Z,j(i)}) - \log \theta_{it} \quad (221)$$

The desired expression comes from substituting Equations 213 and 214 into the above. $\square$

This calculation validates our log-linear regression model in Equation 24. 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 ,*

$$\alpha_{j(i),Z} = \left(1 - \frac{1}{\epsilon} \right)^{-1} \frac{q_{j(i),Z,t} Z_{it}}{q_{it} x_{it}} \exp \left(-\frac{m_{it}}{\epsilon} \right) \quad (222)$$

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 220 and

calculate.

$$\begin{aligned}
\alpha_{j(i),Z} &= \left(1 - \frac{1}{\epsilon}\right)^{-1} \frac{q_{j(i),Z,t} Z_{it}}{q_{it} x_{it}} \exp\left(-\frac{m_{it}}{\epsilon}\right) \\
&= \left(1 - \frac{1}{\epsilon}\right)^{-1} \frac{\exp(\log q_{i,T,t} + \log x_{it} + \log \alpha_{Z,j(i)})}{\exp(\log q_{it} + \log x_{it})} \exp\left(-\frac{m_{it}}{\epsilon}\right) \\
&= \left(1 - \frac{1}{\epsilon}\right)^{-1} \exp\left(\log q_{i,T,t} - \log q_{it} + \log \alpha_{Z,j(i)} - \frac{m_{it}}{\epsilon}\right)
\end{aligned} \tag{223}$$

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 214 and the fact that $\log x_{it} = \log x_{it}^* + m_{it}$:

$$\begin{aligned}
\log q_{it} &= \gamma_i + \frac{1}{\epsilon} \log X_t - \frac{1}{\epsilon} \left(\epsilon \log \left(1 - \frac{1}{\epsilon}\right) + \epsilon \gamma_i + \log X_t - \epsilon \log q_{i,T,t} + m_{it} \right) \\
&= -\log \left(1 - \frac{1}{\epsilon}\right) + \log q_{i,T,t} - \frac{m_{it}}{\epsilon}
\end{aligned} \tag{224}$$

We then substitute the above back into Equation 223 to get

$$\begin{aligned}
\alpha_{j(i),Z} &= \left(1 - \frac{1}{\epsilon}\right)^{-1} \exp\left(\log q_{i,T,t} - \left(-\log \left(1 - \frac{1}{\epsilon}\right) + \log q_{i,T,t} - \frac{m_{it}}{\epsilon}\right) + \log \alpha_{j(i),Z} - \frac{m_{it}}{\epsilon}\right) \\
&= \left(1 - \frac{1}{\epsilon}\right)^{-1} \exp\left((\log q_{i,T,t} - \log q_{i,T,t}) + \log \left(1 - \frac{1}{\epsilon}\right) + \log \alpha_{j(i),Z} + \left(\frac{m_{it}}{\epsilon} - \frac{m_{it}}{\epsilon}\right)\right) \\
&= \alpha_{j(i),Z}
\end{aligned} \tag{225}$$

as desired. $\square$

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 one on average. In practice, our mistakes are zero mean by construction and have a variance of about 0.08 in sample, so this correction is quantitatively negligible.

We finally show, in the theory, how the calculation of Equation 212, net of fixed effects, recovers a re-scaling of TFP. In this subsection's language, that calculation is

$$\log \tilde{\theta}_{it} = \log x_{it} + \log q_{it} - \frac{1}{\mu} \left(\sum_Z \alpha_{Z,j(i)} (\log q_{j(i),Z,t} + \log Z_{it}) \right) \tag{226}$$

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

$$\log \tilde{\theta}_{it} = \left(1 - \frac{1}{\epsilon}\right) \log x_{it} + \gamma_i + \frac{1}{\epsilon} \log X_t - \frac{1}{\mu} \left(\sum_Z \alpha_{Z,j(i)} (\log q_{j(i),Z,t} + \log Z_{it}) \right) \tag{227}$$

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

$$\log \tilde{\theta}_{it} = \frac{1}{\mu} \log \theta_{it} + \gamma_i + \frac{1}{\epsilon} \log X_t - \frac{1}{\mu} \left(\sum_Z \alpha_{Z,j(i)} \log q_{j(i),Z,t} \right) \quad (228)$$

Grouping terms into fixed effects, we obtain:

$$\log \tilde{\theta}_{it} = \frac{1}{\mu} \log \theta_{it} + \tau_{j(i),t} + \gamma_i \quad (229)$$

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

E Dispersion in TFPR and Value Marginal Products

In this Appendix, we theoretically and empirically study the behavior of revenue-TFPR (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

$$\log \theta_{it} = \log x_{it} - (\alpha_{j(i),M} \log M_{it} + \alpha_{j(i),K} \log K_{it} + \alpha_{j(i),L} \log L_{it}) \quad (230)$$

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

$$\log \theta_{it}^R = \log \theta_{it} + \log p_{it} = \log \text{Sales}_{it} - (\alpha_{j(i),M} \log M_{it} + \alpha_{j(i),K} \log K_{it} + \alpha_{j(i),L} \log L_{it}) \quad (231)$$

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 D.3](#), by combining this definition with the input-choice policy functions derived in [Lemma 2](#):

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

$$\log \theta_{it}^R = \tilde{\gamma}_i + \Xi_{j(i),t} - \frac{1}{\epsilon} m_{it} \quad (232)$$

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

$$\log \tilde{\theta}_{it}^R = \log \text{Sales}_{it} - (\alpha_{M,j(i)} \cdot \log \text{MatExp}_{it} + \alpha_{L,j(i)} \cdot \log \text{Empl}_{it} + \alpha_{K,j(i)} \cdot \log \text{CapStock}_{it}) \quad (233)$$

where variable definitions follow the convention of Appendix D. 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 $\log \hat{\theta}_{it}^R$.³⁸ We also calculate the log value marginal product of labor (VMPL) by first calculating

$$\log \tilde{\theta}_{it}^L = \log \text{Sales}_{it} - \log L_{it} \quad (234)$$

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

Figure A8 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 price 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.

F Additional Quantitative 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.

F.1 Sensitivity of Main Results

Parameter Choice. To probe robustness to our choice of elasticity of substitution ϵ and the wage rule slope χ , we re-calibrate the model for alternative choices. We summarize these experiments by considering “high” and “low” deviations for each parameter, holding fixed

³⁸Compare with Equation 212, which resembles Equation 233 but for deflating the input shares by the markup. This is exactly the within-model adjustment for prices p_{it} .

the others at baseline values, and present the proportional difference from the baseline in three summary statistics introduced in Section 6.2: (i) the relative output effect of negative and positive shocks, normalized in $\log \theta$ units such that a positive shock 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 A11. Lowering the elasticity of substitution or increasing the implied average markups can have ambiguous effects because it simultaneously increases the bite of a fixed level of misoptimization on misallocation, productivity, and output, while decreasing the bite of the profit-curvature channel toward cyclical attention. We find numerically that increasing markups or decreasing ϵ , toward the level implied by De Loecker et al. (2020), significantly increases the extent of our predicted asymmetries ($\approx 1.75x$), while decreasing markups or increasing ϵ , toward the level implied by Edmond et al. (2018), decreases them. 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:

$$\mathcal{U}(\{C_{t+j}, L_{t+j}\}_{j \in \mathbb{N}}) = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j \frac{\left(C_{t+j} - \frac{L_{t+j}^{1+\phi}}{1+\phi}\right)^{1-\gamma}}{1-\gamma} \quad (235)$$

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 calibrate the parameter ϕ in the RBC model to replicate an elasticity of 0.094 between log real wages and log real output (see Table 4). This is achieved when $\phi = \frac{\chi}{1-\chi}$, where $\chi = 0.094$ is the calibration target.³⁹

Figure A12 shows our results, in the same format as Figure 6 under the baseline model. We find comparable behavior of the attention wedge and losses from misoptimization and inattention. For business-cycle asymmetries, 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.

F.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.

³⁹To make it easy to compare the RBC and AC models, we fix ϕ in the calibration with inattention, rather than recalibrate it to achieve the desired elasticity of real wages to output. Results under this alternative calibration are essentially identical.

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 ϵ between products, which would need to have been higher in the previous, low-markup era than it is today. In our model, a lower elasticity of substitution or higher markup increases the output cost of a fixed amount of misoptimization dispersion, as it intuitively makes each individual product more essential to the consumed good; and it has *a priori* ambiguous effects on the extent of equilibrium attention cycles.

In our model, we run the following simple experiment. First, we adjust ϵ upward to simulate a 15 percentage point decrease in the aggregate markup, to match the estimate of Demirer (2020) for markups since the 1960s; and second, we adjust ϵ 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 groups of bars of Figure A13).

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.⁴⁰ We find, as shown in panels four and five of Figure A13, stronger effects of attention cycles in the regime with more rigid wages. This underscores the complementarity between attention cycles and the steepness of the Keynesian cross, and suggests a novel pathway by which factor price rigidity can influence patterns of macroeconomic volatility.

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 its 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 λ , might interact with our main business cycle predictions. Proposition 4 showed that increases in uncertainty depress output in our model. These shocks also, according to the results of Proposition 1, increase the *sensitivity* of dispersion to macroeconomic conditions and hence, based on extrapolation of this partial-equilibrium logic, may amplify the extent of misoptimization cycles. For this reason, we might predict that elevated uncertainty is also complementary to the asymmetry and state-dependence generated at the macro level.

⁴⁰In particular, we use the ratio of the 1964-2007 estimate and 2007-2017 estimate in Table 3A of Galí and Gambetti (2019).

We explore this relationship by solving for the model equilibrium under scenarios with depressed and elevated attention costs, and numerically verify the predicted complementarity (panels six and seven of Figure A13). Thus our theory predicts that business cycles caused and/or amplified by background uncertainty-inducing events may induce sharper fluctuations in aggregate volatility due to endogenous reallocation of attention.

G Supplemental Tables and Figures

Table A1: Misoptimization and Firm Performance

	Outcome: Returns R_{it}				Outcome: Profitability π_{it}			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{u}_{it}^2$	-0.232 (0.0212)	-0.227 (0.0212)	-0.0589 (0.0241)	-0.0491 (0.0241)	-0.319 (0.0178)	-0.320 (0.0177)	-0.112 (0.0132)	-0.111 (0.0126)
Sector x Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	Yes	Yes	No	No	Yes	Yes
TFP Control	No	Yes	No	Yes	No	Yes	No	Yes
N	42026	42026	41650	41650	51566	51566	51515	51515
R^2	0.239	0.262	0.384	0.403	0.117	0.131	0.663	0.681

Notes: In columns 1-4, the outcome is the log stock return; in columns 5-8, the outcome is profitability, defined as EBIT divided by lagged variable costs. $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 4.1. In columns 2, 4, 6, and 8, we control for the first difference in log TFP, $\Delta \log \hat{\theta}_{it}$. Standard errors are clustered at the firm level.

Table A2: Dynamic Effects of Misoptimization

	Outcome: TFP Growth $\Delta \log \hat{\theta}_{i,t+h}$			Outcome: Returns $R_{i,t+h}$			Outcome: Profitability $\pi_{i,t+h}$		
	(1) $h = 0$	(2) $h = 1$	(3) $h = 2$	(4) $h = 0$	(5) $h = 1$	(6) $h = 2$	(7) $h = 0$	(8) $h = 1$	(9) $h = 2$
$\hat{u}_{it}^2$	-0.00661 (0.00772)	0.0165 (0.00822)	-0.00615 (0.00907)	-0.232 (0.0212)	-0.248 (0.0245)	-0.247 (0.0263)	-0.319 (0.0178)	-0.287 (0.0178)	-0.263 (0.0184)
Sector x Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	50998	41130	32746	42026	35036	28424	51566	42471	34301
R^2	0.235	0.249	0.267	0.239	0.242	0.249	0.117	0.123	0.126

Notes: Each column is the estimate of a separate projection regression. The outcomes are TFP growth $\Delta \log \hat{\theta}_{i,t+h}$ (first three columns), log stock returns $R_{i,t+h}$ (second three columns), and profitability $\pi_{i,t+h}$ (last three columns). $\hat{u}_{it}^2$ is the squared firm-level misoptimization residual, constructed using the methods described in Section 4.1. Standard errors are clustered at the firm level.

Table A3: Markets Punish Misoptimizations Harder in Low States, Additional Controls

	Outcome: log Stock Return, R_{it}					
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{u}_{it}^2 \times R_t$	0.376 (0.142)	0.345 (0.136)	0.373 (0.140)	0.317 (0.163)	0.336 (0.140)	0.492 (0.230)
Sector x Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Direct Effect of $\hat{u}_{it}^2$	Yes	Yes	Yes	Yes	Yes	Yes
log TFP x Interaction	No	Yes	No	No	No	No
Leverage x Interaction	No	No	Yes	No	No	No
Lag Return x Interaction	No	No	No	Yes	No	No
Industry FE x Interaction	No	No	No	No	Yes	No
Firm FE x Interaction	No	No	No	No	No	Yes
N	42026	42026	41872	35186	42026	41650
R^2	0.239	0.262	0.247	0.240	0.241	0.495

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

Table A4: The Effects of Misoptimization in Levels

	Returns R_{it}		Profitability π_{it}		Sq. Mistakes $\hat{m}_{it}^2$	
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{m}_{it}^2$	-0.0417 (0.0134)	-0.0512 (0.0150)	-0.0212 (0.00911)	-0.0252 (0.00986)		
$\hat{m}_{it}^2 \times R_t$		0.106 (0.0708)		0.0432 (0.0403)		
log MacroAttn $_{it}$					-0.0120 (0.00636)	-0.0210 (0.00774)
Sector x Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	No	Yes
N	44512	44512	58252	58252	35000	34321
R^2	0.383	0.383	0.655	0.655	0.053	0.489

Notes: $\hat{m}_{it}$ is the firm-level misoptimization. These specifications replicate results in Tables 1, 2, 3, and A1 with $\hat{m}_{it}$ in place of $\hat{u}_{it}$. Standard errors are clustered at the firm level.

Table A5: Policy Function Estimation

Panel A: Persistence of Misoptimization				
	Outcome: $\hat{m}_{i,t}^0$			
	Baseline	Adj. Cost	Leverage	Quadratic
$\hat{m}_{i,t-1}^0$	0.696 (0.00837)	0.0153 (0.0106)	0.696 (0.00840)	0.682 (0.00893)
N	53838	45806	53473	53838
R^2	0.499	0.000236	0.499	0.473
Panel B: Policy Functions				
	Outcome: $\log L_{it} - \hat{\rho} \log L_{i,t-1}$			
$\log \hat{\theta}_{it}$	0.421 (0.0150)	0.384 (0.0170)	0.420 (0.0157)	0.466 (0.0164)
$\log \hat{\theta}_{i,t-1}$	-0.0329 (0.0113)	-0.0910 (0.0133)	-0.0264 (0.0118)	0.00531 (0.0124)
$\log L_{i,t-1}$		0.810 (0.0118)		
$\log L_{i,t-2}$		-0.0400 (0.00906)		
Lev_{it}			-0.0287 (0.0155)	
$\text{Lev}_{i,t-1}$			-0.0767 (0.0140)	
$\log \hat{\theta}_{it} \times \text{Lev}_{it}$			0.00182 (0.0114)	
$\log \hat{\theta}_{i,t-1} \times \text{Lev}_{i,t-1}$			-0.0375 (0.0108)	
$(\log \hat{\theta}_{it})^2$				0.0449 (0.00709)
$(\log \hat{\theta}_{i,t-1})^2$				0.0324 (0.00537)
N	52449	44536	52078	52449
R^2	0.896	0.990	0.896	0.904

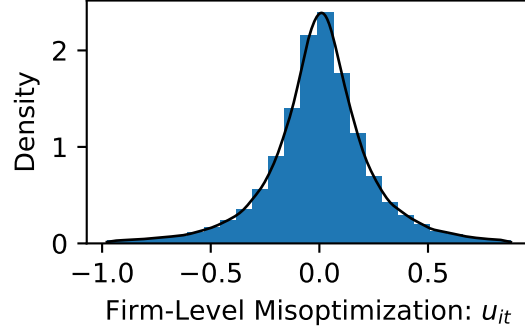
Notes: The four columns correspond to four of our policy-function estimation methods, as described in the main text. The coefficient on $\hat{m}_{i,t-1}^0$ in panel A corresponds in the model to parameter ρ . Standard errors are clustered at the firm level.

Table A6: Comparison of TFP Measures

	Outcome: Cost-Share log TFP		
	(1)	(2)	(3)
Olley and Pakes log TFP	0.892 (0.0232)	0.919 (0.0139)	1.008 (0.0131)
Sector x Time FE	No	No	Yes
Firm FE	No	Yes	Yes
N	29263	28565	28564
R^2	0.280	0.926	0.950

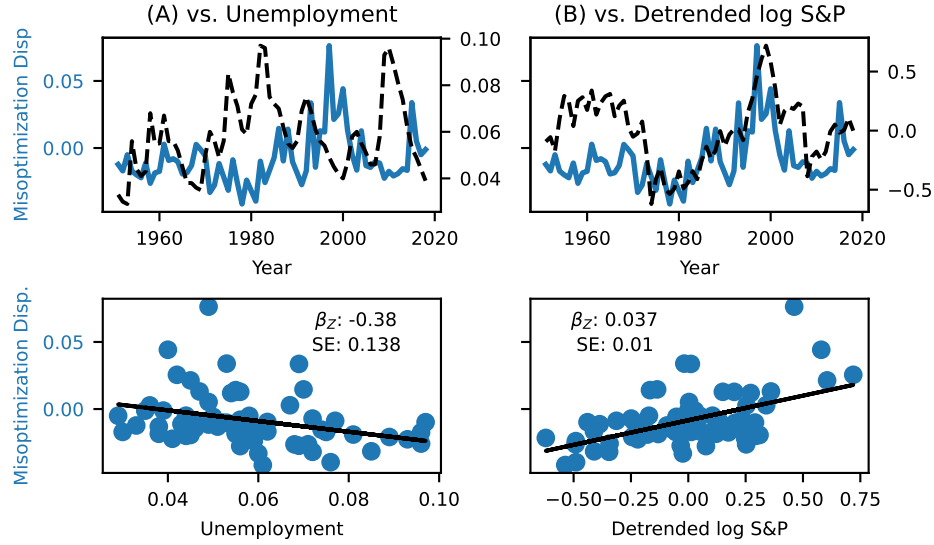
Notes: “Cost-Share log TFP” is our baseline measure of $\log \hat{\theta}_{it}$ used in the main analysis. “Olley and Pakes log TFP” is the alternative measure based on the method of [Olley and Pakes \(1996\)](#). Standard errors are clustered at the firm level.

Figure A1: The Distribution of Misoptimizations



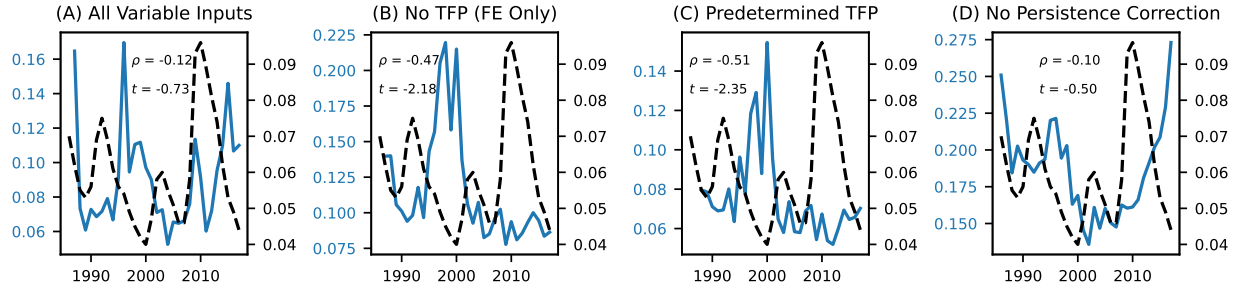
Notes: This figure shows the distribution, across firms i and time periods t , of misoptimizations $\hat{u}_{it}$, as defined in Equation 25. The distribution is truncated at the 1% tails. The solid black line is a Gaussian kernel density estimate.

Figure A2: Misoptimization Dispersion is Pro-Cyclical in the Long Sample



Notes: This Figure replicates Figure 2 using our long-sample (1950-2018) calculation of Misoptimization Dispersion, described in Section 5.2. 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 Under Less Stringent Constructions



Notes: This figure shows the cyclicalities of various constructions of Misoptimization Dispersion that relax key steps of the construction described in Section 4.1. In Panel A, we compute misoptimizations for total variable cost expenditures (total labor plus materials expenditure) in dollar terms. In Panel B, we compute policy functions without any explicit control for TFP—that is, the only controls are firm and industry-by-time fixed effects. In Panel C, we compute policy functions where lagged TFP (predetermined, and observable to the firm) takes the place of contemporaneous TFP. In Panel D, we calculate dispersion in the first-stage residuals $\hat{m}_{it}^0$, which do not account for autocorrelation.

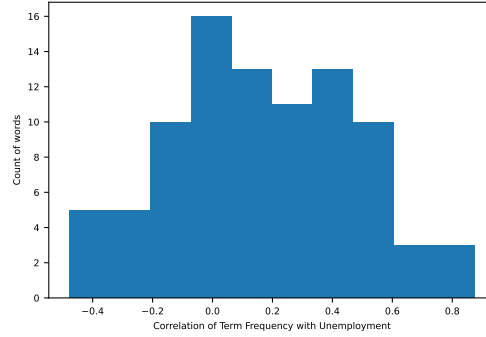
Figure A4: Frequency over Time of Each Word in MacroAttention (Part I)



Figure A4: Frequency over Time of Each Word in MacroAttention (Part II)

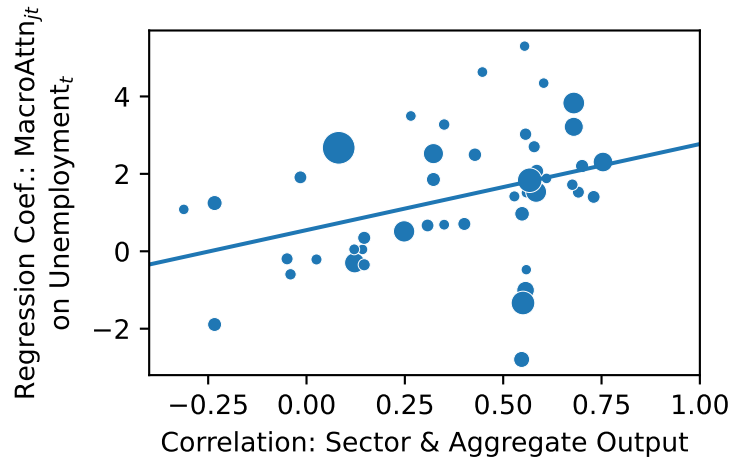


Figure A5: Correlations with Unemployment by Word



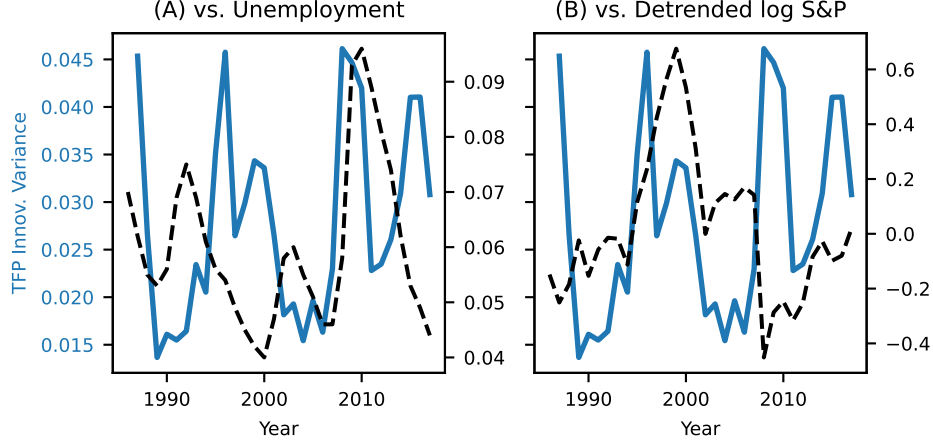
Notes: This figure shows a histogram of the correlations of each individual word in the Macro Attention index (see Figure A4) with the unemployment rate.

Figure A6: Industry-Specific Cyclicalities of Macro Attention



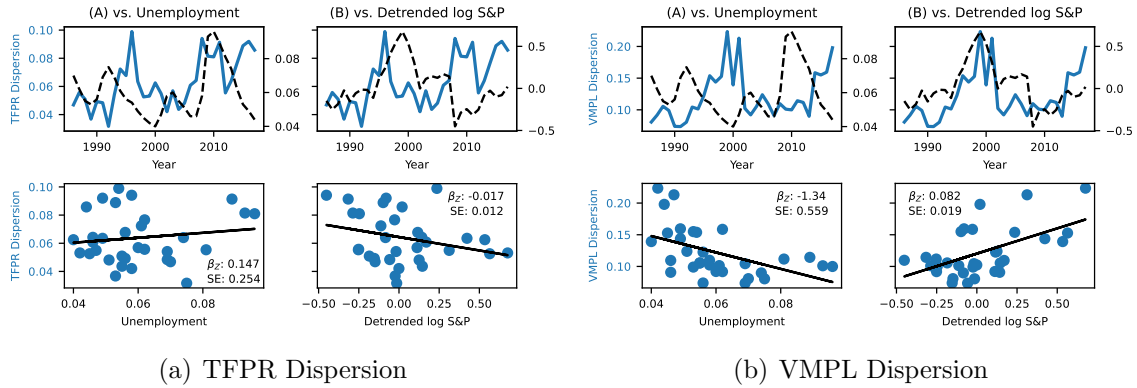
Notes: The horizontal axis is the correlation of sectoral and aggregate nominal GDP growth. The vertical axis is the regression coefficient of log sectoral macro attention, net of quarterly fixed effects, on the US unemployment rate. The dots are sized based on total sales in Compustat in 2015. The blue solid line is a cross-industry linear regression line.

Figure A7: Relationship of TFP Innovation Variance with Macro Variables



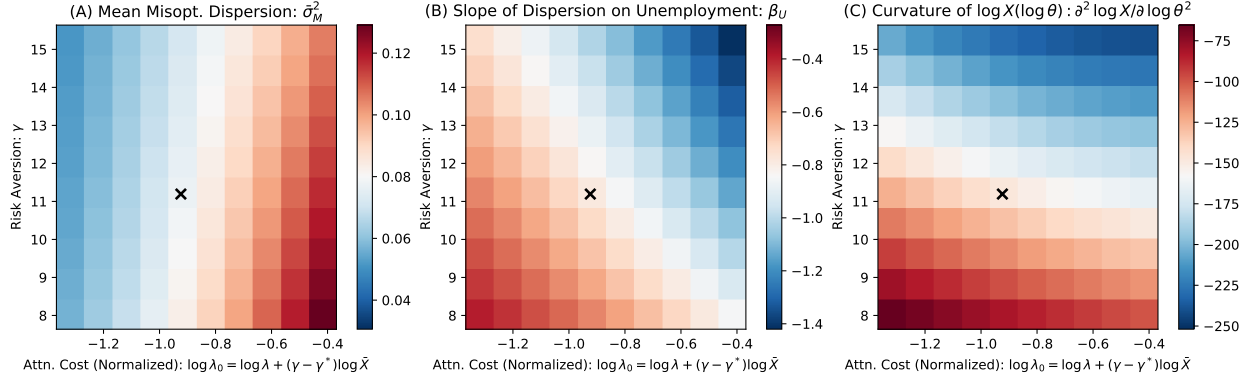
Notes: “TFP Innovations” are the residuals from an AR(1) model for $\log \theta_{it}$ with firm and sector-by-time fixed effects, as described in Section 5.7 of the main text. We calculate their variance, consistent with our main calculations of Misoptimization Dispersion, using optimal-sales weights.

Figure A8: The Cyclical Behavior of TFPR and VMPL Dispersion



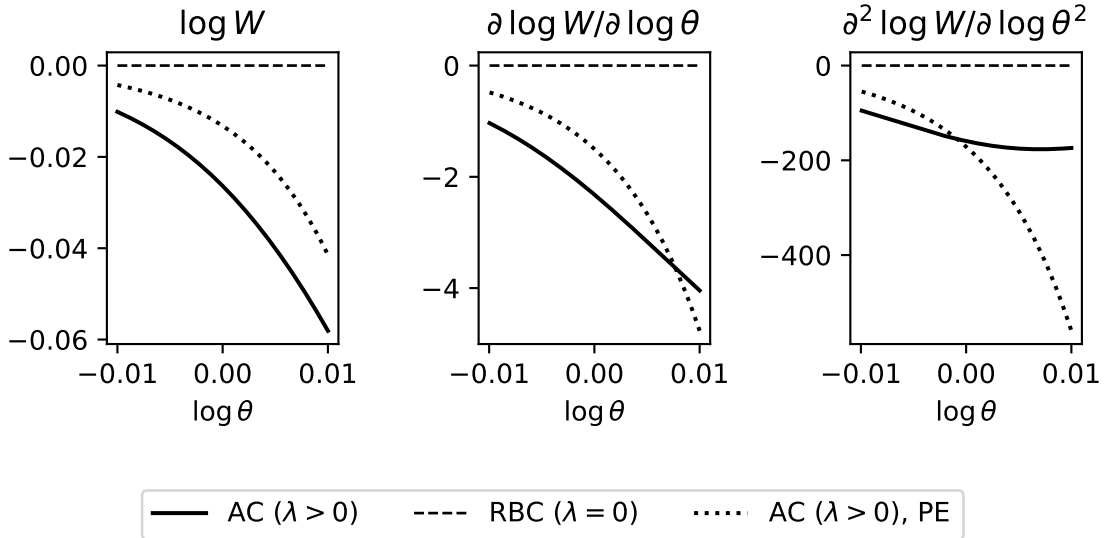
Notes: Each half of this Figure replicates Figure 2 using our measures of TFPR Dispersion (left) and VMPL Dispersion (right). Variable construction is defined in Appendix E. Because of the composition adjustment, the variance metrics can be negative. The standard errors for the linear fits are HAC robust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

Figure A9: Sensitivity of Model Predictions to Parameters



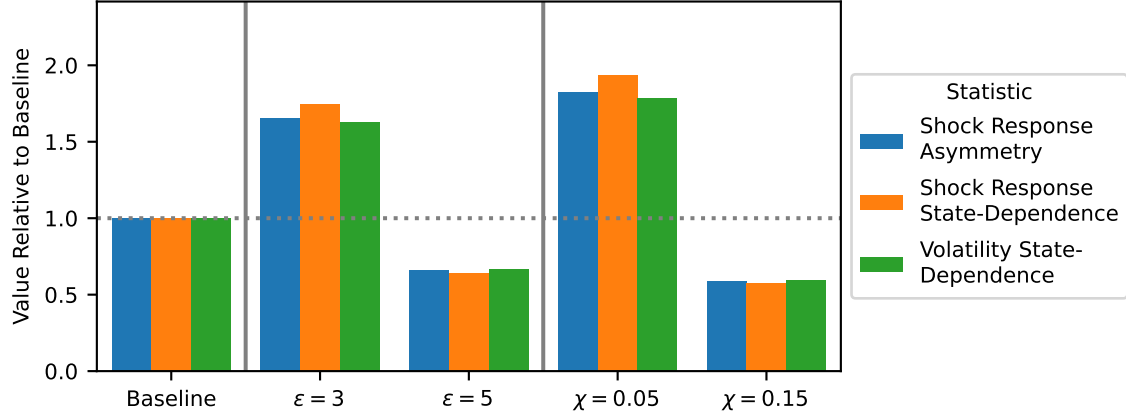
Notes: This Figure reports the sensitivity of model predictions to parameters. In each Panel, the vertical axis varies risk aversion γ . The horizontal axis varies a reparametrized version of the attention cost parameter, $\log \lambda_0 = \log \lambda + (\gamma - \gamma^*) \log \bar{X}$, where $\gamma^* = 11.6$ is the calibrated value of risk aversion and $\bar{X}$ is consumption under the RBC model ($\lambda = 0$) at the median productivity state; this normalizes attention costs by the marginal utility in the median state. All panels hold fixed other model parameters. Panel A shows, as a function of parameters, the mean level of Misoptimization Dispersion (a calibration target). Panel B shows the slope of Misoptimization Dispersion on unemployment, interpreted in the model as $-\log L$ (a calibration target). Panel C shows the curvature of $\log X(\log \theta)$, averaged over the state space (a non-targeted moment). The “X” marks denote the calibrated parameter values.

Figure A10: The Attention Wedge and Its Derivatives



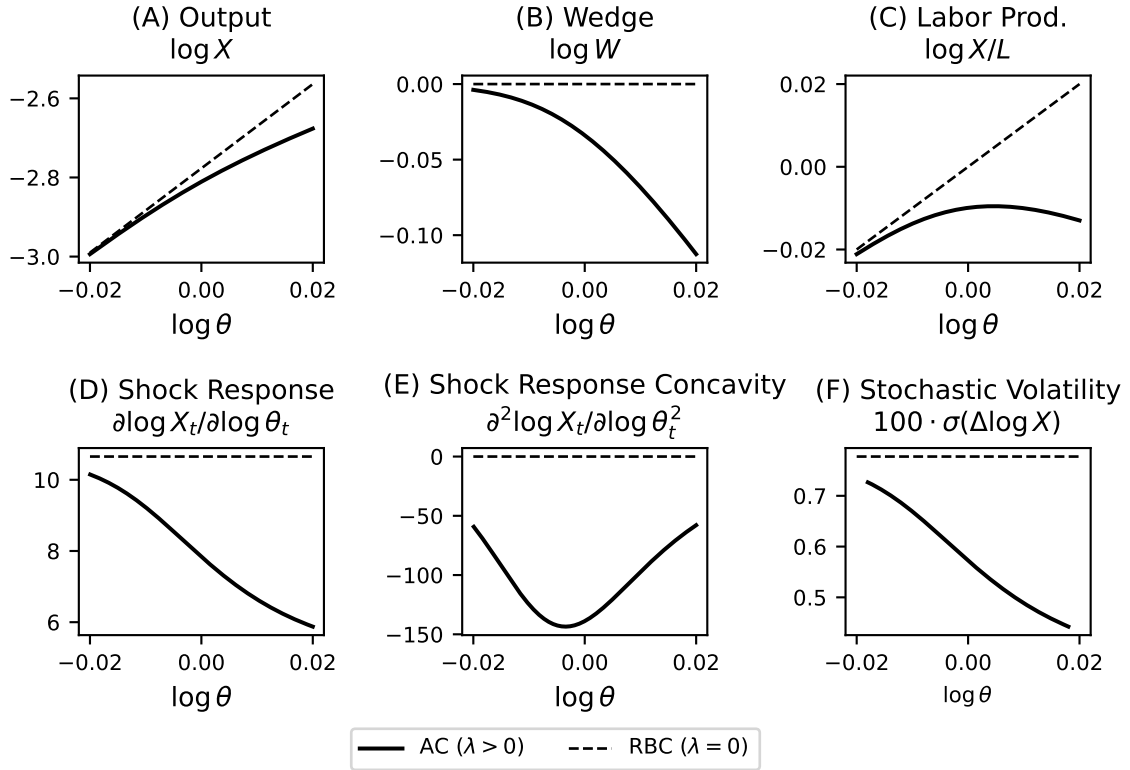
Notes: The panels respectively show the level, first derivative, and second derivative of the log attention wedge in the log state. The “Partial Equilibrium” thought experiment, plotted in each figure as a dotted line, is for firms to best-reply to the output and wages of the counterfactual RBC equilibrium.

Figure A11: Robustness of Numerical Results to Parameter Choices



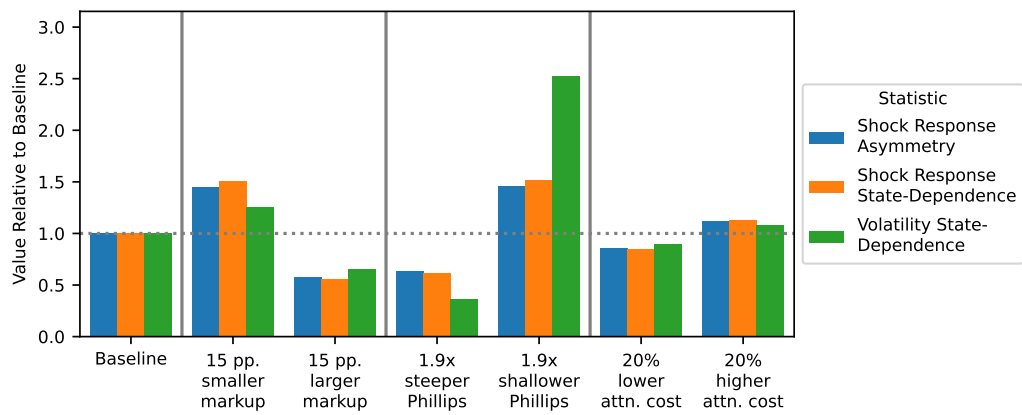
Notes: We re-calibrate the model under each indicated parameter choice. The outcomes and their interpretation are described in Appendix F.1.

Figure A12: Output, Wedge, and Labor Productivity with GHH Preferences



Notes: This recreates Figure 6 from the main analysis in the variant model with Greenwood et al. (1988) preferences, described in Appendix F.

Figure A13: Predictions in Counterfactual Scenarios



Notes: The scenarios are described in the appendix. The outcomes are the same as in Figure A11, and are described in Appendix F.1.

References

- Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3):623–685.
- Caplin, A. and Dean, M. (2013). Behavioral implications of rational inattention with Shannon entropy. Technical report, National Bureau of Economic Research.
- 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.
- 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.