

Heterogeneous Firms, Rational Inattention, and Monetary Policy*

Shengliang Ou[†] Donghai Zhang[‡] Renbin Zhang[§]

September, 2026

Abstract

Firm heterogeneity shapes the distribution of attention across firms and, in turn, monetary transmission. We develop a heterogeneous-firm rational-inattention framework that endogenizes the joint size–attention distribution. A natural prior is that attention rises with firm size because mispricing affects more sales. Yet larger firms face lower demand elasticities and desired pass-through, which reduce the profit loss from mispricing and desired-price sensitivity to nominal shocks. Demand curvature determines which force dominates, generating increasing, decreasing, or hump-shaped attention. Empirically, attention is hump-shaped in size. This distribution shapes monetary policy’s heterogeneous effects across firms and whether nominal shocks raise or lower measured TFP.

JEL Classification: E31; E52; D83; L11; E23.

Keywords: Firm heterogeneity; Rational inattention; Monetary policy; Strategic complementarities; Information frictions

*We thank Klaus Adam, Benjamin Born, Ryan Chahrour, Zhen Huo, and Kris Nimark, seminar participants at the University of Bonn, CREST, and Asia School of Business, and participants at the SMU Macro Workshop, the HKUST–Jinan Workshop, and the WESEAMS Workshop for insightful discussions. Donghai Zhang thanks the Ministry of Education, Singapore, for financial support under Academic Research Fund Tier 2 Grant T2EP40224-0009.

[†]Shanghai Jiao Tong University. Email: shengliang.ou@gmail.com.

[‡]National University of Singapore. Email: donghai.d.zhang@gmail.com.

[§]Shandong University. Email: zhang.renbin.ken@gmail.com.

1. Introduction

Since [Lucas \(1972\)](#) and [Phelps \(1970\)](#), information frictions have been a central explanation for monetary non-neutrality: nominal shocks have real effects even when prices are freely adjustable because firms do not perfectly observe aggregate nominal conditions. Rational-inattention models make this friction endogenous by allowing firms to choose how much information to acquire when attention is costly ([Sims 2003](#); [Mackowiak and Wiederholt 2009](#)). Yet much of this literature studies a representative firm. In reality, firms differ widely in market share and in the demand elasticities, markups, and pass-through that govern their pricing incentives. These differences shape both the profit loss from pricing errors, defined as deviations from full-information profit-maximizing prices, and the sensitivity of desired prices to aggregate nominal conditions. They should therefore matter for the allocation of attention across firms. This leads to three questions: which firms acquire more information about aggregate nominal conditions, do larger firms *necessarily* pay more attention, and what are the aggregate and cross-sectional implications of the resulting allocation of attention?

A natural prior is that larger firms acquire more information. Because a pricing error applies to a larger sales base, the return to information should increase with firm size. Our insight is that this scale argument is incomplete. The value of information depends on both the profit loss from mispricing and how strongly the firm’s desired price responds to aggregate nominal conditions. Larger firms tend to face lower local demand elasticities, which weaken the profit consequences of pricing errors, and lower desired pass-through, which makes their desired prices less responsive to nominal shocks. Together, these forces can overturn the scale effect. The resulting joint distribution of firm size and attention is the paper’s central object. It determines which firms drive monetary transmission and whether nominal shocks raise or lower allocative efficiency and measured TFP.

Theory To characterize this joint distribution, we develop a model of monopolistically competitive firms that differ in permanent productivity and hence in market share. Firms face a general demand system that allows demand elasticities and desired pass-through to vary across firms. Before setting prices, they choose how much information to acquire about a common nominal-demand shock. [Lemma 1](#) decomposes the profit consequences of mispricing into market share, demand elasticity, and super-elasticity. [Proposition 1](#) identifies four channels linking firm size to attention: a positive size channel, a negative elasticity channel, an ambiguous super-elasticity channel, and a negative pass-through channel. We refer to the latter three collectively as the *market-power channels* because they operate through local demand conditions that govern markups and pricing incentives.

The market-power channels can offset the size channel, leaving the general size–attention relationship ambiguous.

We then specialize the model to a Kimball demand system.¹ Its curvature parameter governs how demand elasticity and desired pass-through vary with firm size and hence the strength of the market-power channels. Proposition 2 identifies three regimes. Under constant elasticity of substitution (CES) demand, elasticity and desired pass-through are constant across firms. This shuts down the market-power channels, leaving only the scale effect, so attention rises with size. With moderate curvature, these channels overturn the scale effect only among the largest firms, producing a hump-shaped relationship. With sufficiently strong curvature, they dominate throughout the size distribution, and attention declines with size.

Policy Implications and Quantification We next examine two monetary-policy implications of the size–attention distribution: heterogeneous output responses and supply-side effects. The first concerns which firms account for the aggregate real response to a monetary policy shock. Proposition 3 identifies two margins that shape firms’ output responses. First, a real-rigidity margin operates through desired pass-through. Holding attention fixed, larger firms have lower pass-through and adjust prices less, so more of the shock is absorbed by quantities and their output responses are larger. Second, a nominal-rigidity margin operates through endogenous attention. Holding desired pass-through fixed, more attentive firms adjust prices more strongly, leaving less adjustment to quantities and making their output responses smaller. Under moderate curvature, attention rises with size among small and medium-sized firms but declines among the largest firms. The nominal-rigidity margin therefore offsets the real-rigidity margin over the former range and reinforces it over the latter. The overall size–output relationship depends on the relative strength of these two margins.

The second implication concerns the supply-side effects of monetary policy.² Heterogeneous price adjustments change relative prices and markups, thereby reallocating demand across firms. Proposition 4 decomposes the resulting TFP response into reallocation arising from differences in desired pass-through and differences in attention across firms. The sign of each component depends on how demand elasticities covary with pass-through and attention across firms. Demand curvature shapes these relationships. The same force that governs the allocation of attention therefore determines whether an expansionary nominal shock raises or lowers measured TFP.

¹We use the Kimball specification in [Klenow and Willis \(2016\)](#).

²Following [Baqae et al. \(2024\)](#), we define supply-side effects as changes in measured aggregate TFP arising from changes in misallocation.

We quantify both implications using the lower-curvature economy as our baseline. This calibration is informed by available micro estimates and generates the hump-shaped attention profile documented below. On average, firms’ output responses decline with size, consistent with evidence that small and young firms respond more strongly to monetary policy shocks than large firms.³ The relationship is nevertheless nonmonotone: output responses decline over most of the size distribution but rise again mildly among the very largest firms. Firms in the bottom half of the size distribution account for about 27% of the aggregate output response but only 19% of aggregate sales. Their contribution to aggregate non-neutrality is therefore about 40% larger than their sales share.⁴ An expansionary monetary shock mildly raises measured TFP on impact. In a counterfactual that eliminates the downturn in attention among the largest firms and instead forces attention to continue rising with size, the same shock lowers measured TFP.

Evidence We provide evidence on the shape of the size–attention distribution. Following [Song and Stern \(2025\)](#), we use topic-specific keyword frequencies in annual Form 10-K filings to measure firms’ attention to macroeconomic conditions. Attention rises with firm size through roughly the 70th percentile of the size distribution and declines thereafter. This pattern appears in the unconditional relationship across size deciles and remains after controlling for firm characteristics and industry-by-fiscal-year fixed effects. It therefore also holds among firms in the same industry and fiscal year. The resulting hump shape aligns more naturally with the lower-curvature regime and supports the presence of market-power channels that offset the size effect.

Model Extensions Finally, we show that the forces shaping the size–attention distribution remain operative in richer environments. We extend the framework to multi-product firms, heterogeneous information costs, and idiosyncratic productivity uncertainty in addition to aggregate demand shocks. Multi-product firms can reuse information across operating units, while marginal information costs may fall with firm size. Both forces strengthen the incentives of large firms to acquire information but leave the underlying benefit-cost logic unchanged: size, elasticity, super-elasticity, and pass-through continue to govern the value of information. Under lower curvature, attention remains hump-shaped in size, leaving the baseline policy implications qualitatively unchanged. The qualitative size–attention relationship is also unchanged when firms face idiosyncratic

³See, for example, [Gertler and Gilchrist \(1994\)](#) and [Cloyne et al. \(2023\)](#). [Gnewuch and Zhang \(2025\)](#) further show that the stronger responses of small and young firms persist among firms unlikely to be financially constrained, whether identified by low leverage, high liquidity, or dividend payments.

⁴Under higher curvature, attention instead declines with firm size, shifting the real effects of monetary policy toward larger firms.

productivity uncertainty alongside aggregate demand shocks.

Related literature. The paper relates to several strands of work.

First, the paper contributes to the literature on information frictions as a source of monetary non-neutrality. The literature following [Lucas \(1972\)](#) and [Phelps \(1970\)](#) has studied the amplification and persistence of the real effects of nominal disturbances, emphasizing staggered information updating, higher-order beliefs, and frictional coordination (e.g., [Mankiw and Reis 2002](#); [Woodford 2003](#); [Adam 2007](#); [Angeletos and La'O 2009](#); [Angeletos 2018](#); [Angeletos et al. 2018](#)). [Huo and Pedroni \(2023\)](#) further show how endogenous information aggregation can generate persistence. More recently, [Chahrour and Jurado \(2025b,a\)](#) revisit how information and expectations shape the propagation of aggregate shocks. Methodologically, [Nimark \(2008, 2017\)](#), [Huo and Pedroni \(2020\)](#), and [Huo and Takayama \(2025\)](#) develop tractable representations of economies with higher-order beliefs and dispersed information. We shift the focus from aggregate persistence to the endogenous allocation of attention across heterogeneous firms and its implications for monetary transmission and supply-side effects.

Second, the paper builds on rational-inattention theories of price setting. In this literature, nominal rigidity is an endogenous consequence of firms' information choices ([Sims 2003](#)). [Mackowiak and Wiederholt \(2009\)](#) show that firms devote less attention to aggregate conditions when idiosyncratic conditions are more volatile or matter more for profits. [Maćkowiak and Wiederholt \(2015\)](#) embed this allocation in general equilibrium and generate persistent responses to monetary and technology shocks.⁵ Endogenous attention also changes optimal policy. [Paciello and Wiederholt \(2014\)](#) study optimal policy when firms choose attention to aggregate conditions, while [Ou et al. \(2026\)](#) show how sectoral price stickiness changes attention and optimal monetary policy. [Fang, Huo, Li, and Wu \(2025\)](#) study information acquisition and optimal policy in production networks.⁶ Recent work introduces industrial organization into rational-inattention price setting. [Afrouzi \(2024\)](#) shows that the number of competitors shapes firms' attention to aggregate variables and thereby inflation dynamics and monetary non-neutrality. In [Afrouzi \(2024\)](#), firms are symmetric within each industry, and the focus is on how attention varies with market structure. In contrast, we introduce asymmetry across firms within a market through size, market power, and pass-through. This makes the joint size–attention distribution an equilibrium object.

⁵[Afrouzi and Yang \(2021\)](#) and [Miao et al. \(2022\)](#) develop solution methods for dynamic rational-inattention problems.

⁶[La'O and Tahbaz-Salehi \(2022\)](#) study optimal monetary policy in production networks with exogenous information frictions.

Our paper is related to the literature studying how firm-level uncertainty and information production shape resource allocation across heterogeneous firms over longer horizons. [David et al. \(2016\)](#) examine imperfect information about firm-level fundamentals and quantify its consequences for input misallocation and aggregate productivity. [Asriyan and Kohlhas \(2025\)](#) endogenize data acquisition and show how it shapes firm scale, product choice, pricing, and welfare. We bring endogenous and heterogeneous information acquisition into a business-cycle setting: firms choose attention to a common nominal-demand shock before setting prices. In our framework, market-power channels counteract the size effect, generating increasing, decreasing, or hump-shaped size–attention relationships. The resulting joint size–attention distribution governs monetary transmission across firms and its effects on aggregate output and measured TFP.

The paper proceeds as follows. Section 2 presents the general framework and firms’ information-acquisition problem. Section 3 characterizes the equilibrium size–attention distribution. Section 4 derives its implications for the cross-sectional incidence and supply-side effects of monetary policy. Section 5 quantifies these implications and provides firm-level evidence on the size–attention relationship. Section 6 considers extensions and robustness. Section 7 concludes.

2. The General Framework

This section presents the framework that generates the joint distribution of firm size and attention. Productivity heterogeneity determines firm size, while firms choose how much information to acquire before setting prices. Given the productivity distribution, these information choices endogenously determine how attention is allocated across firms of different sizes. We begin with a general demand system in which firm heterogeneity in size can translate into differences in the local demand conditions that determine market power and pass-through. This setup accommodates CES demand as a benchmark and allows us to specialize later to Kimball demand.

2.1. Households

There is a continuum of identical consumers. Preferences over the aggregate consumption bundle Y and labor L are represented by $U(Y, L) = \log Y - L$. Time is discrete. The baseline is static in the sense that aggregate nominal-demand shocks are independent across periods, allowing us to suppress the time subscript t . Section 6 extends the analysis to persistent aggregate nominal shocks and shows that the central mechanism and the main quantitative patterns remain unchanged.

The consumption bundle Y consists of a continuum of differentiated varieties indexed

by $i \in [0, 1]$. Preferences over these varieties are homothetic, and Y is defined implicitly by

$$\int_0^1 G_i \left(\frac{Y_i}{Y} \right) di = 1,$$

where Y_i denotes consumption of variety i and each G_i is increasing and concave. CES preferences obtain as a special case when $G_i = G$ for all i and G is a power function. More generally, the function G_i , and hence the implied mapping from relative prices to relative demands, may differ across varieties. This specification nests [Kimball \(1995\)](#) preferences as a special case in which the demand mapping is variety-invariant. It is therefore more general than the standard Kimball aggregator.⁷

The representative consumer maximizes utility subject to the budget constraint

$$\int_0^1 P_i Y_i di = WL + \Pi,$$

where W is the wage, L denotes total hours worked, and Π denotes profits. The household's optimal expenditure allocation across varieties yields the inverse demand curve for each variety:

$$\frac{P_i}{P} = G_i' \left(\frac{Y_i}{Y} \right), \quad (2.1)$$

where P^Y denotes the ideal price index and the price aggregator P is defined as $P \equiv P^Y / \int_0^1 G_i' \left(\frac{Y_i}{Y} \right) \left(\frac{Y_i}{Y} \right) di$.⁸ Under CES, the ideal price index coincides with P . To a first-order approximation, the ideal price index also describes the Consumer Price Index (CPI).

Equation (2.1) implies that relative demand for variety i depends only on its relative price, P_i/P . Let z_i denote this relative price:

$$z_i \equiv \frac{P_i}{P}.$$

In equilibrium, larger and more productive firms have lower values of z_i .

Two local demand objects will be central to the analysis: elasticity and super-elasticity.

⁷See [Matsuyama and Ushchev \(2017\)](#) for a more detailed discussion of these differences.

⁸More specifically, $P^Y \equiv \min_{\{Y_i\}} \left\{ \int_0^1 P_i Y_i di : Y = 1 \right\}$.

The price elasticity of demand for variety i , denoted ϵ_i , is

$$\epsilon_i \equiv -\frac{\partial \log \frac{Y_i}{Y}}{\partial \log z_i} = \frac{G'_i\left(\frac{Y_i}{Y}\right)}{-\left(\frac{Y_i}{Y}\right) G''_i\left(\frac{Y_i}{Y}\right)}, \quad (2.2)$$

which can vary across varieties. We have $\epsilon_i > 1$ for all i , implying positive markups.

The super-elasticity, defined as the elasticity of the demand elasticity, is

$$\eta_i \equiv \frac{\partial \log \epsilon_i}{\partial \log z_i} = \frac{z_i \epsilon'_i}{\epsilon_i}, \quad (2.3)$$

where $\epsilon'_i \equiv \partial \epsilon_i / \partial z_i$ measures how the elasticity changes with the relative price z_i .

2.2. Monetary Authority

The monetary authority chooses aggregate nominal demand,

$$Q \equiv PY. \quad (2.4)$$

We interpret movements in Q as monetary policy shocks. Because the equilibrium nominal wage varies with Q , these shocks shift firms' nominal marginal costs.

Letting $q \equiv \log Q$, we assume

$$q \sim N(0, \sigma_q^2). \quad (2.5)$$

2.3. Firms

The economy contains a continuum of monopolistically competitive firms. Firm i has productivity e^{a_i} and produces using labor as the sole input:

$$Y_i = f(e^{a_i}, L_i) = e^{a_i} L_i,$$

where L_i denotes its labor input. Productivity differences generate heterogeneity in firm size and market power. The type parameters $\{a_i\}_{i \in [0,1]}$ are non-stochastic and common knowledge. Aggregate labor is

$$L \equiv \int_0^1 L_i di.$$

Firms do not observe the monetary shock directly but can acquire noisy signals at a cost. First, they choose signal precision. The aggregate shock is then realized, firms ob-

serve their signals, and they set prices conditional on their information sets. Production, consumption, and profits are realized afterward.

Price setting. Each firm chooses its price to maximize expected profits conditional on its information set I_i .

The optimal price solves

$$P_i^* = \arg_{P_i} \max \mathbb{E} (P_i Y_i - W L_i | I_i), \quad (2.6)$$

subject to the demand curve (2.1).

Absent information frictions, the firm sets its price as a markup over marginal cost. The desired markup is

$$\mu_i \equiv \frac{\epsilon_i}{\epsilon_i - 1}.$$

Endogenous Information Acquisition. Because q is the only unobserved payoff-relevant state and is Gaussian, the optimal information structure can be represented by the following scalar signal (Mondria, 2010).⁹

$$s_i = q + e_i, \quad e_i \sim N(0, \sigma_{e,i}^2). \quad (2.7)$$

The firm chooses the variance of the signal noise through its information capacity.

Firm i chooses its information capacity κ_i before shocks are realized to maximize ex ante expected profits net of information-acquisition costs:

$$\kappa_i^* = \arg \max_{\kappa_i} \mathbb{E}[P_i^* Y_i - W L_i] - \lambda f(\kappa_i), \quad (2.8)$$

taking as given the optimal pricing rule in (2.6). The firm's information choice is subject to the information-flow constraint

$$I(\sigma_q^2, \hat{\sigma}_{q,i}^2) = \kappa_i, \quad (2.9)$$

⁹In general, the optimal choice of signal structure is over the firm's full-information profit-maximizing price. In our setting, that price is a function of the aggregate nominal-demand shock q . A signal about q therefore spans the payoff-relevant object and makes the signal structure in (2.7) optimal.

where $\hat{\sigma}_{q,i}^2$ denotes firm i 's posterior variance of q , given by

$$\hat{\sigma}_{q,i}^2 = \left(\frac{1}{\sigma_q^2} + \frac{1}{\sigma_{e,i}^2} \right)^{-1}. \quad (2.10)$$

The information cost is $\lambda f(\kappa_i)$, where κ_i measures the amount of information acquired. The function f is assumed to be twice continuously differentiable on its domain and non-negative, strictly increasing, and convex on $\mathbb{R}_+$. We hold $f(\cdot)$ and λ common across firms to isolate heterogeneity in the benefits of information.

We assume a logarithmic information-flow function. Specifically, under Shannon mutual information, $I(\sigma_q^2, \hat{\sigma}_{q,i}^2) = \frac{1}{2} \log_2(\sigma_q^2 / \hat{\sigma}_{q,i}^2)$, so κ_i measures the reduction in entropy generated by the signal.

After observing its signal, firm i updates its belief about q according to Bayes' rule:

$$\mathbb{E}(q \mid s_i) = K_i s_i, \quad (2.11)$$

where

$$K_i \equiv \frac{\sigma_q^2}{\sigma_q^2 + (\sigma_{e,i}^*)^2} \quad (2.12)$$

is the Kalman gain. Information capacity, posterior variance, signal precision, and the Kalman gain are equivalent measures of attention. Higher κ_i , signal precision, and K_i , together with lower posterior variance, indicate greater attention.

2.4. Linear-Quadratic Approximation

We use a linear-quadratic approximation around the steady state. For any variable X , let x denote its log deviation from the steady state.

Linearized Price Setting. Linearizing the optimal pricing condition in (2.6) gives

$$p_i^* = \mathbb{E}[p_i^\diamond \mid s_i], \quad (2.13)$$

where $p_i^\diamond$ is firm i 's full-information profit-maximizing price (Mackowiak and Wiederholt, 2009). It depends on nominal marginal cost and the aggregate price level:

$$p_i^\diamond = \rho_i q + (1 - \rho_i) p. \quad (2.14)$$

Here ρ_i denotes firm i 's desired partial-equilibrium pass-through, defined as the elasticity of its desired price with respect to nominal marginal cost while holding aggregate conditions fixed. It is given by

$$\rho_i = \frac{1}{1 - \frac{\partial \log \mu_i}{\partial \log z_i}},$$

where μ_i is the desired markup defined above.

Equation (2.13) connects information to pricing: the firm sets its price equal to the conditional expectation of its full-information target. Information frictions therefore operate through forecast errors about $p_i^\diamond$.

Under CES preferences, desired markups are constant and desired pass-through is complete, $\rho_i = 1$. With variable markups, desired pass-through may be incomplete. For brevity, we hereafter refer to ρ_i as the firm's pass-through.

Pass-through is inversely related to strategic complementarity in price setting. Equation (2.14) shows that a lower ρ_i places more weight on the aggregate price level, so the firm's desired price comoves more strongly with competitors' prices. Thus, $(1 - \rho_i)$ measures strategic complementarity arising from variable desired markups. This firm-level force differs from the aggregate strategic complementarities emphasized in earlier work on CES sticky-price models (Adam 2005).

The log-linearized aggregate price indices p and p^Y are

$$p = \int_0^1 \tilde{\epsilon}_i p_i di, \quad p^Y = \int_0^1 u_i p_i di, \quad (2.15)$$

where

$$\tilde{\epsilon}_i \equiv \frac{\bar{P}_i \bar{Y}_i \bar{\epsilon}_i}{\int_0^1 \bar{P}_i \bar{Y}_i \bar{\epsilon}_i di}, \quad u_i \equiv \frac{\bar{P}_i \bar{Y}_i}{\int_0^1 \bar{P}_j \bar{Y}_j dj}.$$

Linear-Quadratic Information Acquisition. A second-order approximation of profits yields the loss from mispricing (Mackowiak and Wiederholt, 2009):

$$\frac{|\pi_{pp,i}|}{2} (p_i - p_i^\diamond)^2, \quad (2.16)$$

We refer to equation (2.16) as firm i 's *profit-loss function*. Its curvature with respect to the firm's log price is $|\pi_{pp,i}|$, the absolute value of the second derivative of profits evaluated at the initial equilibrium. A larger $|\pi_{pp,i}|$ makes pricing errors more costly and raises the value of information. Section 3 derives $|\pi_{pp,i}|$ for heterogeneous firms under non-CES

demand.

Information is more valuable when mispricing is more costly and when the full-information target responds more strongly to the unobserved shock. These two forces are governed by $|\pi_{pp,i}|$ and pass-through, respectively.

The optimal information-acquisition problem in (2.8) can therefore be rewritten as a linear-quadratic problem:

$$\kappa_i^* = \arg \min_{\kappa_i} \mathbb{E} \left[\frac{|\pi_{pp,i}|}{2} (p_i^\diamond - p_i^*)^2 \right] + \lambda f(\kappa_i), \quad (2.17)$$

subject to (2.13), (2.14), and the information-flow constraint (2.9).

The next section characterizes how $|\pi_{pp,i}|$ and ρ_i vary across firms and jointly determine their information choices.

3. The Equilibrium Size–Attention Distribution

This section characterizes the equilibrium size–attention distribution. We first decompose the curvature of the profit-loss function into firm scale and local demand conditions. We then use general properties of demand to identify the forces linking firm size to information acquisition. Finally, we specialize to Kimball demand and show how demand curvature determines which forces dominate.

3.1. Curvature of the Profit-Loss Function

Information is more valuable when the profit-loss function is more curved. The following lemma decomposes its curvature, $|\pi_{pp,i}|$, into three objects.

Lemma 1 (Profit-Loss Curvature). *The curvature of the profit-loss function, $|\pi_{pp,i}|$, is given by*

$$|\pi_{pp,i}| = s(z_i) [\epsilon_i - 1 + \eta_i], \quad (3.1)$$

where $s(z_i) = \frac{p_i Y_i}{p Y}$, and ϵ_i and η_i denote the elasticity and super-elasticity defined in (2.2) and (2.3), respectively.

Proof: see Appendix A.1.

Lemma 1 identifies three determinants of the curvature of the profit-loss function. First, $|\pi_{pp,i}|$ is proportional to the firm’s market share. A given percentage pricing error affects a larger sales base for a larger firm and therefore causes a larger profit loss. We refer to this force as the *size channel*. Because sales shares $s(z_i)$ increase in output shares Y_i/Y , either measure gives the same ranking of firm size.

Second, $|\pi_{pp,i}|$ is increasing in the local price elasticity of demand, ϵ_i . When demand is

more elastic, a small deviation from the optimal price induces a larger quantity response. Profits are therefore more sensitive to pricing errors.

Third, $|\pi_{pp,i}|$ is increasing in super-elasticity, η_i . Super-elasticity measures how the elasticity of demand changes as the firm's relative price or quantity changes; it therefore captures demand curvature beyond the constant-elasticity case. Even if two goods have the same elasticity at the optimum, they may differ in how elasticity changes when the firm moves away from its optimal price. When η_i is high, a deviation from the optimal price moves the firm more quickly into a region in which demand becomes more elastic, so consumers substitute away more strongly than under isoelastic demand. Mispricing therefore produces a larger profit loss.

The decomposition separates firm scale from the local shape of demand. Market share measures the sales exposed to a pricing error, while elasticity and super-elasticity determine the profit loss per unit of sales. The size–attention relationship therefore depends on how these three objects covary with firm size.

The expression also clarifies the CES benchmarks. Under CES demand, $\eta_i = 0$ and $\epsilon_i = \epsilon$ for all firms. With homogeneous firms,

$$|\pi_{pp,i}| = \epsilon - 1.$$

With heterogeneous firm sizes,

$$|\pi_{pp,i}| = s(z_i)(\epsilon - 1).$$

Thus, under CES demand, heterogeneity in the curvature of the profit-loss function arises only through the size channel.

3.2. Heterogeneous Incentives to Acquire Information

To characterize how attention varies with firm size, we must determine how the objects in Lemma 1 vary across firms. Three Marshallian laws of demand provide this mapping.

Marshallian Laws of Demand. We impose three monotonicity conditions on demand. Together, they link a firm's relative price to its market share, elasticity, markup, and pass-through.

Definition 1 (Marshallian Laws of Demand). *We say that the demand system satisfies the Marshallian laws of demand if the following three conditions hold for all i .*

First law:

$$\frac{\partial(Y_i/Y)}{\partial z_i} < 0. \quad (3.2)$$

Second law:

$$\frac{\partial \log \epsilon_i}{\partial \log z_i} > 0 \quad \Leftrightarrow \quad \frac{\partial \log \mu_i}{\partial \log z_i} < 0, \quad (3.3)$$

where $\mu_i \equiv \frac{\epsilon_i}{\epsilon_i - 1}$ denotes the desired markup.

Third law:

$$\frac{\partial}{\partial z_i} \left(\frac{\partial \log \mu_i}{\partial \log z_i} \right) > 0. \quad (3.4)$$

The first law states that relative demand is downward sloping, so firms with lower relative prices have higher relative output. Throughout, we focus on $\epsilon_i > 1$, a stronger restriction commonly imposed in the literature and already assumed above to ensure positive markups. Since $s(z_i) = z_i(Y_i/Y)$, this restriction implies

$$\frac{\partial s(z_i)}{\partial z_i} = \frac{Y_i}{Y}(1 - \epsilon_i) < 0.$$

Thus, firms with lower z_i are larger whether size is measured by relative output or by sales.

The second law states that demand elasticity increases with relative price. Equivalently, the desired markup decreases with relative price. Larger firms have lower z_i and therefore face lower demand elasticities and charge higher desired markups.

The third law governs markup responsiveness. Combined with the second law, it implies that the desired markup responds more strongly to relative-price changes for larger firms. When marginal cost changes, larger firms therefore adjust their markups more aggressively.

This property determines pass-through. Since $\rho_i = (1 - \partial \log \mu_i / \partial \log z_i)^{-1}$, the third law implies $\partial \rho_i / \partial z_i > 0$. Larger firms therefore have lower pass-through. Because their desired markups respond more strongly, they absorb a larger share of marginal-cost shocks through markup adjustment.

The laws give unambiguous mappings from firm size to elasticity and pass-through, but not to super-elasticity. Larger firms have more responsive desired markups but face

lower demand elasticities. These forces have opposing effects on super-elasticity.¹⁰ The following lemma summarizes these implications.

Lemma 2 (Laws of Demand and Firm Heterogeneity). *More productive firms have lower relative prices z_i and are therefore larger. Under the Marshallian laws of demand:*

1. *if the second law of demand holds, larger firms have lower demand elasticities ϵ_i ;*
2. *if the third law of demand holds, larger firms have lower pass-through ρ_i ;*
3. *the relationship between firm size and super-elasticity, $\eta_i = \frac{z_i \epsilon'_i}{\epsilon_i}$, is in general ambiguous.*

Proof. Relative price z_i is inversely related to firm size. The second law implies that ϵ_i increases with z_i , so it decreases with firm size. Since $\rho_i = (1 - \partial \log \mu_i / \partial \log z_i)^{-1}$, the third law implies that ρ_i also increases with z_i and thus decreases with firm size. Finally, $\eta_i = -(\partial \log \mu_i / \partial \log z_i)(\epsilon_i - 1)$. As z_i rises, the first factor falls while the second rises, leaving the relationship between η_i and firm size ambiguous. $\square$

Combining Lemmas 1 and 2 identifies four channels linking firm size to attention. Information is more valuable when the profit-loss function is more curved and when the full-information price target is more responsive to the aggregate shock. The first margin depends on market share, elasticity, and super-elasticity. The second depends on pass-through.

First, the *size channel* raises attention with firm size. A pricing error applies to a larger sales base and therefore causes a larger profit loss. Second, the *elasticity channel* lowers attention with firm size. Larger firms face lower demand elasticities, making profits less sensitive to pricing errors. Third, the *super-elasticity channel* is ambiguous because super-elasticity may rise or fall with firm size. Finally, the *pass-through channel* lowers attention with firm size. Lower pass-through makes the profit-maximizing price less responsive to the aggregate shock, reducing the value of tracking aggregate conditions.¹¹

The following proposition summarizes these channels.

Proposition 1 (Channels). *The following channels determine how information acquisition varies with firm size:*

1. *a positive size channel through market share $s(z_i)$;*

¹⁰Super-elasticity can be written as

$$\eta_i = \frac{z_i \epsilon'_i}{\epsilon_i} = -\frac{\partial \log \mu_i}{\partial \log z_i}(\epsilon_i - 1).$$

¹¹The response of firms' equilibrium prices to q is proportional to ρ_i .

2. a negative elasticity channel through demand elasticity ϵ_i ;
3. an ambiguous super-elasticity channel through η_i ;
4. a negative pass-through channel through ρ_i .

We refer to the elasticity, super-elasticity, and pass-through channels collectively as the *market-power channels* because they operate through local demand conditions that govern markups and pricing incentives. Because the size and market-power channels work in opposing directions, the general size–attention relationship is ambiguous. The next subsection specializes to Kimball demand to determine which force dominates.

3.3. Demand Curvature and Equilibrium Attention

To resolve the ambiguity in Proposition 1, we use the Kimball aggregator in [Klenow and Willis \(2016\)](#). We use this specification for tractability: its curvature parameter maps firm size into elasticity, super-elasticity, and pass-through. This structure allows us to determine whether the size channel or the market-power channels dominate.

3.3.1. Kimball Demand System

Demand System. The Kimball demand aggregator is

$$G(Y_i/Y) = 1 + (\epsilon - 1) \exp\left(\frac{1}{\eta}\right) \eta^{\frac{\epsilon}{\eta}-1} \left[\Gamma\left(\frac{\epsilon}{\eta}, \frac{1}{\eta}\right) - \Gamma\left(\frac{\epsilon}{\eta}, \frac{(Y_i/Y)^{\eta/\epsilon}}{\eta}\right) \right], \quad (3.5)$$

where $\Gamma(\cdot, \cdot)$ denotes the incomplete gamma function. The parameter ϵ is the elasticity of substitution for the average-sized firm in the steady state, evaluated at $\tilde{Y}_i/\tilde{Y} = 1$. The parameter η is the super-elasticity at the same point and governs the departure from CES. CES demand obtains when $\eta = 0$. A higher η makes demand elasticity more sensitive to relative prices. It therefore increases cross-sectional dispersion in elasticities and strengthens strategic complementarity in pricing.

Define the choke price as $\bar{P}_c \equiv \frac{\epsilon-1}{\epsilon} \exp\left(\frac{1}{\eta}\right) P$, above which demand is zero. The household's expenditure-allocation problem then yields

$$\frac{Y_i}{Y} = \left(-\eta \log\left(\frac{P_i}{\bar{P}_c}\right) \right)^{\frac{\epsilon}{\eta}}. \quad (3.6)$$

The firm's position on this demand curve determines both its size and its local demand conditions. Given the firm-size distribution, ϵ and η therefore determine the cross-sectional distributions of elasticity, super-elasticity, and pass-through.

Pricing Equation. The full-information price target retains the general form $p_i^\diamond = \rho_i q + (1 - \rho_i)p$. Under Kimball demand, pass-through is

$$\rho_i = \frac{\eta \log\left(\frac{\bar{p}_i}{\bar{p}_c}\right) + \epsilon}{\eta \log\left(\frac{\bar{p}_i}{\bar{p}_c}\right) + (\epsilon + \eta)}.$$

Pass-through increases with ϵ/η . Higher curvature therefore strengthens strategic complementarity and reduces the sensitivity of desired prices to marginal cost.

The aggregate price indices are defined as in the general case in (2.15), with explicit expressions for u_i and $\tilde{\epsilon}_i$.¹² When $\eta = 0$ and firms are homogeneous, the aggregate price indices coincide: $p = p^Y = \int p_i di$.

Curvature of the Profit-Loss Function. Lemma 1 gives the curvature of the profit-loss function as $|\pi_{pp,i}| = s(z_i)[\epsilon_i - 1 + \eta_i]$. Under the Kimball aggregator, elasticity and super-elasticity are

$$\epsilon_i = \epsilon \left(\frac{\bar{Y}_i}{\bar{Y}} \right)^{-\eta/\epsilon}, \quad \eta_i = \eta \left(\frac{\bar{Y}_i}{\bar{Y}} \right)^{-\eta/\epsilon}.$$

Their ratio is constant across firms:

$$\frac{\eta_i}{\epsilon_i} = \frac{\eta}{\epsilon}.$$

Thus, η/ϵ governs the relative importance of super-elasticity and elasticity at every point in the size distribution. Given the firm-size distribution and the parameters ϵ and η , these expressions determine ϵ_i , η_i , and the curvature of the profit-loss function for each firm.¹³

3.3.2. Who Acquires More Information and When

Kimball demand turns the general ambiguity in Proposition 1 into a comparison governed by η/ϵ . A higher ratio makes local demand conditions vary more strongly across firms of different sizes. It therefore strengthens the market-power channels relative to the size channel.

¹²Specifically, $u_i = \frac{\exp\left(\frac{1}{\eta} - \frac{1}{\eta} \left(\frac{\bar{Y}_i}{\bar{Y}}\right)^{\frac{\eta}{\epsilon}}\right) \frac{\bar{Y}_i}{\bar{Y}}}{\int \exp\left(\frac{1}{\eta} - \frac{1}{\eta} \left(\frac{\bar{Y}_i}{\bar{Y}}\right)^{\frac{\eta}{\epsilon}}\right) \frac{\bar{Y}_i}{\bar{Y}} di}$ and $\tilde{\epsilon}_i = \frac{\left(\frac{\bar{Y}_i}{\bar{Y}}\right)^{\frac{\epsilon-\eta}{\epsilon}} \exp\left(\frac{1}{\eta} - \frac{1}{\eta} \left(\frac{\bar{Y}_i}{\bar{Y}}\right)^{\frac{\eta}{\epsilon}}\right)}{\int \left(\frac{\bar{Y}_i}{\bar{Y}}\right)^{\frac{\epsilon-\eta}{\epsilon}} \exp\left(\frac{1}{\eta} - \frac{1}{\eta} \left(\frac{\bar{Y}_i}{\bar{Y}}\right)^{\frac{\eta}{\epsilon}}\right) di}.$

¹³The admissible parameter space requires $\epsilon_i > 1$ for every firm. Otherwise, desired markups are non-positive or undefined.

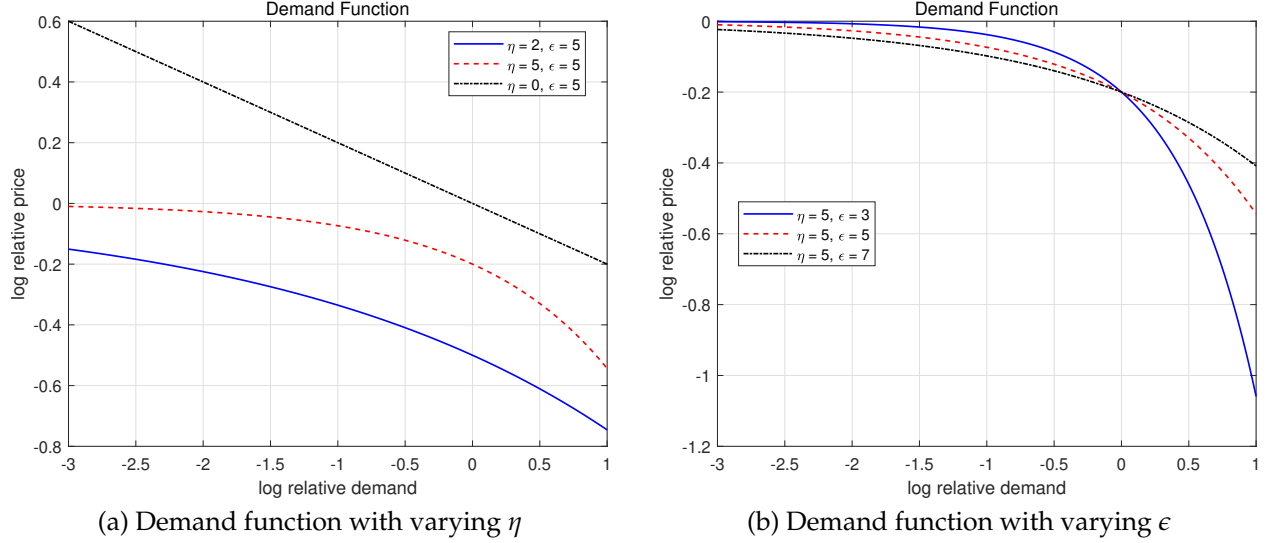


Figure 1: Kimball Curvature and Cross-Sectional Heterogeneity

Notes: The figure illustrates how the Kimball parameters alter the shape of demand. The ratio η/ϵ governs how strongly elasticity and super-elasticity vary across firms located at different points on the demand curve.

Lemma 3 (Demand Curvature and Cross-Sectional Heterogeneity). *Under non-CES Kimball demand, the ratio η/ϵ governs the strength of cross-sectional heterogeneity induced by demand curvature. For a given dispersion in firm size, a larger η/ϵ implies greater dispersion in elasticity and super-elasticity across firms.*

Proof: See Appendix A.2.

Figure 1 illustrates the lemma. Firms of different sizes occupy different points on the same demand curve, and the local slope at each point reflects the elasticity they face. Under CES demand ($\eta = 0$), elasticity is constant across firms. Once $\eta > 0$, elasticity varies with firm size. The left panel shows how a higher η increases this variation, while the right panel shows how ϵ also changes the dispersion in local slopes. Together, the panels illustrate why η/ϵ summarizes the strength of cross-sectional heterogeneity.

As η/ϵ rises, the market-power channels become stronger for larger firms. They first offset and eventually dominate the positive size channel. Proposition 2 characterizes the resulting three regimes.

Proposition 2 (Heterogeneous Information Acquisition). *The relationship between firm size and information acquisition depends on the magnitude of η relative to ϵ :*

1. If $\eta = 0$, that is, under CES demand, larger firms acquire more information.
2. If $0 < \eta < \epsilon$, the relationship between firm size and information acquisition is hump-shaped.
3. If $\eta \geq \epsilon$, larger firms acquire less information.

Proof: See Appendix A.3.

Under CES demand, elasticity and pass-through are constant across firms. The market-power channels therefore disappear, and the size channel makes attention increase with firm size. With moderate curvature, $0 < \eta < \epsilon$, the market-power channels overturn the size channel only among the largest firms. Attention is therefore hump-shaped in size.¹⁴ With sufficiently strong curvature, $\eta \geq \epsilon$, the market-power channels dominate throughout the size distribution, and attention declines with firm size.

Proposition 2 characterizes the equilibrium size–attention distribution as a function of demand curvature. The next section shows how this distribution shapes the effects of monetary policy.

4. Implications for Monetary Policy

This section shows how the joint size–attention distribution shapes monetary policy along two dimensions. First, it determines the heterogeneous effects across firms and which firms account for the aggregate real response. Second, it determines how heterogeneous price adjustment reallocates demand and changes allocative efficiency and measured TFP. Both implications depend on the size–attention regimes characterized in Proposition 2.

4.1. Cross-Sectional Incidence of Monetary Policy

How does monetary policy affect firms differently, and which firms account for its real aggregate effects? The answer depends on two margins. A *real-rigidity* margin operates through partial-equilibrium pass-through, ρ_i , which governs how strongly desired prices respond to aggregate nominal-demand shocks. A *nominal-rigidity* margin operates through information acquisition, κ_i , which governs how closely actual prices track desired prices. Together, these margins determine how much of a monetary shock is absorbed by prices and how much by quantities.

The real-rigidity margin is unambiguous. Larger firms have lower partial-equilibrium pass-through. Holding attention fixed, their prices therefore respond less to monetary shocks, shifting more of the adjustment to quantities. Through this margin alone, firms' output responses increase with firm size.

The nominal-rigidity margin can work in either direction. More attentive firms track aggregate conditions more closely and adjust prices more strongly. If larger firms acquire less information, this margin reinforces the real-rigidity margin, making their output re-

¹⁴In the moderate-curvature case, the proposition characterizes the size–attention schedule over the admissible size domain such that $\epsilon_i > 1$; the realized cross-section is hump-shaped whenever the support of firm sizes spans the unique turning point, as it does in our baseline calibration.

sponses larger. If they acquire more information, it works in the opposite direction.

The following proposition summarizes the two margins.

Proposition 3 (Heterogeneous Real Effects of Monetary Policy). *Under non-CES Kimball demand, aggregate nominal-demand shocks generate heterogeneous price and output responses across firms through two margins:*

1. *a real-rigidity margin, operating through heterogeneous partial-equilibrium pass-through ρ_i ;*
2. *a nominal-rigidity margin, operating through heterogeneous information choices κ_i .*

Moreover:

- (i) *through the real-rigidity margin, firms' output responses are increasing in firm size;*
- (ii) *through the nominal-rigidity margin, firms' output responses are increasing in firm size if $\eta \geq \epsilon$, and U-shaped in firm size if $0 < \eta < \epsilon$.*

Proof. Equation (3.6) implies that output responses move inversely with relative-price responses. The result then follows immediately from the preceding results for pass-through ρ_i and attention K_i . $\square$

Proposition 3 shows that whether output responses rise or fall with firm size depends jointly on pass-through and endogenous attention. To isolate the real-rigidity margin, consider a benchmark in which firms differ in pass-through but not in nominal rigidity.

Corollary 1 (No Heterogeneous Nominal Rigidity). *If firms differ in pass-through but not in nominal rigidity, then firms' output responses to monetary policy shocks are increasing in firm size.*

Because larger firms have lower pass-through, this benchmark predicts that their output responds more strongly to monetary policy. This prediction contrasts with evidence that small and young firms respond more strongly than large firms (Gertler and Gilchrist 1994, Cloyne et al. 2023, Gnewuch and Zhang 2025).

Endogenous information acquisition changes this benchmark. Under CES demand, attention increases with firm size, so the nominal-rigidity margin offsets the real-rigidity margin. Under moderate curvature, $0 < \eta < \epsilon$, attention is hump-shaped. The nominal-rigidity margin then offsets the real-rigidity margin among small and medium-sized firms but reinforces it among the largest firms. The overall size–output relationship is therefore generally nonmonotone. When $\eta \geq \epsilon$, both attention and pass-through decline with firm size, so the two pricing margins work in the same direction. In the quantitative

analysis, we show that, on average, firms' output responses decline with size, although they increase again among the largest firms.

4.2. Supply-Side Effects of Monetary Policy

Following [Baqae et al. \(2024\)](#), we define the supply-side effect of an aggregate nominal-demand shock as the impact response of measured TFP. Proposition 4 decomposes this response into a pass-through reallocation channel and an information-acquisition reallocation channel.

To state the result, for any nonnegative weights x_i with $\int_0^1 x_i di > 0$, define the x -weighted expectation and covariance as

$$\mathbb{E}_x[a_i] \equiv \frac{\int_0^1 a_i x_i di}{\int_0^1 x_i di}, \quad \text{Cov}_x(a_i, b_i) \equiv \mathbb{E}_x[a_i b_i] - \mathbb{E}_x[a_i] \mathbb{E}_x[b_i].$$

The subscripts u and uK refer to weights u_i and $u_i K_i$, respectively. We define the aggregate markup as the sales-weighted harmonic average $\bar{\mu} \equiv \mathbb{E}_u[\mu_i^{-1}]^{-1}$.

With this notation, the TFP response can be decomposed as follows.

Proposition 4 (Supply-Side Effects of Monetary Policy). *Fix a distribution of firms' Kalman gains $\{K_i\}_{i \in [0,1]}$. Following a unit aggregate nominal-demand shock, the induced response of TFP is*

$$\underbrace{\varsigma_\rho \text{Cov}_{uK}(\epsilon_i, \rho_i)}_{\text{pass-through reallocation channel}} + \underbrace{\varsigma_K \text{Cov}_u(\epsilon_i, K_i)}_{\text{information-acquisition reallocation channel}},$$

where $\varsigma_\rho \geq 0$ and $\varsigma_K \geq 0$ are given by

$$\varsigma_\rho \equiv \frac{\bar{\mu} \mathbb{E}_u[K_i] \mathbb{E}_u[1 - K_i]}{\mathbb{E}_u([K_i \rho_i + (1 - K_i)] \epsilon_i)}, \quad \varsigma_K \equiv \frac{\bar{\mu} \mathbb{E}_{uK}[\rho_i]}{\mathbb{E}_u([K_i \rho_i + (1 - K_i)] \epsilon_i)}.$$

Proof: See Appendix A.4.

Proposition 4 has a direct reallocation interpretation. An aggregate nominal-demand shock affects measured TFP when heterogeneous price adjustment changes relative prices and reallocates demand across firms. High-markup firms are inefficiently too small, so shifting demand toward them improves allocative efficiency and measured TFP; shifting demand toward low-markup firms has the opposite effect. In this model, heterogeneous price adjustment has two sources: pass-through and endogenous attention.

The first term captures reallocation generated by heterogeneous pass-through, as in [Baqae et al. \(2024\)](#). When firms with higher demand elasticities also have higher pass-

through, they raise their prices more following an expansionary nominal-demand shock. Expenditure then shifts away from these lower-markup firms and toward higher-markup firms, raising TFP.

The second term captures reallocation generated by endogenous information acquisition. Firms with higher Kalman gains observe aggregate conditions more precisely and adjust prices more strongly after the shock. The allocation of attention therefore determines which firms move their relative prices the most. If more attentive firms also face higher demand elasticities, these lower-markup firms raise prices relatively more, so expenditure shifts toward higher-markup firms and measured TFP rises. If instead they face lower demand elasticities, these high-markup firms raise prices relatively more, shifting expenditure away from them and reducing measured TFP.

Because ζ_ρ and ζ_K are nonnegative, the two reallocation terms govern the sign of the TFP response. Under Kimball demand, curvature shapes both terms through its effects on pass-through and attention.

When η is large, larger firms have lower demand elasticities, lower pass-through, and lower attention. The cross-sectional patterns in ρ_i and K_i are therefore aligned with that in ϵ_i , so both reallocation forces raise TFP. An expansionary monetary shock then improves allocative efficiency by shifting demand toward higher-markup firms and generates a positive supply-side effect through reallocation.

When η is small, the pass-through channel remains positive, but the information-acquisition channel can work in the opposite direction. Larger firms then acquire more information, making high-markup firms adjust prices more strongly. This force shifts expenditure away from high-markup firms, offsetting the improvement in allocative efficiency generated by heterogeneous pass-through and potentially overturning it. Section 5 quantifies the relative importance of these two forces.

5. Quantitative and Empirical Evidence

This section disciplines the joint size–attention distribution and quantifies its monetary-policy implications. First, we calibrate the baseline and comparison economies and quantify their cross-sectional and supply-side effects. Second, we use firms’ Form 10-K filings to test the scale-based prior and determine which size–attention regime better matches the data.

5.1. Quantitative Analysis

Calibration. The calibration has two components. The first governs the attention cost function. Following [Paciello and Wiederholt \(2014\)](#), we assume that the cost of informa-

tion acquisition is exponential, $f(\kappa_i) = 2^{2\kappa_i}$. We choose the cost parameter λ so that the *average* Kalman gain in the endogenous-information equilibrium is 0.5, consistent with the empirical evidence in [Coibion and Gorodnichenko \(2015\)](#).

The second component governs the demand system and the firm-size distribution. The Kimball parameters ϵ and η determine the cross-section of elasticities and pass-through, while the size distribution determines the strength of the size channel. We discipline the cross-sectional distribution of U.S. establishment size using the U.S. Census Bureau’s Business Dynamics Statistics (BDS). In particular, we target the establishment-level employment distribution and map it into the model-implied distribution of sales shares. Figure [A.1](#) plots the resulting cumulative sales share.

We set $\epsilon = 5$ and $\eta = 0.5\epsilon$ in the baseline, consistent with the range of micro estimates reported by [Beck and Levin \(2020\)](#) and with the relatively modest curvature favored by [Bergin and Feenstra \(2001\)](#).

Together, the calibrated firm-size distribution, demand curvature, and average Kalman gain determine the equilibrium size–attention distribution.

Heterogeneous Effects of Monetary Policy. Figure [2](#) reports the heterogeneous effects of monetary policy across firms. The solid (red) lines correspond to endogenous information acquisition. The dashed (blue) lines report a counterfactual in which all Kalman gains are fixed at 0.5. This counterfactual preserves average attention while eliminating cross-sectional heterogeneity in nominal rigidity.

Endogenous attention reverses the average size ranking generated by heterogeneous pass-through alone. Attention is hump-shaped, price responses increase with firm size over most of the distribution, and output responses decline through much of the distribution before rising among the largest firms. When Kalman gains are fixed, price responses instead decline with size and output responses increase. Endogenous heterogeneity in nominal rigidity therefore shifts the real effects of monetary policy toward smaller firms.

The bottom-right panel of Figure [2](#) shows that the bottom 50% of the size distribution accounts for about 27% of the aggregate output response, compared with about 19% of aggregate sales (Figure [A.1](#)). Their share of the aggregate output response is therefore about 40% larger than their sales share. In the fixed-attention counterfactual, the cumulative contribution shifts toward larger firms.

For comparison, we set $\eta = \epsilon$. Stronger curvature and strategic complementarities are common in the macro literature ([Klenow and Willis, 2016](#); [Eichenbaum and Fisher, 2007](#); [Chari et al., 2000](#); [Smets and Wouters, 2007](#)). Figure [A.2](#) reports the results. Under this calibration, attention declines with firm size. Small firms raise prices so aggressively that their output falls after an expansionary aggregate-demand shock, whereas large firms’

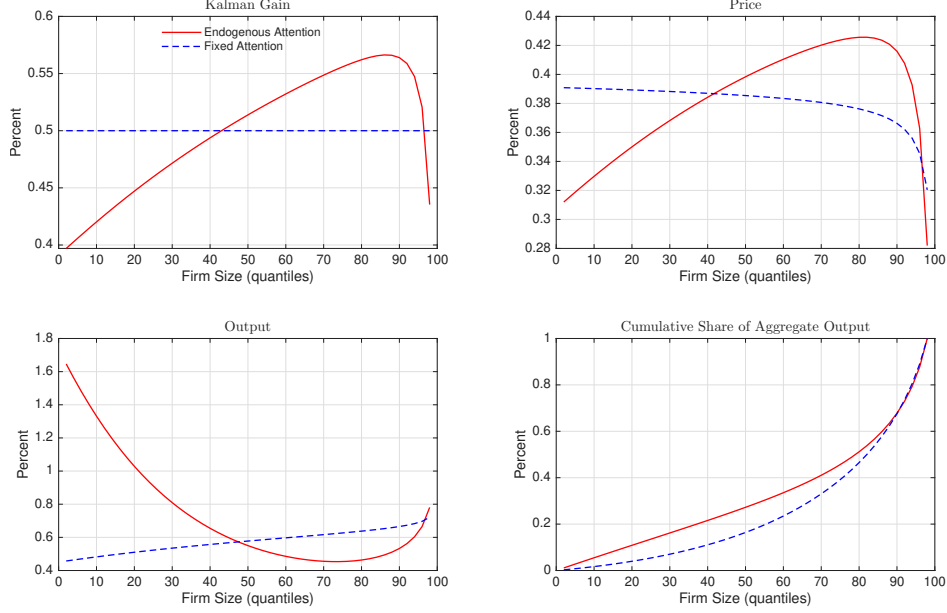


Figure 2: Heterogeneous Effects of Monetary Policy

Notes: Solid (red) lines report the baseline with endogenous attention. Dashed (blue) lines report the fixed-attention counterfactual, which preserves average attention while eliminating its cross-sectional heterogeneity. Firm size is indexed by quantile.

output rises.¹⁵ Firms in the bottom half of the size distribution account for almost none of the aggregate output response.

Supply-Side Effects. Table 1 reports the impact response of measured TFP to a 1% increase in nominal demand. It presents the lower-curvature baseline and the higher-curvature comparison, together with one counterfactual for each.

Table 1: Supply-Side Effects of Monetary Policy

Specification	Calibration	Impact response (%)
Baseline: Lower Curvature	$\eta = 0.5\epsilon$	0.001
<i>Continued-rise counterfactual</i>	$\eta = 0.5\epsilon$, post-peak K_i keeps rising	-0.013
For Comparison: Higher Curvature	$\eta = \epsilon$	0.150
<i>Fixed-attention counterfactual</i>	$\eta = \epsilon$, $K_i = 0.5 \forall i$	0.083

Notes: The table reports the impact response of measured TFP to a 1% nominal-demand expansion. The lower-curvature baseline and higher-curvature comparison feature endogenous attention. The continued-rise counterfactual replaces the hump-shaped attention profile with one that continues to rise after the peak. The fixed-attention counterfactual holds attention constant across firms. Negative values indicate a decline in measured TFP.

The supply-side effect is quantitatively small in the lower-curvature baseline. Mea-

¹⁵Although aggregate demand increases, small firms may contract because they adjust prices more aggressively than large firms. The induced rise in their relative prices shifts demand away from them, and this expenditure-switching effect can dominate the direct expansionary effect of monetary policy.

sured TFP rises by only 0.001% on impact. When we suppress the downturn in attention among the largest firms and force Kalman gains to continue rising with size,¹⁶ TFP instead falls by 0.013%. The shape of the attention profile therefore changes the sign of the response, although its magnitude remains small. Under higher curvature, the TFP response is larger: it rises by 0.150% with endogenous attention and by 0.083% when attention is fixed across firms. Endogenous heterogeneity therefore amplifies the positive supply-side effect in the comparison economy.

5.2. Empirical Evidence

A natural prior is that larger firms acquire more information. Because a pricing error applies to a larger sales base, attention should rise monotonically with firm size. Our model shows why this scale logic can fail. Lower demand elasticities and lower desired pass-through reduce the value of information for larger firms. With moderate demand curvature, these forces dominate only among the largest firms and generate a hump-shaped size–attention relationship. We examine this prediction using firms’ annual Form 10-K filings.

Measurement. We measure firm-level attention from topic-specific keyword frequencies in annual Form 10-K filings, following [Song and Stern \(2025\)](#) and [Flynn and Sastry \(2024\)](#). The measure records the share of each filing devoted to a given macroeconomic topic. We construct the word counts using the textual-analysis procedure of [Loughran and McDonald \(2011\)](#), which removes stop words and standardizes the text using a pre-defined dictionary. The macroeconomic topics and keyword classifications follow [Song and Stern \(2025\)](#).

Our sample covers annual Form 10-K filings from 1994 through 2020. We match each filing to Compustat annual data using the firm identifier (gvkey) and the fiscal year implied by the filing’s fiscal-year-end date. For the figures, we require positive total word counts and positive total assets so that attention density and log assets are defined. These restrictions leave 130,373 filings from 16,096 firms. For the regressions, we further require a successful Compustat match and nonmissing book leverage, firm age, log filing length, and two-digit industry classification. The resulting sample contains 127,027 filings from 15,815 firms. Stacking the nine topics yields 1,143,243 filing-topic observations.

¹⁶We construct this counterfactual by extrapolating the rising segment of the Kalman-gain profile to firms beyond its baseline peak.

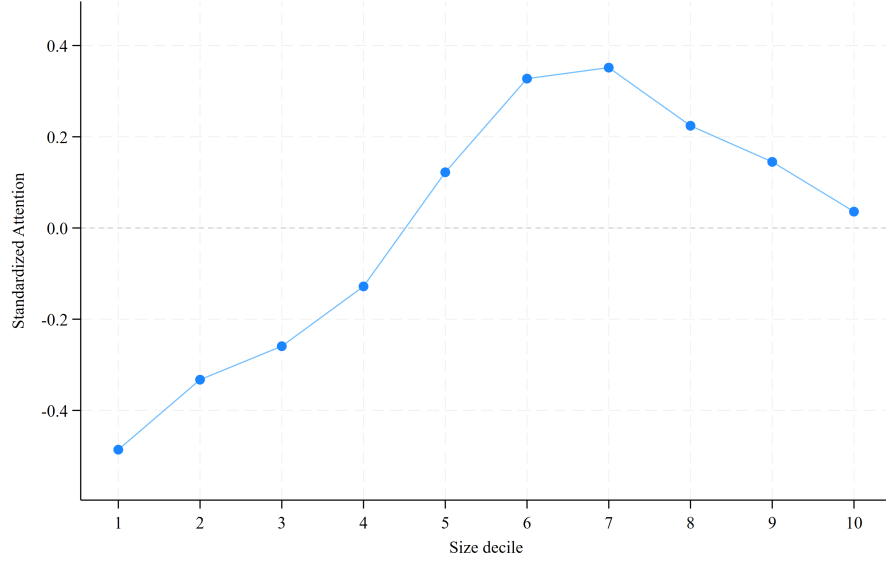


Figure 3: Firm Size and Attention to General Economic Conditions

Notes: Each point reports mean standardized attention to general economic conditions within an annual asset decile. Firms are assigned to deciles based on log total assets within each filing calendar year. Attention is measured as the number of topic-specific words in a firm’s annual Form 10-K filing divided by the filing’s total number of words and standardized within the sample for general economic conditions.

For each firm i , year t , and topic j , attention is defined as

$$\hat{\kappa}_{i,t}^j = \frac{\text{Total Topic Words}_{i,t}^j}{\text{Total Words}_{i,t}}. \quad (5.1)$$

Dividing by total words adjusts for differences in filing length. We standardize $\hat{\kappa}_{i,t}^j$ so that attention is expressed in standard-deviation units. Among the nine topics, “general economic conditions” is the closest empirical counterpart to the aggregate nominal environment in the model. The remaining topics show whether the size–attention relationship extends beyond this measure.

Unconditional Evidence. Figure 3 plots average attention to general economic conditions across annual asset deciles. Attention rises from the first through the seventh decile and then declines among the largest firms. Figure A.3 reports the relationship across the nine topics: general economic conditions, output, employment, consumption, investment, FOMC, housing, inflation, and oil. With the exception of oil-related attention, the size–attention relationship is broadly hump-shaped across topics.¹⁷

¹⁷Oil-related attention also emerges as an outlier in the separate empirical analysis of [Ou et al. \(2026\)](#). Taken together, these findings suggest that oil-related attention may be a special case that warrants further study. Identifying the source of this cross-topic heterogeneity is beyond the scope of this paper.

Regression Evidence. The unconditional pattern could reflect differences in composition across industries and years. We therefore ask whether the hump shape remains within industry-year and after controlling for firm characteristics. We stack the nine topic-specific observations for each firm-year and estimate

$$\tilde{\kappa}_{i,t}^j = \alpha + \beta_1 \text{Size}_{i,t} + \beta_2 \text{Size}_{i,t}^2 + \gamma' \mathbf{X}_{i,t} + FE + \varepsilon_{i,t}^j. \quad (5.2)$$

Here, $\tilde{\kappa}_{i,t}^j$ is standardized attention for firm i , topic j , and year t . The vector $\mathbf{X}_{i,t}$ includes book leverage, firm age, and log filing length. In Columns (1)–(3), $\text{Size}_{i,t}$ is the decile rank of log total assets computed within each filing calendar year. Columns (4) and (5) instead use 50 annual size quantiles. All specifications absorb industry-by-fiscal-year fixed effects and therefore compare firms within the same industry and fiscal year. The model predicts that attention first rises and then falls with firm size. In the regression, this pattern requires $\beta_1 > 0$, $\beta_2 < 0$, and an implied turning point within the observed size distribution. The turning point is the size rank at which the fitted slope becomes zero and is given by $-\hat{\beta}_1 / (2\hat{\beta}_2)$.

Table 2 reports the results. Columns (1)–(3) progressively add topic and industry-by-topic fixed effects. Column (4) replaces size deciles with 50 annual size quantiles, while Column (5) focuses on general economic conditions. Across all five specifications, the linear coefficient is positive and the quadratic coefficient is negative, with both statistically significant at the 1% level. The implied turning points are 6.87 among the 10 size deciles, 32.44 in the pooled 50-quantile specification, and 30.33 for general economic conditions. The conditional evidence therefore reproduces the hump shape in Figures 3 and A.3.

Together, the evidence rejects the scale-only prior and supports the hump-shaped size–attention distribution generated by the lower-curvature regime.

6. Extensions and Robustness

How sensitive are the equilibrium size–attention distribution and its monetary-policy implications to the simplifying assumptions of the baseline model? We consider five extensions. First, we allow for alternative household preferences and persistent aggregate shocks. Second, we introduce two forces that may give large firms an additional advantage in acquiring information: the reuse of a common signal across product lines and lower information costs. Finally, we consider firms to acquire information about idiosyncratic supply shocks. These extensions preserve the benefit–cost logic of the baseline while clarifying how additional aggregate and firm-level forces shift the allocation of attention.

Table 2: Firm Size and Standardized Attention to Macroeconomic Conditions

	(1)	(2)	(3)	(4)	(5)
<i>Firm characteristics</i>					
Size	0.0546*** (0.0133)	0.0546*** (0.0133)	0.0546*** (0.0133)	0.0103*** (0.0025)	0.0506*** (0.0154)
Size ²	-0.0040*** (0.0010)	-0.0040*** (0.0010)	-0.0040*** (0.0010)	-0.0002*** (0.0000)	-0.0008*** (0.0003)
Turning point	6.87 (of 10)	6.87 (of 10)	6.87 (of 10)	32.44 (of 50)	30.33 (of 50)
<i>Fixed effects</i>					
Industry × Fiscal-year FE	Yes	Yes	Yes	Yes	Yes
Topic FE	No	Yes	N.A.	N.A.	N.A.
Industry × Topic FE	No	No	Yes	Yes	N.A.
Observations	1,143,243	1,143,243	1,143,243	1,143,243	127,026
R ²	0.0146	0.1313	0.1625	0.1626	0.3044

Notes: Columns (1)–(4) pool the nine topics and standardize attention across all nonmissing firm-year-topic observations. Columns (1)–(3) use annual size deciles. Column (4) retains the fixed effects in Column (3) but uses 50 annual size quantiles. Column (5) focuses on general economic conditions, standardizes attention within this sample, and uses 50 annual size quantiles. All specifications absorb industry-by-fiscal-year fixed effects and control for book leverage, Compustat-history age, and log filing length. Topic and industry-by-topic fixed effects are included as indicated. Turning points are calculated as $-\hat{\beta}_1 / (2\hat{\beta}_2)$ using unrounded coefficient estimates; the denominator beneath each estimate reports the total number of size groups. Robust standard errors are in parentheses and are clustered by fiscal year and industry.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

6.1. Alternative Preferences

Does the result depend on the baseline preference specification? The assumption $U(Y, L) = \log Y - L$ removes the aggregate strategic complementarity emphasized by Adam (2007). It therefore isolates the firm-level strategic complementarity generated by demand curvature. We relax this assumption by considering CRRA preferences,

$$U(Y, L) = \frac{Y^{1-\sigma}}{1-\sigma} - \frac{L^{1+\phi}}{1+\phi}.$$

Figure A.4 reports the results under the baseline calibration. The size–attention relationship remains hump-shaped, and smaller firms continue to account for a disproportionate share of the real response to monetary policy. Firms below the median size account for about 42% of the aggregate output response, compared with about 27% under the baseline log-linear calibration. Allowing for aggregate strategic complementarity therefore preserves the main mechanism and strengthens the quantitative role of smaller

firms.

6.2. Persistent Aggregate Shocks

Does persistence change the allocation of attention across firms? Suppose aggregate nominal demand follows

$$q_t = \rho_q q_{t-1} + \varepsilon_t^q,$$

and consider separately whether firms observe lagged aggregate conditions.

Observed lagged shocks. If firms observe q_{t-1} when setting prices, the only unknown state is the innovation ε_t^q . Firm i can therefore acquire a scalar signal

$$s_{i,t} = \varepsilon_t^q + e_{i,t},$$

about the current innovation, or equivalently about q_t net of its known component $\rho_q q_{t-1}$. Replacing the i.i.d. shock q with ε_t^q leaves the firm's information problem unchanged. The profit-loss curvature $|\pi_{pp,i}|$ and pass-through ρ_i therefore continue to govern the cross-sectional incentives to acquire information. Persistence does not alter the theoretical or quantitative results when lagged aggregate conditions are public, an assumption commonly used in the incomplete-information literature ([Angeletos and La'O 2009](#), [Lorenzoni 2010](#); [Ou et al. 2021](#)).

Unobserved lagged shocks. When lagged shocks are not observed, firms must infer the history of nominal-demand innovations. The payoff-relevant state is then multidimensional, and the model no longer has a closed-form solution. We solve a quantitative approximation by truncating the shock history at horizon L . Following [Afrouzi and Yang \(2021\)](#), each firm acquires a single signal about the payoff-relevant linear combination of current and lagged innovations. We solve the resulting dynamic rational-inattention problem and iterate on firms' desired-price loadings until the implied aggregate price dynamics are consistent with those loadings. [Online Appendix B.1](#) provides the state-space representation and solution algorithm.

Figure [A.5](#) reports the results under the baseline calibration with $\rho_q = 0.5$. Attention remains hump-shaped in firm size, and monetary policy continues to have larger real effects on smaller firms over most of the size distribution. The i.i.d. assumption therefore simplifies the analysis without driving the paper's main quantitative results.

6.3. Multi-Product Firms

Operating multiple product lines creates an additional scale advantage in information acquisition. A common signal about aggregate conditions can be reused across a firm's

products, so its value rises with product scope. We formalize this force by allowing headquarters to acquire one signal that is shared with decentralized operating units.

Let firm i operate a mass $M_i = \mathcal{M}(a_i)$ of ex ante symmetric operating units. We assume $\mathcal{M}'(a_i) \geq 0$, so more productive firms operate more units. Headquarters acquires one common signal about the aggregate nominal shock, and each unit manager sets prices using that signal. Under symmetry within the firm, each unit's desired price remains

$$p_{ij}^\diamond = \rho_i q + (1 - \rho_i) p,$$

so the unit-level pricing problem is unchanged. The extension isolates the value of reusing information across product lines. Online Appendix B.2 presents the full setup.

Let $\omega_i \equiv \frac{p_{ij} Y_{ij}}{p Y}$ denote the sales share of a representative operating unit, and let

$$\Omega_i \equiv M_i \omega_i$$

denote total firm sales. Then firm i 's quadratic loss from a common pricing mistake is

$$\frac{|\Pi_{pp,i}|}{2} (p_i - p_i^\diamond)^2, \quad |\Pi_{pp,i}| = \Omega_i [\epsilon_i - 1 + \eta_i].$$

Because the same pricing error applies to every unit, the size term now loads on total firm sales Ω_i rather than representative-unit sales ω_i . The four channels in Proposition 1 remain unchanged, but information reuse strengthens the positive size channel.

To characterize this additional force, suppose product scope scales isoelastically with representative-unit sales:

$$M_i = \bar{M} \omega_i^\chi, \quad \chi \geq 0,$$

where $\bar{M} > 0$ is common across firms. Total firm sales then satisfy

$$\Omega_i = \bar{M} \omega_i^{1+\chi}.$$

The parameter χ is the elasticity of product scope with respect to representative-unit sales:

$$\frac{\partial \log M_i}{\partial \log \omega_i} = \chi.$$

A higher χ means that product scope expands more strongly with firm size, increasing the revenue base over which the common signal can be reused.

Proposition 5 (Heterogeneous Information Acquisition with Multi-Product Firms). *Con-*

sider the shared-information extension with firm scope satisfying $M_i = \bar{M}\omega_i^\chi$, where $\chi \geq 0$. Then:

1. if $\eta = 0$, larger firms acquire more information;
2. if $0 < \eta < (1 + \chi)\epsilon$, the relationship between firm size and information acquisition is hump-shaped;
3. if $\eta \geq (1 + \chi)\epsilon$, larger firms acquire less information.

These results hold whether firm size is measured by representative-unit sales ω_i or by total firm sales Ω_i .

Proof: See Appendix A.5.

When $\chi = 0$, Proposition 5 reduces to the baseline result. As χ rises, the cutoff between the hump-shaped and decreasing regimes shifts from $\eta = \epsilon$ to $\eta = (1 + \chi)\epsilon$. Information reuse therefore expands the parameter region in which attention is hump-shaped. Stronger demand curvature is required for attention to decline throughout the firm-size distribution.

6.4. Heterogeneous Information Costs

Information costs may also vary systematically with firm size. Larger firms may have better data infrastructure and forecasting capacity or may spread information-acquisition costs across a broader organization. Lower costs would give them an additional advantage in acquiring information. We capture this force by allowing the attention cost to take the form

$$\lambda_i f(\kappa_i),$$

where the cost shifter λ_i varies across firms.

The information-acquisition problem becomes

$$\kappa_i^* = \arg \min_{\kappa_i} \mathbb{E} \left[\frac{|\pi_{pp,i}|}{2} (p_i^\diamond - p_i^*)^2 \right] + \lambda_i f(\kappa_i).$$

The benefit side is unchanged: $|\pi_{pp,i}|$ and ρ_i continue to determine the value of information. Heterogeneous costs add a separate cost-side channel through λ_i .

To derive a tractable analog of Proposition 2, we impose an isoelastic relationship between information costs and the baseline sales share, $s(z_i) = P_i Y_i / (PY)$:

$$\lambda_i = \bar{\lambda} s(z_i)^{-\psi}, \quad \psi \geq 0,$$

where $\bar{\lambda} > 0$ is common across firms. The parameter ψ governs how information costs covary with firm size and can be summarized by

$$-\frac{\partial \log \lambda_i}{\partial \log s(z_i)} = \psi.$$

A higher ψ makes information costs decline more rapidly with size and therefore strengthens the positive size force.

Proposition 6 (Heterogeneous Information Acquisition with Heterogeneous Information Costs). *Suppose marginal information costs satisfy $\lambda_i = \bar{\lambda}s(z_i)^{-\psi}$, with $\psi \geq 0$. Then:*

1. *if $\eta = 0$, larger firms acquire more information;*
2. *if $0 < \eta < (1 + \psi)\epsilon$, the relationship between firm size and information acquisition is hump-shaped;*
3. *if $\eta \geq (1 + \psi)\epsilon$, larger firms acquire less information.*

Proof: See Appendix A.6.

When $\psi = 0$, Proposition 6 reduces to the baseline result. As ψ rises, the cutoff between the hump-shaped and decreasing regimes shifts from $\eta = \epsilon$ to $\eta = (1 + \psi)\epsilon$. Lower information costs for larger firms therefore expand the parameter region in which attention is hump-shaped. As in the multi-product extension, stronger demand curvature is required for attention to decline throughout the size distribution.

Quantitative Assessment We next assess whether the common-cost assumption matters quantitatively. Rather than impose the isoelastic specification above, we choose a firm-specific cost vector $\{\lambda_i\}_i$ to match the empirical cross section of standardized attention documented in Subsection 5.2. For each firm-size quantile, we choose λ_i so that the model-implied measure

$$\frac{\kappa_i - \bar{\kappa}}{\text{sd}(\kappa_i)},$$

matches its empirical counterpart.¹⁸ We also impose an average Kalman gain of 0.5, preserving the aggregate level of attention in the baseline. The homogeneous-cost comparison recalibrates a common λ to deliver the same average Kalman gain.

Figure 4 shows that allowing information costs to vary across firms preserves the main cross-sectional patterns. Relative to the homogeneous-cost comparison, the aggregate output response rises from 0.6200 to 0.6229, while aggregate price adjustment falls from

¹⁸The empirical target is the 50-quantile counterpart of Figure 3.

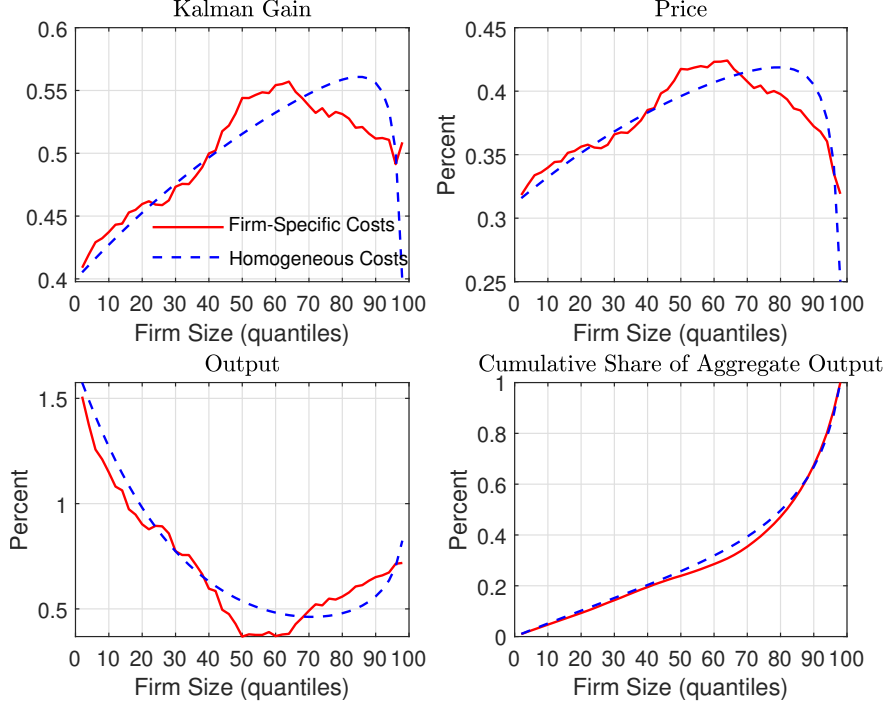


Figure 4: Heterogeneous Information Costs

Notes: Solid (red) lines report the model with firm-specific information costs calibrated to match the empirical cross section of standardized attention. Dashed (blue) lines report the homogeneous-cost comparison, which preserves the average Kalman gain of 0.5. Firm size is indexed by quantile.

0.3800 to 0.3771. The implied TFP response is also slightly larger. These differences are small. The cross-sectional transmission results are therefore robust to flexible firm-level information costs.

6.5. Information Acquisition with Aggregate and Idiosyncratic Shocks

The baseline analysis isolates attention to aggregate nominal demand. We now show that the qualitative size–attention relationship is unchanged when firms also face idiosyncratic productivity shocks. Firm i 's productivity consists of its publicly known permanent type a_i and an unobserved component v_i , where $v_i \sim N(0, \sigma_v^2)$ is independent of the aggregate shock q . We assume that σ_v^2 is common across firms and that the idiosyncratic shocks wash out in the aggregate.

The firm's full-information profit-maximizing price is

$$p_i^\diamond = \rho_i(q - v_i) + (1 - \rho_i)p, \quad (6.1)$$

so both shocks affect the firm's pricing decision through the scalar target $p_i^\diamond$. Under the quadratic approximation, the firm's pricing loss depends on the shocks only through this target. It is therefore optimal for a rationally inattentive firm to direct its information

acquisition toward $p_i^\diamond$; equivalently, its optimal information structure can be represented by a single noisy signal of the full-information target price. Because $p = \varphi q$, the prior variance of this target is

$$\text{Var}(p_i^\diamond) = \sigma_q^2 [\varphi + (1 - \varphi)\rho_i]^2 + \sigma_v^2 \rho_i^2.$$

Thus, the value of information reflects the firm's exposure to both aggregate and idiosyncratic uncertainty. The following proposition shows that adding the second source of uncertainty preserves the three size–attention regimes.

Proposition 7 (Aggregate and Idiosyncratic Shocks: Information Acquisition). *When firms optimally acquire information about $p_i^\diamond$ in the presence of both shocks, the relationship between firm size and information acquisition depends on the magnitude of η relative to ϵ :*

1. *if $\eta = 0$, larger firms acquire more information;*
2. *if $0 < \eta < \epsilon$, the relationship between firm size and information acquisition is hump-shaped;*
3. *if $\eta \geq \epsilon$, larger firms acquire less information.*

Proof: See Appendix [A.7](#).

Proposition 7 establishes that the baseline qualitative relationship between firm size and attention is robust to the inclusion of idiosyncratic uncertainty. Quantitatively, the location of the peak of the hump-shaped attention profile under lower curvature depends on the relative variances of the aggregate and idiosyncratic shocks.

7. Conclusion

A natural prior is that larger firms acquire more information because a pricing error affects a larger sales base. Our results show why this prior can fail. Firm scale raises the value of information, but local demand conditions work in the opposite direction. Larger firms tend to face lower demand elasticities, which weaken the profit consequences of mispricing, and lower desired pass-through, which makes their target prices less responsive to nominal shocks. Demand curvature determines which force dominates. Attention increases with firm size under CES demand, is hump-shaped under moderate curvature, and decreases under sufficiently strong curvature. The resulting joint distribution of firm size and attention is therefore an equilibrium outcome governed by demand curvature.

Using Form 10-K filings, we document a hump-shaped relationship between firm size and attention to macroeconomic conditions: attention rises through the middle of the size

distribution and declines among the largest firms. This pattern rejects the scale-only prior and supports the offsetting market-power channels that we emphasize.

This joint distribution matters for both aggregate transmission and allocative efficiency. For aggregate transmission, endogenous attention changes which firms drive the real effects of monetary policy. In the baseline, attention is hump-shaped in firm size, small and medium-sized firms account for a disproportionate share of the aggregate real response, and removing heterogeneity in attention reverses the cross-sectional ranking of output responses. For allocative efficiency, the same allocation of attention shapes the supply-side effects of monetary policy by changing the dispersion of relative prices and markups. An expansionary nominal shock mildly raises measured TFP in the baseline. When attention is instead forced to continue rising among the largest firms, the same shock lowers measured TFP.

More broadly, the paper provides a tractable framework for rational inattention with heterogeneous firms. By linking information choices to firm size, local demand conditions, and pass-through, it makes the size–attention distribution an equilibrium object. Monetary transmission depends on which firms are informed and how their market shares and pricing incentives differ. The framework also opens questions about optimal policy. With firm-specific employment subsidies that eliminate markup distortions, the divine coincidence holds: full price stabilization implements the first-best allocation. A natural next step is to study how the endogenous size–attention distribution shapes optimal policy when fiscal instruments are incomplete or additional shocks break this coincidence.

The framework also guides further empirical work. Studies such as [Coibion et al. \(2018\)](#), [Asriyan and Kohlhas \(2025\)](#), and [Song and Stern \(2025\)](#) document heterogeneous information frictions across firms. We complement this evidence by documenting a hump-shaped relationship between firm size and attention to macroeconomic conditions. Our theoretical analysis links firms’ information choices to the demand conditions that govern their pricing incentives. Examining information acquisition and pricing jointly would help distinguish the size effect from the offsetting market-power forces. Testing these mechanisms requires linked firm-level measures of information acquisition, product prices, and balance-sheet characteristics. Future research that collects and links such data could assess how market power shapes information choices and how those choices translate into heterogeneous price responses to monetary policy.

A. Proofs of Main Results

This appendix collects the proofs of the formal results not proved in the main text: Lemmas 1 and 3 and Propositions 2, 4, 5, 6, and 7.

A.1. Proof of Lemma 1

Proof. Let $c_i \equiv W/(e^{a_i}P)$ denote real marginal cost. Since $s(z_i) = z_i Y_i/Y$, normalized profits can be written as

$$\pi_i \equiv \frac{\Pi_i}{PY} = s(z_i) \left(1 - \frac{c_i}{z_i}\right).$$

Define the log-relative-price derivative operator $D \equiv z_i \partial / \partial z_i$. Holding P fixed, D is also the derivative with respect to the firm's log price p_i . The definition of elasticity implies

$$D \log s(z_i) = 1 - \epsilon_i, \quad Ds(z_i) = s(z_i)(1 - \epsilon_i).$$

Using $D(c_i/z_i) = -c_i/z_i$, we obtain

$$D\pi_i = s(z_i)H_i, \quad H_i \equiv 1 - \epsilon_i + \frac{c_i}{z_i}\epsilon_i.$$

At the optimum, $H_i = 0$, which gives $c_i/z_i = (\epsilon_i - 1)/\epsilon_i$. Therefore,

$$D^2\pi_i = (Ds)H_i + s(z_i)DH_i = s(z_i)DH_i.$$

By the definition of super-elasticity, $D\epsilon_i = \eta_i\epsilon_i$. Consequently, evaluating DH_i at the optimum gives

$$\begin{aligned} \pi_{pp,i} &= D^2\pi_i = s(z_i) [-\eta_i\epsilon_i + (\epsilon_i - 1)(\eta_i - 1)] \\ &= -s(z_i)(\epsilon_i - 1 + \eta_i). \end{aligned}$$

Taking absolute values proves the result. □

A.2. Proof of Lemma 3

Proof. Measure cross-sectional dispersion by variance. Let $y_i \equiv Y_i/Y$ and $\alpha \equiv \eta/\epsilon > 0$. Under Kimball demand,

$$\log \epsilon_i = \log \epsilon - \alpha \log y_i,$$

and hence

$$\text{Var}_i(\log \epsilon_i) = \alpha^2 \text{Var}_i(\log y_i),$$

which is strictly increasing in α for any non-degenerate firm-size distribution. Similarly, $\log \eta_i = \log \eta - \alpha \log y_i$, so the same logic gives

$$\text{Var}_i(\log \eta_i) = \alpha^2 \text{Var}_i(\log y_i).$$

□

A.3. Proof of Proposition 2

Proof. Write the profit-maximizing price as $p_i^\diamond = \rho_i(q - p) + p$. In an interior information-acquisition equilibrium, $p = \varphi q \leq q$, where $\varphi \in (0, 1]$. Suppress the firm subscript and let $y \equiv Y_i/Y$ and $s \equiv y^{\eta/\epsilon}$. The regularity condition $\epsilon_i > 1$ implies $s \leq \epsilon$. Equation (2.16) then shows that attention is increasing in

$$F(y, \eta) = |\pi_{pp}(y, \eta)| [\rho_i(1 - \varphi) + \varphi]^2.$$

When $\eta = 0$, $\rho_i = 1$ and $F(y, \eta) = |\pi_{pp}(y, \eta)|$, which increases with size. The following four steps consider $\eta > 0$.

Step 1: Target-price sensitivity. Using $\rho = (\epsilon - s)/(\epsilon + \eta - s)$,

$$\varphi + (1 - \varphi)\rho = \frac{\varphi(\epsilon + \eta - s) + (1 - \varphi)(\epsilon - s)}{\epsilon + \eta - s} = \frac{\epsilon + \varphi\eta - s}{\epsilon + \eta - s}.$$

Step 2: The return to information.

$$|\pi_{pp}| = \exp\left(\frac{1}{\eta} - \frac{1}{\eta}y^{\eta/\epsilon}\right) \cdot \left(-y + (\epsilon + \eta)y^{(\epsilon-\eta)/\epsilon}\right)$$

Taking logs and dropping constants,

$$\ln |\pi_{pp}| = \ln y - \frac{s}{\eta} + \ln(\epsilon + \eta - s) - \ln s,$$

hence

$$\ln F = \ln |\pi_{pp}| + 2 \ln\left(\frac{\epsilon + \varphi\eta - s}{\epsilon + \eta - s}\right) = \ln y - \frac{s}{\eta} - \ln s - \ln(\epsilon + \eta - s) + 2 \ln(\epsilon + \varphi\eta - s).$$

Step 3: The effect of firm size. Since $ds/dy = (\eta/\epsilon)(s/y)$, differentiating yields

$$\frac{F_y}{F} = \frac{1}{y} M_F(s; \eta), \quad s \in (0, \epsilon),$$

where

$$M_F(s; \eta, \varphi) = 1 - \frac{s + \eta}{\epsilon} + \frac{\eta s}{\epsilon(\epsilon + \eta - s)} - \frac{2\eta s}{\epsilon(\epsilon + \varphi\eta - s)}.$$

Thus $\text{sign}(F_y) = \text{sign}(M_F)$. Moreover,

$$M_F(0; \eta, \varphi) = 1 - \frac{\eta}{\epsilon}.$$

$$M_F(\epsilon; \eta, \varphi) = 1 - \frac{\eta}{\epsilon} - \frac{2}{\varphi} < 0,$$

Step 4: The three regimes. Differentiate:

$$M'_F(s; \eta, \varphi) = -\frac{1}{\epsilon} + \frac{\eta(\epsilon + \eta)}{\epsilon(\epsilon + \eta - s)^2} - \frac{2\eta(\epsilon + \varphi\eta)}{\epsilon(\epsilon + \varphi\eta - s)^2}.$$

It remains to show that $M'_F(s; \eta, \varphi) < 0$ for all (φ, s) in this domain. For fixed s ,

$$\frac{\partial M'_F}{\partial \varphi} = \frac{2\eta^2(\epsilon + \varphi\eta + s)}{\epsilon(\epsilon + \varphi\eta - s)^3} > 0,$$

so $M'_F(s; \eta, \varphi)$ is increasing in φ and thus

$$M'_F(s; \eta, \varphi) \leq M'_F(s; \eta, 1).$$

Setting $\varphi = 1$ gives

$$M'_F(s; \eta, 1) = -\frac{1}{\epsilon} - \frac{\eta(\epsilon + \eta)}{\epsilon(\epsilon + \eta - s)^2}.$$

This expression is strictly decreasing in $s \in (0, \epsilon)$, so its supremum is obtained as $s \downarrow 0$:

$$\sup_{s \in (0, \epsilon)} M'_F(s; \eta, 1) = -\frac{\epsilon + 2\eta}{\epsilon(\epsilon + \eta)} < 0.$$

Therefore $M'_F(s; \eta, \varphi) < 0$ for all $\varphi \in (0, 1]$ and $s \in (0, \epsilon)$.

If $\eta < \epsilon$, combining $M_F(0; \eta, \varphi) > 0$ with the negative upper-endpoint condition above yields a unique root and hence single-peakedness of F . If $\eta \geq \epsilon$, $M_F(0; \eta, \varphi) \leq 0$. Since $M_F(0; \eta, \varphi) \leq 0$ and M_F is strictly decreasing, we have $M_F(s; \eta, \varphi) < 0$ for all $s \in (0, \epsilon)$, so $F_y < 0$ everywhere and F is strictly decreasing. $\square$

A.4. Proof of Proposition 4

Proof. We follow the approach used to prove Proposition 1 in [Baqaee et al. \(2024\)](#). Specialized to our one-sector economy, their first-order reallocation identity gives

$$d \log \text{TFP} = \frac{\bar{\mu}}{\mathbb{E}_u[\epsilon_i]} \text{Cov}_u(\epsilon_i, p_i) = \bar{\mu}(p - \mathbb{E}_u[p_i]),$$

where the second equality uses $p = \mathbb{E}_u[\epsilon_i p_i] / \mathbb{E}_u[\epsilon_i]$ from (2.15). Following a unit shock, firm i 's price response is

$$p_i = K_i[\rho_i + (1 - \rho_i)p].$$

Let

$$D \equiv \mathbb{E}_u([K_i \rho_i + (1 - K_i)]\epsilon_i).$$

Aggregating firms' price responses and solving the fixed point for p yields

$$p = \frac{\mathbb{E}_u[K_i \epsilon_i \rho_i]}{D}.$$

Substituting these expressions into the reallocation identity and rearranging covariances gives

$$d \log \text{TFP} = \frac{\bar{\mu}}{D} \left[\mathbb{E}_u[K_i] \mathbb{E}_u[1 - K_i] \text{Cov}_{uK}(\epsilon_i, \rho_i) + \mathbb{E}_{uK}[\rho_i] \text{Cov}_u(\epsilon_i, K_i) \right].$$

The two coefficients are precisely ς_ρ and ς_K in the proposition. They are nonnegative because $K_i \in [0, 1]$, $\rho_i > 0$, and $D > 0$. $\square$

A.5. Proof of Proposition 5

Proof. To obtain a tractable characterization, suppose firm scope scales isoelastically with representative-unit sales,

$$M_i = \bar{M} \omega_i^\chi, \quad \chi \geq 0,$$

so total firm sales satisfy

$$\Omega_i = \bar{M} \omega_i^{1+\chi}.$$

The parameter χ is the elasticity of product scope with respect to representative-unit sales. Under the Kimball demand system, let y denote representative-unit relative demand and define $x \equiv y^{\eta/\epsilon}$. Let $F(y, \eta)$ denote the baseline object in Appendix A.3 governing the sign of the size–attention relationship. Because a common signal is reused across M_i

units, product scope multiplies the baseline value of information by $\omega(y)^\chi$:

$$F^{sh}(y, \eta) = \omega(y)^\chi F(y, \eta),$$

where $\omega(y) \propto y \exp\left(\frac{1}{\eta} - \frac{x}{\eta}\right)$ is representative-unit sales. Its log derivative is therefore

$$\frac{F_y^{sh}}{F^{sh}} = \frac{1}{y} M_F^{sh}(x; \eta, \varphi, \chi),$$

with

$$M_F^{sh}(x; \eta, \varphi, \chi) = 1 + \chi - \frac{(1 + \chi)x + \eta}{\epsilon} + \frac{\eta x}{\epsilon(\epsilon + \eta - x)} - \frac{2\eta x}{\epsilon(\epsilon + \varphi\eta - x)}.$$

As in Appendix A.3, M_F^{sh} is strictly decreasing in x , and

$$M_F^{sh}(0; \eta, \varphi, \chi) = 1 + \chi - \frac{\eta}{\epsilon}, \quad M_F^{sh}(\epsilon; \eta, \varphi, \chi) = 1 - \frac{\eta}{\epsilon} - \frac{2}{\varphi} < 0.$$

When $\eta = 0$, elasticity and pass-through are constant across units, while shared information scales the positive size force through Ω_i . Hence attention increases with firm size. For $\eta > 0$, M_F^{sh} governs the derivative. Since M_F^{sh} is strictly decreasing in x , its sign is pinned down by $M_F^{sh}(0; \eta, \varphi, \chi) = 1 + \chi - \eta/\epsilon$. If $0 < \eta < (1 + \chi)\epsilon$, then $M_F^{sh}(0; \eta, \varphi, \chi) > 0$ while $M_F^{sh}(\epsilon; \eta, \varphi, \chi) < 0$, so it crosses zero once and attention is hump-shaped. If $\eta \geq (1 + \chi)\epsilon$, then $M_F^{sh}(0; \eta, \varphi, \chi) \leq 0$, so $M_F^{sh} < 0$ for all x and attention decreases with firm size. Since $\Omega_i = \bar{M}\omega_i^{1+\chi}$ is a monotone transformation of ω_i , the results hold under either size measure. $\square$

A.6. Proof of Proposition 6

Proof. Optimal attention is increasing in the benefit–cost ratio A_i/λ_i , where

$$A_i \equiv \frac{\sigma_q^2}{2} |\pi_{pp,i}| [\rho_i(1 - \varphi) + \varphi]^2, \quad \lambda_i = \bar{\lambda} s(z_i)^{-\psi}.$$

Under the Kimball demand system, let y denote relative demand, define $x \equiv y^{\eta/\epsilon}$, and let $F(y, \eta)$ denote the baseline benefit of information in Appendix A.3. The baseline sales share can be written as $s(z(y)) \propto y \exp\left(\frac{1}{\eta} - \frac{x}{\eta}\right)$. The effective benefit–cost ratio therefore satisfies

$$\frac{A_i}{\lambda_i} \propto s(z(y))^\psi F(y, \eta),$$

and its log derivative is

$$\frac{\partial}{\partial \log y} \log \left(\frac{A_i}{\lambda_i} \right) = M_F^c(x; \eta, \varphi, \psi),$$

where

$$M_F^c(x; \eta, \varphi, \psi) = 1 + \psi - \frac{(1 + \psi)x + \eta}{\epsilon} + \frac{\eta x}{\epsilon(\epsilon + \eta - x)} - \frac{2\eta x}{\epsilon(\epsilon + \varphi\eta - x)}.$$

As in Appendix A.3, M_F^c is strictly decreasing in x . Moreover,

$$M_F^c(0; \eta, \varphi, \psi) = 1 + \psi - \frac{\eta}{\epsilon}, \quad M_F^c(\epsilon; \eta, \varphi, \psi) = 1 - \frac{\eta}{\epsilon} - \frac{2}{\varphi} < 0.$$

When $\eta = 0$, elasticity and pass-through are constant across firms. The benefit–cost ratio A_i/λ_i increases with firm size, so attention also increases. For $\eta > 0$, M_F^c governs the derivative. Since M_F^c is strictly decreasing in x , its sign is pinned down by $M_F^c(0; \eta, \varphi, \psi) = 1 + \psi - \eta/\epsilon$. If $0 < \eta < (1 + \psi)\epsilon$, then $M_F^c(0; \eta, \varphi, \psi) > 0$ while $M_F^c(\epsilon; \eta, \varphi, \psi) < 0$, so it crosses zero once and attention is hump-shaped. If $\eta \geq (1 + \psi)\epsilon$, then $M_F^c(0; \eta, \varphi, \psi) \leq 0$, so $M_F^c < 0$ for all x and attention decreases with size. $\square$

A.7. Proof of Proposition 7

Proof. Since $p = \varphi q$ and $q \perp v_i$, the return to a signal about the full target price is

$$F_i^{\text{both}} = |\pi_{pp,i}| \left\{ \sigma_q^2 [\varphi + (1 - \varphi)\rho_i]^2 + \sigma_v^2 \rho_i^2 \right\},$$

and optimal attention is increasing in F_i^{both} . In the interior equilibrium considered here, $\varphi > 0$. When $\eta = 0$, $\rho_i = 1$ and target-price variance is common across firms, so the size channel makes attention increase with size. When $\eta \geq \epsilon$, the aggregate-shock component decreases by Proposition 2, while the same derivative calculation shows that $|\pi_{pp,i}| \rho_i^2$ decreases with size. Their sum therefore also decreases.

It remains to consider $0 < \eta < \epsilon$. Let $y = Y_i/Y$, $x = y^{\eta/\epsilon}$, $z = \epsilon - x$, $r = \sigma_v^2/\sigma_q^2$, $c = \varphi\eta$, and $C = 1 + r$. Using the Kimball expressions for $|\pi_{pp,i}|$ and ρ_i , and dropping positive constants,

$$F^{\text{both}}(y) \propto e^{-x/\eta} x^{\epsilon/\eta - 1} \frac{\sigma_q^2 (\epsilon + \varphi\eta - x)^2 + \sigma_v^2 (\epsilon - x)^2}{\epsilon + \eta - x}.$$

The sign of its size derivative is the sign of

$$P(z) = Cz^4 + (2c + \eta C)z^3 + [c^2 - \eta C(\epsilon - \eta)]z^2 - [\eta c^2 + 2C\epsilon\eta^2]z + \eta c[(\epsilon - \eta)c - 2\epsilon\eta].$$

The coefficient signs, in descending order, are $+$, $+$, ambiguous, $-$, $-$. Hence Descartes' rule of signs implies that P has exactly one positive root. Since $P(0) < 0 < P(\epsilon)$, this root lies in the feasible interval, so F^{both} first rises and then falls with firm size. Attention is therefore hump-shaped. $\square$

References

- Adam, K. (2005). Optimal monetary policy with imperfect common knowledge. *Journal of Monetary Economics* 52(2), 267–301. [2.4](#)
- Adam, K. (2007). Optimal Monetary Policy with Imperfect Common Knowledge. *Journal of monetary Economics* 54(2), 267–301. [1](#), [6.1](#)
- Afrouzi, H. (2024). Strategic inattention, inflation dynamics, and the non-neutrality of money. *Journal of Political Economy* 132(10), 3378–3420. [1](#)
- Afrouzi, H. and C. Yang (2021). Dynamic rational inattention and the phillips curve. Available at SSRN 3770462. [5](#), [6.2](#), [B.1](#)
- Angeletos, G.-M. (2018). Frictional coordination. *Journal of the European Economic Association* 16(3), 563–603. [1](#)
- Angeletos, G.-M., F. Collard, and H. Dellas (2018). Quantifying confidence. *Econometrica* 86(5), 1689–1726. [1](#)
- Angeletos, G.-M. and J. La'O (2009). Incomplete information, higher-order beliefs and price inertia. *Journal of Monetary Economics* 56(S), S19–S37. [1](#), [6.2](#)
- Asriyan, V. and A. N. Kohlhas (2025, April). The macroeconomics of data: Scale, product choice, and pricing in the information age. Technical report, Working Paper. [1](#), [7](#)
- Baqaei, D. R., E. Farhi, and K. Sangani (2024). The supply-side effects of monetary policy. *Journal of Political Economy* 132(4), 1065–1112. [2](#), [4.2](#), [4.2](#), [A.4](#)
- Beck, G. W. and S. M. Lein (2020). Price elasticities and demand-side real rigidities in micro data and in macro models. *Journal of Monetary Economics* 115, 200–212. [5.1](#)

- Bergin, P. R. and R. C. Feenstra (2001, August). Pricing-to-market, staggered contracts, and real exchange rate persistence. *Journal of International Economics* 54(2), 333–359. [5.1](#)
- Chahrour, R. and K. Jurado (2025a). Expectation response functions in dynamic linear economies. Working Paper. [1](#)
- Chahrour, R. A. and K. Jurado (2025b). Revisiting the forecasts of others. *Journal of Monetary Economics* 155(S), 103844. [1](#)
- Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2000, September). Sticky price models of the business cycle: Can the contract multiplier solve the persistence problem? *Econometrica* 68(5), 1151–1179. [5.1](#)
- Cloyne, J., C. Ferreira, M. Froemel, and P. Surico (2023). Monetary policy, corporate finance, and investment. *Journal of the European Economic Association* 21(6), 2586–2634. [3](#), [4.1](#)
- Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review* 105(8), 2644–2678. [5.1](#)
- Coibion, O., Y. Gorodnichenko, and S. Kumar (2018). How do firms form their expectations? new survey evidence. *American Economic Review* 108(9), 2671–2713. [7](#)
- David, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016). Information, misallocation, and aggregate productivity. *The Quarterly Journal of Economics* 131(2), 943–1005. [1](#)
- Eichenbaum, M. and J. D. M. Fisher (2007, October). Estimating the frequency of price re-optimization in calvo-style models. *Journal of Monetary Economics* 54(7), 2032–2047. [5.1](#)
- Fang, Y., Z. Huo, M. Li, and J. Wu (2025). Inattentive network: Evidence and theory. *Available at SSRN* 5310414. [1](#)
- Flynn, J. P. and K. Sastry (2024, jun). Attention cycles. Technical Report 32553, National Bureau of Economic Research. [5.2](#)
- Gertler, M. and S. Gilchrist (1994). Monetary policy, business cycles, and the behavior of small manufacturing firms. *The Quarterly Journal of Economics* 109(2), 309–340. [3](#), [4.1](#)
- Gnewuch, M. and D. Zhang (2025). Monetary policy, firm heterogeneity, and the distribution of investment rates. *Journal of Monetary Economics* 149, 103721. [3](#), [4.1](#)

- Huo, Z. and M. Pedroni (2020). A single-judge solution to beauty contests. *American Economic Review* 110(2), 526–568. [1](#)
- Huo, Z. and M. Pedroni (2023). Dynamic information aggregation: Learning from the past. *Journal of Monetary Economics* 136, 107–124. [1](#)
- Huo, Z. and N. Takayama (2025). Rational expectations models with higher-order beliefs. *Review of Economic Studies* 92(5), 3138–3173. [1](#)
- Kimball, M. S. (1995). The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit and Banking* 27(4), 1241–1277. [2.1](#)
- Klenow, P. J. and J. L. Willis (2016, July). Real rigidities and nominal price changes. *Economica* 83(331), 443–472. [1](#), [3.3](#), [5.1](#)
- La’O, J. and A. Tahbaz-Salehi (2022). Optimal monetary policy in production networks. *Econometrica* 90(3), 1295–1336. [6](#)
- Lorenzoni, G. (2010, January). Optimal Monetary Policy with Uncertain Fundamentals and Dispersed Information. *The Review of Economic Studies* 77(1), 305–338. [6.2](#)
- Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance* 66(1), 35–65. [5.2](#)
- Lucas, R. (1972). Expectations and the neutrality of money. *Journal of Economic Theory* 4, 103–124. [1](#), [1](#)
- Mackowiak, B. and M. Wiederholt (2009). Optimal Sticky Prices under Rational Inattention. *American Economic Review* 99(3), 769–803. [1](#), [1](#), [2.4](#), [2.4](#)
- Maćkowiak, B. and M. Wiederholt (2015). Business cycle dynamics under rational inattention. *The Review of Economic Studies* 82(4), 1502–1532. [1](#)
- Mankiw, N. G. and R. Reis (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics* 117(4), 1295–1328. [1](#)
- Matsuyama, K. and P. Ushchev (2017). Beyond ces: Three alternative aggregators. *Journal of Political Economy* 125(6), 1797–1841. [7](#)
- Miao, J., J. Wu, and E. R. Young (2022). Multivariate rational inattention. *Econometrica* 90(2), 907–945. [5](#)

- Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory* 145(5), 1837–1864. [2.3](#)
- Nimark, K. (2008). Dynamic Pricing and Imperfect Common Knowledge. *Journal of monetary Economics* 55(2), 365–382. [1](#)
- Nimark, K. P. (2017). Dynamic higher order expectations. CEPR Discussion Paper 11863, Centre for Economic Policy Research. [1](#)
- Ou, S., P. Yin, D. Zhang, and R. Zhang (2026). Nominal rigidities, rational inattention, and the optimal monetary policy. *Journal of Economic Theory* 232, 106129. [1](#), [17](#)
- Ou, S., D. Zhang, and R. Zhang (2021). Information frictions, monetary policy, and the paradox of price flexibility. *Journal of Monetary Economics* 120, 70–82. [6.2](#)
- Paciello, L. and M. Wiederholt (2014). Exogenous Information, Endogenous Information, and Optimal Monetary Policy. *Review of Economic Studies* 81(1), 356–388. [1](#), [5.1](#)
- Phelps, E. S. (1970). Information, expectations, and the stability of unemployment. *American Economic Review* 60, 744–756. [1](#), [1](#)
- Sims, C. A. (2003). Implications of Rational Inattention. *Journal of monetary Economics* 50(3), 665–690. [1](#), [1](#)
- Smets, F. and R. Wouters (2007, June). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review* 97(3), 586–606. [5.1](#)
- Song, W. and S. Stern (2025). Firm Inattention and the Efficacy of Monetary Policy: A Text-Based Approach. *The Review of Economic Studies* 92(5), 3438–3469. [1](#), [5.2](#), [7](#)
- Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. In P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford (Eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics*, pp. 25–58. Princeton, NJ: Princeton University Press. [1](#)

Online Appendix for Heterogeneous Firms, Rational Inattention, and Monetary Policy

Shengliang Ou, Donghai Zhang, and Renbin Zhang

This Online Appendix contains supplementary material for the paper.

Appendix [A](#) reports additional quantitative and empirical figures referenced in Sections [5](#) and [6](#). Appendix [B](#) provides the state-space representation and numerical algorithm for persistent aggregate nominal-demand shocks, together with additional details of the multi-product and heterogeneous-information-cost extensions.

A. Additional Figures

This section collects the supplementary quantitative and empirical figures referenced in Sections 5 and 6.

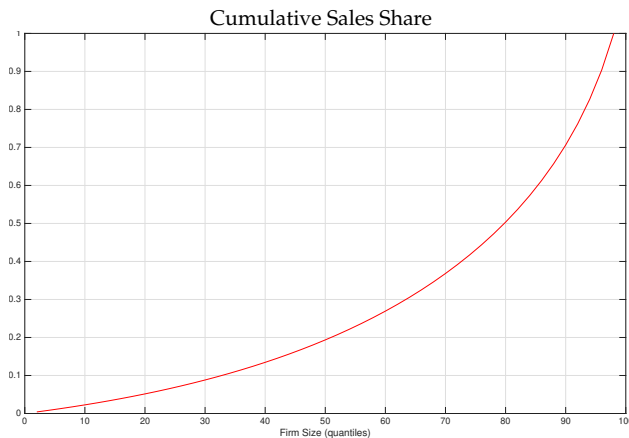


Figure A.1: Calibrated Firm Size Distribution: Cumulative Sales Share

Notes: The figure plots the cumulative sales share implied by the calibrated firm-size distribution. Firm size is indexed by quantile.

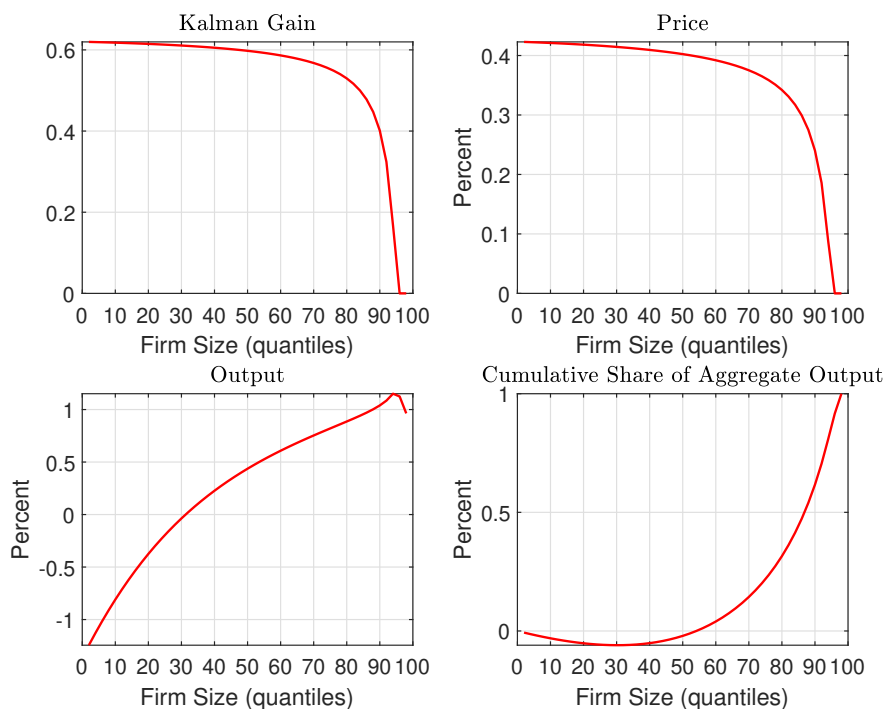


Figure A.2: Heterogeneous Effects of Monetary Policy: Higher-Curvature Comparison

Notes: The figure reports the higher-curvature comparison with endogenous attention. We set $\eta = \epsilon = 5$. Firm size is indexed by quantile.

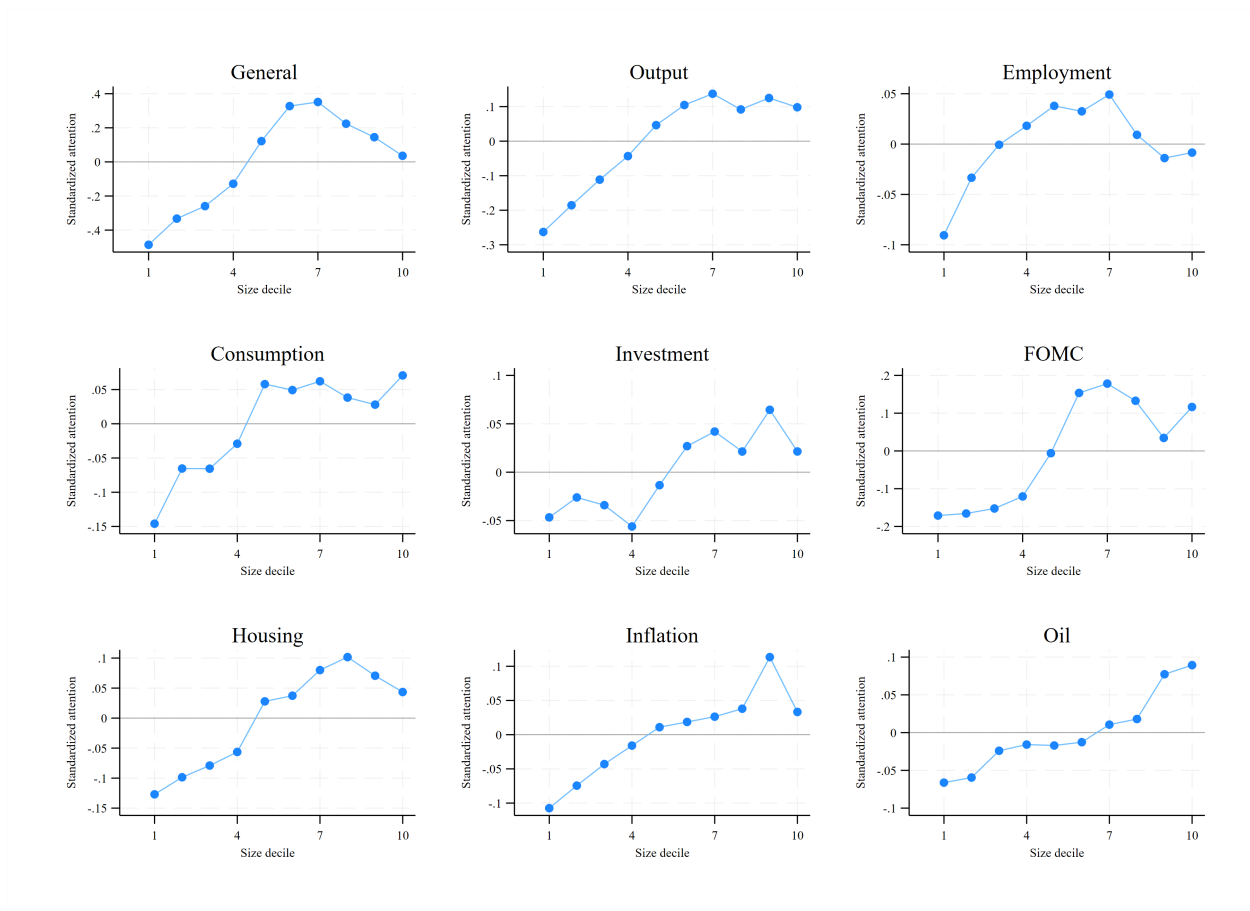


Figure A.3: Firm Size and Attention across Macroeconomic Topics

Notes: Each panel reports mean standardized attention to the indicated macroeconomic topic within an annual asset decile. Firms are assigned to deciles based on log total assets within each filing calendar year. Attention is measured as the share of topic-specific words in each filing and standardized separately by topic.

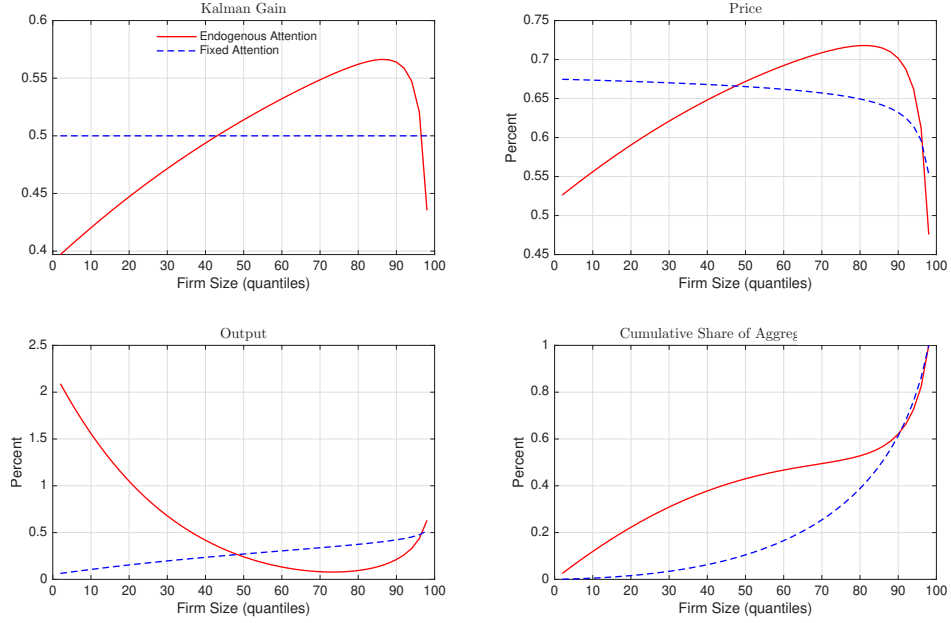


Figure A.4: Heterogeneous Effects of Monetary Policy under CRRA Preferences

Notes: Solid (red) lines report the baseline with endogenous attention under CRRA preferences. Dashed (blue) lines report the fixed-attention counterfactual, which preserves average attention while eliminating its cross-sectional heterogeneity. We set $\sigma = 2$ and $\phi = 1$. Firm size is indexed by quantile.

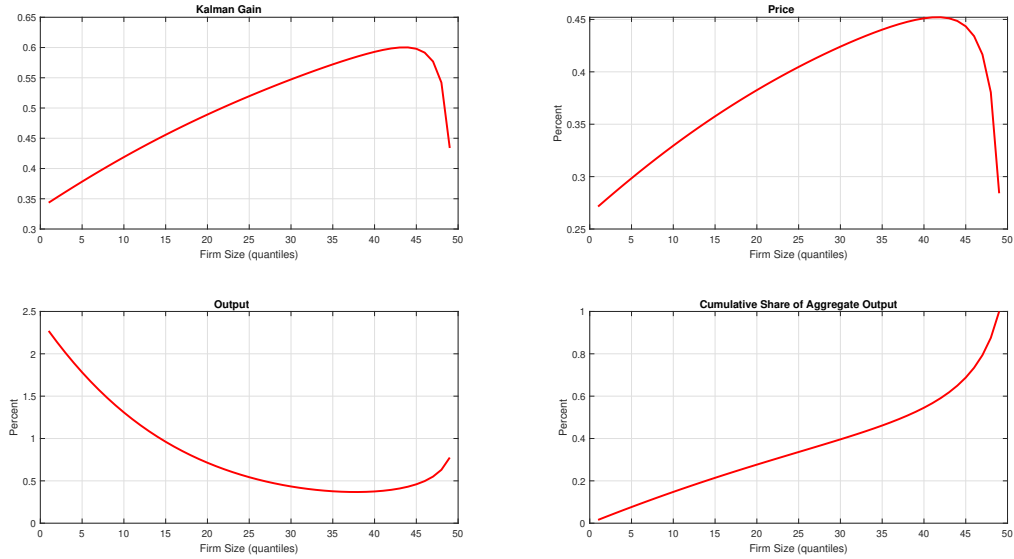


Figure A.5: Heterogeneous Effects of Monetary Policy with Persistent Shocks

Notes: The figure reports firms' responses to an aggregate-demand shock in the persistent-shock extension under the baseline calibration. The persistence of the shock, ρ_q , is set to 0.5. Firm size is indexed by quantile.

B. Details for Extensions and Robustness

This section provides the derivations and numerical details for the extensions in Section 6. It covers persistent aggregate shocks, multi-product firms with shared information, and heterogeneous information costs.

B.1. Persistent Aggregate Shocks

This subsection describes the numerical solution for the case of unobserved lagged shocks. The solution has two stages. For conjectured full-information target-price loadings, we solve each firm's dynamic information problem. We then aggregate the implied price loadings and update the conjecture until the aggregate price dynamics are consistent with firms' pricing decisions.

State-space representation. We approximate the hidden state by truncating the history of nominal-demand innovations at horizon L and define the state vector

$$u^t = (\varepsilon_t^q, \varepsilon_{t-1}^q, \dots, \varepsilon_{t-L+1}^q)'. \quad (\text{B.1})$$

The truncated state evolves according to

$$u^t = Au^{t-1} + Q\varepsilon_t^q, \quad (\text{B.2})$$

where $Q = (1, 0, \dots, 0)'$ is the innovation-loading vector and A is the $L \times L$ shift matrix. We conjecture that the vector H_i maps this state into firm i 's full-information target price:

$$p_{i,t}^\diamond = H_i' u^t. \quad (\text{B.3})$$

Firm-level information choice. Following [Afrouzi and Yang \(2021\)](#), the optimal information structure can be represented by a single signal about the payoff-relevant linear combination of the state,

$$s_{i,t} = \mathbf{y}_i' u^t + e_{i,t}, \quad e_{i,t} \sim N(0, \sigma_{e,i}^2), \quad (\text{B.4})$$

The vector $\mathbf{y}_i$ determines which linear combination of current and lagged innovations the signal measures. The firm observes the signal before setting its price, and its information set consists of the history $s_i^t = (s_{i,t}, s_{i,t-1}, \dots, s_{i,0})$. For a conjectured H_i , firm i chooses its

posterior covariance matrix by solving

$$\min_{\Sigma_{i,t|t} \in \mathcal{S}_+^L} \frac{|\pi_{pp,i}|}{2(1-\beta)} \mathbb{E}_{i,t} (p_{i,t} - p_{i,t}^\diamond)^2 + \lambda \log_2 \left(\frac{|\Sigma_{i,t|t-1}|}{|\Sigma_{i,t|t}|} \right), \quad (\text{B.5})$$

subject to

$$\Sigma_{i,t+1|t} = A \Sigma_{i,t|t} A' + Q Q', \quad (\text{B.6})$$

$$\Sigma_{i,t|t-1} - \Sigma_{i,t|t} \succeq 0, \quad (\text{B.7})$$

$$0 < \Sigma_{i,0|-1} < \infty. \quad (\text{B.8})$$

Here, $\Sigma_{i,t|t-1}$ and $\Sigma_{i,t|t}$ denote the prior and posterior covariance matrices, respectively, and β is the discount factor. The objective trades off expected pricing losses against information costs. The first constraint maps current posterior uncertainty into next period's prior uncertainty. The second requires posterior uncertainty to be weakly smaller than prior uncertainty, and the third specifies a positive and finite initial covariance.

Equilibrium fixed point. The full-information target-price loadings depend on aggregate price dynamics, which in turn depend on firms' information choices and pricing rules. We therefore solve for equilibrium by fixed-point iteration on $\{H_i\}_i$.

1. Initialize the firm-level full-information target-price loadings $\{H_i\}_i$ such that

$$p_{i,t}^\diamond = H_i' u^t. \quad (\text{B.9})$$

2. Given $\{H_i\}_i$, solve each firm's dynamic rational-inattention problem and obtain $(K_i, \mathbf{y}_i)$.

3. For each firm, construct the forecast operator

$$X_i = \sum_{j=0}^{\infty} [(I - K_i \mathbf{y}_i') A]^j K_i \mathbf{y}_i' (A')^j, \quad (\text{B.10})$$

so that the firm's actual price loading is

$$H_{p,i} = X_i' H_i. \quad (\text{B.11})$$

4. Aggregate across firms using the weights z_i to obtain the loading of $p_t - \lambda_t$:

$$H_\lambda = \int z_i H_{p,i} di. \quad (\text{B.12})$$

5. Update each firm's full-information target-price loading using

$$H_i^{\text{new}} = \rho_i N + (1 - \rho_i) H_\lambda. \quad (\text{B.13})$$

6. Repeat until $\{H_i\}_i$ converges.

At convergence, the conjectured target-price loadings coincide with those implied by firms' pricing decisions and the aggregate price dynamics. The resulting loadings determine the impulse responses. We use them to compute firm-level output responses and each firm's cumulative contribution to aggregate output.

B.2. Multi-Product Firms with Shared Information

The multi-product extension introduces an additional scale force: a firm can reuse one signal across its operating units. This subsection presents the full environment, complementing the proof of Proposition 5 in Appendix A.5.

Operating units and demand. Firm i operates a mass $M_i = \mathcal{M}(a_i)$ of ex ante symmetric operating units, with

$$\mathcal{M}'(a_i) \geq 0.$$

More productive firms therefore operate more units. We normalize the total mass of units to one:

$$\int_0^1 M_i di = 1.$$

Units are indexed by (i, j) , where $j \in [0, M_i]$, and each unit is produced with productivity e^{a_i} :

$$Y_{ij} = e^{a_i} L_{ij}.$$

Preferences are defined over unit-level varieties,

$$\int_0^1 \int_0^{M_i} G\left(\frac{Y_{ij}}{Y}\right) dj di = 1,$$

so firm ownership does not directly enter unit-level demand.

Shared information and decentralized pricing. Headquarters chooses one common signal about the aggregate nominal shock,

$$s_i = q + e_i, \quad e_i \sim N(0, \sigma_{e,i}^2),$$

and shares it with all operating units. Each unit manager sets its price taking aggregates as given and does not internalize the effect of its price on the firm's other units. Under within-firm symmetry, all units of firm i choose the same price in equilibrium, and each unit's desired price remains

$$p_{ij}^\diamond = \rho_i q + (1 - \rho_i) p.$$

Firm-level pricing losses. Let

$$\omega_i \equiv \frac{P_{ij} Y_{ij}}{PY}$$

denote the sales share of a representative operating unit, and let

$$\Omega_i \equiv M_i \omega_i$$

denote total firm sales. Since all units of firm i are symmetric and use the same signal, firm profits satisfy

$$\Pi_i = M_i \pi_i^u,$$

where π_i^u denotes representative-unit profit. The quadratic loss from a common pricing mistake is therefore

$$\frac{|\Pi_{pp,i}|}{2} (p_i - p_i^\diamond)^2, \quad |\Pi_{pp,i}| = \Omega_i [\epsilon_i - 1 + \eta_i].$$

The unit-level pricing problem is unchanged, but the value of the common signal now scales with total firm sales Ω_i rather than representative-unit sales ω_i . Information reuse therefore strengthens the positive size force.

Appendix A.5 derives the size–attention relationship and proves Proposition 5.

B.3. Heterogeneous Information Costs

Heterogeneous information costs add a cost-side source of variation in attention to the baseline benefit-side forces. This subsection separates these two sides of the information choice, complementing the proof of Proposition 6 in Appendix A.6.

Information choice with heterogeneous costs. Under the cost specification

$$\lambda_i f(\kappa_i),$$

firm i 's information-acquisition problem can be written as

$$\kappa_i^* = \arg \min_{\kappa_i} \left\{ A_i 2^{-2\kappa_i} + \lambda_i f(\kappa_i) \right\}, \quad A_i \equiv \frac{\sigma_q^2}{2} |\pi_{pp,i}| [\rho_i(1 - \varphi) + \varphi]^2,$$

where A_i collects the baseline benefit of information acquisition. Under convex $f(\cdot)$, optimal attention κ_i^* is increasing in the benefit–cost ratio A_i/λ_i . The cross-sectional allocation of attention therefore depends on both the baseline return to information and the firm-specific cost shifter.

Size-dependent information costs. To obtain a tractable characterization, suppose the cost shifter declines isoelastically with the baseline sales share:

$$\lambda_i = \bar{\lambda} s(z_i)^{-\psi}, \quad \psi \geq 0.$$

The parameter ψ governs how strongly information costs decline with firm size. A higher ψ strengthens the positive size force.

Appendix [A.6](#) derives the size–attention relationship and proves Proposition [6](#).