# Consumer Information Friction and Monetary Non-Neutrality: A Sequential Search Approach *

Xiaojie Liu†

May 27, 2025

Click here to view the latest version

**Abstract**

This paper develops a dynamic general equilibrium model of monetary non-neutrality driven solely by consumer-side search and information frictions, in contrast to standard frameworks that rely on firm-side nominal rigidities. Consumers incur search costs and possess incomplete information about average nominal marginal costs. The key mechanism is that, following a monetary shock, consumers attribute some of the resulting price changes to firm-specific adverse shocks, inducing them to search for alternatives. To dissuade search, firms compress the markup and limit the passthrough of the shock to prices. This mechanism operates through the attenuation of strategic complementarities in pricing due to information frictions. The model is tractable and can be efficiently solved using sequence-space methods. Using impulse-response matching, I estimate the degree of information friction to be substantial, accounting for a significant share of observed monetary non-neutrality. I further extend the model to incorporate aggregate productivity shocks and apply it to the post-pandemic inflation surge. Finally, I provide two pieces of empirical evidence that support the model's core mechanisms.

**Keywords:** Monetary non-neutrality, Price stickiness, Phillips curve, Consumer search, Information frictions, Incomplete passthrough, Strategic Complementarity, Higher-order Beliefs, Sequence-space Jacobian

*I am grateful for comments and discussions with George-Marios Angeletos, Lawrence Christiano, Diego Huerta, Benjamin Jones, Diego Känzig, Jennifer La'O, Jean-Paul L'Huillier, Chen Lian, Sara Moreira, Ryungha Oh, Alessandro Pavan, Harry Pei, Alireza Tahbaz-Salehi, Michael Woodford, Huo Zhen. I would like to thank Youyou Xu for excellent research assistance.

†Kellogg School of Management, Northwestern University. Email: xiaojie.liu@kellogg.northwestern.edu

*"This paper presents a theory that justifies price stickiness, namely, that firms, fearing to upset their customers, attribute a cost to price changes."*

— Rotemberg (1982)

Monetary policy is known to have large real effects on the economy in the short run. Both output and inflation decline following an unexpected increase in interest rate. This pattern has been repeatedly uncovered in the empirical literature.[1] Most of existing theories explain this phenomenon by focusing on frictions on the firm side. Models of price stickiness posit that price adjustments are infrequent due to either exogenous factors (Taylor, 1980; Calvo, 1983) or fixed costs (Mankiw, 1985; Golosov and Lucas, 2007). Another theory which dates at least back to Phelps (1969) and Lucas (1972) suggests that firms set prices based on incomplete information about aggregate shocks. Reis (2006) and Alvarez et al. (2016) further argue that costs of acquiring and processing information contribute to price rigidity endogenously.

Growing evidence suggests that consumers face more severe frictions than firms.For example, consumers systematically misperceive inflation (Binetti et al., 2024; Candia et al., 2023), exhibit selective attention to salient prices (Kumar et al., 2015), and form expectations based on daily shopping experiences (D'Acunto et al., 2021). Moreover, consumer search literature has documented a large price dispersion for an identical product (Stigler, 1961; Kaplan and Menzio, 2015). Theoretical models have shown that such dispersion can arise from consumer search frictions (Burdett and Judd, 1983; Stahl, 1989). These evidence motivates two central questions: First, can monetary non-neutrality be micro-founded solely on the basis of consumer-side frictions? If so, what does the Phillips curve look like? Second, how can empirically observed consumer frictions—such as slow learning and search costs—be mapped into a structural model? What monetary transmission mechanisms emerge from such a model once it is estimated?

To address these questions, I propose a monetary model that places consumer-side frictions at the center. Consumers face two frictions: (i) information friction about firms' nominal marginal costs and (ii) search frictions on the goods market. On the other hand, firms have full information about the model economy and set the prices flexibly. The main mechanism operates as follows. Following a positive monetary shock, the nominal marginal costs increase. Firms tend to increase prices. However, shoppers with incomplete information about the costs attribute much of this price increase to firm-specific adverse shocks. Marginal shoppers who are initially indifferent between purchasing and searching are now incentivized to seek outside options. To dissuade search, firms compress markups, thereby limiting the passthrough of the monetary shock.

[1] Christiano et al. (1999) identify this effect using timing restrictions in VAR. Recently, high-frequency identification approach helps resolve the endogeneity bias in the VAR approach and confirms this finding (Gertler and Karadi, 2015; Bauer and Swanson, 2023). Hazell et al. (2022) estimate the slope of Phillips curve to be very flat using cross-state variation in price indices. Ramey (2016) provides a great summary of this literature.

To show the core mechanism, I begin with a static model featuring three agents: shoppers, firms, and a monetary authority. The monetary authority sets the nominal wage. Following the standard consumer search literature (Wolinsky, 1986; Anderson and Renault, 1999), I assume that shoppers search sequentially and randomly with free recall. Shoppers must incur a search cost to visit a firm and learn the price. I extend the standard framework in two ways. First, firms are heterogeneous in productivity. Second, shoppers have incomplete information about the nominal wage. These two features lead to shoppers' rational confusion between the aggregate (nominal wage) and idiosyncratic (firm productivity) components of the cost. In particular, when observing a price, shoppers who believe the nominal wage is lower will interpret the price as indicative of encountering a firm with lower productivity and vice versa.

I begin by characterizing the full-information rational-expectation equilibrium, where the only friction is search friction. I demonstrate that, under full information about the nominal wage, monetary policy is neutral. The intuition is that although shoppers lack knowledge of individual firms' prices, they understand that, in equilibrium, all prices fully adjust to changes in the nominal wage. As a result, relative prices and search decisions remain unchanged. The key takeaway of this result is that the search friction alone is not sufficient to generate monetary non-neutrality. We need the interplay of search and information frictions.

Next, I present the main theorem that the model economy under incomplete information can be simply characterized by the following equation,

$$\hat{p} = \Gamma \hat{mc} + (1 - \Gamma) E^c \hat{p} \tag{1}$$

where $\Gamma$ is aggregate own-cost passthrough defined similar to Amiti et al. (2019); $\hat{p}$ is the change in price index; $\hat{mc}$ is the nominal marginal cost; and $E^c$ represents shoppers' average expectation. This equation is similar to equation (2.7) in Woodford (2003) in its form. Woodford (2003) characterizes a model with information frictions on the firm side about aggregate money supply. The difference here is that firms have full information. The passthrough of nominal marginal cost shock to the price index is incomplete solely due to incomplete information on the consumer side.

To understand the intuition, first consider an extreme case where shoppers have no information about the nominal wage. In this scenario, a nominal wage shock resembles an idiosyncratic shock to firms, and the passthrough reduces to $\Gamma$. When consumers are partially informed, firms start to respond to other firms' prices. However, this strategic complementarity in pricing is attenuated by consumer-side information frictions. The reason is that shoppers confuse the increase in nominal prices with the increase in relative prices. As a result, firms' demand is more elastic than it is under full information. Firms absorb part of the increase in nominal marginal cost in their markups.

I further show the monotone comparative statics of passthroughs: under mild assumptions on the distribution of match shocks, (i) high-productivity firms charge higher markup and lower prices, and exhibit lower passthroughs of aggregate shocks and (ii) passthroughs of all firms decrease with search and information frictions. I further discuss the intuitions behind symmetry. For positive shocks, Rotemberg (1982) is correct. firms, fearing the loss of marginal consumers, refrain from fully increasing prices. For negative shocks, the appropriate interpretation is that firms do not fully reduce prices in order to extract greater profits from infra-marginal consumers.

In the second part of the paper, I develop a dynamic general equilibrium model that builds on the core mechanism introduced in the static setting. To better align the model with empirical evidence, I also incorporate two other key frictions: habit formation and nominal wage rigidity. The model also features slow learning for shoppers and persistent monetary shocks on interest rate. To allow monetary shocks to influence both nominal marginal costs and nominal aggregate spending while remaining unobserved by shoppers, I introduce a representative household structure. Specifically, the household consists of a single, fully informed worker and a continuum of partially informed shoppers. The worker earns income and allocates it across the household's shoppers. Shoppers make decisions on where to buy goods based on their information. Following Auclert et al. (2020) and Auclert et al. (2024), I specify a simplified information structure for shoppers and recast equation (1) as matrices in sequence space. I compute the model impulse responses using the sequence-space methods.

I calibrate the model parameters to match the moments from the literature and my own empirical finding. In particular, the aggregate own-cost passthrough is a sufficient statistic that summarizes all the parameters related to the search friction. I take its value from Amiti et al. (2019) and Gopinath et al. (2011). The degree of information friction among shoppers is then estimated using impulse response matching. Specifically, I estimate an external IV-SVAR following Gertler and Karadi (2015), using standard macroeconomic variables alongside household inflation expectations from the Michigan Survey of Consumers. The results reveal that household inflation expectations underreact significantly to monetary shocks even 4 years after the shock, suggesting significant information frictions. The average duration of information frictions is estimated to be about 37 quarters. To analyze the monetary transmission mechanism, I switch off the information friction for shoppers in the model. As expected, the change in the price index becomes nearly five times larger, and expectations track the actual price level one-for-one—contrary to the empirical evidence. As a result, the degree of monetary non-neutrality is substantially diminished: the peak output response to a monetary shock declines by over 50 basis points.

Finally, I extend the dynamic model to incorporate aggregate productivity shocks. In contrast to standard New Keynesian models, the inflation response in this framework features a sharp initial increase followed by a rapid decline. This arises because, although nominal wage responses are dampened by wage rigidity and strategic complementarities are attenuated by consumer-side information frictions, firms still pass through a fraction $\Gamma$ of the productivity shock to prices, as implied by equation (1). I further apply this extension to examine the post-pandemic episode of elevated inflation. Beaudry et al. (2024) documents that during this period, household inflation expectations were closely aligned with actual inflation, suggesting minimal information frictions on the consumer side. I compute the model under the assumption of full information, the inflation response almost doubles. This application highlights the importance of consumer-side information friction in determining the slope of Phillips curve.

In the last part of the paper, I present two pieces of empirical evidence that support the key model mechanisms. First, I construct an inflation attention index and show, using state-level inflation and unemployment data, that increased attention to national inflation is associated with a steeper slope in the regional Phillips curve. When the average attention index rises from average 0.26 before 1990 to 0.69 after 1990, the slope of the regional Phillips curve declines by approximately 16%. Second, using large-scale consumer panel data, I first measure search activities as non-routine share of spending in each product category. Then, I find that one standard deviation (51 bp) increase in unanticipated inflation for food and drinks leads to a 26.5 bp increase in the non-routine share of spending, which is highly statistically significant. The direction of the response is consistent with the model prediction.

**Related Literature –** This paper contributes to several strands of literature. The first studies the price stickiness by focusing on the consumer-side mechanisms. The prominent paper in this literature is Matějka (2015). He shows that firms set discrete prices as consumers "hate" price fluctuations, which increases their costs of attention. Rotemberg (2011) and Rebelo et al. (2024) propose behavioral theories of price stickiness.

Second, this paper bridges the Neo-Monetarist and New-Keynesian views by providing a search-theoretic micro-foundation for price stickiness and monetary non-neutrality. In particular, based on Burdett and Judd (1983), Head et al. (2012) and Burdett and Menzio (2018) show that price stickiness can result from the mixed pricing strategy. As nominal price increases, profit can still be maximized despite a fall in real price. In their model, the money is neutral.

In contrast, this paper follows the literature using sequential search models to provide a basis for price stickiness. Bénabou and Gertner (1993) and L'Huillier (2020) focus on the role of individual prices consumers observe during the search as revealing the information about aggregate shocks. Gaballo and Paciello (2021) show a model where consumers are motivated to leave the monopolistic firm and find lower prices in a separate market where firms have perfect competition, when the inflation rises. The price is sticky because some demand shifts to the low-price market. This paper contributes to this literature by integrating search models into a dynamics general equilibrium model with a heterogeneous firm block and incomplete information on the consumer side. I also adopt the sequence-space method to compute the model following the recent development in Auclert et al. (2021).

**Outline** − The rest of the paper is organized as follows. Section 1 presents a static model and establishes the main results on the passthrough. Section 2 presents the dynamic model and the calibration results. Section 3 shows empirical evidences. The last section concludes. Appendix A contains some of the proofs omitted from the text. Appendix B contains the proof of the dynamic model. Appendix C contains details of calibration and estimation. Appendix D contains details of empirical setup and additional empirical evidence.

# 1 Static Model

In this section, I develop a macroeconomic model that has three types of agents: shoppers, firms, and a monetary authority. Shoppers face information frictions about monetary shocks and search frictions when deciding where to buy goods. In contrast, firms are fully informed. I start with the static model to explain the main mechanism.

**Notation** − I use lower case to denote $\log Y$ for any variable $Y$, i.e., $y = \log Y$ and lower case with hat to denote log-deviation from the steady-state value, i.e., $\hat{y} = \log Y - \log \bar{Y}$.

## 1.1 Setup

Time is discrete and infinite $t \in \boldsymbol{N}$. The timeline is as follows. Within a period, the monetary authority first sets the nominal wage. Shoppers are endowed with cash which also serves as the signal about the nominal wage. Firms post prices and shoppers search sequentially. A fraction of shoppers make the purchase in a given round and the remaining keep searching. The period ends until all shoppers make the purchase. I denote the round of search $r = 1, 2, 3, ...$ within a period. All rounds of search happen within one period.

**Firm** − The economy is populated with a unit mass of firms indexed $k \in [0, 1]$, each of which produces a differentiated product using the following production technology,

$$Y_k = A_k L_k \tag{2}$$

where $L_k$ is the amount of labor employed. $A_k$ is the firm's productivity, which is i.i.d. across firms. Specifically, I assume that $\log A_k$ is drawn from the normal distribution $\Phi_a(a) = \mathcal{N}(0, \sigma_a^2)$. The marginal cost is $\frac{W}{A_k}$, where $W$ is the nominal wage. The nominal wage represents the average nominal marginal cost. Labor is supplied outside this economy in the current setup.

Suppose the optimal pricing strategy $p^* : \mathcal{R}^2 \to \mathcal{R}$ exists, which I will explore in Section 2. It maps its states $\{a_k, w\}$ to the price of its good, i.e., $p_k = p^*(a_k, w)$. Heterogeneity in $a_k$ across firms induces an endogenous price distribution $F(p|w)$, which has a density distribution $f(p|w)$. If

$p^*$ is monotone in both arguments, then the inverse mapping from $p_k$ to $a_k$ given $w$ can be denoted by $p^{*-1}(p, w)$. The price distribution conditional on $w$ is given by,

$$F(p|w) = 1 - \Phi_a(p^{*-1}(p, w)) \tag{3}$$

**Monetary Authority** – The monetary authority draws the log nominal wage from the normal distribution $\mathcal{N}(\bar{w}, \sigma_w^2)$, which serves as the prior for shoppers. Denote the monetary shock by $\hat{w} = w - \bar{w}$.

**Shopper** – The economy is populated with a continuum of shoppers indexed by $i \in [0, 1]$. At the beginning of the period, each shopper is endowed with cash $X_i$. It follows,

$$X_i = W \exp\left(\sigma_x \varepsilon_{xi} - \frac{\sigma_x^2}{2}\right) \tag{4}$$

where $\varepsilon_{xi}$ is i.i.d. across shoppers and it follows $\varepsilon_{xi} \sim \mathcal{N}(0, 1)$. It is also independent of any other shocks. The variance adjustment term ensures that the mean of $X_i$ is $W$. The shopper treats $X_i$ as signal about $w$. The expected log nominal wage is then given by $E(w|x_i) = \theta x_i + (1-\theta)\bar{w}$, where $\theta = \frac{\sigma_x^{-2}}{\sigma_w^{-2} + \sigma_x^{-2}}$. It measures the information friction. Higher $\theta$ implies that signals are more informative. Based on the Bayes' rule, shopper $i$'s posterior belief $H(w|x_i)$ follows $\mathcal{N}(E(w|x_i), (\sigma_w^{-2} + \sigma_z^{-2})^{-1})$.

In the rational expectations equilibrium, shoppers know the equilibrium pricing strategy $p^*$. Using the posterior belief about $w$, shoppers can construct their own perceived price distributions, $f(p|x)$, conditional on $x$:

$$f(p|x) = \int f(p|w) h(w|x) dw \tag{5}$$

The above equation yields two implications. First, when information is perfect, i.e., $x = w$, the perceived and objective price distributions coincide. Second, if $h(w|x_i)$ first-order stochastically dominates (FOSD) $h(w|x_j)$ and $p^*$ increases in $w$, then $f(p|x_i)$ FOSD $f(p|x_j)$. Intuitively, if a shopper believes the nominal wage is on average lower, the perceived price distribution is shifted to the left.

Each shopper buys and consumes only one good.[2] The utility (or value) that shopper $i$ gains from consuming good $k$ is given by,

$$\log \frac{X_i}{P_k} + \frac{1}{\lambda} \epsilon_{ik} \tag{6}$$

where $\epsilon_{ik} \sim G$ is match utility between shopper i and good k. $G$ is triple continuously differentiable and its density function is $g$. It captures idiosyncratic consumer preferences for certain goods over

[2]In Section 2.8, I extend the model by allowing shoppers to get access to multiple goods in one search.

others. $\epsilon_{ik}$ are i.i.d across firms and shoppers.[3] Also, following the literature, I assume that $G$ is *log-concave.* [4] The parameter $\lambda$ controls the relative importance between two utilities. A larger $\lambda$ implies more competition on prices. Since $X_i$ only affects the level of utility and does not change the relative utilities across goods, I define the normalized utility as follows,

$$y_{ik} = -p_k + \frac{1}{\lambda}\epsilon_{ik} \tag{7}$$

According to (7), the value distribution of drawing a random good is a convolution of the perceived price distribution and the distribution of match utility $\psi(y|x) = \int f(u-y|x)\lambda g(\lambda u)du$.

Shoppers search sequentially and randomly following Wolinsky (1986) and Anderson and Renault (1999). By incurring a search cost $\kappa > 0$, the shopper can visit a firm to learn both its price and the associated match utility. Each shopper is assigned with the first good for free.[5] Shoppers have free recall, meaning there are no additional costs for purchasing goods from firms they have previously visited. In addition, shoppers do not commit to any plan made before setting out to search.[6] I make the following simplifying assumption such that $f(p|x)$ remains fixed throughout all rounds of search.

**Assumption 1.** *Shoppers do not learn about monetary shocks from individual prices they observe during the search.*

This assumption is plausible if the variance of idiosyncratic productivity shocks is way larger than the variance of the monetary shocks.even when a shopper consistently encounters firms charging high prices, she attributes this to bad luck rather than an increased nominal wage. I make this assumption for tractability. First, optimal prices are non-linear in productivity and nominal wage. Bayesian updating about the aggregate state is intractable. Second, to know which shoppers stop in each round of search, we need to keep track of each shopper's whole search history, which

[3]The price distribution is not degenerated and the optimal prices are related to firm's idiosyncratic states, if the random utility term is match-specific. Otherwise, if it is only shopper-specific, it does not matter for search decisions. Firms compete only on price. The price distribution is degenerated to a single optimal price. If it is only good-specific, the optimal prices are not related to firm's productivity.

[4]Note that some commonly used distribution functions are log-concave, e.g., normal distribution, uniform distribution, Gumbel distribution. See Bagnoli and Bergstrom (2005) for a broad discussion of log-concavity that do and do not satisfy this condition. The assumption of log-concavity ensures that the hazard rate $\frac{g(x)}{1-G(x)}$ is monotonically increasing.

[5]Shoppers always participate the market with free first good.

[6]If the shopper formulated her search plan prior to search and she committed to that plan, then she would take into account the expected total search costs of sampling, and she would stop with a lower quality match if she were unlucky and happened to sample a sequence of firms for which she ill-suited. In the case of a shopper doing sequential search without commitment, she ignores past fixed costs of search as sunk. Burdett and Judd (1983) considers a search problem with commitment and homogeneous firms.

leads to exploding states. As I will show, our main mechanism remains valid as long as information friction exists. It is independent of how shoppers acquire information.

Now I state the shopper's problem. Let $v_{ir}$ denote the maximum value among the firms visited by shopper $i$ up to the $r$th round. By definition, $v_{i1} = y_{i1}$. For $r > 1$, $v_{ir}$ is recursively given by,

$$v_{ir} = \max\{v_{ir-1}, y_{ir}\} \tag{8}$$

In the $r$th round of the sequential search, the state of the shopper is $v_{ir}$. The shopper has the option to stop searching and accept $v_{ir}$ or continue searching. The value function for the shopper, $U : \mathcal{R} \to \mathcal{R}$, in each state $v \in \mathcal{R}$, satisfies,

$$U(v|x) = \max\left\{v, -\kappa + U(v|x)\int_{-\infty}^{v} \psi(y|x)dy + \int_{v}^{\infty} U(y|x)\psi(y|x)dy\right\} \tag{9}$$

The maximum represents that the shopper can either receive the maximum value $v$ up to this round and stop searching, or continue searching by incurring a search cost $\kappa$ and drawing a random good. If the value of that good is lower than $v$, which occurs with probability $\int_{-\infty}^{v} \psi(y|x)\,dy$, the shopper will retain the value $U(v)$, since she has free recall. Otherwise, she will obtain a higher value from the newly drawn good and update $v$ according to (8). The value function $U$ is stationary only when Assumption 1 holds. Otherwise, with the information sets expanding over the rounds of search, the value function $U$ should be indexed by the search round. The shopper's problem is solved in two steps. First, she finds the $U$ function that solves the functional equation (9). Second, she keeps sampling firms until $v$ first exceeds the expression to the right of the comma in (9).

### 1.2 Equilibrium Definitions

The equilibrium concept is Perfect Bayesian Nash equilibrium (PBNE).[7] Formally, I define the equilibrium as follows:

**Definition 1** (Equilibrium)**.** *A Perfect Bayesian Nash equilibrium is a triplet of allocation, prices, and beliefs such that*

1. *Firms choose the optimal pricing strategy $p^*$ to maximize profits given the optimal search strategy.*

2. *Shoppers search without commitment. They do not update beliefs after observing prices during the search. Conditional on the information sets, they combine the optimal pricing strategy*

[7]In the standard search literature, prices are bounded and consumers know the range of prices. Therefore, consumers are able to detect the off-equilibrium prices. The literature imposes that consumers keep the same (passive) beliefs about the distribution of future prices on and off the equilibrium path. In our case, since productivity is assumed to have unbounded support, any positive price is an on-equilibrium price. The regulations on the off-equilibrium belief is not strictly needed in this model.

*$p^*$ and other exogenous distributions to compute $U(v|x)$ for each state $v$. Shoppers' optimal search strategy is then determined by the stopping rule as shown in (9).*

3. *Nominal wage and cash for each shopper are chosen exogenously.*

4. *Goods market clears.*

In addition, I define the full-information equilibrium in which shoppers know the nominal wage. It serves a natural benchmark for the incomplete-information equilibrium.

**Definition 2** (Full-Information Equilibrium)**.** *A full-information equilibrium is the equilibrium defined above, except that shoppers know the nominal wage.*

# 2 Equilibrium Characterization

I now characterize the equilibrium. I proceed in four steps. First, I characterize the search and pricing strategies. Second, I show the existence and properties of full-information equilibrium. Third, I show the monetary neutrality under full information. Finally, I present the optimal strategies under the first-order approximation around full-information equilibrium.

## 2.1 Characterization of Optimal Search Strategy

I first characterize the search strategy. Each shopper needs to first find the $U$ function and then decide when to stop searching. The solution to the shopper's problem is presented in the following proposition,

**Proposition 1.** *Under Assumption 1, the optimal search strategy follows a threshold rule. The threshold is denoted $v^*(x)$. If $v < v^*(x)$, the shopper keeps searching; otherwise, she stops and makes the purchase. The threshold $v^*(x)$ decreases in $x$. It is determined by,*

$$\int \int_{\lambda(v^*(x)+p)}^{\infty} \big(\frac{1}{\lambda}\epsilon - p - v^*(x)\big) g(\epsilon) d\epsilon f(p|x) dp = \kappa \tag{10}$$

*where $\Psi(p|x)$ is the perceived distribution of the value of a random draw.*

*Proof.* See Appendix A. ∎

The optimal search strategy is straightforward. The shopper keeps sampling firms until the target value $v^*(x)$ is reached. Indeed, in Appendix A, I show that the value function $U(v|x)$ is given by,

$$U(v|x) = \max\{v, v^*(x)\} \tag{11}$$

This implies that, regardless of the current state $v$, the value of an additional search is constant, equal to the threshold $v^*(x)$. When $v < v^*(x)$, the value function is always $v^*(x)$, indicating that the shopper opts to continue searching. Conversely, when $v \geq v^*(x)$, the value function equals the state $v$, implying that the shopper accepts $v$.

The left-hand side of (10) represents the marginal utility of additional search at the threshold, i.e., when shopper's state is equal to $v^*(x_i)$. Suppose she samples another firm $k$, she will prefer the new good if $\frac{1}{\lambda}\epsilon_{ik} - p_k > v^*(x_i)$. Since the shopper can return without cost, the marginal utility obtained in this case is $\max\{\frac{1}{\lambda}\epsilon_{ik} - p_k - v^*(x_i), 0\}$. The threshold is achieved when the marginal utility is equal to the search cost. The threshold is decreasing in the shopper's signal $x$: A higher signal about the nominal wage leads to a lower threshold. Intuitively, when a shopper perceives the nominal wage to be lower, they believe the price distribution is shifted leftward. This increases the perceived probability of encountering a low-price firm, thereby raising the reservation value and increasing the likelihood of continued search.

This result generalizes the standard threshold rule found in the literature (see Weitzman, 1979; Wolinsky, 1986) by incorporating both an endogenous price distribution and incomplete information. In the canonical models, firms are homogeneous, implying a degenerate price distribution, and shoppers possess correct beliefs about this common price throughout their search. In contrast, our model features heterogeneous firms and the incomplete information distorts shoppers' perception about price distribution. Proposition 1 demonstrates that the optimal search strategy retains the threshold structure.

## 2.2 Characterization of Optimal Pricing Strategy

I aggregate the optimal search decisions across shoppers and derive the demand allocation across firms and the resulting profits. Since shopper $i$ only purchases the good $k$ if $\frac{1}{\lambda}\epsilon_{ik} - p_k > v^*(x_i)$, the probability of purchasing from firm $k$ is $1 - G\big(\lambda(v^*(x_i) + p_k)\big)$. Since learning from shopping is prohibited, this probability is the same for all rounds of search. The probability of any shopper purchasing from any firm in each search given $w$ is given by,

$$\rho = \int\int \big(1 - G\big(\lambda(v^*(x) + p)\big)\big) d\Phi_x(x) f(p|w) dp \tag{12}$$

Suppose the mass of shoppers who visit any firm in the first round is one. A fraction $\rho$ of these shoppers settle with the firms they visit in the first round. The remaining $1 - \rho$ shoppers search in the second round, a further $(1-\rho)^2$ search in the third round, and so on. Then, the expected number of searches of each shopper is $\rho^{-1}$. Firms are atomistic and take $\rho$ as given when setting prices. Furthermore, the total expenditure spent in firm $k$ after all rounds of search is given by,

$$\omega_k = \frac{1}{\rho}\int X\big(1 - G\big(\lambda(v^*(x) + p_k)\big)\big) d\Phi_x(x) \tag{13}$$

The dispersion in the cash in hand affects the expenditure allocation through the dispersion of thresholds and the covariance between the cash in hand and the threshold, as more cash in hand induces higher expectation of the nominal wage. The profit for firm $k$ is given by,

$$\pi_k = \frac{1}{\rho}\int X\big(1 - G\big(\lambda(v^*(x) + p_k)\big)\big)d\Phi_x(x)\frac{1}{P_k}(P_k - \frac{W}{A_k}) \tag{14}$$

The demand is derived by dividing the total expenditure spent in firm $k$ by $P_k$. The profit is thus the total demand times the profit per sale. Monopolistic firms compete on prices. They maximize the profit in (14) with respect to prices.

**Proposition 2.** *Let $\mu_k$ denote the markup and $e_k$ denote the elasticity of demand. Firm $k$ charges a markup over its marginal cost,*

$$P_k = \mu_k\frac{W}{A_k}; \quad \mu_k = \frac{e_k}{e_k - 1} \tag{15}$$

*The elasticity of demand $e_k$ is determined by,*

$$e_k = \lambda\frac{\int Xg\big(\lambda(v^*(x) + p_k)\big)d\Phi_x(x)}{\int X\big(1 - G\big(\lambda(v^*(x) + p_k)\big)\big)d\Phi_x(x)} + 1 \tag{16}$$

*Proof.* See Appendix A. ■

The optimal price is a markup times the marginal cost. The elasticity of demand depends on the distribution of thresholds. This is a highly non-linear equilibrium. To understand the intuition, we first focus on the full-information benchmark.

### 2.3 Characterization of the Full-Information Equilibrium

In the full-information equilibrium, shoppers know the nominal wage. Let $v^*(w)$ denote the value of threshold in this equilibrium. From Proposition 1, it is given by,

$$\int\int_{\lambda(v^*(w)+p)}^{\infty}\big(\frac{1}{\lambda}\epsilon - p - v^*(w)\big)g(\epsilon)d\epsilon f(p|w)dp = \kappa \tag{17}$$

where $f(p|w)$ is the actual price distribution. The distribution of thresholds is reduced to a single value $v^*(w)$. It is easy to show that firm's profit is given by,

$$\pi_k = \frac{1}{\rho}\Big(1 - G\big(\lambda(v^*(w) + p_k)\big)\Big)\frac{W}{P_k}(P_k - \frac{W}{A_k}) \tag{18}$$

where $\rho = \int\big(1 - G(\lambda(v^*(w) + p))\big)f(p|w)dp$. The optimal pricing strategy is given by,

$$P_k = \frac{e_k}{e_k - 1}\frac{W}{A_k} \tag{19}$$

$$e_k = \lambda\frac{g\big(\lambda(v^*(w) + p_k)\big)}{1 - G\big(\lambda(v^*(w) + p_k)\big)} + 1 \tag{20}$$

Firm's market power is determined by two factors. Larger $\lambda$ implies more competition on prices, inducing higher elasticity.[8] In addition, the search friction naturally gives rise to monopoly power

[8] Anderson et al. (1987) shows that without search frictions, if $G$ is Gumbel distribution, the demand system is exactly CES and the elasticity of substitution is $\lambda + 1$.

(Diamond, 1971). In particular, the effect of search friction on the elasticity is represented by a hazard function. The density $g$ represents the number of marginal shoppers who are indifferent between making purchase and continuing searching that the firm will lose if it increases its price. The survival function $1 - G$ indicates that increasing prices will increase the profit obtained from all infra-marginal shoppers. The ratio between the two captures the trade-off when setting prices. Moreover, since $G$ is log-concave, the hazard function increases in its argument. As a result, the elasticity of demand increases in the threshold and firm's own price. It has two implications. First, when firms face pickier shoppers implied by the higher threshold, demand becomes more elastic. Second, high-productivity firms will set lower price and have higher markup. Next, I establish the existence of the full-information equilibrium.

**Theorem 1** (Existence of the Full-Information Equilibrium)**.** *There exists a unique full-information equilibrium in which consumers search actively.*

*Proof.* See Appendix A. ■

This theorem is proved in two steps. First, the elasticity of demand increases in price. It implies that higher price induces higher elasticity and lower markup. Therefore, the individual prices are uniquely determined by (19) and (20). Second, we can prove that $v^*(w)$ monotonically decreases with the search cost. There is a unique solution to (17) for the price distribution derived from optimal pricing strategy. The equilibrium is then the fixed point of the reservation value and the price distribution. Note that there always exist a continuum of equilibria in which firms charge sufficiently high prices and shoppers do not search. However, there is only one equilibrium in which shoppers search actively. From now on, I refer to the full-information equilibrium where $w = \bar{w}$ as the full-information steady state. The following theorem shows our first main result regarding monetary non-neutrality.

**Theorem 2.** *In the full information equilibrium, monetary policy is neutral. Monetary shocks are fully passed through to firms' prices.*

*Proof.* See Appendix A. ■

This theorem establishes that when shoppers know the monetary shock, the passthrough from the monetary shock to prices is complete. This is true because in equilibrium, each firm understands that, if itself and all its competitors raise their prices one-to-one with a monetary shock, although consumers do not know individual prices of each firm, they understand the relative prices are unchanged and thus there is no gain from searching more or less. In sum, search friction alone is not sufficient to generate monetary non-neutrality. We need the interplay of both search and information frictions.

## 2.4 Approximate Optimal Strategies

Under incomplete information, both the threshold shown in Proposition 1 and the pricing strategy shown in Proposition (2) are highly non-linear. To proceed, I consider monetary shocks $\hat{w}$ drawn from a distribution with small standard deviation, i.e., $\sigma_w \to 0$. At the same time, I keep signal-to-noise ratio $\sigma_w^2/\sigma_x^2$ fixed, which correspond to a fixed level of information friction, $\theta$. Then, I approximate optimal strategies to the first order around full-information steady state given its existence.

Let $\varphi$ denote the passthrough from the monetary shock $\hat{w}$ to prices. I refer to $\varphi$ as correlated-shock passthrough. It is defined in the steady state. Therefore, shoppers know the distribution of correlated-shock passthrough.[9] Unlike the standard demand system where an aggregate demand function is available and then the price index is naturally defined, in the model with search frictions, there is no obvious way to define a price index. However, the standard theory (Hulten, 1973, Hulten, 1978) offers a simple non-parametric formula for the *log change* in price index. If preference is homothetic and stable[10], which is the case here, the log change in the price index is the expenditure share-weighted log changes in all the prices:

$$\hat{p} = \Phi \hat{w} \tag{21}$$

where $\Phi = \int \varphi_k \bar{\omega}_k dk$ and $\bar{\omega}_k = \frac{1}{\bar{\rho}}\big(1 - G\big(\lambda(v^*(\bar{w}) + p_k)\big)\big)$ is the expenditure share in the full-information steady state. The aggregate correlated-shock passthrough $\Phi$ measures the passthrough from the monetary shock to the price index. Shoppers know $\Phi$ as they know the distribution of $\varphi_k$ and $\bar{\omega}_k$. Therefore, the average expectation of the change in price index is $\bar{E}\hat{p} = \Phi \bar{E}\hat{w} = \theta \hat{p}$. The following proposition presents the first-order approximation to the optimal search strategy in (10) and the optimal pricing strategy in (16).

**Proposition 3.** *Fix the variance ratio $\frac{\sigma_z^2}{\sigma_w^2}, \frac{\sigma_s^2}{\sigma_w^2}$. To the first order as $\sigma_w \to 0$,*
***[Part 1]*** *the threshold $v^*(x)$ is given by,*

$$v^*(x) = v^*(\bar{w}) - E(\hat{p}|x) \tag{22}$$

[9] The first-order approximation to the optimal pricing strategy $p^*$ is given by $p^*(a_k, w) = p^*(a_k, \bar{w}) + p^*_w(a_k, \bar{w})\hat{w}$.

[10] Homotheticity requires that the income elasticity of demand must equal one for each good. Stability requires that consumers adjust their spending only in response to changes in income and relative prices. Baqaee and Burstein (2023) show how to generalize the standard theory to more general preferences. In our case, according to 14, the demand for firm $k$ is proportional to the aggregate income $W = \int X_i di$, which implies homotheticity. In addition, there is no other factors, such as taste shocks, state of nature, other than relative prices and income, that can affect the demand.

*where $\Phi$ is the aggregate correlated-shock passthrough .*

***[Part 2]** The elasticity of demand is given by,*

$$e_k = \lambda \frac{g\big(\lambda(v^*(w) + \bar{p}_k + \hat{p}_k - \bar{E}\hat{p})\big)}{1 - G\big(\lambda(v^*(w) + \bar{p}_k + \hat{p}_k - \bar{E}\hat{p})\big)} + 1 \tag{23}$$

*where $\bar{E}\hat{p} = \theta\hat{p}$. In addition, $e_k$ increases in $\hat{p}_k - \bar{E}\hat{p}$.*

*Proof.* See Appendix A. ∎

This proposition shows that the threshold under incomplete information is equal to the threshold under full-information steady state, adjusted downward the expected change in the price index. Consistent with Proposition 1, the threshold is negatively related to $x$. The result takes a step forward by showing that under the first-order approximation, all the changes in perceived price distribution following the shock for shoppers with information set $x$ is encapsulated by $E(\hat{p}|x)$. Furthermore, the average change of thresholds is equal to $\bar{E}\hat{p} = \theta\hat{p}$, which implies that the adjustment in threshold is smaller due to information friction, leading to higher thresholds, i.e., pickier consumers, compared to the full-information benchmark.

The second part of the proposition characterizes the elasticity of demand when thresholds are on average higher than the full-information benchmark. On the first order, the distribution of thresholds and covariance between cash in hand and thresholds vanish to zero and only $\bar{E}\hat{p}$ is retained, which is inherited from the average change in thresholds. The result shows that the elasticity depends on the perceived relative price, $\hat{p}_k - \bar{E}\hat{p}$. Due to the information friction, it is larger than the actual relative price. Consequently, firms behave as though they are competing with rivals setting prices lower than their actual levels, prompting them to reduce their own prices. This increases the elasticity of demand and compressed markups. The intuition is closely related to Lucas (1972), where agents mistake increases in nominal prices for increases in relative prices. However, in this context, the mechanism operates entirely on the household side.

## 2.5 Characterization of Passthroughs

I now characterize passthroughs in the equilibrium. I present the main finding: the aggregate passthrough of a money supply shock to the price index is generically incomplete. The correlated-shock passthrough is composed of two fundamental passthroughs: the own-cost passthrough and the cross-price passthrough. Following Amiti et al. (2019), I define relevant passthroughs as follows.

**Lemma 1.** *The price responds to both firm's own cost shocks and competitors' prices,*

$$\hat{p}_k = \gamma_k(\hat{w} - \hat{a}_k) + \xi_k \hat{p} \tag{24}$$

*where $\gamma_k$ is the own-cost passthrough and $\xi_k$ is the cross-price passthrough.*

$$\gamma_k = (1 - \frac{d \log \mu_k}{dp_k}\Big|_{\hat{w}=0})^{-1}; \xi_k = \frac{d \log \mu_k}{dp}\Big|_{\hat{w}=0} \gamma_k \tag{25}$$

*where $\mu_k$ is the markup. The correlated-shock passthrough of each firm is given by,*

$$\varphi_k = \gamma_k + \Phi \xi_k \tag{26}$$

*The aggregate correlated-shock passthrough $\Phi$ is given by,*

$$\Phi = \frac{\Gamma}{1 - \Xi} \tag{27}$$

*where $\Gamma = \int \gamma_k \bar{\omega}_k dk$, $\Xi = \int \xi_k \bar{\omega}_k dk$.*

*Proof.* See Appendix A. ■

The lemma derives the own-cost and cross-price passthroughs using the markup elasticities, which, in equilibrium, can be obtained from 16. The monetary shock acts as both a shock to marginal costs and a shock to competitors' prices. Accordingly, the correlated-shock passthrough is defined as the combination of these two components. By integrating both sides, the aggregate correlated-shock passthrough is derived. Importantly, the own-cost passthrough $\gamma$ does not depend on the information friction $\theta$ as it is evaluated at the steady state. In contrast, the cross-price passthrough $\xi$ depends on how shoppers perceive the change in price index, which is shock-relevant. I now state our main results on passthroughs under incomplete information.

**Theorem 3.** *The model economy can be summarized by the following equation,*

$$\hat{p} = \Gamma \hat{w} + (1 - \Gamma) \bar{E}^s \hat{p} \tag{28}$$

*where $\bar{E}^s \hat{p}$ represents shoppers' average expectation of $\hat{p}$. The aggregate correlated-shock passthrough is given by,*

$$\Phi = \frac{\Gamma}{1 - \theta(1 - \Gamma)} \tag{29}$$

$\Phi < 1$ *for any* $\theta < 1$ *given* $\Gamma < 1$.

*Proof.* See Appendix A. ■

This is the main theorem of this paper. It first frames the model economy as a beauty contest game, where firms exhibit strategic complementarity in response to *shoppers'* beliefs about the price index. This is distinguished from the classic beauty contest game between firms (Lucas, 1972; Woodford, 2003), where strategic complementarity is in terms of *firms'* beliefs.

The theorem also establishes that the aggregate correlated-shock passthrough is generically incomplete if there exists any information and search friction. It relies on two key parameters in this model: the aggregate own-cost passthrough, which is the sufficient statics for the effect of search friction on price index, and the degree of information friction, which affects the strategic complementarity in pricing across firms. If $\kappa = 0$, this model reduces to the CES benchmark

(Anderson et al., 1987), where consumers observe all prices and match utilities costlessly. Then, $\Gamma = 1$, we achieve monetary neutrality. Similarly, when $\theta = 1$, we have $\bar{E}^s\hat{p} = \hat{p}$, also yielding monetary neutrality. Any case in between gives monetary non-neutrality.

Figure 1 illustrates the three passthroughs. The right panel shows that under full information, the decrease in the own-shock passthrough over productivity is exactly offset by the increase in cross-price passthrough. The resulting correlated-shock passthrough is always one for firms of any productivity. This result aligns with Theorem 2, where I prove the exact result. The decomposition shows the contribution of each passthrough to the correlated-shock passthrough for firms with different productivities. High-productivity firms pass through less of their own cost shocks and more of other firms' prices. Notably, Amiti et al. (2019) establishes that the correlated-shock passthrough is one for broad preferences including nested CES and first-order Kimball demand family. The theorem complements their results and emphasizes that the complete information about the price index is essential. Indeed, the complete knowledge of all prices is often assumed in these commonly used demand systems.

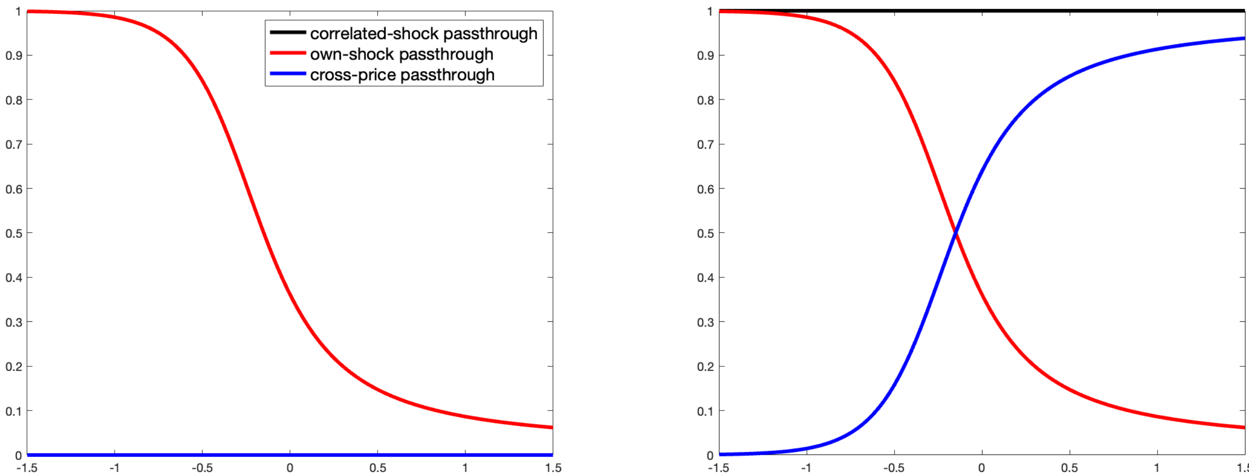
The left panel shows the case where shoppers have *zero* information about the shock. The cross-price passthrough is uniformly *zero* for any firm. As a result, the correlated-shock passthrough equals the own-cost passthrough. The intuition is that when shoppers do not perceive any change in the price index, firms are not able to pass any of the change in other firms' prices to their own prices. In this case, the monetary shock is as if an idiosyncratic cost shock to firms. In the intermediate cases, the increase in cross-price passthrough is still not sufficient to offset the decrease in own-shock passthrough, leading to incomplete correlated-shock passthrough.

Moreover, the individual correlated-shock passthrough is defined as $\varphi_k = \gamma_k + \Phi\xi_k$. The effect of strategic complementarity is further dampened by the aggregate correlated-shock passthrough as $\Phi < 1$. This induces a even smaller individual correlated-shock passthrough. Intuitively, the fact that shoppers understand $\Phi < 1$ reduces their expected change in price index for any given monetary shock. Firms then reduce response in pricing in response, resulting in even lower passthrough. Therefore, the incompleteness of the correlated-shock passthroughs is amplified by the firms' incentive of setting prices close to the price index, which is a form of real rigidity (Klenow and Willis, 2016).

### 2.6 Main Insight: Strategic Complementarity Attenuation

In Proposition 3, we have discussed that elasticity of demand for individual firms increases because shoppers perceive a dampened response of price index. Now, I "dig deeper" to show how such mechanism for *individual* firms relates to the attenuation on the strategic complementarity in pricing on the *aggregate* level. For this purpose, I will borrow heavily from game theory as in

Figure 1: Distribution of three types of passthroughs under incomplete and complete information




*Notes:* The figure plots distributions of passthroughs based on one calibration of the model. The red line represents own-cost passthrough. The blue line represents cross-price passthrough. The black line represents correlated-shock passthrough . The left and right panels show passthroughs in the incomplete and complete information cases, respectively.

Morris and Shin (2002), Woodford (2003), and Angeletos and Lian (2023). The following Corollary presents a way to translate the above economy into a beauty contest game.

**Corollary 1.** *In the rational expectations equilibrium, the change in price index can be express as the infinite sum of higher-order beliefs of $\hat{w}$,*

$$\hat{p} = \Gamma \sum_{h=0}^{\infty} (1-\Gamma)^h \bar{E}^h \hat{w} \tag{30}$$

*where shoppers' h-order average expectation $\bar{E}^h \hat{w} = \bar{E}(\bar{E}^{h-1}\hat{w})$, and $\bar{E}^0 \hat{w} = \hat{w}$. Using the fact that $\bar{E}^h \hat{w} = \theta^h \hat{w}$, the aggregate correlated-shock passthrough is given by,*

$$\Phi = \Gamma \sum_{h=0}^{\infty} (1-\Gamma)^h \theta^h \tag{31}$$

*Proof.* See Appendix A. ■

Equation (30) can be derived by iterating (28) in Theorem 3 forward[11]. We can interpret this infinite sum as the aggregate response of price index unfolding in the notional time within a period. In the first round, when the monetary shock occurs, assume no firm has yet responded to the monetary shock and that firms expect other firms not to adjust their prices. The shock acts as

[11] To write the game into the infinite sum of higher-order beliefs as in (30), we need (i) firms are rational and (ii) shoppers know that firms are rational and (iii) the common knowledge of rationality among shoppers and (iv) firms know the common knowledge of rationality among shoppers. In Appendix A, I show which rationality is required in each step of derivation. The rational expectations equilibrium is a “super” concept that includes all these rationality.

an idiosyncratic shock to each firm, resulting in a price increase of $\Gamma\hat{w}$. In the second round, it becomes common knowledge among firms that the price index has increased by this amount. Due to the strategic complementarity of price-setting, firms pass an additional portion of the shoppers' perceived increase in the price index from the first round onto their prices, specifically $(1-\Gamma)\bar{E}\Gamma\hat{w}$, under the belief that other firms will not further adjust prices. This iterative process continues. In the $h+1$th round, the additional passthrough is $(1-\Gamma)^h\bar{E}^h\hat{w}$. The intuition connects to the literature on level-$k$ thinking (García-Schmidt and Woodford, 2019; Farhi and Werning, 2019).

Moreover, the information friction attenuates the strategic complementarity in pricing. It attenuates more the high-order strategic complementarity, since the higher-order beliefs of shoppers are more anchored to the prior. For $\theta$ close to zero (meaning a sufficiently large departure from common knowledge of the shock), the aggregate correlated-shock passthrough is arbitrarily close to the aggregate own-cost passthrough. But as $\theta$ increases (meaning a higher degree of common knowledge of the shock), the higher-order strategic complementarity rises rapidly and thus increases overall passthrough. Therefore, by varying $\theta$ between 0 and 1, we can thus span all the values between the aggregate own-cost passthrough and the full-information outcome.

**Symmetry** − Theorem 3 relies on first-order approximation. The responses of price index is symmetric for both positive and negative monetary shocks by definition. In this section, I give interpretations about the incomplete passthrough for both cases.

Consider a positive monetary shock, such as a 1% increase in the nominal wage. In response, firms tend to raise prices. If firms increase prices by 1%, they risk losing marginal shoppers who were previously indifferent between making a purchase and continuing their search. Conversely, if firms do not raise prices at all, they forgo profits from infra-marginal shoppers by charging lower markups. To maximize profits, firms choose prices somewhere in between. In this case, Rotemberg (1982) provides the appropriate interpretation: firms, fearing to alienate marginal consumers, limit the passthrough of the shock.

Now consider a negative monetary shock, such as a 1% decrease in the nominal wage. If firms reduce prices by 1%, they attract more shoppers, who perceive themselves as fortunate to have found a high-productivity firm. On the other hand, if firms do not lower prices, they gain more profit from infra-marginal shoppers who are willing to pay higher markups. To maximize profits, firms opt not to fully decrease their prices. In this case, the appropriate interpretation is that firms limit the passthrough to extract greater profits from infra-marginal consumers.

## 2.7 Monotone Comparative Statics

In this section, I first show the cross-sectional properties of individual correlated-shock passthrough and limit results within an equilibrium. Then, I show comparative statics results on search and information frictions across equilibria. I define passthroughs on the primitive: $\varphi_k = \varphi(a_k)$,

$\gamma_k = \gamma(a_k)$ and $\xi_k = \xi(a_k)$. The results also depends on the specific functional assumption on the distribution of match utility, $G$. Following the literature, I assume $G$ is Gumbel distribution.

**Proposition 4.** *In a given equilibrium, markup $\mu(a)$ and cross-price passthrough $\xi(a)$ increases in productivity; own-cost passthrough $\gamma(a)$ and total passthrough $\varphi(a)$ decreases in productivity. Also, the following limit results hold:*

1. $\lim_{a\to\infty} \mu(a) = \infty$; $\lim_{a\to\infty} \varphi(a) = \Phi\theta$; $\lim_{a\to\infty} \gamma(a) = 0$; $\lim_{a\to\infty} \xi(a) = \theta$
2. $\lim_{a\to-\infty} \mu(a) = \frac{\lambda+1}{\lambda}$; $\lim_{a\to-\infty} \varphi(a) = 1$; $\lim_{a\to-\infty} \gamma(a) = 1$; $\lim_{a\to-\infty} \xi(a) = 0$

*Proof.* See Appendix A. ■

High-productivity firms exhibit lower own-cost passthrough because their lower prices attract marginal consumers with lower match utility. Given $G$ follows Gumbel distribution, these consumers are more sensitive to relative prices, leading to a lower passthrough of own-cost shocks. For similar reasons, high-productivity firms have strong incentives to keep up with competitors' prices, resulting in stronger strategic complementarity. In addition, high-productivity firms set lower prices, and higher markups. Similar to oligopolistic CES models, the lower bound of markups is determined by the degree of substitutability $\lambda$ between goods, while the upper bound approaches infinity. These results also align with with empirical evidence highlighted in the literature: (i) more productive firms charge higher markups (Amiti et al., 2014); (ii) more productive firms pass through less exchange rate shocks (Amiti et al., 2019).

Furthermore, because information frictions dampen only the cross-price passthrough, high-productivity firms experience a more pronounced decline in cross-price passthrough and, therefore, correlated-shock passthrough. Coupled with their high expenditure share, these firms disproportionately contribute to the incompleteness of aggregate correlated-shock passthrough. According to the limit results, the correlated-shock passthrough is bounded below by $\Phi\theta$. This implies that $\hat{p}_k - \bar{E}\hat{p}$ is always positive, resulting in uniformly higher perceived relative prices for all firms. Consequently, following a positive monetary shock, shoppers are more likely to search and passthroughs are incomplete for all firms.

Next, I present results on the comparative statics of the aggregate correlated-shock passthrough. I establish that, all else equal, (i) the correlated-shock passthrough is larger when the search cost is smaller for all firms, given $G$ is Gumbel distribution, and (ii) the correlated-shock passthrough is larger when information friction is smaller for all firms. In Appendix A, I show the results for own-cost and cross-price passthroughs.

**Proposition 5.** *If $\varphi(a;\kappa)$ is correlated-shock passthrough for given $\kappa$, then for $\kappa_2 > \kappa_1$, $\varphi(a;\kappa_2) < \varphi(a;\kappa_1)$ for $\forall a$, and $\lim_{\kappa\to\infty} \varphi(a) = 0$, $\lim_{\kappa\to 0} \varphi(a) = 1$; If $\varphi(a;\theta)$ is correlated-shock passthrough for given $\theta$, then for $\theta_2 > \theta_1$, $\varphi(a;\theta_2) > \varphi(a;\theta_1)$ for $\forall a$, and $\lim_{\theta\to 0} \varphi(a) = \gamma(a)$, $\lim_{\theta\to 1} \varphi(a) = 1$.*

*Proof.* See Appendix A. ■

This proposition establishes that, all else equal, correlated-shock passthrough is lower in an economy with higher search costs and information frictions for all firms. Two frictions are mutually reinforcing. higher search costs raise markups, making firms more sensitive to relative price changes and thus reducing both own-cost and correlated-shock passthrough. Conversely, when shoppers are better informed about the aggregate price level, firms respond more strongly to competitors' prices, increasing correlated-shock passthrough. Moreover, the limiting results show that passthrough can range from zero to one, allowing for a wide spectrum of monetary non-neutrality. In the dynamic model, shoppers will learn the shock over time. This proposition shows that if learning leads to the common knowledge of the shock, the monetary policy will be neutral in the long run.

Taken together, Theorem 3 establishes that the passthrough from monetary shocks to the price index is generically incomplete. Corollary 1 identifies the core mechanism: information frictions attenuate strategic complementarities in pricing. Proposition 4 and 5 further show that the incomplete passthrough is primarily driven by high-productivity firms. Plus, aggregate passthrough declines as search and information frictions intensify.

## 2.8 Extension to Multiple Goods

In this section, I discuss one extension. In particular, I allow multiple goods to be produced by one firm. The firm produces goods with different productivity for each good.

$$\log A_{kj} = \log A + a_k + a_{kj}$$

where $a_{kj}$ is the productivity of producing good j by firm k. It is i.i.d following $a_{kj} \sim \mathcal{N}(0, \tilde{\sigma}_{ap}^2)$. I assume that there is no search frictions when shopping within a firm. Shoppers decide which firm to purchase form and then buy the CES aggregation of all the goods in the firm. Let $P_k$ denote the CES price index of multiple goods in firm $k$. We have the following result.

**Proposition 6.** *Each firm charges same markup over all the products it sells.*

$$P_{kj} = \frac{e_k}{e_k - 1} \frac{W}{A_{kj}} \tag{32}$$

*where $e_k$ is the elasticity of demand uniform for all $j$. It is determined by*

$$P_k = \frac{e_k}{e_k - 1} \frac{W}{A_k}; e_k = \lambda \frac{\int X\big(g\big(\lambda(v^*(x) + p_k)\big)\big) d\Phi_x(x)}{\int X\big(1 - G\big(\lambda(v^*(x) + p_k)\big)\big) d\Phi_x(x))} + 1 \tag{33}$$

*where $P_k$ is the CES price index of $P_{kj}$. The passthrough of product-level productivity shocks $a_{kj}$ increases toward one when the number of products increases.*

*Proof.* See Appendix D. ■

This extension is consistent with the result in Hottman et al. (2016) and speaks to the empirical literature on passthrough of exchange rate shock to retail prices. Goldberg and Hellerstein (2013) and Nakamura and Zerom (2010) find complete passthrough of wholesale prices to retail prices for beer and coffee sales in retail stores.[12] The proposition shows that when there are many goods in one store, the passthrough of product-specific idiosyncratic shocks is closed to one. However, this does not mean that the passthrough of aggregate shocks and firm-specific shocks is complete.

# 3 Dynamic General Equilibrium Model

In this section, I present the full-fledged dynamic general equilibrium model, which extends the static model by incorporating a new agent's problem and nominal wage rigidity. I will dissect mechanisms of how each of component affects the impulse responses in detail in Section 4.3.

First, I close the model—where the nominal wage and cash in hand are exogenous—in general equilibrium by introducing a new agent: the worker. In particular, there is a representative household which consists of a single worker and a continuum of shoppers. The worker makes the labor supply and aggregate consumption spending decisions, leaving only the decisions about where to buy particular goods to the shoppers. I assume the worker has full information about the model economy, but there is no communication between the worker and the shoppers. Under this setup, a monetary shock can affect both nominal marginal costs and nominal aggregate demand, yet remain unknown to the shoppers. Second, following Christiano et al. (2005), I introduce the nominal wage rigidity, which is crucial to dampen the response of nominal marginal cost in this model and generate plausible impulse responses that are consistent with the data.

In addition, the dynamic model features persistent monetary shocks to the interest rate and slow learning about the underlying shock on the shopper side. For simplicity, I assume away security markets where shoppers can trade claims that are contingent on the search process and final choices.[13]

**Timeline –** The timeline is as follows. The period is divided into morning and afternoon. In the morning, the monetary authority sets the interest rate, and the worker makes decisions on the labor supply, the bond position, and the aggregate consumption spending. In the afternoon, firms post prices, and shoppers search sequentially and decide where to buy. They are constrained by the amount of cash allocated to them. The worker has full information and shoppers have incomplete information about monetary shocks.

[12] See Gopinath et al. (2011) for a summary of this literature.

[13] For example, a security might provide positive returns if a shopper experiences a long sequence of unfavorable draws or if the final choice only slightly exceeds the threshold. Mongey and Waugh (2024) show that the demand allocations in a standard discrete-choice model without search frictions can be different when the market is complete.

**Monetary Authority –** The monetary authority sets the nominal interest rate. It follows the Taylor rule,

$$i_t = \phi \hat{\pi}_t + \upsilon_{mt} \tag{34}$$

where $\upsilon_{mt}$ follows,

$$\upsilon_{mt} = \rho_m \upsilon_{mt-1} + \sigma_m \varepsilon_{mt} \tag{35}$$

where $\varepsilon_{mt}$ is the monetary shock and it follows $\varepsilon_{mt} \sim \mathcal{N}(0,1)$.

**Firm –** Let $A_{kt}$ denote the firm's productivity. It follows,

$$\log A_{kt} = \rho_a \log A_{kt} + \sigma_a \varepsilon_{akt} \tag{36}$$

where the idiosyncratic productivity shock follows $\varepsilon_{akt} \sim \mathcal{N}(0,1)$.

**Worker –** In each household, the worker maximizes expected discounted utility with a discount factor $\beta \in (0,1)$. Period utility is defined over the integration of values which shoppers will obtain from search in the afternoon, given by

$$\int \left( \log \left( \frac{X_{it}}{P_{k(i)t}} \right) + \frac{1}{\lambda} \epsilon_{ik(i)t} \right) di$$

where $k(i)$ denotes the firm selected by shopper $i$. Due to log utility, the level of consumption spending $X_{it}$ does not influence the shopper's search decision, allowing the worker's and shopper's problems to be separated. Also, I assume that the allocation of the aggregate consumption spending $X_t$ across shoppers is i.i.d across shoppers and time, according to the following rule,

$$X_{it} = X_t \exp\left(\sigma_x \varepsilon_{xit} - \frac{\sigma_x^2}{2}\right)$$

where $\varepsilon_{xit} \sim \mathcal{N}(0,1)$ and $\int \log X_{it} di = \log X_t$ by definition. Therefore, the worker only needs to determine the aggregate consumption $C_t$, implying aggregate consumption spending $X_t = P_t C_t$, and then transfer randomly the individual consumption spending to each shopper for them to decide where to buy goods.

To match the hump-shaped response of output, following Smets and Wouters (2007), I assume that the worker has external habits. The worker can also save in risk-free bonds $B_t$ (in zero net supply) that pay an interest rate of $R_t$. The worker's problem writes,

$$\max_{B_t, X_t, L_t} E_0 \sum_{t=0}^{\infty} \beta^t \left( \int \log(C_t - \lambda_c C_{t-1}) - \frac{L_t^{1+\eta}}{1+\eta} \right)$$
$$\text{s.t.} \quad P_t C_t + B_t = W_t L_t + R_{t-1} B_{t-1} + \Pi_t$$

where $L_t$ is the labor effort, $W_t$ is nominal wage, and $\Pi_t$ is total nominal profits of firms. The parameter $\lambda_c$ represents how consumption sticks to its last-period level. The optimization problem gives the first-order Euler equation as in Smets and Wouters (2007),

$$\hat{y}_t = c_1 \hat{y}_{t-1} + (1 - c1)\hat{y}_{t+1} - c_2(\hat{i}_t - \hat{\pi}_{t+1}) \tag{37}$$

where $c_1 = \frac{\lambda_c}{1+\lambda_c}$ and $c_2 = \frac{1-\lambda_c}{1+\lambda_c}$. Larger $\lambda_c$ not only dampens the response of output to the change in the real interest rate but anchor its level to the last-period output response. In Section 4.4, I consider an alternative setup following Auclert et al. (2020) to generate hump-shaped output response.

**Shopper** − In the afternoon, each shopper receives the consumption spending $X_{it}$ and solve the static problem in (9) period by period. In principle, they can infer the information that workers received in the morning from $X_{it}$. However, since $X_{it}$ is a signal about endogenous variables, incorporating it would significantly increase computational complexity, which is beyond the scope of this paper. For tractability, I set $\sigma_x \to \infty$ and only consider exogenous signals as follows.

Following Auclert et al. (2020) and Angeletos and Huo (2021), I assume that each worker receives noisy information about the monetary shock. In particular, Auclert et al. (2020) model consumer learning about persistent monetary shocks by decomposing the shock into a sequence of independent news shocks $\{\upsilon_{mt}\}_{t=0}^{\infty}$. In period $t$, workers receive independent private signals about $k$-period-ahead news shock $s_{it}^k = \upsilon_{mt}^{(k)} + \sigma_{st}^k u_{it}^k$ where $u_{it}^k \sim \mathcal{N}(0,1)$ for $k = 0, 1, 2, ....$[14] At date-0, suppose there is a $k$-period-ahead shock $\upsilon$. Then, applying standard Bayesian updating, workers' average belief about this shock at date $j \leq k$ is $\bar{\mathbb{E}}_j \upsilon = \frac{\sum_{i=0}^{j} \sigma_{si}^{-2}}{\sigma_m^{-2} + \sum_{i=0}^{j} \sigma_{si}^{-2}} \upsilon$. Following Auclert et al. (2020), I choose the sequence $\sigma_{si}$ such that $\bar{\mathbb{E}}_j \upsilon = 1 - (1-\theta)^j$, for $j = 1, 2, \ldots$, where $\theta$ is a constant representing the degree of information friction for workers at date 1. This time-varying sequence ensures that average beliefs about the monetary shock converge fast. In contrast, using a constant $\sigma_{sj}$ across all $j$ results in unrealistically slow learning. For example, if the signal-to-noise ratio is 1/9, then even at date 100, the average belief remains 9% below the true shock.

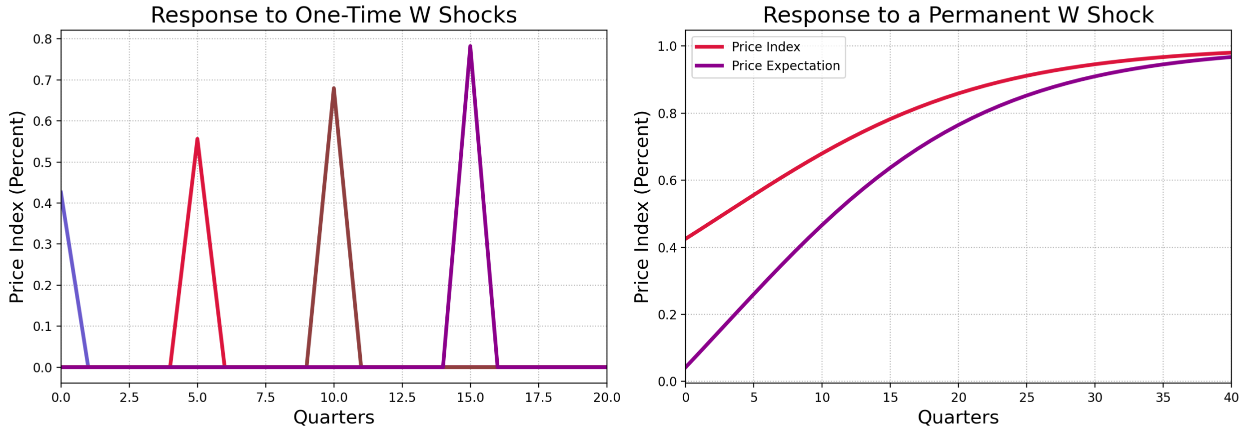
Stacking equation (28) across $t$ into a vector-valued equation and rearranging, we obtain,

$$\hat{\boldsymbol{p}} = \underbrace{\Gamma(\boldsymbol{I} - (1-\Gamma)\boldsymbol{\Theta})^{-1}}_{\boldsymbol{\Psi}} \widehat{\boldsymbol{w}} \tag{38}$$

where $\boldsymbol{\Theta}$ is a diagonal matrix with entries $1 - (1-\theta)^j$ for $j = 0, 1, 2, ....$ Following Auclert et al. (2024), I denote the passthrough matrix as $\boldsymbol{\Psi} = \Gamma(\boldsymbol{I} - (1-\Gamma)\boldsymbol{\Theta})^{-1}$. It characterizes the response of the price level to nominal wage changes.

[14] For tractability, I only consider the exogenous signals. Incorporating endogenous signals would significantly increase computational complexity, which is beyond the scope of this paper. However, I suspect the results will not change significantly.

Figure 2: Properties of the Passthrough Matrix $\boldsymbol{\Psi}$




*Notes:* The figure plots the responses of price index to one-time nominal wage shocks (left panel) and a permanent nominal wage shock (right panel) when $\theta = 0.1$. In addition, the average expectation of price index is plotted in the right panel.

The left panel of Figure 2 presents the responses of price index to one-time nominal wage shocks. Two properties are worth noting. First, this information-based Phillips curve exhibits no forward-looking behavior: firms pass through the nominal wage shock to prices contemporaneously, at the time the shock occurs[15]. Second, for shocks that materialize later, shoppers have more time to acquire information, resulting in a larger price index response. Both properties resemble the firm-side information-based Phillips curve discussed in Mankiw and Reis (2002). The right panel of Figure 2 shows the response of price index and average expectation of price index to a permanent nominal wage shock. When learning is slow, i.e., $\theta_s = 0.1$, the price index converges to 1 in about 10 years. The expectation initially undershoots significantly the price index and gradually converge to it over time. Following Auclert et al. (2024), I then derive the *generalized Phillips curve* by substituting $\hat{\boldsymbol{w}} = \hat{\boldsymbol{w}}^{\boldsymbol{r}} + \hat{\boldsymbol{p}}$, where $\hat{w}^r$ is the log-deviation of real wage, and solving the fixed point,

$$\hat{\boldsymbol{\pi}} = \underbrace{(\boldsymbol{I} - \boldsymbol{L})\Gamma\boldsymbol{\Psi}\left(\boldsymbol{I} - \boldsymbol{\Psi}\right)^{-1}}_{\boldsymbol{K}}\hat{\boldsymbol{w}}^{\boldsymbol{r}} \tag{39}$$

where $\boldsymbol{L}$ is the lag matrix with entries of 1 one line below the diagonal. Equation (39) maps any change in real wage $\hat{\boldsymbol{w}}$ to inflation. Matrix $\boldsymbol{K}$ represents the slope of the *generalized Phillips curve.*

However, the response of the price index to a 1% nominal wage shock is bounded below by $\Gamma = 0.4$. This lower bound is attained in the limiting case where shoppers receive *zero* information

[15]Buettner and Madzharova (2021) find a rapid passthrough of VAT-induced cost changes to prices within four months—two months before and two months after the reform. This finding contradicts the predictions of New Keynesian models, which suggests that firms would begin raising prices at least one year prior to the reform. In contrast, the model here predicts no price changes before the reform and full passthrough upon implementation, as these tax reforms are common knowledge to consumers, resulting in minimal price stickiness.

about the shock. In contrast, nominal rigidity à la Calvo—where only 25% of firms can adjust their prices each period—implies a more muted price index response of just 0.12. To generate a more realistic inflation response in the model, I introduce nominal wage rigidity. As shown in equation (28), dampening the response of nominal wages directly reduces the magnitude of the price index response.

**Labor Union –** Following Erceg et al. (2000) and Smets and Wouters (2007), the worker supplies labor to unions, which differentiate this labor into types of labor services indexed by $L_{jt}$. A competitive labor packer aggregates these differentiated services using a Kimball aggregator (Kimball, 1995), and sells the composite labor input to firms at a nominal wage $W_t$. Firms use the aggregated labor service $L_t$ in production, where aggregation satisfies:

$$\int_j G_w\left(\frac{L_{jt}}{L_t}\right) dj = 1, \tag{40}$$

with $G_w$ denoting the Kimball aggregator. The aggregator satisfies $G'_w(x) = \exp\left\{\frac{1-x^{\frac{\zeta_w}{\varepsilon_w}}}{\zeta_w}\right\}$, where $\varepsilon_w$ is the elasticity of substitution across labor types and $\zeta_w$ is the super-elasticity. Wages are set by unions subject to nominal rigidities à la Calvo: with probability $1-\lambda_w$, unions can re-optimize wages in each period, while otherwise wages are indexed to past inflation. Following Auclert et al. (2021), the union's optimization problem implies, to the first-order, the following wage Phillips curve:

$$\hat{w}^r_t = w_1(\hat{w}^r_{t-1} + \pi_{t-1}) + w_2(\mathbb{E}_t\hat{w}^r_{t+1} + \pi_{t+1}) - \pi_t - \kappa_w(\hat{w}^r_t - \eta\hat{l}_t - \hat{y}_t), \tag{41}$$

where $w_1 = \frac{1}{\beta}$, $w_2 = \frac{\beta}{1+\beta}$, and the slope parameter $\kappa_w = \frac{(1-\beta\lambda_w)(1-\lambda_w)}{\lambda_w}\frac{\varepsilon_w}{\varepsilon_w-1+\zeta_w}$. The slope $\kappa_w$ governs the extent to which the gap between the real wage and the marginal rate of substitution between leisure and consumption is reflected in the real wage.

**Reallocation and Aggregate Productivity –** A positive (negative) monetary shock induces more (less) search under incomplete information, with larger information frictions leading to even greater search intensity. Shoppers are more likely to be drawn to high-productivity, low-price firms, leading to an expansion in demand for these firms and a contraction for low-productivity firms. This reallocation of demand improves allocative efficiency and enhances aggregate productivity (Baqaee et al., 2024; Baqaee and Farhi, 2020). To see this formally, the first-order approximation of the aggregate production function implies:

$$\hat{y}_t = \hat{l}_t + (\hat{v}_t - \hat{p}_t) \tag{42}$$

where $\hat{v}_t = \hat{w}_t \int \bar{\alpha}_k \varphi_{kt}\, dk$ and $\bar{\alpha}_k = \frac{\bar{\omega}_k \bar{\mu}_k^{-1}}{\int \bar{\omega}_k \bar{\mu}_k^{-1}\, dk}$ represents the steady-state labor demand share, where $\bar{\mu}_k$ represents steady-state markup. Importantly, markups are not uniform across firms, so the change in aggregate productivity is nonzero.

To determine its sign, since high-productivity firms have higher markups, the weight $\bar{\alpha}_k$ effectively down-weights the incomplete passthrough $\varphi_{kt}$ for high-productivity firms and up-weights it for low-productivity firms. Given that $\varphi_{kt}$ is lower for more productive firms, it follows that:

$$\hat{v}_t/\hat{w}_t > \hat{p}_t/\hat{w}_t$$

Thus, a positive monetary shock induces an increase in aggregate productivity through endogenous reallocation of demand. Higher productivity dampens the response of price index and amplifies the monetary non-neutrality. Following the same method as computing $\hat{p}_t$, $\hat{v}_t$ is given by,

$$\hat{\boldsymbol{v}} = \underbrace{\tilde{\Gamma}(\boldsymbol{I} - (1-\tilde{\Gamma})\boldsymbol{M})^{-1}}_{\tilde{\boldsymbol{\Psi}}} \hat{\boldsymbol{w}} \tag{43}$$

where $\tilde{\Gamma} = \int \bar{\alpha}_k \gamma_k dk$ and $\tilde{\Gamma} > \Gamma$. Then, the aggregate production function is given by,

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{l}} + (\tilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi})\hat{\boldsymbol{w}} \tag{44}$$

**Equilibrium** – In the limit of small monetary shock, the equilibrium dynamics of variables $\{\hat{\pi}_t, \hat{p}_t, \hat{w}_t^r, \hat{w}_t, \hat{v}_t, \hat{c}_t, \hat{i}_t\}$ is governed by equations (37), (41), (44), (34), (43) and two equations that connect nominal and real variables: $\hat{w}_t = \hat{w}_t^r + \hat{p}_t$ and $\hat{\pi}_t = \hat{p}_t - \hat{p}_{t-1}$.

# 4 Calibration and Estimation

In this section, I first calibrate the aggregate own-cost passthrough and estimate the parameters governing search frictions. I also quantify the effect of demand reallocation on aggregate productivity. Next, I estimate the information friction by matching the empirical impulse responses from a structural VAR (SVAR) with those generated by the model.

## 4.1 Search Friction

The search cost in the model is described as a utility cost, which is hard to measure in the data. Fortunately, the aggregate own-cost passthrough is a sufficient statistic for "deep" parameters related to the search friction, i.e., search cost $\kappa$, relative importance of match utility $\lambda$, and standard deviation and persistence of idiosyncratic productivity $\sigma_a$ and $\rho$. Another property of aggregate own-cost passthrough is that it is evaluated at the steady state and it does not depend on the information friction. Therefore, we can separate the calibration of these two frictions.

Amiti et al. (2019) estimate the aggregate own-cost passthrough to be approximately 0.5 using comprehensive data on Belgian manufacturing firms. They use this value to calibrate the super-elasticity parameter in Kimball preferences. However, a caveat of their estimate is that buyer-seller relationships in tradable sectors often involve firms on both sides of the market, who are generally better informed about prices than households. In contrast, search and information frictions are

Table 1: Baseline Calibration of the Model

| Parameter | Description | Value |
|---|---|---|
| $\kappa$ | Search cost | 0.25 |
| $\lambda$ | Relative importance of match utility | 4.19 |

*Notes:* The table reports the calibrated values for parameters that are related to the aggregate own-cost passthrough.

Table 2: Model Fit

| | Moment | Model | Data | Source |
|---|---|---|---|---|
| **M1** | Average markup | 1.67 | 1.67 | DellaVigna and Gentzkow (2019) |
| **M2** | Aggregate own-cost passthrough | 0.4 | 0.4 | Amiti et al. (2019); Gopinath et al. (2011) |

*Notes:* The table summarizes the moments, model and data values of these moments, and the sources of the empirical values of these moments.

more pronounced in non-tradable sectors—such as retail and broad services—where firms directly engage with households.[16] Gopinath et al. (2011) use data from a retail chain containing wholesale cost information and report substantial heterogeneity in own-cost passthrough, with a median of approximately 0.5 for U.S. stores and 0.25 for Canadian stores. Since they cannot control for the pricing behavior of *all* the other firms, as Amiti et al. (2019) do, their estimates may be upward biased due to unobserved strategic complementarities from firms not in this dataset. Based on these findings, I calibrate $\Gamma = 0.4$. Robustness checks using alternative values of $\Gamma$ are presented in Appendix B.

Next, I quantify the effect of demand reallocation and also provide a sanity check of the implied values of the parameters governing the search friction. To proceed, I use the estimate of elasticity of substitution from DellaVigna and Gentzkow (2019). They use NielsenIQ Retailer Scanner database and find that the average elasticity of substitution across stores and products is 0.25, implying a markup of 1.67. In the data, one retailer sell a wide range of products. In one of the extension, I generalize the model to the setup where one firm sell multiple products and there is no search friction within the store. I show that retailers charge the same markup for all products they sell, which is consistent with Hottman et al. (2016). In addition, the idiosyncratic productivity dispersion $\sigma_a$ and persistence are set to be consistent with Decker et al. (2020), i.e., $\sigma_a = 0.3$, $\rho_a = 0.6$.

Table 1 summarizes the calibrated parameters, and Table 2 presents the fit of the model to data. Despite its parsimonious structure, the model is successful in matching key moments in the

[16] Amiti et al. (2019) show that, in their dataset, the sum of own-cost and cross-price passthroughs is not statistically different from one, suggesting that firms are more informed than households. This implies that the framework in this paper may not be suitable for such contexts. Gopinath and Itskhoki (2008) argue that in these settings, price rigidities are more plausibly driven by contractual frictions.

data. They have three implications. First, although the targeted elasticity of demand is 2.5, the implied $\lambda$ is just 4.2. Remember $\lambda + 1$ represents the elasticity of demand when $\kappa \to 0$ as shown in Proposition 4. This suggests that the search friction accounts for a substantial bulk of the market power. Second, based on the calibrated values of deep parameters, we can infer the average search per shopper $\rho^{-1}$ equal to 1.37. This suggests a relatively large search cost and about 73% shoppers make the purchase in the first round of search. Therefore, Assumption 1 is plausible, as shoppers only visit a limited number of firms on average. Finally, the implied $\tilde{\Gamma}$ based on calibrated parameters is 0.426, which is larger than $\Gamma$ as expected. However, the magnitude is very small. The reallocation effect is therefore second-order.

## 4.2 Information Friction

I first introduce data and present the empirical findings. Then, I perform impulse response matching to estimate the information friction.

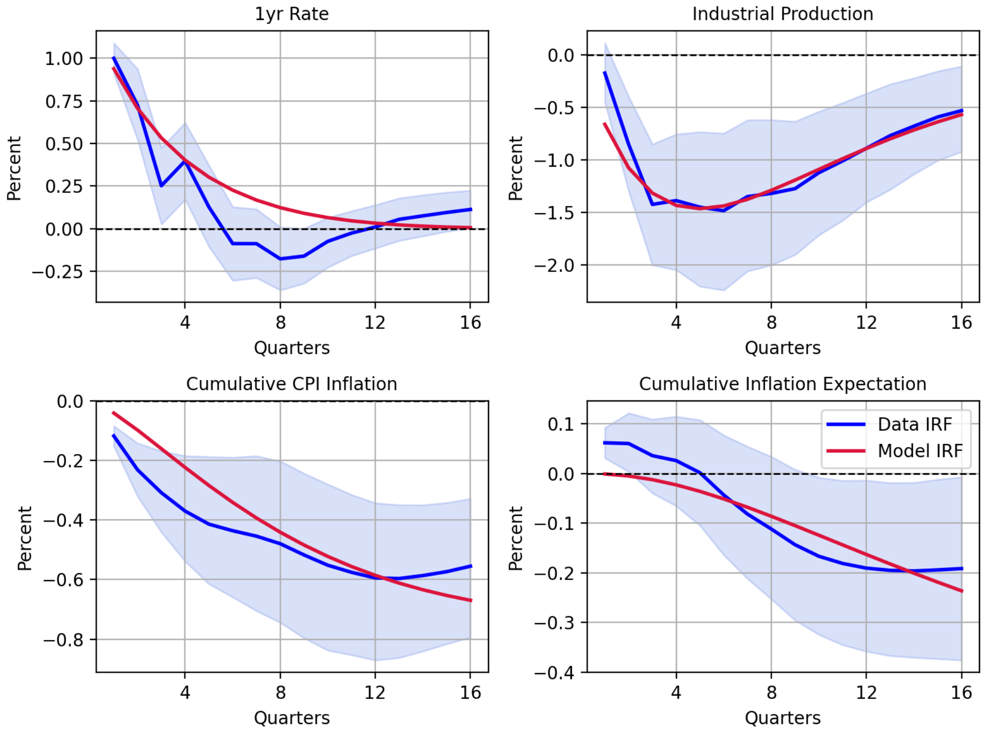
**Data and Method –** The dataset used in the main specification consists of monthly observations on five key macroeconomic variables: the industrial production index; the consumer price index (CPI) including food and energy; consumer inflation expectations from the Michigan Survey of Consumers; the excess bond premium; and the 1-year government bond rate. The sample spans from January 1978—the earliest date for which inflation expectation data are available—through December 2019.

I include the CPI measure that incorporates food and energy prices, motivated by a large body of evidence showing that households form inflation expectations based on salient and frequently encountered prices. For example, Kumar et al. (2015) show that households are more attentive to highly visible prices, such as those for gasoline. Similarly, D'Acunto et al. (2021) find that consumers learn about inflation through their retail shopping experiences. To ensure consistency between the actual inflation series and what households observe and internalize in forming expectations, I therefore use the CPI measure that includes food and energy components. I exclude real wage and hours from the empirical specification because the model abstracts from investment and intermediate inputs, making the model-implied measures of real wage and hours not directly comparable to their empirical counterparts.

The Michigan survey of Consumers only reports the yearly inflation expectation. To approximate the monthly expectation, I divide this annual figure by 12, which yields the average expected monthly inflation over the next year. Taking the cumulative sum of this monthly series provides an estimate of the expectation of log price index.[17] Similarly, I use the actual inflation to construct

[17] Note that this constructed series is not identical to the expected log price index, as it relies on the average monthly inflation expectation rather than the true expected monthly inflation rates.

Figure 3: Empirical and Baseline Model Impulse Responses




*Notes:* The figure plots the impulse responses for 4 variables following a 100 bp shock on the 1-year government bond rate. The sample period is 1978M1–2019M12. The blue lines represent the empirical impulse responses and the red lines represent the model counterparts. The shaded areas are 68% confidence intervals derived from wild bootstrap. The red line represents the impulse responses implied in Section 3. The x-axis denotes quarters from the shock, starting at 0. The y-axis denotes percent.

an estimate of the log price index. I will use the constructed cumulative inflation and inflation expectation in the SVAR.[18]

I begin by estimating the reduced form of a monthly five-variable VAR, using twelve lags for each variable. The five reduced-form residuals are modeled as linear combinations of five underlying structural shocks. The goal is to identify how the structural shocks to monetary policy affect the contemporaneous reduced form residuals. Once these shocks are identified, the VAR can be used to trace out their dynamic effects on the macroeconomic variables. To identify exogenous variation for the monetary shocks, I employ external instruments for the one-year government bond rate based on high-frequency surprises in futures market prices around FOMC announcements following the approach in Gertler and Karadi (2015) and Bauer and Swanson (2023). I use the constructed monetary policy shock series provided by Bauer and Swanson (2023).

[18]Appendix B shows that the impulse responses of log price index and this estimate are nearly indistinguishable.

**Empirical Findings –** Figure 3 presents the main results. The blue lines depict the empirical impulse responses of four macroeconomic variables following a 100 basis point increase in the one-year government bond rate. Industrial production exhibits a hump-shaped response, consistent with existing literature (Gertler and Karadi, 2015), declining by approximately 150 bps within the first year. Cumulative inflation falls on impact by about 20 bps and continues to decline, reaching a total reduction of 60 bps over four years. In contrast, cumulative inflation expectations rise by 8 bps on impact[19], then gradually decline, falling by 20 bps over the same horizon. These empirical patterns have two key implications. First, inflation expectations systematically undershoot realized inflation, supporting the core assumption of the model—namely, that consumers possess incomplete information about monetary shocks. Second, consumer learning appears remarkably sluggish: even after four years, consumers have internalized only about one-third of the total decline in cumulative inflation. This implies the presence of substantial information frictions, reflected in a low value of the parameter $\theta$. A simple back-of-the-envelope calculation illustrates this: $1 - (1-\theta)^{16} = 1/3$, which yields $\theta = 0.022$. I will show in what follows that this provides a useful reference point for the structural estimate of $\theta$ obtained through impulse response matching.

**Impulse Response Matching –** Following Christiano et al. (2005), I use impulse response matching to estimate 4 parameters: nominal wage rigidity $\lambda_w$, shoppers' average information friction $\theta_s$, workers' average information friction $\theta_w$, and persistence of monetary shock $\rho_m$. Collecting them in the vector $\Lambda = \{\lambda_w, \theta_w, \theta_s, \rho_m\}$. Let $\boldsymbol{J}, \widehat{\boldsymbol{J}}$ denote model and empirical impulse responses. The estimator $\hat{\Lambda}$ solves,

$$\min_{\Lambda} \left(\boldsymbol{J}(\Lambda) - \widehat{\boldsymbol{J}}\right)' \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{J}(\Lambda) - \widehat{\boldsymbol{J}}\right)$$

$\boldsymbol{\Sigma}$ is a diagonal matrix that has confidence interval of empirical impulse responses in its entries. I also give 5 times more weights to inflation expectation as it is crucial in estimating $\theta$.

Table 3 summarizes the calibrated parameters. The aggregate own-cost passthrough as the associated passthrough matrix that describes the effect of demand reallocation are taken from Section 4.1. In addition, I follow the standard procedure in the literature (Smets and Wouters, 2007) of assuming a wage markup of 1.5 which implies the elasticity across labor types equal to 3, and a Kimball superelasticity for wages of $\zeta_w = 10$.

The model-implied impulse responses at the estimated parameters are shown in red in Figure 3. Overall, the model provides a close fit to the empirical counterparts. In Appendix B I present results from estimating a Calvo model, where the wage rigidity and the persistence of the monetary policy shock are fixed at the values reported in Table 3. The Calvo parameter is estimated to be around 0.9. The model yields a good fit for the inflation and output responses. However, it fails to

[19] The fact that the inflation expectation increases at the beginning upon an negative monetary shock is a puzzle. Providing an explanation to it is beyond scope of this paper.

Table 3: Calibrated and Estimated Parameters.

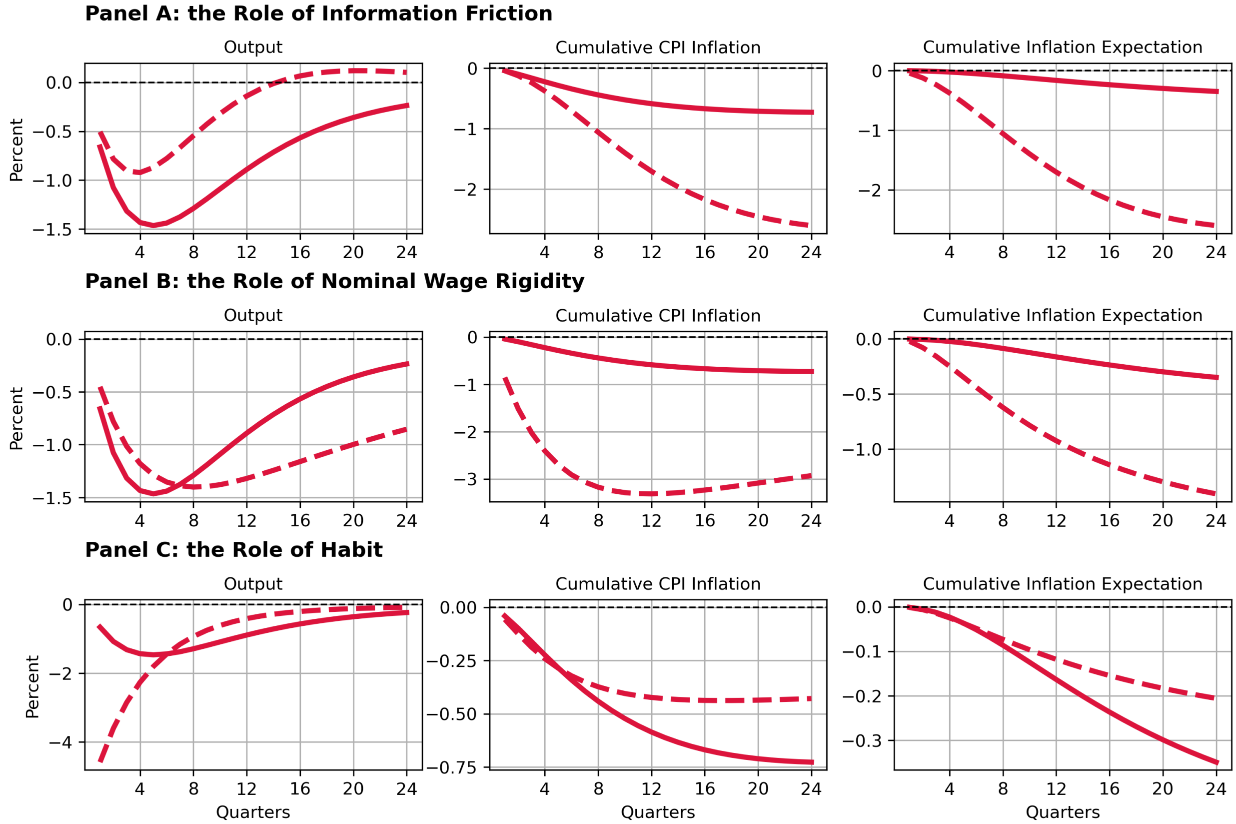
| Panel A: Calibrated parameters | | | Panel B: Estimated parameters | | |
|---|---|---|---|---|---|
| **Param.** | **Description** | **Value** | **Param.** | **Description** | **Value** |
| $\sigma$ | EIS | 1 | $\theta$ | Information friction | 0.027 |
| $\eta$ | Frisch elasticity | 0.5 | $\lambda_c$ | Workers' habit | 0.852 |
| $\beta$ | Discount factor | 0.998 | $\lambda_w$ | Calvo wage stickiness | 0.932 |
| $\phi$ | Taylor rule coefficient | 1.5 | $\rho_m$ | Persistence of monetary shock | 0.792 |
| $\varepsilon_w$ | Elasticity across labor types | 3 | | | |
| $\lambda_w$ | Kimball Superelasticity | 10 | | | |
| $\Gamma$ | Aggregate own-cost passthrough | 0.4 | | | |
| $\tilde{\Gamma}$ | Demand reallocation (Table 1) | 0.426 | | | |

capture the response of inflation expectations. This is expected: under full information, inflation expectations move one-for-one with actual inflation. As such, the Calvo model lacks the mechanism necessary to generate the observed wedge between expected and actual inflation. In contrast, the ability of our model to jointly match both series is clearly a success. The estimated information friction parameter for shoppers show a large degree of information friction, $\theta = 0.027$, implying an average duration of incomplete information of 37 quarters. This high level of friction is necessary for the model to replicate the muted response of inflation expectations, which empirically decline by only about one-third of the change in actual inflation in 4 years. This is consistent with the simple back-of-envelop calculation derived before.

The wage rigidity reported in Table 3 is $\lambda_w = 0.932$, indicating a higher degree of stickiness than that found in Christiano et al. (2005) and Smets and Wouters (2007). However, this parameter is not identified by directly fitting the impulse response of real wages, since marginal cost in this model reflects not only labor costs but also broader components of production costs, e.g., capital. Moreover, in the absence of capital and variable capital utilization, the estimated wage rigidity also captures the stickiness of other prices not explicitly modeled. Instead, it is estimated by matching the response of cumulative CPI inflation. Its high value reflects the relatively muted inflation response observed in the data. This estimate is consistent with Auclert et al. (2020), who calibrate wage and price rigidities to align with empirical impulse responses to monetary shocks identified using external instruments from Romer and Romer (2004). These externally identified responses tend to be more dampened compared to those derived from traditional Cholesky decompositions, thereby requiring greater nominal rigidity in the model to achieve a good empirical fit.

In sum, wage rigidity is primarily identified from the response of cumulative inflation; information frictions are disciplined by the relative responses of cumulative inflation and inflation expectations; habit formation is essential for generating the hump-shaped output response; and

Figure 4: Decomposing the Transmission Channels of Monetary Policy




*Notes:* For all three panels, the solid line plots the baseline impulse responses at the estimated parameters in Table 3 as the benchmark. In Panel A, the dashed line represents impulse responses when $\theta = 1$, holding all other parameters at their estimated values. In Panel B, the dashed line represents impulse responses when $\theta_w = 1$. In Panel C, the dashed line represents impulse responses when $\lambda_c = 0$.

the persistence of the monetary shock governs both the persistence and magnitude of the output response.

## 4.3 Monetary Policy Transmission Channels

In this section, I discuss monetary policy transmission channels in the dynamic general equilibrium model. In particular, I focus on the role of information friction, nominal wage rigidity and habit formation in contributing to the model-implied impulse responses.

**The Importance of Information Friction** − To understand the importance of information friction, I switch off the information friction for shoppers in the model by setting $\theta_s = 1$, holding all other parameters at their estimated values. Panel A in Figure 4 presents the results. The solid line represents the impulse response generated from the baseline model. The dashed line shows the responses when shoppers have full information. As expected, the decline in the price index becomes substantially larger—nearly five times greater than in the baseline model. Moreover, expectations of the price index now track the actual price level exactly. As a result, the degree of monetary non-neutrality is significantly reduced, with the output response shrinking by more than 50 bps in its peak.

**The Importance of Wage Rigidity –** To assess the role of wage rigidity, I turn off wage rigidity in the model by setting $\lambda_w = 0$. As shown in Panel B, the absence of wage rigidity leads to an immediate and pronounced decline in the price index, about 100 bps on impact, despite the presence of large information friction. It keeps declining to 300 bps and after about two years, inflation turns positive, causing the price level to rebound and the real interest rate to rise. As a result, the decline in output is even more severe after two years under full wage flexibility.

The underlying mechanism is as follows. When wages are flexible, nominal marginal costs fall sharply in response to the shock. However, shoppers' expectations of the price level adjust slowly and remain largely unchanged. As shown in equation 28, firms have an incentive to set prices close to the expected price index. Therefore, as the initial deflationary force dissipates, firms begin to raise prices in response to the inertia in price expectations. In contrast, with wage rigidity, nominal marginal costs decline more gradually, giving shoppers' expectations time to adjust. This reduces the divergence between expected and actual prices, thereby smoothing out the inflation dynamics. These results confirm the role of wage rigidity in preventing a sharp change in marginal costs (Christiano et al., 2005) and generating reasonable inflation responses.[20]

**The Importance of Habit Formation –** As is well known, habit formation not only generates a hump-shaped response of output but also contributes to greater persistence in its dynamics. To isolate this effect, I shut down habit formation in the worker's problem by setting $\lambda_c = 0$. As shown in Panel C, the absence of habit leads to a more pronounced decline in output on impact, followed by a rapid and exponential decay. Inflation also becomes less persistent, resulting in a more muted path for cumulative inflation and faster convergence of inflation expectations. Importantly, habit formation does not interfere with the model's core mechanisms related to information frictions. Its sole purpose is to help the model replicate the empirically observed hump-shaped output response.

Taken together, information friction, wage rigidity, and habit formation generate substantial monetary non-neutrality, and jointly match the impulse responses of macroeconomic variables and inflation expectation.

[20] Christiano et al. (2016) derive wage inertia from their specification of how firms and workers negotiate wages. They argue that this micro-foundation outperforms a variant of the standard New Keynesian Calvo sticky wage model.

Table 4: Robust Calibration of the Model

| Parameter | Description | Value |
|---|---|---|
| $\theta$ | Shoppers' information friction | 0.026 |
| $\theta_w$ | Workers' information friction | 0.104 |
| $\lambda_w$ | Calvo wage stickiness | 0.931 |
| $\rho_m$ | Persistence of monetary shock | 0.856 |

### 4.4 Robustness

In this section, I incorporate the information friction in the worker's problem to account for the hump-shaped response of output as in Auclert et al. (2020) and re-estimate the model. I show the results are similar for the key parameters.

There is a continuum of households which consists of a worker and a shopper. Following Auclert et al. (2020), I assume workers face information frictions about monetary shocks as well, when making decisions on consumption, labor supply and bond position. In principle, the worker and the shopper within one household can have different information sets. The information structure of workers are similar to shoppers described in Section 3, which is characterized by a single parameter $\theta_w$. The optimization problem gives the usual first-order Euler equation with information friction-enhanced MPC (marginal propensity to consume) matrix. Following the sequence-space Jacobian approach developed by Auclert et al. (2021), we have,

$$\hat{\boldsymbol{y}} = \boldsymbol{M^r}\hat{\boldsymbol{r}} + \boldsymbol{M^y}\hat{\boldsymbol{y}} \tag{45}$$

where $\hat{\boldsymbol{y}}$ is vector of output over time. $\hat{\boldsymbol{r}}$ is vector of real interest rate over time, where $\hat{r}_t = \hat{i}_t - \hat{\pi}_{t+1}$. Matrix $\boldsymbol{M^r}$ and $\boldsymbol{M^y}$ are information-friction modified Jacobians, which measures partial and general equilibrium responses of output to real interest rate shocks. In Appendix C, I discuss these Jacobians in detail. Other parts of the setup are the same as in the baseline model.

The hump-shaped output response arises in this setup because, under incomplete information, workers underestimate the true increase in the real interest rate. In a representative-agent setting, where output responds only through intertemporal substitution, such misperception leads to muted adjustments in consumption spending and thus output. As workers gradually learn about the shock, output begins to decline more; however, by that point, the shock dissipates, giving rise to a hump-shaped response.

Table 4 reports the calibrated parameters under the new model specification. As shown in Figure C.1 in Appendix C, the model continues to provide a good fit to the empirical impulse responses. Notably, the parameters governing shoppers' information frictions and nominal wage rigidity remain nearly identical to those in the baseline calibration. The reason for this stability is that for any given path of output, the dynamics of the real wage are pinned down solely by the degree of wage rigidity, as implied by equation (41).[21]. Then, from equation (39), inflation dynamics are directly determined by the path of the real wage. Consequently, regardless of how the model replicates the path of output, if it can provide a perfect fit to the empirical impulse response ideally, the estimated wage rigidity and shoppers' information friction should not change. Again, they jointly determine the dynamics of inflation and inflation expectation.

The estimated information friction parameter implies a 10-quarter average duration of information friction, which is close to that reported by Auclert et al. (2020), who similarly estimate household attention within a sticky expectation framework (Carroll et al., 2018) and find an average duration of inattention of about 14 quarters. Additionally, the estimated persistence of the monetary policy shock is higher than in the baseline specification. This is because habit formation amplifies the persistence of output responses, while worker-side information frictions primarily delay the timing of the response.

To know the role of information friction for workers, I switch off the information friction for workers in the model by setting $\theta_w = 1$. The results are shown in Figure C.2 Panel B. In this case, the output declines sharply on impact, and returns exponentially. Moreover, eliminating worker-side information frictions nearly doubles the responses of both cumulative inflation and inflation expectations. This highlights the additional role of $\theta_w$ in dampening inflation dynamics. However, the mechanisms by which the two types of information frictions affect inflation are different. Worker-side frictions reduce inflation primarily by weakening the intertemporal substitution channel. In contrast, shopper-side frictions directly dampen the degree of strategic complementarity in firms' pricing decisions, which has a quantitatively larger effect on inflation dynamics, as demonstrated in both Panel A and Panel B in Figure C.2.

# 5 Aggregate Productivity Shocks

In this section, I extend the model to incorporate aggregate productivity shocks. I show the superior properties of inflation and output responses generated from this model compared to the standard Calvo model with wage rigidity and habit formation.

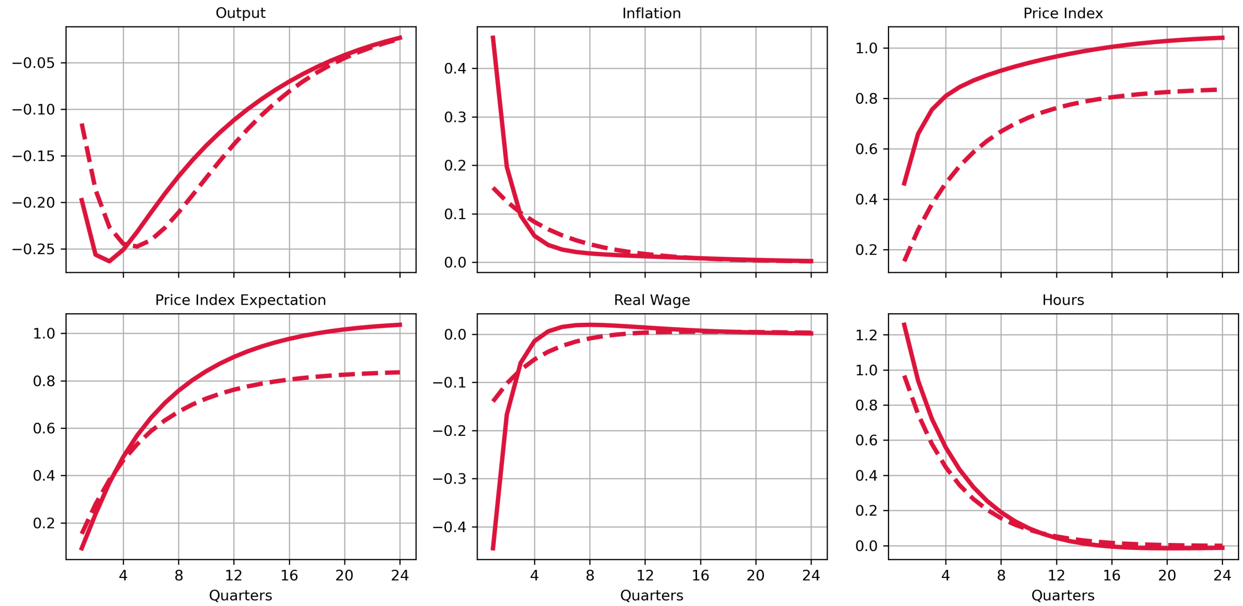
**Setup and Impulse Responses –** The framework can also incorporate the aggregate supply shocks. In particular, in the static model, the firm's productivity has two components,

$$\log A_k = \log A + \sigma_a \varepsilon_{ak} \tag{46}$$

where $A$ is the aggregate productivity shock. Let $a = \log A$. It draws from $\mathcal{N}(0, \sigma_A^2)$. Shoppers receive noisy signals that can inform them about the aggregate productivity shock.

[21]The demand allocation effect is small. Therefore, $\hat{l}_t \approx \hat{y}_t$

Figure 5: Model Impulse Responses After a Productivity Shock




*Notes:* The figure plots the impulse responses for 6 variables following a 100 bp shock on the aggregate productivity. The solid lines represent the impulse responses calibrated to values in Table 1 and $\theta = 0.2$, $\rho_a = 0.8$. The dashed lines represent the Calvo model with wage rigidity and habit formation in preferences. The x-axis denotes quarters from the shock, starting at 0. The y-axis denotes percent.

Since both the nominal wage and the aggregate productivity affect prices only through the marginal costs, the optimal pricing strategy is homogeneous of degree zero in $\{W, A\}$, which implies,

$$p^*(a_k, w, a) = p^*(a_k, w - a) \tag{47}$$

Any positive change in aggregate productivity acts equivalently to a proportional decrease in the nominal wage. Therefore, I can similarly define $\hat{p}_k = -\varphi_k \hat{a}$ and aggregate correlated-shock passthrough $\Phi$.

The intuition mirrors that of monetary shocks. Following a negative productivity shock—such as an increase in input prices due to supply chain disruptions—shoppers with incomplete information may misattribute the resulting price increases to idiosyncratic productivity shocks. This misperception induces them to search more intensively for alternative firms. In response, firms face a trade-off: raising prices aggressively risks losing marginal consumers, while holding prices steady sacrifices profits on infra-marginal consumers. The resulting optimal pricing strategy, therefore, features incomplete passthrough of the productivity shock.

The setup for the dynamic general equilibrium model under aggregate productivity shocks is detailed in Appendix D. Figure 5 displays the impulse responses following a 100 basis point negative productivity shock, holding wage rigidity and habit formation fixed at the values reported in Table 1. Motivated by evidence from Kumar et al. (2015), who document that consumers are highly responsive to fluctuations in salient prices such as gasoline and food, I calibrate the shoppers' information friction parameter to $\theta = 0.2$, implying 5-quarter average duration of information friction. For illustration, I also set the persistence of productivity shock to $\rho_a = 0.8$. The solid lines in Figure 5 represent this baseline calibration. For comparison, I also consider an enhanced Calvo pricing model incorporating wage rigidity and habit formation, with a Calvo parameter of $\lambda_p = 0.25$, implying 4-quarter average duration of price cell. Under this specification, the output and hours responses are broadly similar to those in the baseline model.

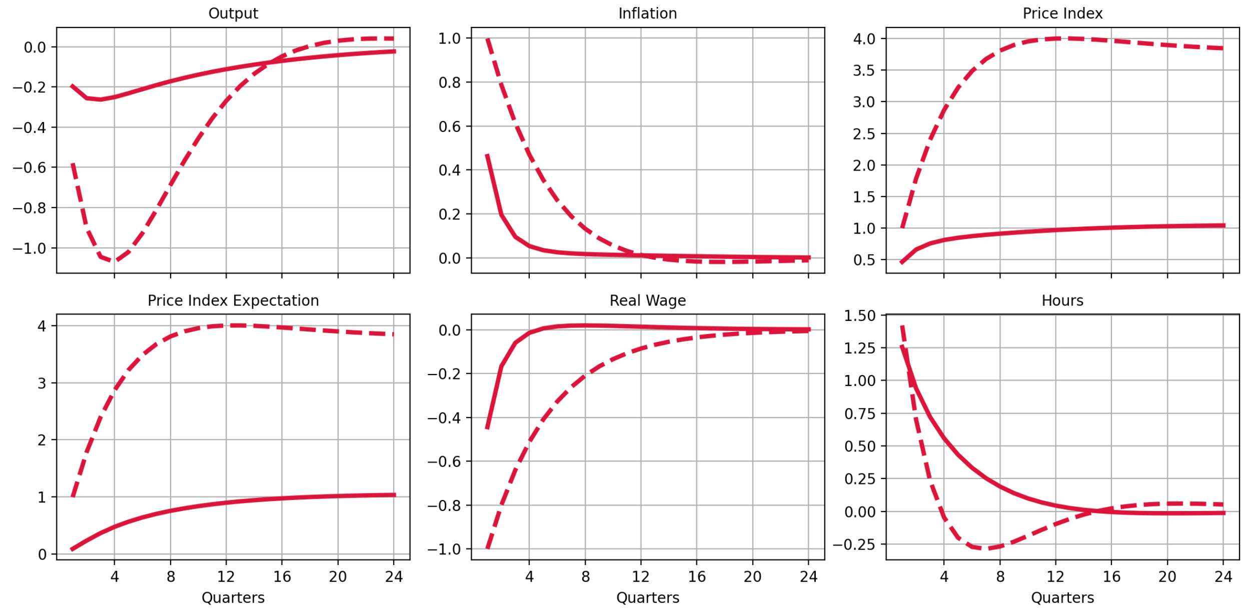
However, inflation dynamics differ notably between the two models. In the baseline case, inflation rises sharply by 50 bps on impact but quickly returns to zero. In contrast, the Calvo model generates a more muted inflation response—about 15 bps—followed by a slower decline. This reflects differences in the transmission of marginal cost shocks to prices. In the baseline model, information frictions reduce strategic complementarities in pricing, while wage rigidity dampens only part of the nominal marginal cost. Consequently, even if wages remain unchanged, the overall nominal marginal cost still rises by 100 bps due to the exogenous productivity shock, leading to a price increase of roughly 40 bps—consistent with what is observed in Figure 5. By contrast, the Calvo model directly attenuates the passthrough of the entire marginal cost, including both wage and productivity shock, to prices, resulting in a much muted inflation response. In addition, since firms anchor their prices around expected price index, the muted response of consumers' expectation of price index results in the faster decline in inflation as the shock dissipates in the baseline model.

Taken together, these results highlight the flexibility of the baseline model in jointly matching two empirically observed regimes: (i) dampened responses of inflation and inflation expectations, alongside significant monetary non-neutrality following monetary shocks; and (ii) pronounced but less persistent inflation response, and muted output responses, following productivity shocks.

**Application to the Post-Pandemic Inflation** – A puzzle has emerged in the dynamics of inflation and unemployment in the post-pandemic era. Inflation, likely triggered by supply chain disruptions, surged immediately after the pandemic and remained persistently elevated—contrary to the predictions of a flat Phillips curve estimated using pre-COVID data Hazell et al., 2022. Several theories have been proposed to reconcile these observations. For instance, Benigno and Eggertsson (2023) introduces a non-linear New Keynesian Phillips curve in which labor becomes substantially more costly when the vacancy-to-unemployed ratio crosses a critical threshold. In this paper, I offer an alternative explanation rooted in the framework developed herein.

Specifically, the historically high levels of media coverage and public discourse surrounding supply chain disruptions and inflation likely reduced information frictions, enabling households to

Figure 6: Model Impulse Responses After a Productivity Shock




*Notes:* The figure plots the impulse responses for 6 variables following a 100 bp shock on the aggregate productivity. The solid lines represent the impulse responses calibrated to values in Table 1 and $\theta = 0.2$, $\rho_a = 0.8$. The dashed lines represent the Covid case by setting $\theta = 1$. The x-axis denotes quarters from the shock, starting at 0. The y-axis denotes percent.

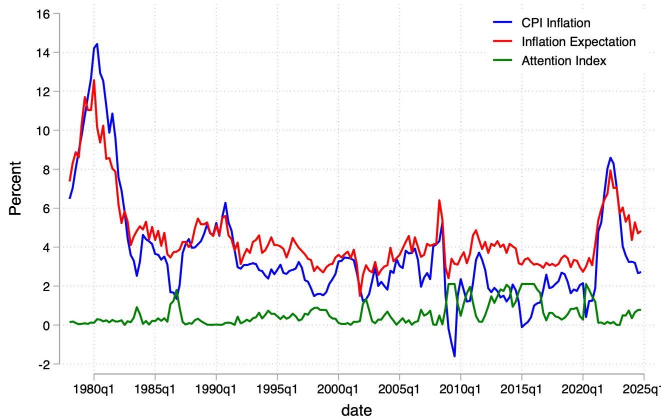
more rapidly internalize the rising production costs faced by firms. Supporting this view, Beaudry et al. (2024) and Hazell (2024) show that realized inflation and household inflation expectations from the Michigan Survey of Consumers moved in close alignment during this period, suggesting that households had relatively full information about the underlying supply shock.[22] To account for this possibility, I consider a counterfactual calibration in which shoppers are assumed to have full information, i.e., $\theta = 1$.

Figure 6 illustrates the impulse response when information in the model is highly precise. The increase in inflation is almost doubled, which aligns with the drastic inflation surge at the onset of the supply chain disruptions. The output also drops more than 5 times of the baseline level. However, the output did not decline significantly in reality compared to the model's predictions, likely due to the concurrent fiscal stimulus (Blanchard and Bernanke, 2023).

The main policy lesson from this application is that the slope of the Phillips curve is endogenous to the level of information available on the consumer side. Conventional Phillips curves, derived from reduced-form assumptions such as infrequent price adjustments and menu costs, are vulnerable

[22] Hazell (2024) further demonstrate that other inflation expectation measures—such as those from the Survey of Professional Forecasters, one-year-ahead inflation expectations derived from swaps, and backward-looking adaptive expectations (e.g., the fourth lag of inflation)—do not exhibit the same degree of sensitivity as household expectations.

Figure 7: Inflation, Inflation Expectation and Inflation Attention Index




to the Lucas Critique. Certain shocks, in particular, may trigger a sequence of public events, media reports, discussion in the public domain, and policy announcements, which significantly increase households' awareness of the inflation, resulting in a steeper slope of Phillips curve. The monetary authority should adopt a more aggressive stance when household inflation expectation is more accurate, as prices are more responsive to the shock in such environments.

# 6 Empirics

In this section, I provide two pieces of empirical evidences to support the key mechanisms of the model: (i) more attention to inflation increases the slope of Phillips curve and (ii) unanticipated inflation increases consumer search activities.

## 6.1 Information Frictions and the Slope of Phillips Curve

Inflation and attention to inflation are inherently endogenous: elevated inflation tends to attract greater public attention, whereas low and stable inflation often results in diminished attention. To circumvent this endogeneity, this section estimates the spillover effect of exogenous shifts in attention to national inflation on the local Phillips curve after controlling for the time fixed effect.

**Inflation Attention Index** – To construct the inflation attention index, I take the quarterly CPI including food and energy prices from FRED and inflation expectations from the Michigan Survey of Consumers from 1978Q1 to 2024Q4 as in Section 4. Inflation attention index captures the distance between inflation and inflation expectation,

$$\text{Index}_t = \left| \frac{\text{actual inflation}_t - \text{expected inflation}_t}{\text{actual inflation}_t} \right| \tag{48}$$

A lower value of this index indicates more household attention to inflation. However, the index may yield disproportionately large values when actual inflation is close to zero, as occurred during the global financial crisis and around 2015. To address this issue, I winsorize the index at the 95th percentile, so the index remains roughly between 0 to 2. Table 5 shows the descriptive statistics of the index. The inflation attention index after 1990 is approximately three times higher than its pre-1990 level for both mean and median.

Table 5: Descriptive Statistics of Inflation Attention Index

| Period | Mean | Median | Std. Dev. |
|---|---|---|---|
| 1978–2019 | 0.56 | 0.34 | 0.58 |
| 1978–1990 | 0.26 | 0.17 | 0.34 |
| 1990–2019 | 0.69 | 0.48 | 0.62 |

*Notes:* This table reports summary statistics for the inflation attention index over different time periods. The index measures the absolute percentage deviation between actual and expected inflation. A lower value indicates greater attention to inflation. The data span from 1978Q1 to 2019Q4.

Figure 7 displays the time series of actual inflation, expected inflation, and the resulting attention index. As shown, inflation expectations track actual inflation relatively closely during periods of elevated inflation, such as the late 1970s and early 1980s, as well as during the recent post-pandemic surge. The inflation expectation does not align well with inflation for periods in between. This is especially true during the financial crisis. The inflation is negative but the inflation expectation does not decrease that much. In addition, the inflation expectation is often higher than inflation in these periods. The index interprets it as inattention to inflation.[23]

**Empirical Specification –** I use state-level inflation and unemployment rate data collected in Hazell et al. (2022). I closely follow and extend the empirical specifications in their paper. Let $\kappa$ denote the slope coefficient of forward sum (truncated at T = 20) of unemployment rates in the regional Phillips curve for nontradeables, and let $\zeta$ denote the coefficient on the interaction between this forward sum and the inflation attention index. The baseline specification is given by:

$$\pi_{it}^N = \alpha_i + \gamma_t - \kappa(\sum_{j=0}^{T} \beta^j u_{i,t+j}) - \zeta(\text{Index}_t \times \sum_{j=0}^{T} \beta^j u_{i,t+j}) - \lambda(\sum_{j=0}^{T} \beta^j \hat{p}_{i,t+j}^N) + \tilde{\omega}_{it}^N + \eta_{it}^N \tag{49}$$

where $\pi_{it}^N$ is the regional inflation of nontradeables in state $i$ at quarter $t$; $\gamma_t$ is the time fixed effect; $\hat{p}_{i,t}^N$ is the relative price of nontradeables. $\eta_{it}^N$ is an expectations error. I set the discount factor $\beta = 0.99$ as in Hazell et al. (2022). Hazell et al. (2022) propose two identification strategies, which I adopt and extend. The first strategy instruments the forward-looking sums of unemployment and

[23]An alternative interpretation is that households may report higher inflation expectations because their consumption baskets differ from the CPI-weighted basket. However, this explanation is less compelling in light of the close tracking observed during high-inflation periods, suggesting that the degree of attention rather than basket composition is the more plausible driver.

relative prices using their respective four-quarter lags: $u_{i,t-4}$ and $\hat{p}^N_{i,t-4}$, respectively. The second strategy instruments the forward sum of unemployment rates using tradeable demand spillovers. I extend both approaches by also instrumenting the interaction term $\text{Index}_t \times \sum_{j=0}^{T} \beta^j u_{i,t+j}$ with the interaction of the inflation attention index and the respective instrument. For instance, under the first approach, the interaction term is instrumented with $\text{Index}_t \times u_{i,t-4}$.[24]

The identification is achieved as follows. First, under rational expectations, both the lagged state-level unemployment rate and tradeable demand shocks are orthogonal to the state-level expectation error. Second, the inclusion of time fixed effects plays a dual role: it controls for the evolution of long-run inflation expectations, as emphasized in Hazell et al. (2022), and more critically, it accounts for the influence of contemporaneous macroeconomic conditions on the inflation attention index. This is particularly important given the high correlation between inflation and attention to inflation as shown in Figure 7. Consequently, conditional on time fixed effects, the coefficient $\zeta$ can be interpreted as the causal spillover effect of national-level inflation attention on the slope of the regional Phillips curve.

Supporting this interpretation, Baker et al. (2022) document that a substantial share of the nearly 3,500 local newspapers in their database regularly report on both state and national economic conditions. This media structure implies that heightened attention to national-level inflation—especially during salient inflationary episodes—can generate spillover effects to state-level inflation awareness, as households consume both sets of information from a common source.

**Results** − Table 6 presents the regression results. Columns (1) and (3) replicate the baseline estimates from Table 1 of Hazell et al. (2022), while columns (2) and (4) report results from our extended specification in equation (49). Across both identification strategies, the interaction term estimates are negative, indicating that a higher value of the inflation attention index—corresponding to lower household attention to inflation—is associated with a flatter Phillips curve. The estimate based on lagged unemployment instruments is statistically significant, whereas the estimate using the tradeable demand iv is not. This difference is likely due to data availability: in the latter case, all states are missing data during the early years (1978–1980), a period that may contain important variation for identifying the effect of attention. Consequently, I focus on the estimate from the first approach thereafter.

To interpret the economic magnitude, consider two scenarios. First, applying the estimates to the pre- and post-1990 periods—when the average attention index rises from 0.26 to 0.69—the slope of the regional Phillips curve declines by approximately 16%. Second, applying the same logic to the post-pandemic period, during which the attention index declined from 2 to 0, implies a 75%

[24] Appendix D reports results from an alternative specification in which unemployment and relative prices follow AR(1) processes, as described in Equation (19) of Hazell et al. (2022).

Table 6: The Effect of Attention to Inflation on The Regional Phillips Curve

| | Lagged unemployment | | Tradeable demand IV | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\kappa$ | 0.0062** | 0.0088*** | 0.0062** | 0.0072** |
| | (0.0028) | (0.0032) | (0.0025) | (0.0034) |
| $\zeta$ | | -0.0033** | | -0.0015 |
| | | (0.0016) | | (0.0027) |
| State FEs | $\surd$ | $\surd$ | $\surd$ | $\surd$ |
| Time FEs | $\surd$ | $\surd$ | $\surd$ | $\surd$ |
| Observations | 4490 | 4490 | 4241 | 4241 |

*Notes:* The table reports the estimates in specification (49). Each observation is at the state×quarter level covering from 1978Q1 to 2019Q4. The coefficient represents the slope of regional Phillips curve and the effect of attention to national-level inflation to the slope. State and time fixed effects are controlled. Standard errors are calculated using two-sample 2SLS developed by Chodorow-Reich and Wieland (2020). Standard errors are clustered at the state level. *Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level.

increase in the slope. While this effect may appear modest in absolute terms given that $\kappa$ is already small, it is important to emphasize that the estimate captures the spillover effect of attention to *national-level* inflation on the slope of *state-level* Phillips curves—not the direct effect of attention to *state-level* inflation. Therefore, the estimated effect can be considered economically meaningful.

## 6.2 Unanticipated Inflation and Search Activities

In this section, I provide evidence on the effect of unanticipated inflation on search activities. I first briefly describe the data, then propose a new measure of search activity, and lastly I present the results.

**Data –** The data source is the NielsenIQ Consumer Panel Data set.[25] The sample period is 2006Q1 - 2019Q4. NielsenIQ tracks the shopping behavior of average 55,000 households every year. Each household uses in-home scanners to record purchases. Households also record any deals used that may affect the price. These households represent a demographically balanced sample of households in 49 states and about 3,000 counties in the United States. Each household stays in the panel for 30 quarters on average. The dataset has over 1,000 NielsenIQ-defined product modules

[25] Researcher's own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

that are organized into over 100 product groups and covers around 30% of total US expenditures on food and beverages and 2% of the total household consumption.[26]

The dataset contains information about each shopping trip the household takes, such as the retailer, the spending on each product defined as a barcode, the product module and group that the product belongs to, and the date of transaction. Moreover, the data includes households' demographic information such as age, education, employment, marital status, which are updated annually. I aggregate the dataset to the household-quarter level. I only consider households live in the Metropolitan Statistical Area.[27]

**Measurement of Search Activities –** The search protocol described and characterized in the model simplifies the actual shopping process. In practice, shoppers tend to make the majority of their purchases at a primary retail store, while visiting other stores for specific needs. For instance, during the sample period, shoppers allocate an average of 65% of their total spending on a given product group to the store they visit the most frequently. Additionally, shoppers often purchase multiple items in a single trip, and most stores offer a wide range of product groups. As a result, search activities may manifest as a reallocation of spending across stores for the same product group. For example, a shopper may initially buy milk from store A and cheese from store B but, after a shock, begin purchasing both milk and cheese from store A while reducing cheese purchases from store B. This suggests that simply counting the number of trips and the number of distinct stores shopper visits may not fully capture the search effort, even though it is the definition of a search in the model.

To account for the empirical patterns of shopping, I construct the measure of search effort as follows. First, for each consumer and each product group, I find the store she visits most frequently in the entire sample period. I call that store the *routine store*. Second, if she does not buy from the routine store or she buys from multiple stores for this product group, I record the total spending that she spends on the stores other than the routine store. Repeat this process for all the product groups. Finally, I define the non-routine share of spending as the ratio of all the non-routine spending over all groups and the total spending in the given quarter. The non-routine share captures the *extra* search effort to look for the products that shoppers usually buy from the routine store. Figure **??** in Appendix D shows the time trend of the non-routine share of spending. It is highly correlated with the average shopping time derived from American Time Use Survey. This makes us confident that this measure captures correctly the search activities.

[26]For further discussion of the NielsenIQ data, see Broda and Weinstein (2010) and Jaravel (2019).

[27]People living in the countryside may exhibit different search behaviors compared to those in urban areas. They are more likely to engage in "one-stop" shopping due to limited store availability.

I also construct other two measures of search activities. The first is the number of trips each household takes in a quarter. Second, I measure the number of distinct retail stores visited by consumers in a quarter. Both measures captures some aspects of our primary measure. Table 1 presents the summary statistics for these search measures. On average, consumers' non-routine share of spending is 35.6%. They make about 3.5 shopping trips per week, visit approximately 12 different retailers per quarter. The variance of these measures is substantial. This is consistent with our model in which search is primarily driven by idiosyncratic match utility shocks and productivity shocks, which may have substantial variance.

Table 7: Descriptive Statistics of Household Search Behavior

| | Mean | S.D. | 10th Percentile | 90th Percentile | Observations |
|---|---|---|---|---|---|
| Non-routine Share(%) | 35.6 | 25.2 | 3.23 | 71.1 | 2,846,354 |
| Number of Trips | 40.5 | 29.3 | 11 | 80 | 2,846,354 |
| Number of Distinct Retailers | 11.7 | 7.27 | 4 | 22 | 2,846,354 |

*Notes:* The table reports summary statistics for key household search behavior variables in the sample. *Non-routine Share* is the share of non-routine spending in a quarter. *Number of Trips* is the total number of shopping trips per quarter per household. *Number of Retailers* is the distinct retailers visited by a household per quarter.

**Unanticipated Inflation –** Following the literature, I assume that consumers use historical data to forecast the current inflation. In particular, consumers estimate a simple OLS regression of inflation on four lags of the inflation for food and drinks and unemployment rate. The residual from this regression is the unanticipated inflation, which has a mean of -7.8 bp and a standard deviation of 51 bp in the sample period. I focus on inflation for food and drinks because the majority of NielsenIQ products fall into this category.[28] Additionally, I consider the measure based on the inflation for overall goods and services. The unanticipated inflation in this case has a mean of -7.5 bp and a standard deviation of 76 bp. I normalize the two series of unanticipated inflation, so their units are standard deviation in our sample. Appendix C provides further details of regression and discusses several robustness checks.

**Empirical Specification –** To assess the changes in search behavior after an unanticipated inflation shock, our baseline specification is:

$$y_{it+1} = \lambda_i + \beta\tilde{\pi}_t + X_{it} + e_{it} \tag{50}$$

where $t$ is time; $i$ represents consumer. $\lambda_i$ is the consumer fixed effect. $\tilde{\pi}_t$ is the unanticipated inflation. $\beta$ is the coefficient of interest. It measures the magnitude of the correlation between

[28] I do not use the main inflation shock since it is not correlated with the inflation in food and drinks. If any, there is insignificant negative relationship between these two time series. See Appendix C for details.

Table 8: Non-routine Share of Spending and Unanticipated Inflation

| *Dep. var.*: Non-routine Share | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Unanticipated inflation (F&D) | 0.265*** | | 0.253*** | 0.234*** |
| | (0.023) | | (0.023) | (0.023) |
| Unanticipated inflation (overall) | | 0.022 | | |
| | | (0.018) | | |
| Number of trips | | | 0.033*** | -0.028*** |
| | | | (0.002) | (0.003) |
| Number of distinct stores | | | | 0.327*** |
| | | | | (0.011) |
| Observations | 2,660,735 | 2,660,735 | 2,660,735 | 2,660,735 |
| Consumer fixed effect | √ | √ | √ | √ |
| Consumer varying effect | √ | √ | √ | √ |

*Notes:* The table reports the estimates in specification (50). Each observation is at the consumer×quarter level covering from 2006 Q1 to 2019 Q4. The coefficient represents the corresponding change in different measures of search behavior after a standard deviation increase in unanticipated inflation. Consumer fixed and time-varying effects are controlled. Standard errors are clustered at the consumer level. *Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level.

the unanticipated inflation and the consumers' search behavior. $y_{it+1}$ is the non-routine share of spending in the next quarter. I use next-period value for two reasons. First, it avoids reverse causality because inflation and consumer search behavior are jointly determined in theory. Second, it may take time for consumers to change their shopping habits. $X_{it}$ is the time-varying consumer controls. These controls include consumer age, employment, education, marital status, having children or not, and consumer i's total spending in time t. As pointed out by Aguiar and Hurst (2007), these variables have large affect on the pattern of shopping behavior.

The results are presented in Table 8. The first column indicates that one standard deviation (51 bp) increase in unanticipated food and drink inflation leads to a 26.5 bp increase in the non-routine share of spending. This suggests that consumers respond to higher prices by engaging in more active search, shifting purchases to stores outside their routine stores. The magnitude of the response is modest, which is about a half (26.5 bp/51 bp). Given the information friction in Table 1, small magnitude may indicates large search frictions.

The second column uses the unanticipated overall inflation. It shows a much smaller increase in the non-routine share and not statistically significant. The effect is smaller because the NielsenIQ data primarily covers food and beverages, and search behavior is more sensitive to inflation in these sectors. The third and fourth columns introduce controls for the number of shopping trips and the number of distinct stores visited. The coefficients decrease only slightly, suggesting these variables capture a very limited portion of search effort. This suggests that consumers allocate most of their search efforts to substituting products within the same categories across the stores they have already visited, rather than increasing trips or visiting new stores.

The estimated coefficient may be biased downward for several reasons. First, consumers only record purchases from stores included in the NielsenIQ dataset, which predominantly covers large retail stores. As a result, our measure may not fully capture the non-routine share of spending if consumers switch to stores not included in the dataset or to online purchases. Second, substantial substitution within a product group could contribute to the bias. For example, consumers may trade down to lower-quality goods within the same store (Jaimovich et al., 2019) after an increase in inflation. However, we do not observe the time spent in each shopping trip. Finally, the inflation shock may not be entirely passed on to retail prices, as suppliers may absorb part of the shock in their wholesale costs. This incomplete pass-through could further dampen the observed relationship between inflation and non-routine spending.

Overall, this evidence supports a key aspect of the main mechanism: as prices rise after an aggregate shock, consumers are incentivized to search for alternatives. The evidence indicates the response of search activities to unanticipated inflation is statistically significant. There are potentially two reasons for the magnitude of the response being modest. One is that the search friction is large, and another is that the estimate is biased downwards.

# 7 Conclusion

This paper develops a new framework for understanding monetary non-neutrality, driven entirely by consumer-side frictions. At the heart of the model is the information asymmetry about nominal marginal costs between consumers and firms. Specifically, the framework integrates a heterogeneous firm block and incomplete consumer information into a standard sequential search model. When consumers observe a price increase, they attribute it to adverse productivity shocks rather than increases in nominal wages, prompting them to search for outside options. Firms, in turn, internalize this consumer search behavior, limiting the passthrough of cost changes to prices. The framework also accommodates aggregate supply shocks, providing a toolbox for analyzing a wide range of shocks as in the standard New-Keynesian model. Future research can investigate the effect of fiscal policy in this model.

I further present a dynamic general equilibrium model that incorporates both habit formation and wage rigidity. To match the extremely gradual adjustment in household inflation expectations observed in the data, the estimated model features significant information frictions. These frictions amplify monetary non-neutrality and generate much dampened inflation responses. The model can be extended to include aggregate productivity shocks. In response to such shocks, the inflation path in the model displays a sharp initial increase followed by a rapid decline in contrast with the predictions of standard New-Keynesian models. Overall, the model successfully replicates key empirical features of impulse responses to both monetary and productivity shocks.

This paper also provides preliminary empirical validation of the model's key mechanisms. I construct an inflation attention index and show, using state-level inflation and unemployment data, that increased attention to national inflation is associated with a steeper slope in the regional Phillips curve. Additionally, using large-scale consumer panel data, I document that unanticipated inflation significantly increases consumers' search intensity, consistent with the model's predictions.

Several further topics of inquiry are left for future research. First, although this paper focuses on final goods markets, the framework can be extended to any market with many sellers and buyers, such as upstream and downstream firms in supply chains. An interesting extension would be embedding this model into production network models. Second, applying the model to the labor market could yield valuable insights. Workers' incomplete information about the average posted wage could influence their job-search decisions as documented in the recent literature. Finally, to better calibrate the model and assess the quantitative importance of the mechanism, incorporating both firm- and consumer-side frictions into the model is necessary.

## Bibliography


Aguiar, M. and Hurst, E. (2007). Life-cycle prices and production. *American Economic Review*, 97(5):1533–1559.

Alvarez, F. E., Lippi, F., and Paciello, L. (2016). Monetary shocks in models with inattentive producers. *The Review of Economic Studies*, 83(2):421–459. Published: 29 October 2015.

Amiti, M., Itskhoki, O., and Konings, J. (2014). Importers, exporters, and exchange rate disconnect. *American Economic Review*, 104(7):1942–1978.

Amiti, M., Itskhoki, O., and Konings, J. (2019). International shocks, variable markups, and domestic prices. *The Review of Economic Studies*, 86(6):2356–2402.

Anderson, S. P., Palma, A. D., and Thisse, J.-F. (1987). The ces is a discrete choice model? *Economics Letters*, 24(2):139–140.

Anderson, S. P. and Renault, R. (1999). Pricing, product diversity, and search costs: A bertrand-chamberlin-diamond model. *The RAND Journal of Economics*, 30(4):719–735.

Angeletos, G.-M. and Huo, Z. (2021). Myopia and anchoring. *American Economic Review*, 111(4):1166–1200.

Angeletos, G.-M. and Lian, C. (2023). Dampening general equilibrium: Incomplete information and bounded rationality. In Handbook of Economic Expectations, V. ., editor, *Handbook of Economic Expectations*, pages 613–645. Elsevier.

Auclert, A., Bardóczy, B., Rognlie, M., and Straub, L. (2021). Using the sequence-space jacobian to solve and estimate heterogeneous-agent models. *Econometrica*, 89(5):2375–2408.

Auclert, A., Rigato, R., Rognlie, M., and Straub, L. (2024). New pricing models, same old phillips curves? *The Quarterly Journal of Economics*, 139(1):121–186.

Auclert, A., Rognlie, M., and Straub, L. (2020). Micro jumps, macro humps: Monetary policy and business cycles in an estimated hank model. Technical Report 26647, National Bureau of Economic Research.

Baker, S. R., Davis, S. J., and Levy, J. A. (2022). State-level economic policy uncertainty. *Journal of Monetary Economics*, 132:81–99.

Baqaee, D. R. and Burstein, A. (2023). Welfare and output with income effects and taste shocks. *The Quarterly Journal of Economics*, 138(2):769–834.

Baqaee, D. R. and Farhi, E. (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, 135(1):105–163.

Baqaee, D. R., Farhi, E., and Sangani, K. (2024). The supply-side effects of monetary policy. *Journal of Political Economy*, 132(4).

Bauer, M. D. and Swanson, E. T. (2023). A reassessment of monetary policy surprises and high-frequency identification. *NBER Macroeconomics Annual*, 37(1):87–155.

Beaudry, P., Hou, C., and Portier, F. (2024). The dominant role of expectations and broad-based supply shocks in driving inflation. Working Paper 32322, National Bureau of Economic Research.

Benigno, P. and Eggertsson, G. B. (2023). It's baaack: The surge in inflation in the 2020s and the return of the non-linear phillips curve. Working Paper 31197, National Bureau of Economic Research.

Binetti, A., Nuzzi, F., and Stantcheva, S. (2024). People's understanding of inflation. Working Paper 32497, National Bureau of Economic Research. Revision Date June 2024.

Blanchard, O. J. and Bernanke, B. S. (2023). What caused the us pandemic-era inflation? Working Paper 31417, National Bureau of Economic Research.

Broda, C. and Weinstein, D. E. (2010). Product creation and destruction: Evidence and price implications. *American Economic Review*, 100(3):691–723.

Buettner, T. and Madzharova, B. (2021). Unit sales and price effects of preannounced consumption tax reforms: Micro-level evidence from european vat. *American Economic Journal: Economic Policy*, 13(3):103–134.

Burdett, K. and Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica*, 51(4):955–969.

Burdett, K. and Menzio, G. (2018). The (q,s,s) pricing rule. *The Review of Economic Studies*, 85(2):892–928.

Bénabou, R. and Gertner, R. (1993). Search with learning from prices: Does increased inflationary uncertainty lead to higher markups? *The Review of Economic Studies*, 60(1):69–93.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.

Candia, B., Coibion, O., and Gorodnichenko, Y. (2023). The macroeconomic expectations of firms. In Manski, C. F. and Thomas, F., editors, *Handbook of Economic Expectations*, pages 321–353. Elsevier, Amsterdam.

Carroll, C. D., Crawley, E., Slacalek, J., Tokuoka, K., and White, M. N. (2018). Sticky expectations and consumption dynamics. Working Paper 24377, National Bureau of Economic Research.

Chodorow-Reich, G. and Wieland, J. (2020). Secular labor reallocation and business cycles. *Journal of Political Economy*, 128(6):2245–2287.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1, Part A of *Handbook of Macroeconomics*, pages 65–148. Elsevier, Amsterdam.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.

Christiano, L. J., Eichenbaum, M. S., and Trabandt, M. (2016). Unemployment and business cycles. *Econometrica*, 84(4):1523–1569.

D'Acunto, F., Malmendier, U., Ospina, J., and Weber, M. (2021). Exposure to grocery prices and inflation expectations. *Journal of Political Economy*, 129(5):1615–1639.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2020). Changing business dynamism and productivity: Shocks versus responsiveness. *American Economic Review*, 110(12):3952–3990.

DellaVigna, S. and Gentzkow, M. (2019). Uniform pricing in u.s. retail chains. *The Quarterly Journal of Economics*, 134(4):2011–2084.

Diamond, P. A. (1971). A model of price adjustment. *Journal of Economic Theory*, 3(2):156–168.

Erceg, C. J., Henderson, D. W., and Levin, A. T. (2000). Optimal monetary policy with staggered wage and price contracts. *Journal of Monetary Economics*, 46(2):281–313.

Farhi, E. and Werning, I. (2019). Monetary policy, bounded rationality, and incomplete markets. *American Economic Review*, 109(11):3887–3928.

Gaballo, G. and Paciello, L. (2021). Spending allocation under nominal uncertainty: a model of effective price rigidity. CEPR Discussion Paper DP16101, Centre for Economic Policy Research (CEPR).

García-Schmidt, M. and Woodford, M. (2019). Are low interest rates deflationary? a paradox of perfect-foresight analysis. *American Economic Review*, 109(1):86–120.

Gertler, M. and Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Goldberg, P. K. and Hellerstein, R. (2013). A structural approach to identifying the sources of local currency price stability. *The Review of Economic Studies*, 80(1):175–210.

Golosov, M. and Lucas, R. E. (2007). Menu costs and phillips curves. *Journal of Political Economy*, 115(2):171–199.

Gopinath, G., Gourinchas, P.-O., Hsieh, C.-T., and Li, N. (2011). International prices, costs, and markup differences. *American Economic Review*, 101(6):2450–2486. †.

Gopinath, G. and Itskhoki, O. (2008). In search of real rigidities. *NBER Macroeconomics Annual*, 23:261–310.

Hazell, J. (2024). Comment on "the dominant role of expectations and broad based supply shocks in driving inflation". In Leahy, J. V., Eichenbaum, M. S., and Ramey, V. A., editors, *NBER Macroeconomics Annual 2024, Volume 39*. University of Chicago Press. This is a preliminary draft and may not have been subjected to the formal review process of the NBER. This page will be updated as the chapter is revised.

Hazell, J., Herreño, J., Nakamura, E., and Steinsson, J. (2022). The slope of the phillips curve: Evidence from u.s. states. *Quarterly Journal of Economics*, 137(3):1299–1344.

Head, A., Liu, L. Q., Menzio, G., and Wright, R. (2012). Sticky prices: A new monetarist approach. *Journal of the European Economic Association*, 10(5):939–973.

Hottman, C. J., Redding, S. J., and Weinstein, D. E. (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291–1364.

Hulten, C. R. (1973). Divisia index numbers. *Econometrica*, 41(6):1017–1025.

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *The Review of Economic Studies*, 45(3):511–518.

Jaimovich, N., Rebelo, S., and Wong, A. (2019). Trading down and the business cycle. *Journal of Monetary Economics*, 102:96–121.

Jaravel, X. (2019). The unequal gains from product innovations: Evidence from the us retail sector. *The Quarterly Journal of Economics*, 134(2):715–783.

Kaplan, G. and Menzio, G. (2015). The morphology of price dispersion. *International Economic Review*, 56(4):1207–1238.

Kimball, M. S. (1995). The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit and Banking*, 27(4, Part 2):1241–1277.

Klenow, P. J. and Willis, J. L. (2016). Real rigidities and nominal price changes. *Economica*, 83(331):443–472.

Kumar, S., Afrouzi, H., Coibion, O., and Gorodnichenko, Y. (2015). Inflation targeting does not anchor inflation expectations: Evidence from firms in new zealand. NBER Working Paper 21814, National Bureau of Economic Research.

L'Huillier, J.-P. (2020). Consumer imperfect information and endogenous price rigidity. *American Economic Journal: Macroeconomics*, 12(2):94–123.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of economic theory*, 4(2):103–124.

Mankiw, N. G. (1985). Small menu costs and large business cycles: A macroeconomic model of monopoly. *Quarterly Journal of Economics*, 100(2):529–537.

Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics*, 117(4):1295–1328.

Matějka, F. (2015). Rigid pricing and rationally inattentive consumer. *Journal of Economic Theory*, 158:656–678. Part B.

Mongey, S. and Waugh, M. E. (2024). Discrete choice, complete markets, and equilibrium. Staff Report 656, Federal Reserve Bank of Minneapolis.

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, 92(5):1521–1534.

Nakamura, E. and Zerom, D. (2010). Accounting for incomplete pass-through. *The Review of Economic Studies*, 77(3):1192–1230.

Phelps, E. S. (1969). The new microeconomics in inflation and employment theory. *American Economic Review*, 59(2):147–160.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2 of *Handbook of Macroeconomics*, pages 71–162. Elsevier, Amsterdam.

Rebelo, S., Santana, M., and Teles, P. (2024). Behavioral sticky prices. Working Paper 32214, National Bureau of Economic Research. Revised October 2024.

Reis, R. (2006). Inattentive producers. *The Review of Economic Studies*, 73(3):793–821. Published: 01 July 2006.

Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, 94(4):1055–1084.

Rotemberg, J. J. (1982). Sticky prices in the united states. *Journal of Political Economy*, 90(6):1187–1211.

Rotemberg, J. J. (2011). Fair pricing. *Journal of the European Economic Association*, 9(5):952–981. https://www.jstor.org/stable/41305841.

Smets, F. and Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 97(3):586–606.

Stahl, D. O. (1989). Oligopolistic pricing with sequential consumer search. *American Economic Review*, 79(4):700–712.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69(3):213–225.

Taylor, J. B. (1980). Aggregate dynamics and staggered contracts. *Journal of Political Economy*, 88(1):1–23.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica*, 47(3):641–654.

Wolinsky, A. (1986). True monopolistic competition as a result of imperfect information. *The Quarterly Journal of Economics*, 101(3):493–511.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. In Aghion, P., Frydman, R., Stiglitz, J., and Woodford, M., editors, *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, pages 25–58. Princeton University Press, Princeton, NJ.

# Appendix

## A Proofs and Derivations in Static Model

**Proof of Proposition 1:**

*Proof.* We consider the case where the first draw does not need the search cost. Shoppers are randomly assigned a firm for free in the first round of search. This guarantees that they always participate the market.

I prove that for given $x$, if $\int U(v|x)\psi(y|x)dy = -\infty$, then $v^*(x) = -\infty$; if $\int U(v|x)\psi(y|x)dy > -\infty$, there exists a unique threshold defined as follows.

$$v^*(x) = -\frac{\kappa}{1-\Psi(v^*(x)|x)} + \frac{\int_{v^*(x)}^{\infty} y\psi(y)dy}{1-\Psi(v^*(x)|x)} \tag{51}$$

First, given $x$, I consider two limits. When $v \to -\infty$,

$$-\kappa + U(v|x)\int_{-\infty}^{v}\psi(y|x)dy + \int_{v}^{\infty}U(y|x)\psi(y|x)dy = -\kappa + \int U(y|x)\psi(y|x)dy$$

Then, if $\int U(v|x)\psi(y|x)dy > -\infty$, $U(v|x) > v$. Intuitively, when the search cost is lower than the value of search when the agent has the worst value whatsoever, the agent should choose to continue searching. In the other limit where $v \to \infty$,

$$-\kappa + U(v|x)\int_{-\infty}^{v}\psi(y|x)dy + \int_{v}^{\infty}U(y|x)\psi(y|x)dy = -\kappa + U(v|x)$$

Then, $U(v|x) = v$. Intuitively, when the agent has the best value whatsoever, the agent should stop search because she cannot get any other good that offers better value.

Second, I claim that if $U(v|x) > v$, then $U'(v|x) = 0$. To prove the claim, suppose there exist values of $v$ on its support (at least at $-\infty$) such that $U(v) > v$. Then, $U(v)$ satisfies,

$$U(v|x) = -\kappa + U(v|x)\int_{-\infty}^{v}\psi(y|x)dy + \int_{v}^{\infty}U(y|x)\psi(y|x)dy \tag{52}$$

Equivalently,

$$U(v|x) = -\frac{\kappa}{1-\Psi(v|x)} + \frac{\int_{v}^{\infty}U(y|x)\psi(y|x)dy}{1-\Psi(v|x)} \tag{53}$$

It is easy to show that

$$U'(v|x) = \frac{\psi(v|x)}{(1-\Psi(v|x))^2}\left\{\int_{v}^{\infty}U(y|x)\psi(y|x)dy - U(v|x)(1-\Psi(v|x)) - \kappa\right\} = 0 \tag{54}$$

The second equation holds due to (53). Then, it is easy to see that $U(v|x)$ is a constant in $(-\infty, v^*]$, which is equal to $v^*(x)$, and then $U(v) = v$ in $(v^*, \infty)$. $v^*$ is unique. In addition, for $v < v^*$,

$$-\kappa + U(v|x)\int_{-\infty}^{v}\psi(y|x)dy + \int_{v}^{\infty}U(y|x)\psi(y|x)dy = -\kappa + v^*(x)\Psi(v^*(x)) + \int_{v^*(x)}^{\infty}y\psi(y|x)dy = v^*(x) \tag{55}$$

The search problem in (9) is therefore simplified to

$$U(v|x) = \max\{v, v^*(x)\} \tag{56}$$

This implies that the value of an additional search does not depend on the state $v$. No matter what state the shopper has, she always compare the state with $v^*(x)$.

Now, we consider the case in which $\int U(v|x)\psi(y|x)dy = -\infty$ for any $x$. This happens when firms charge arbitrarily high prices, which implies $\psi(y|x)$ is a Dirac function at $y = -\infty$. Then, it is optimal to accept any firm in the first round and the resulting $v^*(x) = -\infty, \forall x$.

Now, we prove the second part: $v^*(x)$ decreases in $x$. First, notice that the threshold is determined alternatively by the following,

$$\int \int_{\lambda(v^*(x)+p)}^{\infty} \big(\frac{1}{\lambda}\epsilon - p - v^*(x)\big)g(\epsilon)d\epsilon f(p|x)dp = \kappa \tag{57}$$

For $x_1 < x_2$, $f(p|x_1)$ is FOSD over $f(p|x_2)$. Since the inner integral in LHS decreases in $p$, that implies

$$v^*(x_1) > v^*(x_2)$$

■

**Proof of Proposition 2:**

*Proof.* We take first-order condition of firm's profit in (14) with respect to $P_k$. It is easy to show that we can always express the pricing strategy as in (15) and define the elasticity of demand as:

$$e_k = -\frac{\partial \log D(P_k)}{\partial \log P_k} = -\frac{\partial \log \Big(\int X\big(1 - G\big(\lambda(v^*(x) + p_k)\big)\big)d\Phi_x(x)\Big)}{\partial p_k} + 1 \tag{58}$$

■

**Proof of Lemma ??:**

*Proof.* We are interested in the case in which consumers search actively. We first define the markup elasticity,

$$\Lambda_k = -\frac{d \log \mu_k}{dp_k} = \frac{1}{e_k(e_k - 1)}\frac{\partial e_k}{\partial p_k}$$

Since $G$ is log-concave, $\frac{\partial e_k}{\partial p_k} > 0$. Therefore, the markup elasticity is positive. We can further write out the definition,

$$\Lambda_k = \frac{\frac{g'}{g}\frac{1-G}{g} + 1}{\frac{1}{\lambda}\frac{1-G}{g} + 1}$$

Here, I omit the argument of functions for simplicity. The argument is $\lambda(v^*(w) + p_k)$. Note that equilibrium price is a function of $v^*(w)$ and $a_{kt}$. I now rewrite the LHS of (17) in terms of the integral over productivity distribution, which is exogenous,

$$\int\int_{\lambda(v^*(w)+p^*(v^*(w),a))}^{\infty} \big(\frac{1}{\lambda}\epsilon - p^*(v^*(w),a) - v^*(w)\big) g(\epsilon)d\epsilon\phi_a(a)da = \kappa \tag{59}$$

where $\phi_a(a)$ is the pdf of productivity distribution. Fix $p^*(v^*(w),a)$, the LHS is decreasing in $v^*(w)$. However, higher $v^*(w)$ also decreases optimal prices. To know the net effect of these two forces, we need to derive $\frac{\partial p(v^*(w),a_k)}{\partial v^*(w)}$. From now on, we use $x_k(v^*(w))$ to denote $x(v^*(w),a_k)$ for any variable $x$.

$$\frac{\partial p_k(v^*(w))}{\partial v^*(w)} = \frac{\partial mu_k(v^*(w))}{\partial v^*(w)} = -\frac{1}{e_k(v^*(w))(e_k(v^*(w))-1)}\frac{\partial e_k(v^*(w))}{\partial v^*(w)} \tag{60}$$

It is easy to show that

$$\frac{\partial e_k(v^*(w))}{\partial v^*(w)} = \lambda^2\frac{g'(1-G)+g^2}{(1-G)^2}\left(1+\frac{\partial p_k(v^*(w))}{\partial v^*(w)}\right) \tag{61}$$

Combine the above equation with (60), we have:

$$\frac{\partial p_k(v^*(w))}{\partial v^*(w)} = -\frac{\Lambda_k}{1+\Lambda_k} \tag{62}$$

Since $\Lambda_k > 0$, $\frac{\partial p_k(v^*(w))}{\partial v^*(w)} \in (-1,0)$. This implies that $\frac{\partial(v^*(w)+p_k(v^*(w)))}{\partial v^*(w)} \in (0,1)$. Therefore, the LHS of (17) decreases in $v^*(w)$. ■

**Proof of Theorem 1:**

*Proof.* First, notice that since $G$ is log-concave, $\mu_k$ decreases with $P_k$. From $P_k = \mu_k\frac{W}{A_k}$, we know that there is a unique solution for each optimal price $P_k$ given the full-information threshold $v^*(w)$. Then, from Lemma **??**, we know that there exists a unique solution $v^*(w)$ to 17 given that the prices are computed through the first-order conditions as in (20). The equilibrium in which shoppers search actively exists and is unique because $v^*(w)$ exists and is unique.

As Diamond (1971) points out famously, there are always a continuum of equilibria where shoppers do not search and firms charge very high prices, i.e., $v^*(w) = -\infty$ and $P_k = \infty, \forall k$. ■

**Remarks on Computation of the Full-information Equilibrium:**

**Remark 1** (Remark on Computing Steady-State Equilibrium)**.** *The proof of Theorem 1 provides insights on the computational method for the full-information equilibrium.*

1. *First guess a* $v^*(w)$
2. *Calculate the optimal price distribution given* $v^*(w)$
3. *Plug guessed* $v^*(w)$ *and derived price distribution into 17 and check if LHS is equal to the given search cost* $\kappa$*.*

4. *Increases guessed $v^*(w)$ if LHS is larger than search cost, according to Lemma* **??**. *Vice versa.*

5. *Loop the procedure 1-4 until the difference between the LHS and the RHS of 17 is smaller than the given tolerance*

**Proof of Theorem 2:**

*Proof.* Suppose the nominal wage increases from $w$ to $w'$. Notice $\Delta w = w' - w$ does not need to be small. We guess that all the optimal prices increase proportionally, i.e., $p_k' = p_k + \Delta w$. Then the price distribution $f(p|w)$ shifts to the right and becomes $f(p - \Delta w|w')$. The threshold is determined by,

$$\int \int_{\lambda(v^*(w')+p)}^{\infty} \big(\frac{1}{\lambda}\epsilon - p - v^*(w')\big) g(\epsilon) d\epsilon f(p - \Delta w|w') dp = \kappa \tag{63}$$

Let $z = p - \Delta w$. Then we can rewrite the LHS of the above equation,

$$\int \int_{\lambda(v^*(w')+\Delta w+z)}^{\infty} \big(\frac{1}{\lambda}\epsilon - z - (\Delta w + v^*(w'))\big) g(\epsilon) d\epsilon f(z|w') dz = \kappa \tag{64}$$

This implies that $v^*(w') = v^*(w) - \Delta w$. According to (20), the elasticity of demand is given by,

$$\begin{aligned} e_k' &= \lambda \frac{g\big(\lambda(v^*(w') + p_k')\big)}{1 - G\big(\lambda(v^*(w') + p_k')\big)} + 1 \\ &= \lambda \frac{g\big(\lambda(v^*(w) - \Delta w + p_k + \Delta w)\big)}{1 - G\big(\lambda(v^*(w) - \Delta w + p_k + \Delta w)\big)} + 1 \\ &= e_k \end{aligned}$$

Since the elasticity of demand does not change, the optimal prices is given by,

$$p_k' = \log(\frac{e_k}{e_k - 1}) + w' - a_k = p_k + w' - w = p_k + \Delta w \tag{65}$$

Therefore, we verify the guess that the optimal prices increase proportionally with the nominal wage. Based on Theorem 1, the equilibrium is unique. Therefore, the above constructed equilibrium is the only equilibrium when the nominal wage is $w'$. ■

**Proof of Proposition 3:** Here, I provide separate proofs for two parts of Proposition 3. I start with the proof of the first part.

***Proof of Part 1***. First, we can rewrite 10 in terms of integrating over the exogenous productivity distribution,

$$\int \int \int_{\lambda(v^*(x)+p^*(a,w))}^{\infty} \big(\frac{1}{\lambda}\epsilon - p^*(a,w) - v^*(x)\big) g(\epsilon) d\epsilon \phi_a(a) da h(w|x) dw = \kappa \tag{66}$$

where $v^*(x)$ is implicitly determined by the above equation. On the first order, the posterior belief of the nominal wage collapses to a Dirac function at $E(w|x)$. Also, combining the shopper's expected price conditional on $x$, on the first order, the above equation becomes,

$$\int \int_{\lambda(v^*(x)+\bar{p}_k+\varphi(a)E(\hat{w}|x))}^{\infty} \big(\frac{1}{\lambda}\epsilon - \bar{p}(a) - \varphi(a)E(\hat{w}|x) - v^*(x)\big)g(\epsilon)d\epsilon\phi_a(a)da = \kappa \tag{67}$$

Take the total derivative on $E(\hat{w}|x)$ on both sides,

$$\int \Big(\varphi(a) + \frac{dv^*(x)}{dE(\hat{w}|x)}\Big)\big(1 - G(\lambda(v^*(x) + \bar{p}(a) + \varphi(a)E(\hat{w}|x)))\big)\phi_a(a)da = 0$$

On the first order, it can be written as follows,

$$\frac{dv^*(x)}{dE(\hat{w}|x))} = -\int \varphi(a)\bar{\omega}(a)\phi_a(a)da = -\int \varphi_k\bar{\omega}_k dk \tag{68}$$

where $\bar{\omega}(a)$ is the expenditure share of firms with productivity $a$ in the full-information equilibrium where $w = \bar{w}$. Then, the threshold, on the first order, is given by,

$$v^*(x) = v^*(\bar{w}) - \frac{\partial v^*(x)}{\partial v^*(x)}v^*(x) = v^*(\bar{w}) - \Phi E(\hat{w}|x)$$

Recall $\hat{p} = \Phi\hat{w}$. We have the result. ■

***Proof of Part 2*** . First, plug the result in Part 1 into the elasticity of demand in (16), the elasticity becomes,

$$e_k = \lambda\frac{\int Xg\big(\lambda(v^*(\bar{w}) - \Phi E(\hat{w}|x) + p_k)\big)d\Phi_x(x)}{\int X\big(1 - G\big(\lambda(v^*(\bar{w}) - \Phi E(\hat{w}|x) + p_k)\big)\big)d\Phi_x(x)} + 1 \tag{69}$$

with some abuse of notation, $\Phi$ is the aggregate correlated-shock passthrough and $\Phi_x(x)$ is the cdf of information sets. Let $y = \lambda\Phi\big(E(\hat{w}|x) - \bar{E}(\hat{w})\big)$ and its pdf is $\phi_y(y)$, which is a Gaussian distribution with mean zero and standard deviation $\sigma = \lambda\Phi\theta\sigma_s$. Further, we denote $z = \lambda(v^*(w) - \Phi\bar{E}(\hat{m}) + p_k)$. The elasticity is rewritten as follows,

$$e_k = \lambda\frac{\int Xg(z+y)\phi_y(y)dy}{\int X\big(1 - G(z+y)\big)\phi_y(y)dy} + 1 \tag{70}$$

It is easy to show that the first-order approximation to the ratio is equivalent to separately approximating numerator and denominator and then combining them. Following this result, we first expand the numerator.

$$g(z+y) = g(z) + g'(z)y + \frac{g''(z)}{2}y^2 + \mathcal{O}(y^3)$$

Substitute this expansion into the integral and also notice that $X \propto \exp(y + \bar{w})$,

$$\begin{aligned}\int \exp(y) g(z+y)\phi_y(y)dy &= \int (1+y)\Big(g(z) + g'(z)y + \frac{g''(z)}{2}y^2 + \mathcal{O}(y^3)\Big)\phi_y(y)dy \\ &= g(z) + \int \Big(\big(g'(z) + \frac{g''(z)}{2}\big)y^2 + \mathcal{O}(y^3)\Big)\phi_y(y)dy \\ &= g(z) + \big(g'(z) + \frac{g''(z)}{2}\big)\sigma^2 + \mathcal{O}(\sigma^3)\end{aligned}$$

Therefore, $\int g(z+y)\phi_y(y)dy \to g(z)$ on the order of $\sigma_s^2$, which is second-order term. Similarly, $\int \big(1 - G(z+y)\big)\phi_y(y)dy \to 1 - G(z)$ on the order of $\sigma_s^2$. On the other hand, the average expectation of monetary shock $\bar{E}(\hat{w})$ approaches zero on the order of $\sigma_s$. Therefore, the elasticity, on the first order, is given by,

$$e_k = \lambda \frac{g(z)}{1 - G(z)} + 1 = \lambda \frac{g(\lambda(v^*(w) - \Phi \bar{E}(\hat{w}) + p_k))}{1 - G(\lambda(v^*(w) - \Phi \bar{E}(\hat{w}) + p_k))} + 1 \tag{71}$$

■

**Proof of Lemma 1:**

*Proof.* First, log price is given by,

$$p_k = \log \mu_k + w - a_k$$

Total differentiate on both sides,

$$\hat{p}_k = \frac{\partial \log \mu_k}{\partial p_k}\hat{p}_k + \frac{\partial \log \mu_k}{\partial p}\hat{p} + \hat{w} \tag{72}$$

We further define $\gamma_k = (1 - \frac{d \log \mu_k}{dp_k}\Big|_{\hat{w}=0})^{-1}$ and $\xi_k = \frac{d \log \mu_k}{dp}\Big|_{\hat{w}=0}\gamma_k$. Then, we have,

$$\hat{p}_k = \gamma_k \hat{w} + \xi_k \hat{p} \tag{73}$$

Since $\hat{p} = \Phi\hat{w}$, the correlated-shock passthrough for individual firm satisfies,

$$\varphi_k = \gamma_k + \Phi\xi_k$$

Integrate on both sides,

$$\Phi = \frac{\Gamma}{1 - \Xi}$$

where $\Gamma = \int \gamma_k \bar{\omega}_k dk$, $\Xi = \int \xi_k \bar{\omega}_k dk$. ■

**Proof of Theorem 3:** I first prove the “Incompleteness” part of the Theorem. Then I prove an important lemma and finally I prove the comparative statics results.

***Proof of Incompleteness*** . First, notice that for any function $f(w)$, the first-order approximation to its derivative is,

$$f'(w) = f'(\bar{w}) + f''(\bar{w})\hat{w} \tag{74}$$

Then, $f'(w)|_{\hat{w}=0} = f'(\bar{w})$ on the first order. I claim that the approximation gives the same result if we reverse the order of the operations. Let's first take the first-order approximation to $f(w)$,

$$f(w) = f(\bar{w}) + f'(\bar{w})\hat{w} \tag{75}$$

Then if we take derivative w.r.t. $\hat{w}$ and then make $\hat{w} = 0$, we get the same result.

Now, following this result, $\frac{\partial e_k}{\partial p_k}\Big|_{\hat{w}=0}$ and $\frac{\partial e_k}{\partial p}\Big|_{\hat{w}=0}$ can be computed by first first-order approximating $e_k$ as we did in Proposition 3 and then take derivative to $\hat{p}_k$ and $\hat{p}$. For simplicity, I omit the argument of the functions, in particular, $z = z(v^*(\bar{w}) + \bar{p}_k)$ for any function $z$. It is straightforward to show that, on the first order,

$$\frac{\partial e_k}{\partial p_k}\Big|_{\hat{w}=0} = \lambda^2 \frac{g'(1-G) + g^2}{(1-G)^2} \tag{76}$$

We can derive that $\gamma_k$ does not depend on $\theta$. In addition, we have

$$\frac{\partial e_k}{\partial p}\Big|_{\hat{w}=0} = -\theta\lambda^2 \frac{g'(1-G) + g^2}{(1-G)^2} \tag{77}$$

Then, we have $\frac{\partial \log \mu_k}{\partial p} = -\theta \frac{\partial \log \mu_k}{\partial p_k}$. According to Lemma 1, we have

$$\xi_k = \theta(1 - \gamma_k) \tag{78}$$

I claim that $\varphi_k < 1, \forall k$ and notice that to prove it, it is sufficient to prove $\gamma_k + \xi_k < 1, \forall k$. We can write out $\gamma_k + \xi_k$,

$$1 - (\gamma_k + \xi_k) = (1 - \gamma_k)(1 - \theta) \tag{79}$$

Since $-\frac{d \log \mu_k}{dp_k} > 0$, $\gamma_k < 1$. ■

Before turning to the result on "Composition", I first prove the following lemma,

**Lemma A-1.** *Let* $\Lambda_k = -\frac{d \log \mu_k}{dp_k}\Big|_{\hat{w}=0}$ *denote the steady-state markup elasticity. When* $G$ *is Gumbel distribution, it has the following property,*

$$\frac{d\Lambda(y)}{dy} < 0$$

*In addition,* $\Lambda(y) \to 0$ *as* $y \to \infty$ *and* $\Lambda(y) \to \infty$ *as* $y \to -\infty$.

*Proof.* To simplify notation, we denote $g(y_k) = g(\lambda(v^*(\bar{w}) + \bar{p}_k))$ and $G(y_k) = G(\lambda(v^*(\bar{w}) + \bar{p}_k))$. We assume that $G$ follows standard type-I extreme-value distribution, i.e., $G(y) = \exp(\exp(-y))$, $g(y) = \exp(-y - \exp(-y))$. It also follows that $\frac{g'}{g} = e^{-y} - 1$.

First, since $G$ is log-concave, $g$ is also log-concave. This implies the following property,

$$(\frac{g'}{g})' < 0; \quad (\frac{1-G}{g})' < 0$$

Additionally, using $G$ is Gumbel, we have $\frac{1-G}{g} \to 1$ as $y \to \infty$ and therefore $\frac{1-G}{g} > 1$.

Second, we write out markup elasticity,

$$\Lambda = \frac{\frac{g'}{g}\frac{1-G}{g} + 1}{\frac{1}{\lambda}\frac{1-G}{g} + 1} \tag{80}$$

Then the derivative of markup elasticity w.r.t $y$ is given by,

$$\begin{aligned}
\frac{d\Lambda(y)}{dy} &= \left(\frac{g'}{g}\right)'\left(\frac{1-G}{g}\right)^2\frac{1}{\lambda} - \left(\frac{1-G}{g}\right)'\frac{1}{\lambda} + \left(\frac{g'}{g}\frac{1-G}{g}\right)' \\
&= -e^{-y}\left(\frac{1-G}{g}\right)^2\frac{1}{\lambda} + \frac{1}{\lambda}(1 + \frac{g'}{g}\frac{1-G}{g}) + \left(\frac{g'}{g}\frac{1-G}{g}\right)' \\
&< -e^{-y}\left(\frac{1-G}{g}\right)\frac{1}{\lambda} + \frac{1}{\lambda}(1 + (e^{-y} - 1)\frac{1-G}{g}) + \left(\frac{g'}{g}\frac{1-G}{g}\right)' \\
&= \frac{1}{\lambda}\left(1 - \frac{1-G}{g}\right) + \left(\frac{g'}{g}\frac{1-G}{g}\right)'
\end{aligned}$$

The inequality is due to $\frac{1-G}{g} > 1$. The first term of the last equation is negative. We need to deal with the second term. Notice that

$$\left(\frac{g'}{g}\frac{1-G}{g}\right)' = \left(\frac{g'}{g}\right)'\left(\frac{1-G}{g}\right) + \left(\frac{g'}{g}\right)\left(\frac{1-G}{g}\right)'$$

It is straightforward that when $y < 0$, $\frac{g'}{g} = e^{-y} - 1 > 0$. Then, $\left(\frac{g'}{g}\frac{1-G}{g}\right)' < 0$. We now focus on the case where $y > 0$.

$$\begin{aligned}
\left(\frac{g'}{g}\frac{1-G}{g}\right)' &= -e^{-y}\frac{1-G}{g} + (1 - e^{-y}) - (1 - e^{-y})^2\frac{1-G}{g} \\
&< -e^{-y}\frac{1-G}{g} + (1 - e^{-y}) - (1 - e^{-y})\frac{1-G}{g} \\
&= (1 - e^{-y}) - \frac{1-G}{g} \\
&< 0
\end{aligned}$$

The first inequality is due to $1 - e^{-y} \in (0, 1)$ when $y > 0$. The second inequality is due to $\frac{1-G}{g} > 1$. Combine all together, we have

$$\frac{d\Lambda(y)}{dy} < 0$$

For limit results, first recall $\frac{1-G}{g} \to 1$ as $y \to \infty$. Plug into (80), we get $\Lambda \to 0$ when $y \to \infty$. Second, since $\frac{1-G}{g} \to \infty$ when $y \to -\infty$, $\Gamma \approx \lambda g'/g \to \infty$ since $g'/g \to \infty$. ■

Next, we turn to the result on composition and comparative statics.

***Proof of Composition and MCS.*** First, notice that to show $\varphi_k$ decreases in $a_k$, it is sufficient to show that $\gamma_k$ decreases in $a_k$ according to (79).

From Lemma 1, $\gamma_k = (1 + \Lambda_k)^{-1}$. Since $\Lambda_k$ decreases in $P_k$, it implies $\gamma_k$ increases in $P_k$. Since $P_k$ decreases in $a_k$, it implies $\gamma_k$ decreases in $a_k$.

Next, we prove the comparative statics results. We first consider the comparative statics on $\kappa$. From Lemma **??**, an increase in the search cost $\kappa$ implies lower threshold $v^*(\bar{w})$. According to Lemma A-1, the steady-state markup elasticity decreases in $v^*(\bar{w})$. This implies a uniform decrease in $\gamma_k$. As a result, $\varphi_k$ decreases in $\kappa$ for any $k$ according to (79). When $\kappa \to 0$, $v^*(\bar{w}) \to \infty$ and therefore $\Lambda_k \to 0$, resulting in $\gamma_k \to 1$ for any $k$. This implies $\varphi_k = 1$ for any $k$. When $\kappa \to \infty$, $v^*(\bar{w}) \to -\infty$, and therefore $\Lambda_k \to \infty$, resulting in $\gamma_k \to 0$ for any $k$. This implies $\varphi_k = 0$ for any $k$.

Then, we consider the comparative statics on $\theta$. An increase in the information friction, which means a decrease in $\theta$, induces lower $\varphi_k$, according to (79). When $\theta \to 0$, $\Phi = \Gamma$ according to Lemma 1. When $\theta \to 1$, $\varphi_k = 1$ according to (79). ■

**Proof of Proposition 1:**

*Proof.* From (78), and Lemma 1, we have,

$$\hat{p}_k = \gamma_k \hat{w} + (1 - \gamma_k)\theta\hat{w} \tag{81}$$

$$= \gamma_k \hat{w} + (1 - \gamma_k)\bar{E}\hat{w} \tag{82}$$

Aggregate, we have

$$\hat{p} = \Gamma\hat{w} + (1 - \Gamma)\bar{E}\hat{w} \tag{83}$$

The aggregation requires every firm is individually rational, so they know (82). Suppose every shopper believe that firms and other shoppers are rational. Then every shopper believes the above condition holds. Shoppers' average expectation of $\hat{p}$ satisfies,

$$\bar{E}\hat{p} = \Gamma\bar{E}\hat{w} + (1 - \Gamma)\bar{E}^2\hat{p} \tag{84}$$

where $\bar{E}^2[\cdot] = \bar{E}[\bar{E}[\cdot]]$ denotes the second-order belief. Iterating ad infinitum, the change in the actual price index $\hat{p}$ can be expressed in terms of the higher-order beliefs of the monetary shock $\hat{w}$:

$$\hat{p} = \Gamma \sum_{h=0}^{\infty} (1 - \Gamma)^h \bar{E}^h \hat{w} \tag{85}$$

"Iterating ad infinitum" amounts to imposing common knowledge of rationality. The first iteration requires that shoppers know that firms and other shoppers are rational, the second iteration requires that shoppers know that others know they are rational and firms are rational, and so on. ■

**Proof of Proposition 4:**

*Proof.* Since the markup $\mu_k$ decreases in $p_k$, it increases in $a_k$. It is easy to show that $\lim_{a\to-\infty} e_k = 1-\lambda\frac{g'}{g} = 1+\lambda$ and $\lim_{a\to\infty} e_k = 1$. Then, $\lim_{a\to-\infty} \mu_k = \frac{\lambda+1}{\lambda}$ and $\lim_{a\to\infty} \mu_k = \infty$.

From Lemma 1, $\gamma_k = (1+\Lambda_k)^{-1}$. Also, we know $\xi_k = \theta(1-\gamma_k)$. Since $\gamma_k$ decreases in $a_k$ as shown in the proof of Theorem 3, $\xi_k$ increases in $a_k$. Also, in the limit when $a \to \infty$, $p \to -\infty$. Then $\Lambda \to \infty$ and $\gamma \to 0$, $\xi \to \theta$. In the opposite limit, $p \to \infty$, and $\Lambda \to 0$, implying $\gamma \to 1$ and $\xi \to 0$. ■

**Proof of Proposition ??:**

*Proof.* See Proof of Theorem 3. ■

**Proof of Proposition ??:**

*Proof.* See Proof of Theorem 3. ■

**Proof of Proposition ??:**

*Proof.* In the case of monetary shock, recall the labor supply condition,

$$\hat{p} + \hat{c} = \hat{w} \tag{86}$$

Combine with the definition of the aggregate passthrough,

$$\hat{p} = \frac{\Phi}{1-\Phi}\hat{c} \tag{87}$$

Plug in the condition in Proposition 1, we have

$$\hat{p} = \frac{\Gamma}{1-\Gamma}\frac{1}{1-\theta}\hat{c} \tag{88}$$

In the case of aggregate supply shock, two equations are as follows,

$$\hat{p} + \hat{c} = \hat{w}$$

$$\hat{p} = \Phi(\hat{w} - \hat{a})$$

Substitute $\hat{w}$, we achieve $\hat{p} = \frac{\Phi}{1-\Phi}(\hat{c} - \hat{a})$. Again, leveraging Proposition 1, we have the result.

■

# B Proofs of Dynamic Model

# C Additional Details on Calibration and Estimation

# D Additional Details on Empirics