Firms' Inflation Expectations: Determinants and Macroeconomic Implications

Rafael Gonçalves   Ina Hajdini

This version: March 20, 2025

We develop a rational inattention model of price-setting firms within a production network to study the determinants and macroeconomic implications of heterogeneity in firms' inflation expectations. In equilibrium, firms' inflation expectations are determined by (i) their attention to industry-specific marginal costs, (ii) the degree of comovement between marginal costs and aggregate inflation, and (iii) the realized marginal cost. Heterogeneity in attention arises endogenously and alters the aggregate effects of monetary and productivity shocks, hinging on the correlation between attention levels and responsiveness of marginal costs to those shocks. Moreover, changes in the production network have ambiguous effects on the sectoral Phillips curves' slopes, as price stickiness is endogenous to the input-output structure. Finally, combining a comprehensive dataset of US firms' inflation expectations with industry-level information, we show that industry-specific upstreamness, the volatility of sector-specific productivity, and realized marginal costs influence firms' attention to inflation. A preliminary version of our calibrated model qualitatively replicates the empirical findings.

Keywords: inflation expectations, rational inattention, production networks, inflation dynamics.
JEL codes: E31, E32, E71.

Gonçalves: Princeton University, rgoncalves@princeton.edu. Hajdini: Federal Reserve Bank of Cleveland, Ina.Hajdini@clev.frb.org. We thank Karthik Sastry and Gianluca Violante for helpful comments. The views expressed here are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

1. Introduction

Modern macroeconomic theory attributes a crucial role to firms' expectations about inflation in influencing the dynamics of both nominal and real variables. However, the limited availability of data on firms' expectations has hindered our understanding of the joint distribution of firms' price decisions, their beliefs, and macroeconomic outcomes. In this context, the full-information rational expectations (FIRE) framework has provided the paradigm for modeling expectations and their influence on macroeconomic outcomes for the past decades.

A recent empirical literature has documented substantial heterogeneity and inaccuracy in firms' inflation expectations, deviating from the predictions of the FIRE paradigm (Coibion, Gorodnichenko, and Kumar 2018; Candia, Coibion, and Gorodnichenko 2024, among others). Notably, firms' "inattention" to inflation seems related to industry-specific characteristics, such as experienced input price changes (Andrade et al. 2022) and the number of competitors (Afrouzi 2024). However, the role that other relevant industry attributes, such as an industry's importance in the production network, the industry-level degree of price rigidity, or the volatility of sectoral conditions, play for firms' expectations remains unexplored.

These facts raise fundamental questions: How do industry attributes influence firms' inflation expectations? What are the macroeconomic implications of the interplay between industry characteristics and endogenous attention to inflation? From an applied perspective, how does a model accounting for a realistic production structure and the observed heterogeneity in firms' inflation expectations improve our understanding of inflation dynamics?

This paper addresses these questions both theoretically and empirically. On the theoretical front, we develop a rational inattention model of price-setting firms operating within a production network. The model sheds light on the role of industry characteristics in shaping firms' inflation expectations and the macroeconomic implications of heterogeneous beliefs. In the model, firms' expectations about the aggregate price level are determined by three factors: (i) firms' attention to their industry-specific marginal cost, (ii) the co-movement between marginal cost and aggregate inflation, and (iii) the realized marginal costs. Furthermore, the interplay between rational inattention and production networks generates a downstream compounding of inattention and alters the aggregate effects of shocks.

On the empirical front, we explore a novel and comprehensive dataset on US firms' inflation expectations and link it with various data sources on industry characteristics. We use the data to test our model's predictions and uncover new facts about the relationship between firms' inflation expectations and industry attributes. We show that firms' attention to inflation, measured by the accuracy of their forecasts, is positively associated with certain industry characteristics, particularly the degree of their industries' "upstreamness" and the volatility of industry-specific productivity shocks. Additionally, we find supporting evidence for the model's prediction on how firms form their inflation expectations based on their noisy observation of marginal costs.

Model. We develop a simple general-equilibrium model of price-setting firms operating within a production network under rational inattention to aggregate demand and industry-specific productivity shocks. Firms flexibly choose which pieces of information to focus on, considering the costs associated with acquiring and processing information. The information friction is the source of both inaccurate expectations (inattention) and nominal rigidities in the model.

Equilibrium attention. To set prices optimally, firms need to forecast its industry-specific marginal cost while considering the costs of acquiring information about the state of the economy. Equilibrium firms' attention is determined by (i) the cost of obtaining information, (ii) the profit losses due to a pricing mistake of a given size, and (iii) the volatility of marginal costs. Importantly, the second and third components are industry-specific, implying that firms in different sectors have varying incentives to acquire information.

Downstream compounding of inattention. At the heart of the model, there is a fixed point between marginal costs' responsiveness to shocks and attention levels. Intuitively, low attention in a certain industry results in dampened responses of that industry's prices to aggregate and sectoral shocks. This leads to a reduction in the volatility of marginal costs in downstream sectors, thus reducing the benefits of acquiring information and encouraging further inattention within those sectors. This intuition leads to our first key theoretical result: inattention compounds downstream due to strategic complementarities in pricing and information acquisition decisions.

Determinants of CPI expectations. Our framework offers an insightful decomposition of the determinants of firms' inflation expectations, or the Consumer Price Index (CPI), in our simple model. Precisely, firms' CPI expectations are shaped by (i) attention to their industry-specific marginal cost, (ii) the comovement between marginal costs and the aggregate price level, and (iii) the realized marginal cost. This decomposition hinges on the fact that, in equilibrium, firms primarily acquire information about their marginal cost, which is the relevant variable for their price-setting decision. This information is used by firms to form beliefs about macroeconomic outcomes, including inflation. Yet, firms in different industries have distinct incentives to attend to marginal costs, influencing consequently the precision of their inflation forecast. Moreover, the accuracy of firms' inflation will depend on the extent to which movements in an industry's marginal cost are informative about fluctuations in overall inflation.

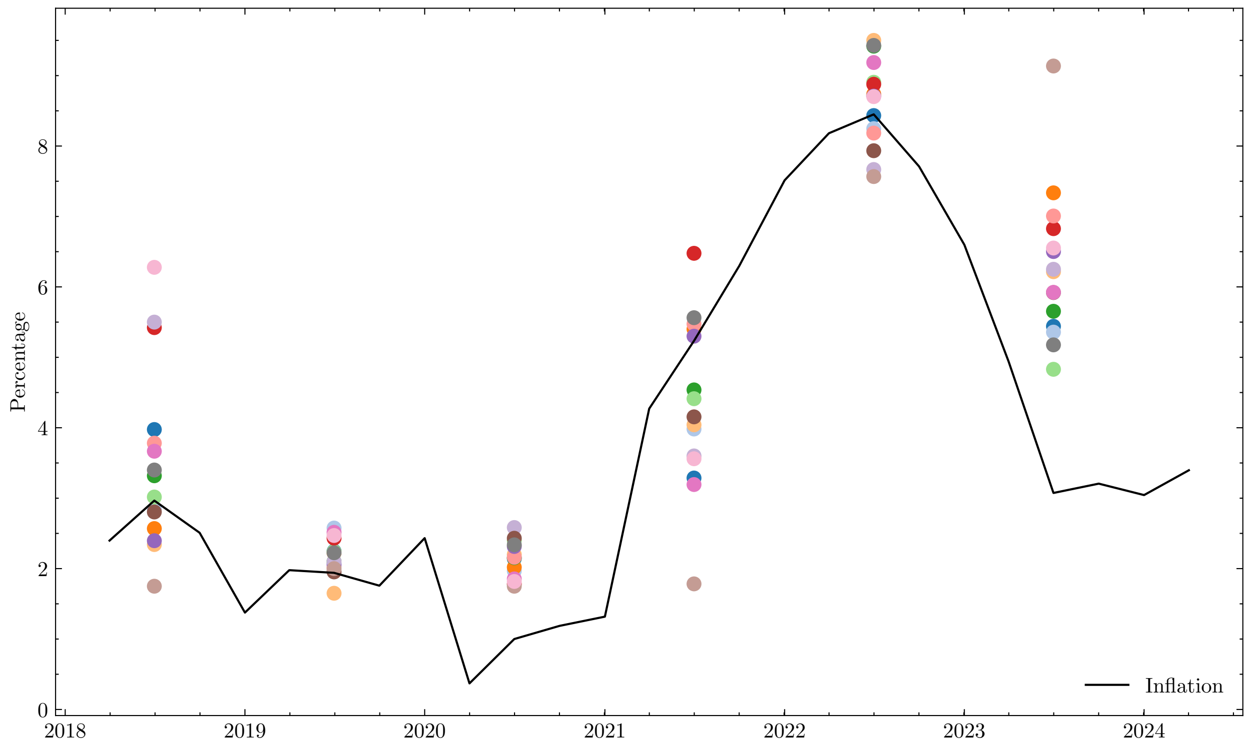
Heterogenous attention and the aggregate effects of shocks. After examining the determinants of heterogeneity in firms' inflation expectations in the model, we explore its macroeconomic implications. We show that heterogeneous attention influences the aggregate effects of shocks to the extent that industry attention levels correlate with the responsiveness of marginal costs to a particular shock. For instance, monetary policy has a larger effect on the aggregate price level when marginal costs respond more to monetary policy shocks in less attentive industries.

The slope of the Phillips curve. Our final theoretical contribution revisits an important result in the literature on inflation and networks, namely that the Phillips curve is flatter in economies with stronger input-output linkages (Basu 1995; Rubbo 2023). We demonstrate that this finding may not apply in our framework, where the degree of nominal rigidities, shaped by equilibrium attention levels, is influenced by the input-output structure of the economy.

Evidence. Our empirical analysis begins with the Survey of Firms' Inflation Expectations (SoFIE), a relatively novel and nationally representative dataset of US firms' inflation expectations. Leveraging this micro-level data, we measure firms' attention to inflation using the absolute value of forecast errors for both current and one-year-ahead inflation projections. Figure 1 reveals substantial dispersion and inaccuracy in inflation perceptions (nowcasts) across industries.

We combine the expectations data with other US data sources containing industry characteristics, such as measures of industry importance in the economy, marginal costs, and the volatility of industry-specific productivity series. As implied by our model, industry characteristics correlate with our measure of firms' attention to inflation. Specifically,

FIGURE 1. Perceived inflation across industries



Note: The black line is the real-time inflation rate from the Federal Reserve Bank of Philadelphia Real-Time Data Research Center. The colored dots are average perceived inflation rates within the following industries: Basic Metals, Electrical, Financial Intermediation, Food and Drink, Hotels and Restaurants, Mechanical Engineering, Other Manufacturing, Other Services, Post and Telecommunication, Renting and Business Activities, Timber and Paper, Transport Manufacturing, Transport and Storage Services. Sample from 2Q2018 to 1Q2024.

we find that firms operating in more upstream industries and those facing more volatile productivity shocks have less accurate inflation forecasts. In a final quantitative analysis, we calibrate our model to match U.S. data and are able to qualitatively replicate the empirical findings.

Related literature. Our paper is related to several strands of the literature.

We build on the literature on rational inattention in price-setting problems (Sims 2003; Maćkowiak, Moench, and Wiederholt 2009; Mackowiak and Wiederholt 2009; Afrouzi 2024; Afrouzi, Flynn, and Yang 2024). Previous research in this area has primarily focused on simplified representations of production that overlook input-output linkages.[1] The

[1]The main exception is Afrouzi (2024), who study how competition affects firms' expectations and inflation dynamics in an oligopolistic competition model with multiple sectors.

primary contribution of our model is to examine the consequences of rational inattention to price setting in multi-sector economies. We show that the interaction of these two model ingredients generates downstream compounding of (in)attention and modifies the effects of aggregate shocks.

In addition, we extend the literature on incomplete information in production networks, which predominantly investigates how information frictions affect the propagation of aggregate and sectoral shocks throughout the production network (Angeletos and Huo 2021; Bui et al. 2022; Nikolakoudis 2024). The analysis most closely aligned with ours is Fang et al. (2024), who study optimal monetary in a price-setting model where firms operate within a production network under rational inattention. However, compared to these studies, our emphasis lies in understanding how the interaction between network structures and rational inattention accounts for the dispersion in firms' inflation expectations and inflation dynamics.

Our paper also relates to a growing body of literature examining inflation and monetary policy within production networks (Pastén, Schoenle, and Weber 2024; Rubbo 2023, among others). Our model particularly builds upon La'O and Tahbaz-Salehi (2022)'s approach, which introduces nominal rigidities to a multi-sector economy through incomplete information in a static setting. However, our novel exploration of the interaction between endogenous attention and input-output linkages yields results that differ from some established findings in this field. For instance, we show that the conventional wisdom of flatter Phillips curves in economies with stronger intermediate input shares may not apply when the degree of nominal rigidity is endogenous to the production structure.

Our work also contributes to the empirical literature documenting departures of firms' inflation expectations from full-information rational expectations (FIRE) predictions and exploring their determinants (Coibion, Gorodnichenko, and Kumar 2018; Andrade et al. 2022; Candia, Coibion, and Gorodnichenko 2024; Weber et al. 2025). Notably, Candia, Coibion, and Gorodnichenko (2024), also using the SoFIE, finds systematic differences in firms' forecasts across industries, while finding no systematic relationship between firm size and forecast accuracy. To the best of our knowledge, our paper is the first to link micro-level data on firms' inflation expectations to sources of industry attributes, guided by new theory.

Outline. The paper proceeds as follows. In section 2, we introduce a simple rational inattention, multi-sector model of price setting. In section 3, we present the model's theoretical insights on the determinants and macroeconomic implications of heterogeneity in firms' inflation expectations. Section 4 presents our data and measurement. In section 5, we document new empirical patterns relating US firms' inflation expectations to industry characteristics and test the model's predictions. Section **??** shows the results of a quantitative model about the effects of monetary policy on inflation. Section 7 concludes.

2. Model

In this section, we build a simple multi-sector model of rationally inattentive price-setting firms operating within a production network.

The network framework follows La'O and Tahbaz-Salehi (2022) more closely, who develop a multi-sector New Keynesian model where nominal rigidities stem from incomplete information in a static setting. The rational inattention (RI) component of the model draws on the theoretical insights by Afrouzi (2024), who extended classical RI results to a setting with multiple agents and potential strategic considerations in information acquisition.

2.1. Environment

The model is a general equilibrium model of price-setting firms operating within a production network under rational inattention. Time is discrete and indexed by $t \in \{0, 1, \ldots\}$. Firms are subject to i.i.d. aggregate demand and industry-specific productivity (TFP) shocks, and they make their pricing and information acquisition decisions under incomplete information about these shocks.

Intermediate firms. There are n industries in the economy, indexed by $i \in I \equiv \{1, 2, \ldots, n\}$. Each industry comprises a unit mass of monopolistically competitive firms indexed by $f \in [0, 1]$. The output of each industry can be either consumed by households or used as an intermediate input for production by firms in all industries, including the same industry.

The monopolistically competitive firms within each industry set prices and use a common constant-returns-to-scale technology F_i to transform labor and intermediate inputs into differentiated goods. Specifically, the production function of firm $f \in [0, 1]$ in industry i is given by a Cobb-Douglass function:

$$
\begin{gathered}
Y_{ift} = Z_{it} F_i(L_{ift}, X_{ift,1}, \ldots, X_{ift,n}) = Z_{it} \zeta_i L_{ift}^{\alpha_i} \prod_{k=1}^{n} X_{ift,k}^{a_{ik}}, \\
\alpha_i + \sum_{k=1}^{n} a_{ik} = 1,
\end{gathered} \tag{1}
$$

where Y_{ift} is the firm's output, L_{ift} is its labor input, and $X_{ift,k}$ is the amount of sectoral commodity $k \in I$ purchased by the firm. Additionally, Z_{it} is the unobserved industry-level productivity shock, and $\zeta_i \equiv \alpha_i^{-\alpha_i} \prod_k a_{ik}^{-a_{ik}}$ is a normalization constant that simplifies the calculations.

We assume that industry-level productivities are i.i.d. log-normally distributed according to

$$\log \boldsymbol{Z}_t \sim \mathcal{N}(0, \Sigma), \tag{2}$$

where $\boldsymbol{Z}_t \equiv (Z_{it})_{i \in I}$ and Σ is an arbitrary covariance matrix allowing for correlated industry-specific productivity shocks.

The economy's input-output linkages are summarized by the matrix A and industry labor shares by the vector α:

$$A \equiv [a_{if}], \quad \boldsymbol{\alpha} \equiv (\alpha_i)_{i \in I}. \tag{3}$$

The nominal profits of firm f in industry i are given by

$$\Pi_{ift} = (1 - \tau_i) P_{ift} Y_{ift} - W_t L_{ift} - \sum_{k=1}^{n} P_k X_{ift,k}, \tag{4}$$

where P_{ift} is the nominal price charged by the firm, P_{kt} is the nominal price of industry k's sectoral output, W denotes the sticky nominal wage set by the labor union sector (to be introduced later), and τ_i is an industry-specific revenue tax (or subsidy) levied by the government.

Differentiated products produced by the unit mass of firms in each industry i are aggregated into a sectoral good using a constant-elasticity-of-substitution (CES) production technology with elasticity of substitution θ_i:

$$Y_{it} = \left(\int_0^1 Y_{ift}^{(\theta_i - 1)/\theta_i} df \right)^{\theta_i/(\theta_i - 1)}.$$

It follows that firm (i, f)'s demand schedule is given by

$$Y_{ift}^d = \left(\frac{P_{ift}}{P_{it}} \right)^{-\theta_i} Y_{it}. \tag{5}$$

Households. The representative household has standard Golosov and Lucas (2007)'s preferences defined over a consumption aggregate C_t and total labor supply L_t:

$$U(C_t, L_t) = \log C_t - L_t, \tag{6}$$

The final consumption basket is given by $C_t = \mathcal{C}(C_{1t}, \ldots, C_{nt})$, where C_{it} is the household's consumption of the good produced by industry i and $\mathcal{C}$ is a homogeneous function of degree one. We assume the consumption aggregator $\mathcal{C}$ is a Cobb-Douglass function expressed as follows

$$\mathcal{C}(C_{1t}, \ldots, C_{nt}) = \prod_{i=1}^{n} C_{it}^{\beta_i}, \quad \sum_{i=1}^{n} \beta_i = 1.$$

The representative household's budget constraint is given by

$$P_t C_t = \sum_{i=1}^{n} P_{it} C_{it} \leq W_t^F L_t + \Pi_t + T_t,$$

where $P = \mathcal{P}(P_1, \ldots, P_n)$ is the nominal price of the household's consumption bundle implied by the consumption aggregator $\mathcal{C}$, Π_t represents total firms' profits, W^F is the flexible wage, and T represents government transfers to households.

To model aggregate demand fluctuations, we sidestep microfoundations and impose a cash-in-advance constraint on households:

$$P_t C_t \leq M_t, \tag{7}$$

where M_t is the money supply or nominal aggregate demand.

The imposed Golosov and Lucas (2007)'s preferences imply that the flexible nominal wage equals nominal demand:

$$W_t^F = M_t. \tag{8}$$

Labor union and sticky wages. As in Rubbo (2023), we model wage stickiness by introducing a labor union sector. The labor unions are treated like any other production sector, and in particular, there is a unit mass of unions, subject to rational inattention. As a result, there are many wages: the flexible wage W_t paid by the unions to the worker, and the prices charged by the various unions to the rest of the economy. The worker owns all the labor unions, and hence, she earns the profit rebates so that, in practice, labor income coincides with the average sticky price charged by the unions. The fact that unions are rationally inattentive generates wage stickiness.

The degree of wage stickiness is denoted by δ_w and measures the elasticity of the nominal wage paid by firms W_t with respect to the flexible wage paid by the labor union: $\log W_t = \delta_w \log W_t^F$. It is given by the optimal attention level chosen by labor unions in equilibrium, which will be determined later.

Government. The government's fiscal instrument is a collection of industry-specific taxes (or subsidies) on firms' revenues, with the resulting tax collection then rebated to the representative household as a lump-sum transfer. The government's budget constraint is given by

$$T_t = \sum_{i=1}^{n} \tau_i \int_0^1 P_{ift} Y_{ift} df. \tag{9}$$

We assume that M_t —which can be interpreted as either money supply or nominal aggregate demand- follows a log-normal distribution:

$$\log M_t \sim \mathcal{N}(0, \sigma_m^2). \tag{10}$$

Information structure. Firms are rationally inattentive about nominal demand M and sectoral productivities $\boldsymbol{Z}_t$. Given their correct prior given by (2) and (10), they choose which signals to observe from a rich set of available signals, $\mathbb{S}$, subject to an information processing constraint. Firms acquire an attention capacity $\kappa_{ift} \geq 0$ subject to a linear cost $\omega \times \kappa_{ift}$. Here, $\omega > 0$ is the information cost parameter.

After firms make their information choices, all shocks and signals are drawn, and each firm observes the realization of its signals. Firms then choose their prices conditional on their information sets, after which demand for each variety is realized. Firms then hire enough labor and intermediate inputs to satisfy demand according to the production function (1). While nominal prices are set under incomplete information, we assume that firms and the representative household make their quantity decisions after observing the prices and the shocks' realizations.

Formally, a strategy for firm (i, f) is to choose an information processing capacity κ_{ift}, a set of signals to observe $S_{ift} \subset \mathbb{S}$, and a pricing strategy that maps its information set to pricing decision, $P_{ift} : S_{ift} \to \mathbb{R}_+$. Given a strategy for all the other firms in the economy, firm (i, f) maximizes its profits given its prior:

$$\max_{S_{ift} \subset \mathbb{S}, P_{ift}(S_{ift}), \kappa_{ift}} \mathbb{E}\left[(1-\tau_i)P_{ift}Y^d_{ift} - W_t L_{ift} - \sum_{k=1}^{n} P_{kt}X_{ift,k} - \omega\kappa_{ift}\right]$$
$$\text{s.t.} \tag{11}$$
$$Y^d_{ift} = P_{ift}\left(\frac{P_{ift}}{P_{it}}\right)^{-\theta_i} Y_{it} \quad \text{(demand schedule)}$$
$$\mathcal{I}(S_{ift}; (M_t, (P_{kt})_{k\in I}, \boldsymbol{Z}_t)) \leq \kappa_{ift} \quad \text{(information processing constraint)}$$

where $\mathcal{I}(\cdot;\cdot)$ is Shannon's mutual information function and measures the amount of information that signals S_{ift} contain about the variables that are relevant to firms' pricing decision, namely, nominal demand M_t, other industries' prices $\boldsymbol{P}_t$, and sectoral productivities $\boldsymbol{Z}_t$.

Equilibrium. An equilibrium for this economy is an allocation for the household,

$$\Omega_H \equiv \{(C_{it})_{i\in I}, L_t : \ t \in \{0, 1, \ldots\}\},$$

a strategy profile for firms given an initial set of signals

$$\Omega_F \equiv \left\{\left(\kappa_{ift}, S_{ift} \subseteq \mathbb{S}, P_{ift}, L_{ift}, (X_{ift,k})_{k\in I}, Y_{ift}\right)_{i\in I, f\in[0,1]} : \ t \in \{0, 1, \ldots\}\right\}$$

and a set of prices

$$\{(P_{ift})_{i\in I, f\in[0,1]}, W_t, W^F_t : \ t \in \{0, 1, \ldots\}\},$$

such that: (a) given prices and Ω_F, Ω_H solves the household's utility maximization problem; (b) given prices and Ω_H, no firm has an incentive to deviate from Ω_F; (c) $\{M_t = P_t C_t\}$ satisfies the monetary policy rule (10) ; (d) the government budget (9) is balanced; (e) labor and goods markets clear.

2.2. Firms' price-setting and information acquisition decisions

In every period t, firms confront a decision-making problem comprising four choices: (i) determination of attention capacity κ_{ift}, (ii) selection of signals to observe S_{ift}, (iii) price setting for their differentiated product p_{ift}, and (iv) optimization of production inputs

$(\boldsymbol{X}_{ift}, L_{ift})$. We will solve the firms' problem backward, focusing on their information acquisition and pricing decisions.

Inputs choice. When choosing inputs, firms operate under complete information, making decisions once all signals and prices are realized. Their objective at this stage is to minimize costs while meeting the demand for their products Y^d_{if}, taking wages and input prices as given:

$$\min W_t L_{ift} + \sum_{k=1}^{n} P_{kt} X_{ift,k} \quad s.t.: \quad F_i(L_{ift}, X_{ift,k}) = Y^d_{ift}.$$

Firm (i, f)'s demand for labor and intermediate inputs is given by

$$L_{ift} = Z_{it}^{-1} \alpha_i \left(\prod_{j=1}^{n} P_j^{a_{ij}} \right) W_t^{\alpha_i - 1} Y_{ift}$$
$$X_{ift,k} = Z_{it}^{-1} a_{ik} \left(\prod_{j=1}^{n} P_{jt}^{a_{ij}} \right) \frac{W_t^{\alpha_i}}{P_{kt}} Y_{ift}.$$

Price setting. After solving for the input choice, firm (i, f)'s profit function can be expressed as follows

$$\Pi(P_{ift}, \mathbf{P}_t, Y_{it}, W_t, Z_{it}) = (1 - \tau_i) P_{ift} \left(\frac{P_{ift}}{P_{it}} \right)^{-\theta_i} Y_{it} - Z_{it}^{-1} \left(\prod_{k=1}^{n} P_{kt}^{a_{ik}} \right) W_t^{\alpha_i} \left(\frac{P_{ift}}{P_{it}} \right)^{-\theta_i} Y_{it},$$

where $\mathbf{P}_t \equiv (P_{it})_{i \in I}$ is the vector of sectoral prices.

To circumvent the complexity of solving rational inattention models, we adopt the usual approach of taking a second-order approximation to the firms' information acquisition and price-setting problems. We derive this second-order approximation around the full-information equilibrium.[2] The approximated firms' profit function is

$$\pi(p_{ift}, \mathbf{p}_t, y_{it}, w_t, z_{it}) = \Pi_1 p_{ift} + \frac{\Pi_{11}}{2} p_{ift}^2 + \Pi_{12}' \mathbf{p}_t p_{ift} + \Pi_{13} y_{it} p_{ift} + \Pi_{14} w_t p_{ift} + \Pi_{15} z_{it} p_{ift}$$
$$+ \text{terms independent of } p_{ift},$$

where all the derivatives of the original profit function are evaluated at the full-information equilibrium. A small letter denotes the log-deviation of the corresponding variable from its steady-state value.

[2] The full-information equilibrium is efficient since prices are fully flexible and marginal costs are equalized across industries. Hence, $P_{ift} = P_{it} = P_t = M_t$ and $Y_{ift} = Y_{it} = 1$ for all $i \in I$ and $f \in [0, 1]$.

After computing the derivatives of the original profit function at the full-information equilibrium, it follows that a firm's optimal price p^*_{ift} is equal to its industry-specific marginal cost:

$$p^*_{ift} = \alpha_i w_t + \sum_{k=1}^{n} a_{ik} p_{kt} - z_{it}. \tag{12}$$

Since all firms within an industry have the same marginal cost (or the same optimal price), hereafter, we will drop the f subscript and denote it by p^*_{it}.

Information acquisition. Following the second-order approximation and using the solution to firms' input choice, we can rewrite the firms' problem (11) as

$$\begin{aligned} \min_{\kappa_{ift}, S_{ift}, p_{ift}(S_{ift})} \quad & \frac{1}{2} B_i \mathbb{E}_{ift}\left[(p_{ift}(S_{ift}) - p^*_{it})^2\right] + \omega \kappa_{ift} \\ \text{s.t.}: \quad & p^*_{it} = \alpha_i \delta_w m_t + \sum_{k=1}^{n} a_{ik} p_{kt} - z_{it}, \end{aligned} \tag{13}$$

where $B_i \equiv |\Pi_{11}| = \theta_i$ is the curvature of the profit function around the full-information equilibrium, which equals the elasticity of substitution across varieties within industry i, θ_i. Note that we also used the equilibrium condition $w_t = \delta_w m_t$.

It can be shown that in all equilibria, each firm observes only one signal, collinear with their price, as proved in single-agent rational inattention settings (Maćkowiak, Matějka, and Wiederholt 2018; Afrouzi and Yang 2021). We call the strategies with this property as recommendation strategies: $p_{ift} = S_{ift}, \;\; S_{ift} \in \mathbb{S}$. The idea is that all alternative strategies are weakly dominated by feasible recommendation strategies. The equilibrium relationship for signals and prices is given by:

$$p_{ift} = \lambda_{if}\left(\alpha_i \delta_w m_t + \sum_{k=1}^{n} a_{ik} p_{kt} - z_{it}\right) + \nu_{ift}, \quad \nu_{ift} \perp \left(m_t, z_{it}, (S_{i',f',t})_{(i',f') \neq (i,f)}\right). \tag{14}$$

Here, $\lambda_{ift} \equiv 1 - e^{-2\kappa_{ift}}$ represents the Kalman gain on the signal, where κ_{ift} is the attention capacity. We call λ_{ift} firm (i, f)'s attention level (to its own marginal cost). The term ν_{ift} represents the noise in prices induced by rational inattention and has the following stochastic properties:

$$\mathbb{E}[\nu_{ift}] = 0, \quad Var[\nu_{ift}] = \lambda_{ift}(1 - \lambda_{ift}) Var[p^*_{ift}]. \tag{15}$$

A proof for the characterization of equilibrium prices above, as well as the feasibility and feasibility of recommendation strategies, is provided in Afrouzi (2024).

The information acquisition problem (13) is time-invariant since shocks are i.i.d., and identical among all firms within the same industry. This implies that $\lambda_{ift} = \lambda_i$ for all firms $f \in [0, 1]$ in each industry $i \in I$ and every period t.

Attention capacity. As we have solved firms' signal selection problem for a given attention capacity κ_i, we can rewrite firms' information problem just in terms of κ_i as follows:

$$\min_{\kappa_i \geq 0} \left\{ \frac{1}{2} e^{-2\kappa_i} B_i V_i^* + \omega \kappa_i \right\},$$

where $V_i^* \equiv Var(p_{it}^*) = Var\left(\alpha_i \delta_w m_t + \sum_{k=1}^n a_{ik} p_{kt} - z_{it}\right)$ is the variance of firm (i, f)'s optimal price, or of its industry-specific marginal cost. Since this problem is identical to all firms in the same industry and time-invariant, we have dropped the f and t subscripts. The solution to this problem is:

$$\kappa_i = \frac{1}{2} \max \left\{ 0, \ln \left(\frac{B_i V_i^*}{\omega} \right) \right\}.$$

We focus on the case where all firms in the economy acquire some positive amount of information, i.e. $\kappa_i > 0$ for all $i \in I$. This holds when $\omega < \theta_i(\alpha_i^2 \delta_w^2 \sigma_m^2 + \sigma_{z,i}^2)$ for all $i \in I$, i.e., the cost of information is low enough for all firms operating in the various industries. Then, industry i's equilibrium attention is given by

$$\lambda_i = 1 - \frac{\omega}{\theta_i V_i^*}, \tag{16}$$

where we used the fact that $B_i = \theta_i$. This formulation captures the trade-off firms face when choosing the optimal attention level. They balance the benefits of more precise information, which depend on the curvature of the profit function with respect to price mistakes, B_i, and the volatility of marginal costs V_i, against the cost of attention, ω.

For the labor union, its optimal attention problem determines the degree of wage stickiness δ_w in the economy:

$$\delta_w = \lambda_U = 1 - \frac{\omega}{\theta_U \sigma_m^2}. \tag{17}$$

2.3. Equilibrium characterization

Rewriting conditions (14) and (16) in terms of industry outcomes in matrix and vector forms, we can characterize the equilibrium as follows:

PROPOSITION 1. *In equilibrium, sectoral prices and attention levels are given by*

$$\boldsymbol{p}_t = \Lambda(I - A\Lambda)^{-1}(\boldsymbol{\alpha}\delta_w m_t - \boldsymbol{z}_t) \tag{18}$$

$$\Lambda = I - \omega diag(\theta)^{-1}\left(Var(\Lambda^{-1}\boldsymbol{p}_t) \odot I\right)^{-1}, \tag{19}$$

where $\boldsymbol{p} \equiv (p_i)_{i \in I}$, $\Lambda \equiv diag(\boldsymbol{\lambda})$ *and* $\odot$ *denotes the Hadamard product (element-wise matrix multiplication).*

PROOF. See Appendix A.1. □

The key fixed point in our model is the relationship between price responses and attention levels. Intuitively, lower attention in a specific industry leads to dampened price responses to aggregate and sectoral shocks, which decreases the volatility of marginal costs in other sectors and encourages further inattention within those sectors.

This is a highly non-linear system of equations, with no closed-form solution for a general input-output structure $(A, \boldsymbol{\alpha})$. Still, we are able to derive many insightful conclusions regarding the determinants and macroeconomic consequences of heterogeneous firms' expectations for general production networks, which will be explored in the next section.

First, we prove equilibrium existence and uniqueness for small information costs, adapting the fixed-point methodology from Fang et al. (2024) for models with endogenous information acquisition and strategic complementarities.

PROPOSITION 2. *There exists a unique fixed point* Λ *that satisfies conditions* (18) *and* (19) *if the information cost* ω *satisfies:*

$$0 < \omega < \min_{i \in I}\left\{\theta_i(\alpha_i^2\delta_w^2\sigma_m^2 + \sigma_{z,i}^2)\right\}. \tag{20}$$

PROOF. See Appendix A.2. □

3. Heterogeneous firms' expectations: causes and consequences

In this section, we examine the determinants and macroeconomic implications of firms' heterogeneous attention in our framework. The model provides four theoretical insights:

two regarding the determinants of heterogeneous attention to inflation and two concerning its aggregates consequences.

On the determinants of attention, we show that inattention compounds downstream in the production network due to strategic complementarities in information acquisition decisions. Moreover, we decompose firms' expectations about the price level (or inflation) into three components: first, firms' attention to their own marginal cost; second, the degree of comovement between marginal costs and the price level; and third, variation in the industry-specific marginal cost.

On the consequences of heterogeneous attention, we show that it changes the propagation of shocks –either amplifying or dampening them – to the extent that attention levels covary with marginal costs' responses across industries. Finally, we revisit an important result in the literature on inflation and networks, namely that the Phillips curve is flatter in economies with stronger input-output linkages (Basu 1995; Rubbo 2023). We show that this result hold may not hold in our framework, where the degree of nominal rigidities, determined by the equilibrium attention levels, is endogenous to the input-output structure of the economy.

3.1. Downstream compounding of inattention

Our first result shows how inattention in a particular sector affects attention in other sectors.

LEMMA 1. *Information acquisition decisions display strategic complementarities. Concretely, for any industries $i, j \in I$,*

$$\frac{d\lambda_i}{d\lambda_j} > 0,$$

whenever $(I - A\Lambda)^{-1}_{ij} > 0$.

PROOF. See appendix A.3. □

This result shows that inattention compounds downstream in the production network. The (i, j)-th element of the matrix $(I - A\Lambda)^{-1}$ corresponds to the elasticity of sector i's marginal cost with respect to sector j's marginal cost, directly, and indirectly through the input-output network. Therefore, the optimal attention choice by industry j will influence industry i's information acquisition decision as long as those industries are connected through the network structure of the economy directly or indirectly.

3.2. The determinants of firms' inflation expectations

We now provide a decomposition of firms' expectations about the consumer price index (CPI).

PROPOSITION 3. *Industry-average CPI forecasts are given by*

$$\bar{\mathbb{E}}_i[p_t^{CPI}] = \underbrace{\frac{Cov(p_{it}^*, p_t^{CPI})}{Var(p_{it}^*)}}_{\text{comovement between mc and CPI}} \quad \underbrace{\lambda_i}_{\text{attention to own mc}} \quad \underbrace{p_{it}^*}_{\text{realized mc}}, \tag{21}$$

where $p^{CPI} \equiv \boldsymbol{\beta}' \boldsymbol{p}$ *is the log of the CPI.*

PROOF. See appendix A.4. □

The intuition behind the above result rests on the fact that rationally inattention firms acquire information about their industry-specific marginal costs, which is the key variable for their profit maximization problem. Thus, firms have more precise CPI expectations when marginal costs exhibit a stronger comovement with the price index. Also, industries optimally choose different attention levels, due to heterogeneous incentives to acquire information. At the firm level, CPI forecasts can be decomposed into two components: an industry-average component detailed in (3) and an idiosyncratic noise term.

Our decomposition yields a testable prediction, which will be empirically validated using data on firms' beliefs and industry-level marginal costs in section 5.

In sum, firms learn optimally about the shocks that mostly impact them. If certain shocks significantly affect a specific industry and turn out to be key drivers of inflation, firms within that industry will have more accurate forecasts. Therefore, an important question arises: which shocks explain fluctuations in inflation the most? Intuitively, three main types of shocks stand out: (i) shocks that impact sectors with high consumption shares, (ii) shocks that have a greater propagation effect, such as those affecting upstream sectors or aggregate shocks, and (iii) more volatile shocks.

3.3. Heterogeneous attention and the aggregate effects of shocks

We now investigate how the heterogeneity in attention alters the transmission of shocks on the aggregate price level.

LEMMA 2. *Let* $p_t^{\phi} \equiv \phi' \boldsymbol{p}_t$ *be an aggregate price index for any weight* ϕ *in the simplex* Δ^n*, and* $\tilde{\delta}^x \equiv d\boldsymbol{p}_t^*/dx$ *be the vector of marginal costs responses to a shock* $x \in \{m, \boldsymbol{z}\}$*. The effect of a shock* *to x on the aggregate price level is given by*

$$\frac{dp_t^\phi}{dx_t} = \underbrace{\mathbb{E}_\phi[\lambda]\mathbb{E}_\phi\left[\tilde{\delta}^x\right]}_{\textit{representative agent channel}} + \underbrace{Cov_\phi(\lambda, \tilde{\delta}^x)}_{\textit{heterogeneous attention channel}} .$$

PROOF. See Appendix A.5. □

The first term captures the representative agent channel: the average attention and the average response of marginal costs to the shock shape its propagation. The second term captures the insight that the aggregate effect of shocks will be larger if the more attentive sectors are the ones whose marginal costs respond more to the shock. For instance, there is more monetary non-neutrality if marginal costs are more responsive to changes in aggregate nominal demand in less attentive industries.

3.4. The slope of industry-level Phillips curves

Let $\tilde{y}$ be the economy's output gap defined as the log deviation of the equilibrium aggregate output to its flexible-price counterpart. We leverage equation (18) to derive sector-level "Phillips curves", linking industry price levels to the aggregate output gap and sectoral productivities.

LEMMA 3. *The reduced-form sector-level "Phillips curves" are given by*

$$\begin{gathered} \boldsymbol{p}_t = \mathcal{B}\tilde{y}_t - \mathcal{V}\boldsymbol{z}_t, \\ \textit{where} \quad \mathcal{B} \equiv \frac{\Lambda(I - A\Lambda)^{-1}\boldsymbol{\alpha}\delta_w}{1 - \boldsymbol{\beta}'\Lambda(I - A\Lambda)\boldsymbol{\alpha}\delta_w}, \\ \mathcal{V} \equiv [\Lambda(I - A\Lambda)^{-1} - \mathcal{B}((I - A)^{-1}\boldsymbol{\beta} - \boldsymbol{\beta}'\Lambda(I - A\Lambda)^{-1})](I - A). \end{gathered} \tag{22}$$

PROOF. See Appendix A.6. □

The vector $\mathcal{B}$ denotes the slope of industry-level Phillips curves. It is determined by two terms. The term $\Lambda(I - A\Lambda)^{-1}$ represents the elasticity of prices with respect to nominal wages, which combines the direct of nominal wages on marginal costs, given by the labor shares $\boldsymbol{\alpha}$, with the indirect effects shaped by the input-output linkages represented by $(I_A\Lambda)^{-1}$, and the response of prices to marginal costs, given by the attention levels Λ. The second term, $(1 - \boldsymbol{\beta}'\Lambda(I - A\Lambda)\boldsymbol{\alpha})^{-1}$, captures the elasticity of nominal wages with respect to the output gap.

All industry-level Phillips curves are flatter in less attentive economies, as inattention generates price stickiness. However, our framework provides a qualification for an important result of the literature on inflation and networks. Specifically, several papers examining price setting and inflation dynamics in production networks (Basu 1995; Rubbo 2023, among others,) find that Phillips curves are flatter in economies with larger intermediate input shares, as network linkages compound nominal frictions present in each industry. In our setting, the effects of changes in the input-output linkages of the economy on the Phillips curves' slope are ambiguous, as equilibrium attention levels are endogenous to the input-output structure. Corollary 1 formalizes these insights.

COROLLARY 1. *The slope of sectoral Phillips curves, $\mathcal{B}$, is weakly increasing in each attention level $(\lambda_i)_{i \in I}$. However, the effects of changes in the input-output structure of the economy $(A, \boldsymbol{\alpha})$ on $\mathcal{B}$ are ambiguous.*

PROOF. See Appendix A.7. □

4. Data and measurement

We start by discussing our dataset and emphasizing its features relevant for measurement purposes.

4.1. Data

The Survey of Firms' Inflation Expectations (SoFIE). At the center of our empirical analysis is the firm-level data on inflation expectations sourced from the Survey of Firms' Inflation Expectations (SoFIE).

The SoFIE is a nationally representative, quarterly survey of inflation expectations conducted by the Federal Reserve Bank of Cleveland. It elicits inflation expectations from CEOs and other top executives and has been conducted since the second quarter of 2018. The survey relies on the recurring participation of a panel of firms, with between 300 and 700 firms participating each quarter.

The dataset provides limited information about each respondent's firm, including the sector in which it operates (services or manufacturing), its industry, and its size, measured by the number of employees. Firms are selected randomly from the manufacturing and services sectors to accurately represent the underlying structure of each sector's contribution to aggregate gross value added.[3]

[3]Within the manufacturing sector, companies are classified into food and drink, textiles and clothing,

The survey includes two questions regarding firms' inflation expectations each quarter. The first question, which remains constant across all waves, asks respondents to provide their expectations for the Consumer Price Index (CPI) inflation rate over the upcoming twelve-month period. Specifically, the question is:

> *What do you think will be the inflation rate (for the Consumer Price Index) over the next 12 months? Please provide an answer in an annual percentage rate.*

For the second question, there is a rotation across four different questions. Each question is asked once per year. In the first quarter, firms are asked about the probability that inflation one year later will exceed 5%. In the second quarter, they are asked what inflation rate they think the Federal Reserve is targeting on average. In the third quarter, they are asked what they think the inflation rate has been over the last twelve months. In the fourth quarter of each year, they are asked about what they think inflation will be over the next five years on average.

Candia, Coibion, and Gorodnichenko (2024) highlights several appealing attributes of this survey for academic research. First, it covers a wide range of firms across various sizes and industries, allowing the construction of a time series of firms' inflation expectations representative of the U.S. economy. Second, it offers a quantitative measure of inflation expectations, in contrast to the primarily qualitative assessments found in other surveys. Third, it enhances the time series of inflation expectations with additional metrics on perceived inflation, inflation uncertainty, long-term expectations, and knowledge of the Fed's inflation target. Together, these elements provide the most comprehensive view of what U.S. firms understand and anticipate regarding inflation and monetary policy (Candia, Coibion, and Gorodnichenko 2024).

On the other hand, the SoFIE has two key limitations worth noting. First, it provides only basic information about firm characteristics, such as industry and company size, to preserve their anonymity. As a result, we cannot link the expectations data to firms' actual outcomes or administrative data. Second, the survey focuses exclusively on various measures of CPI inflation expectations. Data regarding expectations about firm- or industry-level outcomes, including beliefs about firms' own price changes or expected average price changes within their industry, would be extremely valuable for testing our model.

---

electrical, chemicals and plastics, transport, timber and paper, basic metals, mechanical engineering, and other manufacturing. Within the services sector, the companies are classified into hotels and restaurants, transport and storage, post and telecommunication, financial intermediation, renting and business activities, and other services.

Industry-level data sources. We use three data sources with industry-level information to empirically study how industry characteristics explain differences in firms' inflation expectations.

First, we use the 2022 input-output tables constructed by the Bureau of Economic Analysis (BEA) to compute different measures of industry importance within the production network or the overall economy. Our second data source is from Pastén, Schoenle, and Weber (2024), who compute measures of the probability of price adjustments at the industry level using the microdata that underlies the Bureau of Labor Statistics (BLS) producer price index (PPI). They construct this measure by taking the ratio of the number of price changes to the number of sample months. Third, we use the BEA/BLS Integrated Industry-Level Production Account (ILPA) data, which provides estimates of industry-specific productivity from 1987 to 2021.

4.2. Measurement

Forecast errors. We calculate the accuracy of firms' inflation forecasts as the absolute difference between one-year-ahead forecasts and the actual (12-month-ahead) realizations of US consumer price index inflation. We start by constructing a series of firms' inflation forecast errors as the difference between the actual CPI inflation rate and the respondent's forecast:

$$\text{FE}_{fit}[\pi_t] \equiv \pi_t - \mathbb{E}_{fit-4}[\pi_t]. \tag{23}$$

Thus, a positive forecast error corresponds to an underestimation of the variable. We then compute the accuracy of inflation forecasts as the absolute value of forecast errors.

Figure A1 in the appendix shows that the time series of industry inflation forecasts exhibit similar dynamics in levels, closely aligning with inflation fluctuations. However, some industries exhibit significantly more volatile expectations compared to others.

Industry importance. Measures of industries' importance in the economy are constructed at the two-digit ISIC industry classification level for consistency with the SoFIE dataset. Our calculations follow the construction of direct input and labor requirements in Rubbo (2023).

Specifically, we compute industries' total sales as a fraction of GDP, also known as Domar weights, and a measure of industry "upstreamness", which captures the importance of a particular industry as a *direct and indirect* supplier to all industries in the economy.

Formally, let A be the input-output matrix and $L \equiv (I - A)^{-1}$ be the Leontief inverse matrix. The entry (i, j) of the Leontief inverse matrix L quantifies the direct and indirect dependencies of sector i on sector j. The upstreamness measure is given by

$$u_i \equiv \sum_{k=1}^{n} \ell_{ki} - \mathbb{1}\{i = k\}, \tag{24}$$

i.e., it is the i-th column sum of the Leontief inverse (Carvalho and Tahbaz-Salehi 2019).

Volatility of marginal costs. We compute the volatility of industry-level productivities after linearly detrending the productivity series from ILPA for each industry.

Marginal costs. We use the BEA input-output data to compute approximate measures of industry-specific marginal costs. Under Cobb-Douglass technologies, Gagliardone et al. (2023) show that an industry's nominal log-marginal cost equals the log of its nominal average variable costs plus a term representing the degree of returns to scale:

$$mc_{it} = \ln(\mathrm{TVC}_{it}/Y_{it}) + \ln(1 + \zeta_i). \tag{25}$$

Although we assume constant returns to scale in our model, we allow for a potentially arbitrary scale of production in our empirical design by controlling for industry fixed effects.[4] Using data from the BEA input-output accounts, we compute total variable costs TVC_{it} as the sum of intermediate input and labor costs. We construct an industry-specific quantity index by dividing industry sales by its corresponding industry price index.

Industry-specific price stickiness. We aggregate Pastén, Schoenle, and Weber (2024)'s estimates of industry-level probabilities of price readjustment to match the two-digit industry classification in the SoFIE data by calculating a simple average based on the more granular industry classification they have.

5. Empirical results

The model makes two predictions about how firms form beliefs relevant to their price-setting decision. First, firms' expectations are influenced by the characteristics of their respective industries. Second, as outlined in Proposition 3, firms' CPI expectations are

[4]For more general production functions, the return to scale term may vary over time.

TABLE 1. Industry characteristics and inflation expectations

	(1)	(2)	(3)	(4)	(5)
	Dependent variable:				
	$\lvert FE_{ift}[\pi_{t+4}]\rvert$			$\lvert FE_{ift}[\pi_t]\rvert$	
Industry attribute					
Upstreamness	-0.491**	-0.472**	-0.489***	-0.47*	-0.49**
	(0.213)	(0.220)	(0.210)	(0.200)	(0.215)
Sales/GDP	0.13	0.12	0.14	0.15	0.11
	(1.02)	(1.03)	(1.01)	(1.00)	(1.04)
TFP volatility	1.41***	1.38***	1.42***	1.38***	1.43***
	(0.40)	(0.41)	(0.39)	(0.42)	(0.38)
Price stickiness	0.21	0.20	0.22	0.19	0.23
	(0.12)	(0.13)	(0.11)	(0.14)	(0.10)
Time FE		✓	✓		✓
Firm FE			✓		
Observations	11,253	11,253	11,253	3,234	3,234
R-squared	0.172	0.281	0.395	0.162	0.312

Note: Table shows the results of regression (26). For columns 1-4, the outcome $|FE_{ift}[\pi_{t+4}]|$ is the absolute value of the firms' 12-month-ahead inflation forecast error. For columns 5-6, the outcome $|FE_{ift}[\pi_t]|$ is the absolute value of the firm's nowcast error. Standard errors are clustered at the industry level. ***, **, * indicate statistical significance at 1, 5, and 10 percent, respectively.

tightly linked to their perceptions of their industry-specific marginal costs. In this section, we will test both predictions of the model.

5.1. Industry characteristics and inflation forecasts

We first estimate how industry characteristics correlate with the accuracy of firms' inflation forecasts, a measure of firms' "attention" to inflation, by running the following regression:

$$|\mathrm{FE}_{ift}[\pi_t]| = \chi_f + \chi_t + \boldsymbol{\gamma}'\boldsymbol{X}_i + \mathrm{error}_i, \tag{26}$$

where $|\mathrm{FE}_{ift}[\pi_t]|$ is the absolute inflation forecast error of a firm f in industry i in quarter t, and χ_f and χ_t are firm and time fixed effects, respectively. $\boldsymbol{X}_i$ is a vector that stores the following time-invariant industry characteristics: industry upstreamness as previously explained, the industry sales-to-GDP ratio or its size in the economy, the industry-specific probability of price readjustment, and the volatility of industry-specific productivity shocks. γ are the coefficients of interest.

Table 1 summarizes the results of this regression.

The coefficient on the industry upstreamness measure in regression (26) is negative and statistically significant at the 5% level, indicating that industries with higher price stickiness are more attentive to inflation. Under the lens of our theory, this result is consistent with the fact, that upstream industries attain higher attention levels regarding their marginal cost), as inattention compounds downstream in the production network, as demonstrated in Lemma 1. Consequently, lower attention levels lead to less accurate inflation forecasts, as shown in our decomposition of inflation expectations derived in Proposition 3.

The coefficient on the volatility of sector-specific TFP shocks is positive and statistically significant at the 1% significance level, indicating that firms operating in industries subject to more volatile productivity shocks have more inaccurate inflation expectations. This result aligns with the observation that the comovement between marginal costs in those industries and aggregate inflation is low, as industry-specific conditions primarily determine the former while aggregate factors mainly drive the latter. In other words, although firms facing more volatile industry-specific conditions may have precise information about their marginal cost, they do not learn much about the aggregate price given their noisy observation of marginal costs.

It is important to note that our model does not provide a clear prediction regarding the sign of these coefficients; therefore, regression (26) is not a straightforward test of our theory. However, we can compare our empirical estimates with the model counterparts after simulating a calibrated version of our model.

5.2. Marginal costs and inflation forecasts

The theory presented in sections 2 and 3 makes a sharp prediction regarding the firms' belief formation process about the price level, given by:

$$\mathbb{E}_{fit}[p_t^{CPI}] = \lambda_i \frac{Cov(p_{it}^*, p_t^{CPI})}{Var(p_{it}^*)} p_{it}^* + u_{fit}, \quad u_{fit} \perp p_{it}^*. \tag{27}$$

The equation above yields two predictions. First, firms' inflation expectations should strongly comove with their marginal costs. Second, the coefficient should vary across industries according to differences in their attention levels and the comovement between marginal costs and the CPI.

Based on the theoretical result above, we regress firms' inflation expectations on our

TABLE 2. Marginal costs and inflation expectations

	(1)	(2)	(3)	(4)	(5)
	Dependent variable:				
	1-year ahead forecast			Nowcast	
mc_{it}	0.021***	0.020***	0.016**	0.019***	0.018***
	(0.0053)	(0.0072)	(0.008)	(0.0067)	(0.0071)
Time FE		✓	✓		✓
Firm FE			✓		
Observations	11,253	11,253	11,253	3,234	3,234
R-squared	0.321	0.395	0.486	0.310	0.385

Note: Table shows the results of regression (28). For columns 1-4, the outcome is the firms' 12-month-ahead inflation expectations. For columns 5-6, the outcome is the firm's inflation nowcast. Standard errors are clustered at the industry level. ***, **, * indicate statistical significance at 1, 5, and 10 percent, respectively.

measure of industry-specific marginal costs:

$$\mathbb{E}_{fit}[\pi_t] = \chi_f + \chi_t + \beta \mathrm{mc}_{it} + \mathrm{error}_{fit}. \tag{28}$$

We guard against potential misspecifications in our model by including firm and time fixed effects.

The results indicate a robust positive relationship between firms' marginal costs and their inflation expectations across both 12-month-ahead forecasts and nowcasts. Marginal costs are statistically significant at the 1% level in all specifications, with coefficients ranging from 0.016** to 0.021*** for short-term expectations and 0.018***–0.019*** for nowcasts. This suggests that a 1-unit increase in marginal costs raises firms' inflation expectations by approximately 1.6–2.1 percentage points, with the effect slightly stronger for longer-term forecasts. The diminishing magnitude when firm fixed effects are introduced hints that unobserved firm-level heterogeneity may partially mediate the relationship, though the persistence of significance underscores the structural role of cost pressures in shaping expectations.

6. Quantitative analysis

The theory presented in sections 2 and 3 makes a sharp prediction regarding the firms' belief formation process about the price level, given by:

$$\mathbb{E}_{fit}[p_t^{CPI}] = \lambda_i \frac{Cov(p_{it}^*, p_t^{CPI})}{Var(p_{it}^*)} p_{it}^* + u_{fit}, \quad u_{fit} \perp p_{it}^*. \tag{29}$$

TABLE 3. Calibration

Parameter	Description	Value	Source
α, A, β	IO params		BEA IO 2022
δ_w	Wage flexibility	0.30	Beraja, Hurst, and Ospina (2019)
σ_m^2	Variance of nominal demand	0.037	US NGDP quarterly growth
Σ	Cov. of sectoral TFP		ILPA/KLEMS
θ	CES	6	Standard value

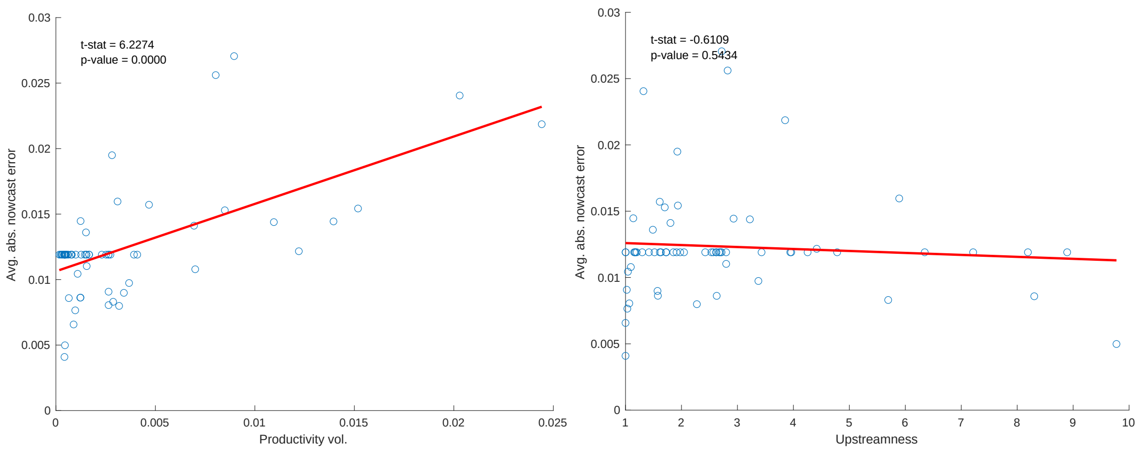
The equation above yields two predictions. First, firms' inflation expectations should strongly comove with their marginal costs. Second, the coefficient should vary across industries according to differences in their attention levels and the comovement between marginal costs and the CPI.

Calibration. In this section, we calibrate our model using the production network across 66 industries.[5] For our baseline exercise, we assume that all industries share the same cost of acquiring information, $\omega = 0.037$, to match an average absolute nowcast error of about 2.2%. The labor shares α, input-output matrix A, and consumption shares β are calibrated to match the BEA input-output accounts from 2022. Wage flexibility is set equal to 0.3, in line with Beraja, Hurst, and Ospina (2019). The variance of the monetary shock is set to 0.037 to match the standard deviation of the US nominal GDP quarterly growth rate. The covariance matrix of sectoral TFP shocks is based on the BEA/BLS Integrated Industry-Level Production Account (ILPA) data. Finally, the elasticity of substitution among goods produced within an industry is set equal to 6 for all industries as in La'O and Tahbaz-Salehi (2022).

Based on the calibration provided in Table 3, we compute the relationship between the model-implied nowcast errors and TFP volatility as well as industry upstreamness at the industry level. The left panel in Figure 2 scatter plots the industry-specific absolute value of nowcast errors against TFP volatility. Consistent with the evidence presented in Table 1, the calibrated model implies that the relationship is positive. On the other hand, the right panel of the same figure scatter plots the industry-specific absolute value of nowcast errors against industry upstreamness. In this first calibration of the model, we find a small, yet negative, relationship between upstreamness and the absolute value of industry-specific nowcast errors.

[5] We rely on the BEA input-output tables at the sumamry level and exclude industries corresponding to federal, state, and local governments.

FIGURE 2. Model-implied relationship between nowcast errors and volatility vs upstreamness



Note: The left panel plots the industry-specific absolute value of nowcast errors against TFP volatility. Industry-specific TFP volatility is extracted from the diagonal of the Σ matrix. Similarly, the right panel plots the industry-specific absolute value of nowcast errors against upstreamness. Upstreamness is computed as described in subsection 4.2. We plot the fitted lines in red.

Importance of the production network. In a one-sector economy, higher TFP volatility would yield more attention and smaller nowcast errors, implying a counterfactual negative relationship between nowcast errors and TFP volatility. By contrast, in a production network higher TFP volatility affects nowcasts through two opposing forces. On one hand, it generates more endogenous attention and hence smaller errors. On the other hand, the marginal cost of firms operating in industries that faces more volatile TFP shocks provides less precise informative about the aggregate price level, yielding thus higher errors. Whether one force is more prominent than the other is an empirical question and our results suggest strong evidence in favor of the latter channel being the dominant one.

Effects of monetary shocks. Finally, we use our calibrated model to quantify the effects of a monetary shock on the aggregate price level. Specifically, from Lemma 2 the effect of a monetary policy shock on the aggregate price level is given by

$$\frac{dp_t^{CPI}}{dm_t} = \mathbb{E}_\beta[\boldsymbol{\lambda}] \times \mathbb{E}_\beta\left[\frac{d\boldsymbol{p}_t^*}{m_t}\right] + Cov_\beta\left(\boldsymbol{\lambda}, \frac{d\boldsymbol{p}_t^*}{m_t}\right). \tag{30}$$

Table 4 summarizes the effect of the monetary shock on each of the components of dp_t^{CPI}/dm_t in equation (30). Our calibrated model reveals that the effect of a monetary shock is about 27% higher when compared to a model with homogeneous attention or no

TABLE 4. Decomposition of monetary shock effects

$\mathbb{E}_\beta[\boldsymbol{\lambda}]$	$\mathbb{E}_\beta\left[\frac{d\boldsymbol{p}_t^*}{dm_t}\right]$	$\mathrm{Cov}_\beta\left(\boldsymbol{\lambda}, \frac{d\boldsymbol{p}_t^*}{dm_t}\right)$	dp_t^{CPI}/dm_t
0.126	0.30	0.014	**0.052**

IO linkages. This result highlights that heterogeneity in firms' attention levels propagates the transmission of macroeconomic shocks to the aggregate price level.

7. Conclusion

This paper studies the determinants and macroeconomic consequences of firms' heterogenous inflation expectations, with a particular focus on the role of industry differences. We build a simple rational inattention model of price-setting firms operating within a production network.

The model provides four theoretical insights: two regarding the determinants of heterogeneous attention to inflation and two concerning its aggregate consequences. On the drivers of the heterogeneity in expectations, we show that inattention compounds downstream in the production network due to strategic complementarities in price setting and information acquisition decisions across industries. Our second result offers an insightful decomposition of firms' inflation expectations into three factors: (i) firms' attention to their own marginal cost, (ii) the degree of comovement between marginal costs and the price level, and (iii) variation in industry-specific marginal costs.

On the consequences of heterogeneous attention, we show that heterogeneous attention modulates the aggregate effects of shocks on the prices, as endogenous attention correlates with the responsiveness of marginal costs to them across industries. Finally, we derive sectoral Phillips curves and show that changes in the production network affect equilibrium attention levels and, consequently, the degree of nominal rigidities in the economy. Therefore, the relationship between the slope of the aggregate Phillips curves and the strength of input-output linkages is ambiguous, contrary to what other studies have found.

Empirically, we explore a novel and comprehensive dataset on US firms' inflation expectations and merge it with other data sources with industry characteristics to test our model. We find that the accuracy of firms' inflation forecasts is systematically associated with industry attributes. Specifically, firms in more downstream industries and those facing more volatile productivity shocks have less accurate inflation forecasts. Moreover, we find supportive evidence for the model's prediction regarding how firms form inflation expectations based on the realized industry-specific marginal costs.

Our framework is a suitable laboratory for understanding the role of inflation expectations under imperfect information for inflation dynamics and monetary policy when considering a more realistic representation of how firms operate interdependently and process information. In ongoing work, we are building a dynamic, more realistic version of our model to quantify the relevance of the theoretical insights outlined in the paper. This extended model will allow us to extract firms' attention levels from empirical data and assess the model's ability to replicate observed patterns of sectoral price stickiness.

References

Afrouzi, Hassan. 2024. "Strategic Inattention, Inflation Dynamics, and the Nonneutrality of Money." *Journal of Political Economy* 132 (10): 3378–3420.

Afrouzi, Hassan, Joel P Flynn, and Choongryul Yang. 2024. "What Can Measured Beliefs Tell Us About Monetary Non-Neutrality?" Working Paper 32541, National Bureau of Economic Research.

Afrouzi, Hassan, and Choongryul Yang. 2021. "Dynamic Rational Inattention and the Phillips Curve." CESifo Working Paper 8840, CESifo Group Munich.

Andrade, Philippe, Olivier Coibion, Erwan Gautier, and Yuriy Gorodnichenko. 2022. "No firm is an island? How industry conditions shape firms' expectations." *Journal of Monetary Economics* 125: 40–56.

Angeletos, George-Marios, and Zhen Huo. 2021. "Myopia and Anchoring." *American Economic Review* 111 (4): 1166–1200.

Basu, Susanto. 1995. "Intermediate Goods and Business Cycles: Implications for Productivity and Welfare." *The American Economic Review* 85 (3): 512–531.

Beraja, Martin, Erik Hurst, and Juan Ospina. 2019. "The Aggregate Implications of Regional Business Cycles." *Econometrica* 87 (6): 1789–1833.

Bui, Ha, Zhen Huo, Andrei A Levchenko, and Nitya Pandalai-Nayar. 2022. "Noisy Global Value Chains." Working Paper 30033, National Bureau of Economic Research.

Candia, Bernardo, Olivier Coibion, and Yuriy Gorodnichenko. 2024. "The inflation expectations of U.S. firms: Evidence from a new survey." *Journal of Monetary Economics* 145: 103569.

Carvalho, Vasco M., and Alireza Tahbaz-Salehi. 2019. "Production Networks: A Primer." *Annual Review of Economics* 11 (Volume 11, 2019): 635–663.

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

Fang, Yue, Zhen Huo, Minghao Li, and Jieran Wu. 2024. “Inattentive Network: Evidence and Theory.” Working paper, Zhejiang University, Yale University, Peking University.

Gagliardone, Luca, Mark Gertler, Simone Lenzu, and Joris Tielens. 2023. “Anatomy of the Phillips Curve: Micro Evidence and Macro Implications.” Working Paper 31382, National Bureau of Economic Research.

Golosov, Mikhail, and Robert E. Lucas. 2007. “Menu Costs and Phillips Curves.” *Journal of Political Economy* 115 (2): 171–199.

La’O, Jennifer, and Alireza Tahbaz-Salehi. 2022. “Optimal Monetary Policy in Production Networks.” *Econometrica* 90 (3): 1295–1336.

Mackowiak, Bartosz, and Mirko Wiederholt. 2009. “Optimal Sticky Prices under Rational Inattention.” *American Economic Review* 99 (3): 769–803.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2018. “Dynamic rational inattention: Analytical results.” *Journal of Economic Theory* 176: 650–692.

Maćkowiak, Bartosz, Emanuel Moench, and Mirko Wiederholt. 2009. “Sectoral price data and models of price setting.” *Journal of Monetary Economics* 56: S78–S99.

Nikolakoudis, George. 2024. “Incomplete Information in Production Networks.”, Princeton University.

Pastén, Ernesto, Raphael Schoenle, and Michael Weber. 2024. “Sectoral Heterogeneity in Nominal Price Rigidity and the Origin of Aggregate Fluctuations.” *American Economic Journal: Macroeconomics* 16 (2): 318–52.

Rubbo, Elisa. 2023. “Networks, Phillips Curves, and Monetary Policy.” *Econometrica* 91 (4): 1417–1455.

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

Weber, Michael, Bernardo Candia, Hassan Afrouzi, Tiziano Ropele, Rodrigo Lluberas, Serafin Frache, Brent Meyer, Saten Kumar, Yuriy Gorodnichenko, Dimitris Georgarakos, Olivier Coibion, Geoff Kenny, and Jorge Ponce. 2025. “Tell Me Something I Don’t Already Know: Learning in Low- and High-Inflation Settings.” *Econometrica* 93 (1): 229–264.

Appendix A. Proofs

A.1. Proof of Proposition 1

PROOF. Writing the condition determining marginal costs (12) in vector form yields:

$$\boldsymbol{p}_t^* = \boldsymbol{\alpha}\delta_w m_t + A\boldsymbol{p}_t - \boldsymbol{z}_t \tag{A1}$$

The equilibrium condition for industry prices in vector form is $\boldsymbol{p}_t = \Lambda \boldsymbol{p}_t^*$ by (14). Substituting that in the marginal cost condition above, it follows that

$$\boldsymbol{p}_t = \Lambda(I - A\Lambda)^{-1}(\boldsymbol{\alpha}\delta_w m_t - \boldsymbol{z}_t). \tag{A2}$$

Writing the optimal attention condition (16) in matrix form gives

$$\Lambda = I - \omega \text{diag}(\theta)^{-1}(\text{Var}(\Lambda^{-1}\boldsymbol{p}_t) \odot I)^{-1}. \tag{A3}$$

□

A.2. Proof of Proposition 2

PROOF. TBC. □

A.3. Proof of Lemma 1

PROOF. The covariance matrix of industry-specific marginal costs is given by:

$$Var(\boldsymbol{p}_t^*) = \tilde{L}(\boldsymbol{\alpha}\boldsymbol{\alpha}'\delta_w^2\sigma_m^2 + \Sigma)\tilde{L}', \tag{A4}$$

where $\tilde{L} \equiv (I - A\Lambda)^{-1}$.

The elements of $\tilde{L}$ are increasing in attention levels $\boldsymbol{\lambda}$. Specifically, for any industries $i, j \in I$, if $(I - A\Lambda)^{-1}_{ij} > 0$, then an increase in λ_j increases the magnitude of $(I - A\Lambda)^{-1}_{ij}$, which in turn increases λ_i as firms respond to higher volatility in marginal costs by acquiring more information as highlighted by equation (16). □

A.4. Proof of Proposition 3

PROOF. Since the equilibrium is a recommendation strategy, each firm's price equals its signal at time t. Projecting the CPI onto each firm's price, it follows that

$$p_t^{CPI} = \zeta_{if} p_{ift} + u_{ift}.$$

where u_{ift} is a residual term orthogonal to p_{ift} and $\zeta_{if} \equiv Cov(p_{ift}, p_t^{CPI})/Var(p_{ift})$. Using the equilibrium price condition (14), it follows that

$$\begin{aligned}\zeta_{if} &= \frac{Cov(\lambda_i p_{it}^* + \nu_{ift}, p_t^{CPI})}{Var(\lambda_i p_{it}^* + \nu_{ift})} \\ &= \frac{Cov(p_{it}^*, p_t^{CPI})}{Var(p_{it}^*)} \equiv \zeta_i^*,\end{aligned}$$

where we used the fact that $\mathbb{E}[\nu_{ift}] = 0$, $Var(\nu_{ift}) = \lambda_i(1-\lambda_i)Var(p_i^*)$ and $Cov(\nu_{ift}, p_{it}^*) = 0$. Therefore, the firm-level CPI forecast can be expressed as:

$$\mathbb{E}_{ift}[p_t^{CPI}] = \zeta_{if} p_{ift} \qquad = \frac{Cov(p_{it}^*, p_t^{CPI})}{Var(p_{it}^*)}\lambda_i p_{it}^* + \frac{Cov(p_{it}^*, p_t^{CPI})}{Var(p_{it}^*)}\nu_{ift},$$

where ν_{ift} is a term orthogonal to p_{it}^*. To obtain the industry-average CPI forecast, we take the expectation over all firms within industry i. Since ν_{ift} is orthogonal to p_{it}^* and has zero mean, it drops out when taking expectations:

$$\bar{\mathbb{E}}_{it}[p_t^{CPI}] = \frac{Cov(p_{it}^*, p_t^{CPI})}{Var(p_{it}^*)}\lambda_i p_{it}^*.$$

□

A.5. Proof of Lemma 2

PROOF. The effect of a shock $x \in \{m, \boldsymbol{z}\}$ on CPI is given by

$$\frac{dp_t^{CPI}}{dx_t} = \beta' \Lambda \frac{d\boldsymbol{p}_t^*}{dx_t} = \mathbb{E}_\beta\left[\lambda \frac{d\boldsymbol{p}_t^*}{dx_t}\right] = \mathbb{E}_\beta[\lambda] \times \mathbb{E}_\beta\left[\frac{d\boldsymbol{p}_t^*}{dx_t}\right] + \mathrm{Cov}_\beta\left(\lambda, \frac{d\boldsymbol{p}_t^*}{dx_t}\right).$$

□

A.6. Proof of Lemma 3

PROOF. Combining equilibrium prices (14) with the fact that the output gap is given by $\tilde{y}_t = m_t - \boldsymbol{\beta}' \boldsymbol{p}_t$ yields:

$$(I - \Lambda(I - A\Lambda)^{-1} \boldsymbol{\alpha} \boldsymbol{\beta}' \delta_w) \boldsymbol{p}_t = \Lambda(I - A\Lambda)^{-1} (\boldsymbol{\alpha} \delta_w \tilde{y}_t - \boldsymbol{z}_t)$$

Inverting the coefficient on inflation on the left-hand side yields the result (22). □

A.7. Proof of Corollary 1

PROOF. The matrix $(I - A\Lambda)^{-1}$ is increasing element-wise in attention levels $\boldsymbol{\lambda}$. Therefore, the slope of the sectoral "Phillips curves", $\mathcal{B}$, is increasing in attention levels as well. □

Appendix B. Additional empirical results

This section complements the empirical sections 4 and 5 by providing more details on the data and additional empirical results.

B.1. Data

Figure A1 plots the evolution of industry-average inflation forecast errors over the time sample.

FIGURE A1. Time-series of one-year-ahead inflation forecast errors