Information Rigidity and Elastic Attention: Evidence from Japan*

Cheng Chen Takahiro Hattori Yulei Luo†

February 2023

Abstract

Recent empirical studies have found substantial information rigidities faced by consumers and firms, when they forecast macro variables (Coibion and Gorodnichenko (2015) and Coibion et al. (2018)). In this study, we examine how information rigidities behave differently when it comes to jointly forecasting macro-, industry- and firm-level variables. Using a firm-level panel dataset that contains *quantitative* forecasts of the (macro) inflation rate, the industry-specific inflation rate, and firm sales, we show that the information rigidity associated with forecasting (macro) inflation is more pervasive than those associated with forecasting the other two variables. We then calibrate the model and find that the RI model with elastic attention matches the data reasonably well.

Keywords: information rigidity, expectations formation, macro and industry-specific shocks

JEL Classification: C53, D83, D84, E13, E37

*We thank the Economic and Social Research Institute of the Cabinet Office for providing us with the Annual Survey of Corporate Behavior. We also thank the Ministry of Finance in Japan for providing us with the Business Outlook Survey. This paper is financially supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant-in-Aid (21K13321).

†Chen: Clemson University, chencheng1983613@gmail.com. Hattori: University of Tokyo, hattori0819@gmail.com. Luo: University of Hong Kong, yulei.luo@gmail.com.

1 Introduction

How agents process information and form expectations are central to numerous economic questions ranging from households' portfolio choices to firm growth. Economists have paid ample attention to these research questions by focusing on the implications of information frictions on individual decisions and economic dynamics.[1] Recently, the empirical literature on information frictions and expectations formation began to take off, owing to the increasing availability of expectations survey data on macro variables.[2] Two common findings from the literature are: (1) there exist pervasive information rigidities (i.e., information frictions) faced by consumers and firms when they forecast macro variables; and (2) the information rigidity, which leads to a systematic mis-forecasting of future macroeconomic outcomes, affects real economic decisions such as firms' hiring and investment decisions.

Surprisingly, existing literature mainly focuses on the information rigidity concerning forecasting *macro* variables. This focus, in our opinion, is due to data constraints, as datasets used by papers in this literature usually contain firms' (quantitative) forecasts of macro variables only.[3] This focus does not mean that only information rigidities associated with forecasting macro variables matter for firms' decisions. In fact, firms also have to process information concerning industry- and firm-level variables when making decisions, as shocks to these variables and macro shocks are far from being perfectly correlated. Moreover, several papers (e.g., Boivin et al. (2009), Maćkowiak and Wiederholt (2015) and Andrade et al. (2020)) argue that firms respond to macro- and industry-level shocks differently and possibly face different degrees of information rigidities. In this paper, we present a more complete picture of information rigidities faced by firms *at three levels*, which existing research has not yet explored.[4]

[1] Earlier contributions include Muth (1960), Muth (1961), Lucas (1972), Lucas (1973), among others.

[2] Seminal works include those by Coibion and Gorodnichenko (2012), Coibion and Gorodnichenko (2015), Coibion et al. (2018), and among others.

[3] Some papers such as Coibion et al. (2018) also investigate information rigidities concerning forecasting *industry-level* variables.

[4] Although several recent papers (e.g., Barrero (2020), Chen et al. (2020), Ma et al. (2020)) investigate information rigidities associated with forecasting firm-level variables such as firm sales and profits, they do

In this study, we empirically investigate the information rigidities faced by firms concerning forecasting three target variables. We do this by merging two panel datasets obtained from the Japanese government over the period of $2004 - 2018$. The first dataset is called the Annual Survey of Corporate Behavior (ASCB). This mandatory survey targets firms publicly listed at stock exchanges in Japan, generating a panel sample of around $1,000$ firms per year. This survey asks each firm to *quantitatively* forecast nominal and real GDP growth rates for the next fiscal year (in January).[5] The firm is also asked to quantitatively forecast the nominal and real output growth rate of the industry it belongs to. The second dataset is a quarterly survey called the Business Outlook Survey (BOS), conducted at the beginning of each quarter. This survey is also mandatory and targets big firms and randomly selected medium-sized and small firms. The second survey asks firms to report realized and expected sales for each semi-year (i.e., April–September and October–March). We merge two datasets by matching firms' Japanese names and their locations, which produces a dataset that contains around 740 firms per year.

In order to guide our empirical analysis, we present a rational inattention (RI) model with elastic attention first. In RI theory, agents have limited information about the state of the world and learn slowly because they cannot process unlimited information.[6] In this paper, a typical firm needs to forecast the macro and industry-specific inflation rates as well as the change in firm-specific demand in order to adjust output. An individual firm in the model does not observe the values of the three target variables perfectly due to limited attention. Instead, it receives noisy signals of these variables every period and thus optimally filter (and forecast) the true values of these variables. Following Sims (2010), Paciello and Wiederholt

not study how firms acquire and process information on industry and macro variables.

[5]The fiscal year begins on April 1. and ends on March 31.

[6]The key innovation relative to standard noisy rational expectations models (e.g., Muth (1960), Lucas (1972), Lucas (1973)) is that the RI hypothesis permits agents to design the distribution of noise terms by focusing limited attention on certain variables at the expense of others. Under RI, agents respond to changes in the true underlying state slowly because it takes time for them to learn exactly what the new state is; they cannot learn without error because the information flow required to describe the state perfectly is larger than what their "Shannon channel"permits. Therefore, the distribution of the RI-induced noise is an outcome of optimal choice and will adapt to changing circumstances in the economy.

(2014), and Luo et al. (2017), we assume that the firm chooses the optimal degree of channel capacity to minimize the conditional variance of the forecast error (FE), given the marginal cost of acquiring and processing information. As a result, it equalizes the marginal benefit of reducing the variance of the FE and the marginal information cost. In addition, our model generates a variable-specific sensitivity (i.e., importance) parameter of the firm's payoff with respect to a reduction in the variance of the FE.[7] Depending on the information cost and the sensitivity parameters, the model yields different values of the channel capacity for the three target variables.

In the empirical part of the study, we first follow the literature (e.g., Coibion and Gorodnichenko (2012), Andrade and Le Bihan (2013) and Ryngaert (2017)) by using the estimated serial correlation of the FE to infer the degree of information rigidity. As agents know information perfectly in the full-information-rational expectation (FIRE) model, ex-post FEs are random and therefore, serially uncorrelated. However, a positive serial correlation of the FEs exists in RI models, as the agent is facing informational constraints and thus absorbs new information *gradually and with delay.* In particular, the first-order serial correlation of the FE equals the persistence of the AR(1) shock process, multiplied by the factor of one minus the Kalman gain. Therefore, the degree of information rigidity, which is inversely measured by the Kalman gain, becomes higher when the serial correlation of the FE is higher.

The regression analyses show that the (positive) serial correlation is the highest for the FE of (macro) inflation and the lowest for the FE of firm sales, while it is in the middle for the FE of the industry-specific inflation rate. The regression results also identify that (1) the process of the firm's demand is more persistent than the (macro) inflation process, which, in turn, is more persistent than the process of industry-specific inflation, and (2) innovations to firm-specific demand are more volatile than those to the industry-specific inflation process, which, in turn, are more volatile than innovations to the process of macro inflation. In total, we find that the Kalman gain is the largest when the firm forecasts its demand, while its value

[7]The sensitivity parameters depend on the two elasticities of substitution of the demand system and the persistence of the firm-specific demand process.

is smaller when the firm forecasts the macro inflation rate and the industry-specific inflation rate.[8] Therefore, the degree of information rigidity faced by firms concerning forecasting macro inflation is more pervasive than the ones associated with forecasting individual firms' demand and industry-specific inflation.

Guided by our model, we can also back out the constant marginal cost of acquiring and processing information and the importance parameters of forecasting the target variables via a calibration exercise. The moments we target are the level of channel capacity allocated to the three target variables obtained from the data. Interestingly, the calibrated importance parameter is bigger for the macro inflation than for the other two target variables, which is probably counter-intuitive. This finding can be rationalized by the fact that firms in our sample are the largest firms in Japan (i.e., public firms) and thus probably care about the macro condition much more than an average Japanese firm.

In spite of the finding on the importance parameters, we show that removing information rigidities concerning industry-specific inflation and firm-specific demand increases the firm's payoff substantially more than removing the information rigidity associated with macro inflation. The contribution of removing the information rigidity to the overall gain depends on two factors: (i) the importance of forecasting the variable (correctly), and (ii) the variance of the FE of that variable. In the data, the variance of the FE of the industry-specific inflation rate (and firm sales) is much larger than that of the macro inflation rate. This pattern explains why the gain from removing the information rigidity concerning the macro inflation rate contributes little to the overall gain from removing information rigidities.

Literature Review Our paper builds on a large body of literature that studies the expectations formation of economic agents and how these expectations affect their optimal forecasts. This literature mainly focuses on empirically testing theories of information rigidities using survey data. For instance, Mankiw et al. (2003) use the cross-sectional dis-

[8] A recent paper by Meyer et al. (2021) shows that firms pay more attention to the evolution of their unit costs rather than aggregate inflation. Our results *quantify* the difference in attention allocation among macro, industry-specific, and firm-specific variables.

tribution of forecasts to infer the degree of inattentiveness (i.e., the frequency of updating expectations). Using survey data, Coibion and Gorodnichenko (2012) study the conditional responses of forecasts to aggregate shocks and disagreements among forecasters in order to disentangle the sticky-information specification proposed by Mankiw and Reis (2002) and the noisy-information specification proposed in Sims (2003). Coibion and Gorodnichenko (2015) propose a new approach to quantify the degree of information rigidity using the US and international data of professional forecasters and other agents. Our paper complements this literature by empirically showing that the information rigidity concerning forecasting macro inflation is more pervasive than the ones associated with forecasting the firm's own demand and industry-specific inflation. Moreover, we show that although forecasting macro inflation is revealed to be important in our data, reducing the cost of processing industry-level and firm-specific information can lead to larger gains.

Our paper is closely related to a recent paper by Andrade et al. (2020) who study how firms respond to macro and industry-level shocks by adjusting their expectations and prices. Our paper complements their study by providing *quantitative* estimates of the information rigidities *and* the cost of acquiring and processing information at different levels. We also quantify the payoff losses due to the existence of information costs.

Our paper is also related to the literature on RI proposed in Sims (2003) and Sims (2010) (e.g., Luo (2008), Luo and Young (2014), Paciello and Wiederholt (2014), Luo and Young (2016), Baker et al. (2020), Afrouzi and Yang (2021), and Miao et al. (2022), etc). For example, Luo and Young (2014) find that RI models with elastic attention better replicate different consumption behaviors in emerging and developing small open economies. Paciello and Wiederholt (2014) show that optimal monetary policies under fixed attention and elastic attention differ significantly because the monetary authority can influence firms' decisions on how much attention they devote to aggregate conditions. We contribute to this literature by presenting evidence that allocated attention is heterogeneous across the macro-, industry- and firm-specific target variables.

2 A Simple Model with Elastic Attention

In order to guide our empirical analysis, we present a simple model that a firm needs to forecast changes in macro-, industry- and firm-level variables in order to adjust its output.[9] In the model, the firm does not observe the macro inflation rate, the industry-specific inflation rate, and its demand shifter perfectly. Instead, it receives noisy signals of these three variables every period and thus has to filter the true values of these variables at the end of each period. Accordingly, the firm forecasts the changes in these three variables (from the current period to the next period) based on their values filtered at the end of the current period.

Following the assumptions made in RI models, we assume that the firm in our model economy chooses the channel capacity in order to minimize the variance of the FE subject to a constant marginal cost of acquiring and processing information. As a result, the firm equalizes the marginal benefit of reducing the variance of the FE (by increasing the channel capacity) and the marginal cost of acquiring and processing information. In RI models, the relationship between the marginal cost of acquiring and processing information and the optimal channel capacity (and the Kalman gain in the filtering problem) is a one-to-one mapping. In order to back out several key parameters of the model, we perform a calibration exercise by matching the implied channel capacities from the model with the estimated channel capacities obtained from the empirical section. In addition, we evaluate the performance of our calibrated model in terms of matching non-targeted moments.

2.1 Environment

2.1.1 Demand and Supply

There are N industries in the economy. Each firm produces a differentiated variety within an industry. In the economy, the representative consumer has the following nested-CES

[9] As our macro-level and industry-level forecasts represent changes, we assume that the firm forecasts the change in its demand to make micro-level forecasts consistent with macro-level as well as industry-level forecasts.

preferences, where the first nest is among the composite goods produced by firms from different industries, indexed by i,

$$U_t = \left(\sum_{i=1}^{N} Q_{it}^{\frac{\delta-1}{\delta}} \right)^{\frac{\delta}{\delta-1}},$$

and the second nest is among the varieties $\omega \in \Omega_{it}$ produced by firms from each industry i,

$$Q_{it} = \left(\int_{\omega \in \Omega_{it}} e^{\frac{a_t(\omega)}{\sigma}} q_t(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}.$$

In the first nest, δ is the elasticity between goods produced by firms from different industries. In the second nest, σ is the elasticity between different varieties within the same industry, and $a_t(\omega)$ is the demand shifter for variety ω. We assume that firms differ in their demand shifters, $a_t(\omega)$, and need to have a higher elasticity of substitution within the industry than between industries (i.e., $\sigma > \delta$) in order to have an interior solution for the representative consumer's optimization problem. After denoting consumers' total (nominal) expenditure as Y_t, we can express the demand for a particular variety, ω, as:

$$q_t(\omega) = Y_t P_t^{\delta-1} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)} p_t(\omega)^{-\sigma}, \tag{1}$$

where P_t is the aggregate price index for all goods, and $P_{i,t}$ is the ideal price index of industry i. After substituting the real consumption $C_t \equiv Y_t/P_t$ into equation (1), we obtain

$$q_t(\omega) = C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)} p_t(\omega)^{-\sigma}.$$

The ideal price index of goods industry i can be expressed as

$$P_{i,t} \equiv \left(\int_{\in \Omega_{i,t}}^{\omega} e^{a_t(\omega)} p_t(\omega)^{1-\sigma} d\omega \right)^{1/(1-\sigma)},$$

and the aggregate price index can be written as

$$P_t \equiv \left(\sum_{i=1}^{N} P_{i,t}^{1-\delta} \right)^{1/(1-\delta)} .$$

Each variety ω is produced by a firm whose production function is simply

$$q_t(\omega) = l_t(\omega), \tag{2}$$

where $l_t(\omega)$ is the amount of labor it hires and $q_t(\omega)$ is its real output. Firms hire labor in a perfectly competitive labor market and sell output in monopolistically competitive goods markets.

We assume that the firm chooses output at the beginning of each period in order to maximize the expected profit. Specifically, the objective function is

$$\max_{q_t(\omega)} \quad q_t(\omega) \left[\left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)} \right)^{\frac{1}{\sigma}} q_t(\omega)^{-\frac{1}{\sigma}} - w_t \right], \tag{3}$$

where w_t is the prevailing wage rate. If the firm knew all the information concerning various variables in equation (3) *perfectly*, solving the above optimization problem would be straightforward. However, the assumption we make is that firms do not know the aggregate price level, P_t, the industry-specific price level, $P_{i,t}$, and the demand shifter, $a_t(\omega)$ perfectly. We now discuss how various variables evolve over time and the information environment concerning those variables.

2.1.2 Dynamic Processes

We have four dynamic processes that show up in the profit function in equation (3). First, we have estimated the persistence of the real GDP growth rate of Japan at the annual level, which turns out to be extremely low (0.075) and statistically insignificant. Therefore, we assume that the growth rate of real consumption (i.e., GDP) follows a random walk:

$$g_{t+1}^{c} = \epsilon_{c,t+1},$$

where $g_{t+1}^{c} \equiv \log(C_{t+1}) - \log(C_t)$. $\epsilon_{c,t+1}$ is the innovation to the (logarithm of) real consumption and distributed normally with the mean $\bar{g}_c$ and variance $\sigma_{g_c}^2$ where $\bar{g}_c > 0$.[10]

Second, we assume the two inflation rates (macro and industry-specific), $\pi_{t+1} \equiv \log(P_{t+1}) - \log(P_t)$ and $\pi_{t+1}^{i} \equiv \log(P_{i,t+1}) - \log(P_{i,t})$, all follow the AR(1) process:

$$x_{t+1} = \rho_x x_t + \epsilon_{x,t+1},$$

where $x \in \{\pi, \pi^i\}$ and $\epsilon_{x,t+1}$ is an independently and identically distributed (iid) innovation and distributed normally with mean 0 and variance σ_x^2, which implies that the long-run variance of x is $\sigma_x^2 / (1 - \rho_x^2)$. In Section 4.1, we will show that the persistence of both processes is below one.

Third, we assume that the firm-specific demand shifter $a_t(\omega)$, follows an AR(1) process as well:

$$a_{t+1}(\omega) = \rho_a a_t(\omega) + \epsilon_{a(\omega),t+1}, \tag{4}$$

where $\epsilon_{a(\omega),t+1}$ is an iid innovation and distributed normally with mean 0 and variance σ_a^2.

2.1.3 Information Environment

The key assumption in this paper is that individual firms cannot observe the target variables perfectly due to limited information-processing capacity. Specifically, with finite capacity $\kappa \in (0, \infty)$, a random variable $\{x_t\}$ following a continuous distribution cannot be observed without an error and thus, the information set at time $t+1$, denoted $\mathcal{I}_{t+1}$, is generated by the entire history of noisy signals $\left\{x_j^*\right\}_{j=0}^{t+1}$. Following the literature, we assume the noisy

[10]In the model, there are no investments, government expenditure and net exports. Therefore, consumption equals GDP.

signal takes the following additive form:

$$x^*_{t+1} = x_{t+1} + \eta_{t+1},$$

where η_{t+1} is the endogenous noise caused by the finite capacity. We further assume that η_{t+1} is an iid idiosyncratic shock and is independent of the fundamental shocks affecting the economy. The reason why the RI-induced noise is idiosyncratic is that the endogenous noise arises from the firm's own internal information-processing constraint. Firms with finite capacity choose a new signal $x^*_{t+1} \in \mathcal{I}_{t+1} = \left\{x^*_1, x^*_2, \cdots, x^*_{t+1}\right\}$ that reduces the uncertainty about the variable x_{t+1} as much as possible. Formally, this idea can be described by the information constraint

$$\mathbb{H}\left(x_{t+1}|\mathcal{I}_t\right) - \mathbb{H}\left(x_{t+1}|\mathcal{I}_{t+1}\right) \leq \kappa, \tag{5}$$

where κ is the firm's information channel capacity, $\mathbb{H}\left(x_{t+1}|\ \mathcal{I}_t\right)$ denotes the entropy of the state *prior to* observing the new signal at $t+1$, and $\mathbb{H}\left(x_{t+1}|\ \mathcal{I}_{t+1}\right)$ is the entropy *after* observing the new signal. κ imposes an upper bound on the amount of information flow—that is, the change in the entropy—that can be transmitted in any given period. In this paper, we assume that the prior distribution of x_{t+1} is Gaussian.

In the linear–quadratic–Gaussian (LQG) framework, as has been shown in Sims (2003) and Sims (2010), the true state under RI also follows a normal distribution $s_t|\mathcal{I}_t \sim N\left(\mathbb{E}\left[s_t|\mathcal{I}_t\right], \Sigma_t\right)$, where $\Sigma_t = \mathbb{E}_{\omega,t}\left[\left(x_t - \widehat{x}_t\right)^2\right]$ and $\widehat{x}_t = \mathbb{E}_{\omega,t}\left[x_t\right]$. In addition, given that the noisy signal takes the additive form $x^*_{t+1} = x_{t+1} + \eta_{t+1}$, the noise $\eta_{t+1} \sim N\left(0, \Lambda\right)$ will also be Gaussian. In this case, equation (5) is reduced to

$$\log\left(|\Psi_t|\right) - \log\left(|\Sigma_{t+1}|\right) \leq 2\kappa,$$

where $\Psi_t = \mathbb{E}_{\omega,t}\left[\left(x_{t+1} - \mathbb{E}_{\omega,t}\left[x_{t+1}\right]\right)^2\right]$and Σ_{t+1} are the conditional variances prior to and after observing the new signal, respectively. As more information about the state becomes available in single-agent models, this constraint will be binding.[11] The conditional variance is updated according to the following standard formula in the steady state:

$$\Lambda^{-1} = \Sigma^{-1} - \Psi^{-1}.$$

The evolution of the estimated state, $\widehat{x}_t = \mathbb{E}_{\omega,t}[x_t]$ is governed by the Kalman filtering equation:

$$\widehat{x}_{t+1} = (1 - G_x)\,\rho_x \widehat{x}_t + G_x x^*_{t+1},$$

where $G_x = \Sigma\Lambda^{-1}$ is the Kalman gain,

$$x_t - \mathbb{E}_{\omega,t}[x_t] = x_t - \widehat{x}_t = \frac{(1 - G_x)\,\epsilon_{x,t}}{1 - (1 - G_x)\rho_x \cdot \mathbb{L}} - \frac{G_x \eta_t}{1 - (1 - G_x)\rho_x \cdot \mathbb{L}} \tag{6}$$

is the filtering (estimation) error with $\mathbb{E}_{\omega,t}[x_t - \widehat{x}_t] = 0$, and $\mathbb{L}$ is the standard lag operator. For the forecast error, we define it follows:

$$\begin{aligned} \mathbb{FE}_{\omega,t}(t+1) &= x_{t+1} - \mathbb{E}_{\omega,t}[x_{t+1}] \\ &= \rho_x (x_t - \widehat{x}_t) + \epsilon_{x,t+1}. \end{aligned} \tag{7}$$

Now, we discuss the information environment concerning the growth rate of real GDP and the wage rate. First, as the growth rate of real GDP, g^c_t is an iid random variable, the firm's optimal forecast is simply $\bar{g}_c$ which is the prior mean of g^c_t. Therefore, whether the firm knows the past noisy signals of g^c_t is irrelevant. Next, we assume that the firm knows the wage rate, w_t, perfectly, as it is the firm itself that sets the wage rate and pays wages to workers.

Finally, we discuss the time assumption. First, the firm decides output at the beginning of each period based on its forecasts made (at the end of last period). At the same time, it

[11] By "better" we mean that conditional on the draws by nature for the true state, the expected utility of the agent increases if information about that state is improved.

chooses the channel capacity for each of the three target variables. Then, the goods are sold in the market, which leads to the realized price. The realized price differs from the expected price in general. Finally, the firm receives signals of the three target variables in the current period. As a result, the firm filters the current three target variables and forms its forecasts for those three target variables in the next period.

2.1.4 Discussions of Modeling Choices

In the model, we assume that firms differ in the demand shifters that they need to learn, while firms can also differ in their cost shifters in reality. We make such a modeling choice for two reasons. First, recent literature on firm heterogeneity reveals that it is mainly the demand-side factors that lead to firm heterogeneity (see Hottman et al. (2016)). Second, we think that the demand shifter is *more* likely to be exogenous to the firm compared to the cost shifter and the wage rate, as demand is determined by consumers' tastes which are out of the control of the firm in many circumstances. On the contrary, the firm can invest in its production technology (to reduce costs) and decides the wage rate it offers to the employees. Thus, the firm probably knows more about its costs and the wage rate than its demand shifter. Reasonably, the macro and industry-specific inflation processes are probably exogenous to the firm as well. To some extent, forecasting the industry-specific inflation rate is similar to forecasting the firm's competitors' pricing strategy (in the same industry). We believe that comparing the FE of the macro inflation rate and that of the industry-specific inflation rate is fair, as both targets are out of the control of the focal firm. In total, we think the three target variables that the firm knows imperfectly are more or less out of the control of the firm.

We assume the firm chooses output (instead of setting the price) in the model, since information frictions matter in such a case. If the firm were to set the price, the markup rule implies that the optimal price is the product of a constant (i.e., the markup) and the firm's marginal cost (i.e., the wage rate). As the marginal cost is constant (i.e., one) and the firm knows the wage rate, the firm would not need to know macro-, industry- and firm-level information in order to set the price. Therefore, we assume that the firm chooses output.[12]

2.2 Optimal Forecasting and Elastic Attention

In this subsection, we describe the firm's optimization problem, which can be divided into two steps. First, we solve the filtering/forecasting problem of the three target variables *given* the channel capacity allocated. Second, we solve for the optimal allocation of attention to each of the three target variables.

2.2.1 Filtering/Forecasting

For macro inflation and industry-specific inflation, the filtering problem is standard. Specifically, we have the following updating rule when firm ω minimizes the variance of the filtering (or forecasting) error:

$$\widehat{x}_{t+1} = (1 - G_x)\,\rho_x \widehat{x}_t + G_x \left(x_{t+1} + \eta^{\omega}_{x,t+1} \right),$$

where $\widehat{x}_t = \mathbb{E}_{\omega,t}[x_t]$ (x can be either π^i or π) is the filtered state. The associated Kalman gain is

$$G_x = 1 - e^{-2\kappa_x}, \tag{8}$$

where κ_x ($x \in \{\pi, \pi^i\}$) is the channel capacity.[13] The forecast is simply

$$\mathbb{E}_{\omega,t}[x_{t+1}] = \rho_x \mathbb{E}_{\omega,t}[x_t] = \rho_x \widehat{x}_t.$$

[12] This assumption is somewhat different from existing work that assumes firms choose prices and thus inflation expectations affect their pricing choices. However, this difference in our setup is purely due to data constraints. In an alternative model, we can assume that the firm does not observe its marginal cost and the wage rate. Thus, it needs to filter their processes. In such an environment, information rigidities concerning macro inflation, real GDP growth and the marginal cost would matter for pricing. However, we do not have information on forecasted and realized marginal (or total) cost in the data. Therefore, we cannot quantify the degree of information rigidity concerning the firm's cost. Since the focus of this paper is the attention allocation at various levels (macro v.s. firms), we chose to our current setup in which we can back out the degree of information rigidity concerning firm demand from forecasted and realized firm sales.

[13] The unit of channel capacity in this paper is "nats".

As a result, the conditional variance of the FE is

$$\Psi_x \equiv \mathbb{V}_t\left(x_{t+1} - \mathbb{E}_{\omega,t}\left[x_{t+1}\right]\right) = \rho_x^2 \Sigma_x + \sigma_x^2 = \frac{e^{2\kappa_x}\sigma_x^2}{e^{2\kappa_x} - \rho_x^2}, \tag{9}$$

where x can be either π or π^i, $\Sigma_x \equiv \mathbb{V}_t\left(x_t - \mathbb{E}_{\omega,t}\left[x_t\right]\right) = \sigma_x^2 / \left(e^{2\kappa_x} - \rho_x^2\right)$ is the variance of the filtering error, and σ_x^2 is the variance of the fundamental shock. As the conditional variance of the FE (Ψ_x) is a linear function of the conditional variance of the filtering error (Σ_x), minimizing Ψ_x is equivalent to minimizing Σ_x.

As the macro and industry-specific inflation rates are changes in the price levels, we assume that firm ω forecasts the change in its demand shifter, $a_{t+1}(\omega) - a_t(\omega)$, in order to make the three target variables comparable. A rationale for this is that the firm needs to know by how much it should adjust the output (and employment) between two adjacent periods. Formally, we have the following forecasting problem for firm ω:

$$\min_{G_a} \mathbb{V}_t\left[\left(a_{t+1}(\omega) - a_t(\omega)\right) - \left(\mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right] - \mathbb{E}_{\omega,t}\left[a_t(\omega)\right]\right)\right],$$

where G_a is the associated Kalman gain. Given the AR(1) structure of the demand process, we can rewrite the above problem as

$$\min_{G_a} \mathbb{V}_t\left[\left(\rho_a - 1\right)\left(a_t(\omega) - \mathbb{E}_{\omega,t}\left[a_t(\omega)\right]\right) + \epsilon_{a(\omega),t+1}\right] \text{ or } \min_{G_a} (1-\rho_a)^2 \mathbb{V}_t\left[a_t(\omega) - \mathbb{E}_{\omega,t}\left[a_t(\omega)\right]\right] + \sigma_a^2, \tag{10}$$

where we have used the result that $\mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right] = \rho_a \mathbb{E}_{\omega,t}\left[a_t(\omega)\right]$. The problem in equation (10) is the same as minimizing the variance of the conditional filtering error of $a_t(\omega)$. Therefore, the usual RI techniques apply and we have

$$\Psi_a \equiv \mathbb{V}_t\left[\left(a_{t+1}(\omega) - a_t(\omega)\right) - \left(\mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right] - \mathbb{E}_{\omega,t}\left[a_t(\omega)\right]\right)\right] = \frac{(1-\rho_a)^2 \sigma_a^2}{e^{2\kappa_a} - \rho_a^2} + \sigma_a^2. \tag{11}$$

2.2.2 Attention Allocation

Now, we discuss how the firm allocates its attention optimally in the first stage. To determine the optimal level of attention/capacity devoted to monitoring the three target variables, we make the following assumptions for our model:

Assumption 1 *Individual firms face a constant and homogeneous marginal cost of acquiring and processing information concerning each variable when choosing the optimal level of attention (or channel capacity).*

The above assumption of a constant and *homogeneous* marginal cost of acquiring and processing information is a standard assumption in rational inattention models with attention allocation (see Mackowiak and Wiederholt (2009) and Maćkowiak and Wiederholt (2015)). With a constant and homogeneous information-processing cost, the agent is allowed to adjust the optimal level of attention in such a way that the marginal cost of information-processing for the problem at hand remains constant. The optimal forecasting problem for the typical firm can thus be written as:

$$\min_{\{\kappa_\pi,\kappa_{\pi^i},\kappa_a\}} \left\{(w_\pi\Psi_\pi + \lambda\kappa_\pi) + (w_{\pi^i}\Psi_{\pi^i} + \lambda\kappa_{\pi^i}) + (w_a\Psi_a + \lambda\kappa_a)\right\}.$$

Note that the constant marginal cost of acquiring and processing information, λ, is the same across various target variables. Variables Ψ_π, Ψ_{π^i}, and Ψ_a are defined in equations (9) and (11), respectively. Additionally, w_π, w_{π^i}, and w_a are the three sensitivity parameters. This minimization problem demonstrates the optimizing firm's trade-off between reducing the uncertainty of the perceived state and the cost attached to the reduction in the perceived uncertainty. The following proposition summarizes the solution:

Proposition 1 *At optimum, the individual firm equalizes the marginal benefit of reducing the variance of the ex-post forecast errors and the constant marginal cost of information* *acquisition and processing:*

$$\lambda = w_\pi \left| \frac{\partial \Psi_\pi}{\partial \kappa_\pi} \right| = w_{\pi^i} \left| \frac{\partial \Psi_{\pi^i}}{\partial \kappa_{\pi^i}} \right| = w_a \left| \frac{\partial \Psi_a}{\partial \kappa_a} \right|. \tag{12}$$

Proof. The proof is straightforward. ■

It is worth noting that this result is consistent with the concept of "elastic" capacity proposed in Kahneman (1973). In addition, in a dynamic setting, the marginal cost of information-processing might also be constant over time. In contrast, the optimal degree of attention/capacity can be time-varying. For example, Coibion and Gorodnichenko (2015) used the SPF forecast survey data to test the degree of information rigidity due to both noisy information and sticky information and found that the degree of information rigidity decreased with the volatility of macroeconomic conditions. Specifically, they found that the incidence of information rigidity decreased from the late 1960s to the start of the Great Moderation ($1983-1984$) and had continued to decline since then. They argued that one should be wary of treating the degree of information rigidity as a structural parameter because it responds to changes in macroeconomic conditions.

In our model, the sensitivity parameters are different across the target variables because the variances of FEs of the three target variables play different roles in affecting the firm's payoff. In what follows, we derive the expression for the sensitivity parameters through the lens of our model. First, the optimal output level under full information is

$$\log q_t^{full}(\omega) = \sigma \log \left(\frac{\sigma - 1}{\sigma} \right) + \log C_t + \delta \log P_t + (\sigma - \delta) \log P_{i,t} + a_t(\omega) - \sigma \log (w_t).$$

The (actual) output choice under information rigidities is

$$\log q_t(\omega) = \sigma \log \left(\frac{\sigma - 1}{\sigma} \right) + \sigma \log \left(\mathbb{E}_{\omega,t-1} \left(C_t P_t^\delta P_{i,t}^{\sigma-\delta} e^{a_t(\omega)} \right)^{1/\sigma} \right) - \sigma \log (w_t). \tag{13}$$

Following the standard approach used in RI models, we use Taylor expansion (up to the second order) to approximate the profit function under information rigidities. Specifically, we calculate the deviation of the profit function in our model from that in the full-information case.[14] As a result, the expected loss in firm profits (due to information frictions) is proportional to

$$\mathbb{E}_{\omega,t-1}\left(\log q_t(\omega) - \log q_t^{full}(\omega)\right)^2 = \sigma^2 \mathbb{E}_{\omega,t-1}\left[\log\left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma} - \log\left(\mathbb{E}_{\omega,t-1}\left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma}\right)\right]^2$$

which is related to the variance of filtering/forecasting error of C_t, P_t, $P_{i,t}$ and $a_t(\omega)$. Based on equation (13), the *change* in the optimal output under full information can be written as

$$g_{q,t}^{full}(\omega) = \log\left(\frac{q_t^{full}(\omega)}{q_{t-1}^{full}(\omega)}\right) = g_t^c + \delta\pi_t + (\sigma-\delta)\pi_t^i + \widetilde{a}_t(\omega) - \sigma\left(\log(w_t) - \log(w_{t-1})\right),$$

where $\widetilde{a}_t = a_t(\omega) - a_{t-1}(\omega)$. In reality, firms face not only information rigidities but also adjustment costs of labor. Therefore, they have incentives to minimize the expected loss due to changes in their labor (i.e., output) choices. Specifically, the firms want to minimize the following expected loss due to the change in the output:

$$\begin{aligned}
L_{t+1}(\omega) &= \mathbb{E}_{\omega,t}\left[g_{q,t+1}(\omega) - g_{q,t+1}^{full}(\omega)\right]^2 \\
&= \mathbb{V}_t\left(g_{t+1}^c - \mathbb{E}_{\omega,t}\left[g_{t+1}^c\right]\right) + \delta^2 \mathbb{V}_t\left(\pi_{t+1} - \mathbb{E}_{\omega,t}\left[\pi_{t+1}\right]\right) + (\sigma-\delta)^2 \mathbb{V}_t\left(\pi_{t+1}^i - \mathbb{E}_{\omega,t}\left[\pi_{t+1}^i\right]\right) \\
&\quad + \sigma^2 \mathbb{V}_t\left[\left(\frac{1}{\rho_a} - 1\right)\left(\frac{a_{t+1}(\omega)}{\sigma} - \frac{\mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right]}{\sigma}\right)\right] + \frac{1}{\rho_a}\left(2 - \frac{1}{\rho_a}\right)\sigma_a^2 \\
&= \Psi_g + \delta^2\Psi_\pi + (\sigma-\delta)^2\Psi_{\pi^i} + \sigma^2\left(\frac{1}{\rho_a} - 1\right)^2 \Psi_{a/\sigma} + \frac{1}{\rho_a}\left(2 - \frac{1}{\rho_a}\right)\sigma_a^2, \qquad (14)
\end{aligned}$$

where $g_q(\omega)$ is the optimal change in output under information rigidities and we have used the results that $a_{t+1}(\omega) = \rho_a a_t(\omega) + \epsilon_{a(\omega),t+1}$ and $\mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right] = \rho_a \mathbb{E}_{\omega,t}\left[a_t(\omega)\right]$.[15] Note that the variance of FE of the real GDP growth rate is simply $\sigma_{g^c}^2$, which cannot be reduced by

[14] See Online Appendix 6.3 for the detail.

[15] Note that $\left[\mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right] - \mathbb{E}_{\omega,t}\left[a_t(\omega)\right]\right] - \left[a_{t+1}(\omega) - a_t(\omega)\right] = (1 - 1/\rho_a)\left(\mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right] - a_{t+1}(\omega)\right) - \epsilon_{a(\omega),t+1}/\rho_a$. Thus, we have $\mathbb{V}_t\left[\mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right] - \mathbb{E}_{\omega,t}\left[a_t(\omega)\right]\right] = \sigma^2 \mathbb{V}_t\left[(1 - 1/\rho_a)\left(a_{t+1}(\omega)/\sigma - \mathbb{E}_{\omega,t} a_{t+1}(\omega)/\sigma\right)\right] + \left[1/\rho_a^2 + 2\left(1 - 1/\rho_a\right)/\rho_a\right]\sigma_a^2$ which leads to the expression in equation (14).

allocating more attention to it (as its persistence is zero). For the remaining three terms in equation (14), we can infer the sensitivity parameters as follows:

$$w_\pi = \delta^2,\ w_{\pi^i} = (\sigma - \delta)^2, \text{ and } w_a = \sigma^2 \left(\frac{1}{\rho_a} - 1 \right)^2, \tag{15}$$

In summary, the sensitivity parameters are related to the elasticity of substitution, both between and within industries, and the persistence of the process of the demand shifter. The derivation of the three parameters is new to the literature, as we consider an economy where the firm facing a two-layer CES demand function needs to filter three target variables (at different levels). In Section 4, we will obtain the values of these three parameters by calibration (for δ) and estimating the demand process of the firm.

2.3 Testable Implications

In this subsection, we derive the model's testable implications. In particular, we focus on the serial correlation regression of the FE at the individual firm level; that is, we regress FE in period t on its one-period lag and a set of fixed effects. As our dataset is at the annual frequency and spans 15 years, we have to exploit cross-sectional variations of the forecasts and FEs. As shown in Ryngaert (2017), the serial correlation regression can be run at the individual-firm (or forecaster) level.

We discuss the regression of serial correlation first. Calculation shows that

$$\begin{aligned} \mathbb{FE}^{\pi}_{\omega,t}(t+1) &= \pi_{t+1} - \mathbb{E}_{\omega,t}\left[\pi_{t+1}\right] \\ &= \epsilon_{\pi,t+1} + \frac{\rho_\pi (1-G_\pi)\epsilon_{\pi,t}}{1-\rho_\pi(1-G_\pi)\cdot \mathbb{L}} - \frac{\rho_\pi G_\pi \eta^{\omega}_{\pi,t}}{1-\rho_\pi(1-G_\pi)\cdot \mathbb{L}}. \end{aligned}$$

This yields the following regression equation:

$$\mathbb{FE}^{\pi}_{\omega,t}(t+1) = \rho_\pi(1-G_\pi)\mathbb{FE}^{\pi}_{\omega,t-1}(t) + error^{\omega}_{\pi,t+1}, \tag{16}$$

where $error^{\omega}_{\pi,t+1} = \epsilon_{\pi,t+1} - \rho_{\pi} G_{\pi} \eta^{\omega}_{\pi,t}$ is the error term of the regression which is uncorrelated to $\mathbb{FE}^{\pi}_{\omega,t-1}(t)$. Similarly, we have

$$\mathbb{FE}^{\pi^i}_{\omega,t}(t+1) = \pi^i_{t+1} - \mathbb{E}_{\omega,t}\left[\pi^i_{t+1}\right] = \epsilon_{\pi^i,t+1} + \frac{\rho_{\pi^i}(1-G_{\pi^i})\epsilon_{\pi^i,t}}{1-\rho_{\pi^i}(1-G_{\pi^i})\cdot\mathbb{L}} - \frac{\rho_{\pi^i} G_{\pi^i}\eta^{\omega}_{\pi^i,t}}{1-\rho_{\pi^i}(1-G_{\pi^i})\cdot\mathbb{L}},$$

and

$$\mathbb{FE}^{\pi^i}_{\omega,t}(t+1) = \rho_{\pi^i}(1-G_{\pi^i})\mathbb{FE}^{\pi^i}_{\omega,t-1}(t) + error^{\omega}_{\pi^i,t+1}, \tag{17}$$

where $error^{\omega}_{\pi^i,t+1} = \epsilon_{\pi^i,t+1} - \rho_{\pi^i} G_{\pi^i} \eta^{\omega}_{\pi^i,t}$.

Although our data only contain the forecast of total sales, we can infer the FE of the demand shifter from the FE of sales. As the firm chooses $q_{\omega,t+1}$ at the beginning of period $t+1$, it knows its output (i.e., quantity) when forecasting its sales at the beginning of period $t+1$. Thus, the (logarithm of) FE of sales equals

$$\mathbb{FE}^{log,sales}_{\omega,t}(t+1) \equiv \log\left(R_{\omega,t+1}\right) - \log\left(\mathbb{E}_{\omega,t}\left[R_{\omega,t+1}\right]\right) = \log\left(p_{\omega,t+1}\right) - \log\left(\mathbb{E}_{\omega,t}\left[p_{\omega,t+1}\right]\right),$$

where $p_{\omega,t+1}$ is the price charged in period $t+1$. Therefore, we can rewrite the FE of sales as

$$\begin{aligned} \mathbb{FE}^{log,sales}_{\omega,t}(t+1) &= \frac{\delta}{\sigma}\left[\log(P_{t+1}) - \mathbb{E}_{\omega,t}\left[\log(P_{t+1})\right]\right] + \left(\frac{\sigma-\delta}{\sigma}\right)\left[\log(P_{i,t+1}) - \mathbb{E}_{\omega,t}\left[\log(P_{i,t+1})\right]\right] \\ &+ \frac{1}{\sigma}\left[\log(C_{t+1}) - \mathbb{E}_{\omega,t}\left[\log(C_{t+1})\right]\right] + \frac{1}{\sigma}\left[a_{t+1}(\omega) - \mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right]\right] + const, \end{aligned} \tag{18}$$

where the term *const* includes (subjective) variances of $\log(P_{t+1})$, $\log(P_{i,t+1})$ and $a_{t+1}(\omega)$ which are not stochastic. In order to tease out the component of $\left(a_{t+1}(\omega) - \mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right]\right)/\sigma$ from the FE of sales, we regress the forecast error of sales on firm-size-bin-industry-year fixed effects and obtain the residual term as the counterpart of $\left(a_{t+1}(\omega) - \mathbb{E}_{\omega,t} a_{t+1}(\omega)\right)/\sigma$. The logic here is that firms with similar sizes and from the same industry have similar forecast errors of macro-level and industry-level variables such as the inflation rates and the real GDP growth rate. Therefore, we have

$$
\begin{aligned}
& \mathbb{FE}_{\omega,t}^{res,sales}(t+1) \\
= \; & \frac{a_{t+1}(\omega) - \mathbb{E}_{\omega,t}\left[a_{t+1}(\omega)\right]}{\sigma} \\
= \; & \frac{1}{\sigma}\left[\epsilon_{a(\omega),t+1} + \frac{\rho_a(1-G_a)\epsilon_{a(\omega),t}}{1-\rho_a(1-G_a)\cdot\mathbb{L}} - \frac{\rho_a G_a \eta_{a(\omega),t}}{1-\rho_a(1-G_a)\cdot\mathbb{L}}\right],
\end{aligned}
$$

and

$$
\mathbb{FE}_{\omega,t}^{res,sales}(t+1) = \rho_a(1-G_a)\mathbb{FE}_{\omega,t-1}^{res,sales}(t) + \text{error}_{\,a(\omega),t+1}, \tag{19}
$$

where $\text{error}_{\,a(\omega),t+1} = \left(\epsilon_{a(\omega),t+1} - \rho_a G_a \eta_{a(\omega),t}\right)/\sigma$ which is uncorrelated with $\mathbb{FE}_{\omega,t-1}^{res,sales}(t)$. In summary, the coefficient obtained from the serial correlation regression is the product of the persistence parameter and one minus the Kalman gain. When the degree of information rigidity increases, the Kalman gain shrinks, which results in a larger coefficient obtained from the regression.

Following Coibion and Gorodnichenko (2012, 2015) and Andrade and Le Bihan (2013), we also use the standard deviation of the FEs to measure the degree of forecaster disagreement. Using equations (6) and (7), it is straightforward to show that

$$
\mathbb{FD}_t^x(t+1) = \text{std}\left(\mathbb{FE}_t^x(t+1)\right) = \frac{(1-G_x)\,\rho_x^2}{1-(1-G_x)\,\rho_x^2}\sigma_x^2 + \sigma_x^2,
$$

where x ($x = \pi$, π^i, or $a(\omega)$), and we use the result that $\mathbb{V}_t\left(x_t - \widehat{x}_t\right) = (1-G_x)\,\sigma_x^2/\left[1-(1-G_x)\,\rho_x^2\right]$.

3 Data

In this section, we introduce the datasets we use in our empirical analysis. The first dataset employed is the Annual Survey of Corporate Behavior (ASCB), conducted by the Economic and Social Research Institute in the Cabinet Office of Japan.[16] Each year, the survey ques-

[16]This is the same dataset as the one used in Tanaka et al. (2019).

tionnaire was sent to all listed firms on the Tokyo and Nagoya Stock Exchanges. A total of 2000 firms, on average, were surveyed during these years. Of them, 50% on average responded to the survey each year. The survey is conducted annually in January. Respondents are required to answer the questions regarding their quantitative forecasts for the (real and nominal) GDP growth rate, the growth rate of (real and nominal) industrial output, and the expected average (percentage) change in their input and output prices for the next fiscal year.[17]

The second dataset we use is called the BOS, implemented by Japanese MOF every quarter. The survey covers all big firms (i.e., firms with registered capital of more than 2 billion JPY or, equivalently, 20 million USD) and a representative sample of medium-sized and small firms.[18] We have obtained the second dataset from 2004/Q2 to 2018/Q4. The average response rate of this survey is 75%, which results in a panel sample of roughly $11,500$ firms per quarter. The second survey asks firms to report forecasted sales and operating profits for each semi-year ahead (i.e., April–September and October–March). It also asks the firm to report realized sales and operating profits for the past two half-year periods.

We merge the BOS conducted at the beginning of every second quarter (April) with ASCB, as the timing of the two surveys is close, and there are a large number of firms that report their forecasted sales and operating profits in BOS conducted in April. We use this merged dataset as the main data to conduct our analysis. On average, we are able to match 73% of observations (around 740 firms per year) in the ASCB datasets with the observations in the BOS dataset conducted in April and the matching rate is relatively stable over the years as shown in Table 1.[19]

[17] A fiscal year in Japan (nendo in Japanese) spans from April/1 of the current year to March/31 of the next year.

[18] For firms with registered capital between 0.5 billion JPY and 2 billion JPY, 50% of them are randomly sampled every quarter. For firms with registered capital between 0.1 billion JPY and 0.5 billion JPY, 10% of them are randomly sampled every quarter. For firms with registered capital less than 0.1 billion JPY, 1% of them are randomly sampled every quarter. The random sample is redrawn at the beginning of every fiscal year; that is, as long as a medium-size or small firm is selected for the survey in a given fiscal year, it appears in the survey for all four quarters of that fiscal year.

[19] As BOS data end in 2018/Q4 (i.e., realized variables are unavailable for the fiscal year of 2018), we end up with a panel dataset that contains forecast errors made between April 2004 and April 2017 over 14 years

Table 1: Percentage of successful matching of the main dataset

year	obs. in ASCB	matched obs.	percentage
2004	1,243	794	63.9%
2005	1,031	679	65.9%
2006	1,123	780	69.5%
2007	1,042	756	72.6%
2008	1,035	711	68.7%
2009	1,027	721	70.2%
2010	1,032	756	73.3%
2011	863	629	72.9%
2012	890	674	75.7%
2013	815	631	77.4%
2014	867	672	77.5%
2015	982	766	78.0%
2016	1,062	826	77.8%
2017	1,168	901	77.1%
2018	1,107	858	77.5%
Total	15,287	11,154	73.0%

Notes: The number of observations in ASCB dataset is reported in the second column, and the number of observations in the matched dataset is reported in the third column. Note that in each year t, forecasters in the ASCB dataset are reported in January (i.e., the first quarter), while forecasts in the BOS dataset are reported in April or early May (i.e., the second quarter). Both forecasts are made for the fiscal year of t, and the fiscal year begins in April.

Ideally, we would want to merge BOS conducted at the beginning of every first quarter with ASCB, as both are conducted in January. However, there are fewer firms that report their forecasted sales and operating profits in BOS conducted in January than the one conducted in April.[20] We merge BOS conducted at the beginning of every first quarter with ASCB to create an alternative dataset for our analysis. However, due to many missing values of forecasts in BOS conducted in January, we only use this alternative dataset for robustness checks. We will show our findings are robust to using this alternative dataset.[21]

Although the two datasets contain quite detailed information on quantitative forecasts, both of them have limited firm-level production and financial information. For instance, both datasets do not have information on firm-level employment, capital stock and various types of investment and input purchases. In addition, neither datasets contains information

(in terms of the timing of forecasting).

[20]Roughly 40% firms that answered the survey reported their forecasted sales and operating profits in January, while roughly 75% firms that answered the survey reported their forecasted sales and operating profits in April.

[21]As BOS data starts from 2004/Q2 and ends in 2018/Q4, the alternative dataset contains forecast errors made between January 2005 and January 2017 over 13 years (in terms of the timing of forecasting).

on firm assets and liabilities. Moreover, firms in our merged dataset are by far the biggest firms in Japan. Due to these data constraints, it is hard to explore heterogeneity across firms in terms of information rigidities. Also, we cannot include many firm-level control variables into the regressions that are going to be run in the next section.

3.1 Forecasts and FEs

We construct the FE of macro and industry-specific inflation rates as follows. First, we obtain the time-series data of the (nominal and real) GDP growth rate and that of the (nominal and real) growth rate of industrial output from ESRI's website. Second, we calculate the macro inflation rate by taking the difference between the nominal GDP growth rate and the real GDP growth rate. We implement the same exercise for the industry-specific inflation rate. Then, we define the FE of the macro-level inflation rate and that of industry-specific inflation rate as

$$\mathbb{FE}^{\pi}_{\omega,t-1}(t) \equiv \pi_t - \mathbb{E}_{\omega,t-1}\left[\pi_t\right], \tag{20}$$

and

$$\mathbb{FE}^{\pi^i}_{\omega,t-1}(t) \equiv \pi^i_t - \mathbb{E}_{\omega,t-1}\left[\pi^i_t\right], \tag{21}$$

where ω indicates the firm, t denotes the year, and i refers to the industry the firms belong to. The macro and industry-specific inflation rates are denoted by π and π^i, respectively.

Next, as firms report both realized and forecasted sales, we define the percentage FE and the logarithm of FE as follows:

$$\mathbb{FE}^{pct,sales}_{\omega,t-1}(t) \equiv \frac{R_{\omega,t} - \mathbb{E}_{\omega,t-1}\left[R_{\omega,t}\right]}{\mathbb{E}_{\omega,t-1}\left[R_{\omega,t}\right]}, \tag{22}$$

$$\mathbb{FE}^{log,sales}_{\omega,t-1}(t) \equiv \log\left(R_{\omega,t}\right) - \log\left(\mathbb{E}_{\omega,t-1}\left[R_{\omega,t}\right]\right), \tag{23}$$

where $R_{\omega,t}$ and $\mathbb{E}_{\omega,t-1}\left[R_{\omega,t}\right]$ are realized and forecasted sales of firms ω respectively.

In Table 2, we present the summary statistics of the forecasted and realized inflation rate both at the macro level and at the industry level. Several observations are worth mentioning. First, Japan had experienced deflation during our sample period, as the average inflation rate is negative from 2004 − 2018. This can be seen from the average realized (macro) inflation rate in Table 2. Second, the average realized (industry-specific) inflation rate is slightly positive, which is higher than the average realized (macro) inflation rate. This is possible as the (industry-specific) inflation rate is the price change of industrial output (i.e., not value added).[22] Third, the variation of (industry-specific) inflation rates is larger than that of macro inflation rates, which substantiates the fact that industry-level shocks (to inflation) are more volatile than macro shocks (to inflation). In Table 3, we report the summary statistics of the FEs defined as above. In this table, we observe that Japanese firms had over-predicted the macro inflation rate and under-predicted the industry-specific inflation rate over the period 2004 − 2017. Note that the standard deviation of the FE of the industry-specific inflation rate is larger than that of the macro inflation rate. This is true, even when we use the residual FE of the industry-specific inflation rate where we have removed the aggregate component and the size effect from the original FE. For FE of sales, the average is close to zero, while its standard deviation is much larger than that of the two inflation rates.

Table 2: Summary statistics of the inflation rates

	Obs.	mean	std. dev.	median
realized macro-level inflation rate	10296	−0.40%	1.13%	−0.70%
forecasted macro-level inflation rate	8881	−0.12%	0.80%	0.00%
realized industry-specific inflation rate	10296	0.38%	3.12%	0.39%
forecasted industry-specific inflation rate	7770	−0.04%	0.77%	0.00%

Notes: Realized macro-level and industry-specific inflation rates (23 industries) are obtained from the website of the Economic and Social Research Institute (ESRI) within the Cabinet Office and refer to the fiscal year (April to March). To exclude outliers, we trim the top and bottom one percent of observations of the forecasts. Time span: 2004-2017 (fiscal years).

[22] In the data, average inflation rate of manufacturing industries is higher than that of service sectors. Since we have 13 manufacturing industries out of 23 industries in our dataset, inflation rates of manufacturing industries are over-represented in the calculation of the average (industry-specific) inflation rate. This also explains why the average realized (industry-specific) inflation rate is higher than the average macro inflation rate.

Table 3: Summary statistics of forecast errors

	Obs.	mean	std. dev.	median
forecast error of macro-level inflation rate	8151	−0.25%	1.00%	−0.40%
forecast error of industry-specific inflation rate	7128	0.56%	2.83%	0.40%
(percentage) forecast error of sales	5910	−0.90%	8.70%	−0.45%
(logarithm) forecast error of sales	5911	-0.0131	0.0896	-0.00457
residual forecast error of industry-specific inflation rate	7126	−0.03%	2.43%	−0.13%

Notes: Realized macro-level and industry-specific inflation rates (23 industries) are obtained from the website of the Economic and Social Research Institute (ESRI) within the Cabinet Office and refer to the fiscal year (April to March). The forecast error is defined as the difference between the realized value from the forecasted value. To construct the residual forecast error of the industry-specific inflation rate, we tease out the aggregate component and the size effect from the original forecast error of industry-specific inflation rate. To exclude outliers, we trim the top and bottom one percent of observations of the FEs. Time span: 2004-2017 (fiscal years).

In Table 4, we present the list of industries included in our dataset and the number of observations (of the industry-specific inflation forecast) that belong to each industry. All firms are grouped into 23 (broad) industries, and more than half of them (13) are manufacturing industries. The fact that the ASCB dataset has broad industry classifications helps firms answer the survey, as most firms in our dataset are large firms with businesses across several small industries. It is clear from the table that manufacturing firms are over-represented in the sample (compared with their contribution to the GDP of Japan), as more than half of the observations are from manufacturing industries. However, several non-manufacturing industries such as construction, wholesale/retail, finance, and transportation also have many observations.

Tables 8 and 9 in Online Appendix 6.1 show the same statistics for the alternative dataset (i.e., the BOS conducted in January). Naturally, the standard deviation of the forecast error of sales is larger than that of the main dataset, as firms are asked to forecast their sales three months earlier than the timing of the main dataset.

Table 4: Number of observations from each industry

industry name	obs. of industry-specific inflation forecasts
Fisheries and Agriculture	76
Mining	42
Construction	819
Food	442
Textiles	299
Pulp and paper	108
Chemicals	1025
Coals and oil	61
Ceramics products	318
Primary metal	612
Metal products	282
General machineries	868
Electronic machineries	1052
Transportation equipments	455
Precision machineries	148
Other manufacturing	356
Wholesale/retail	1690
Finance	717
Real estate	185
Transportation	647
Information and Communication	310
Electricity and Gas	194
Other services	448

Notes: This table presents the number of observations of industry-specific inflation expectations for 23 industry in merged dataset. The industry-specific inflation is the inflation rate of industrial output (i.e., not value added).

4 Empirical Results

4.1 Estimations of Shock Processes

We implement our estimation of various dynamic process using data at the annual frequency, as firms' forecasts are made at the annual frequency. We estimate the AR(1) processes of the three variables first. For the process of the (annual) macro inflation rate and that of the industry-specific inflation rate, the estimation results are reported in Table 5. Note that we use the data for $2004-2017$ to implement the estimation, as the sample period of realized inflation rates and firm sales in our dataset is $2004-2017$. In addition, we include both the year and industry fixed effects into the AR(1) regression of the industry-specific inflation rate. We do so, as we want to allow for different long-run average inflation rates across industries and to tease out the correlated part between the macro inflation and the industry-specific inflation from the process of the industry-specific inflation. In other words, the estimated serial correlation of the industry-specific inflation rate is the correlation of the industry-specific inflation (that is unrelated to the macro inflation). The estimated persistence is 0.643 for the macro inflation rate and 0.455 for the industry-specific inflation rate.[23] The estimated standard deviation of innovations (to the inflation process) is 0.91 percentage points for the macro one and 2.76 percentage points for the industry one, which implies that the macro inflation process is less volatile than the industry-specific inflation process.[24]

Now, we turn to the estimation of the firm's demand shifter. Although we only have the information on sales in the dataset, the structure of our model enables us to back out the process of the demand shifter by utilizing the information on the expected sales and output

[23]We have also estimated the persistence of the real GDP growth rate and investigated the serial correlation of its forecast errors. It turns out both the persistence is extremely low (0.075) and statistically insignificant. Although we find that the serial correlation of the forecast errors of real GDP growth is also low and statistically insignificant, we cannot conclude that the degree of information rigidity is low for this variable (as the process itself is not persistent).

[24]We have also used the data of the annual (macro) inflation rate for the period of 2000-2018 (and 1995-2018) to run the AR(1) regression. The estimated coefficients are 0.69 (s.e.: 0.162) and 0.632 (s.e.: 0.132) respectively, which are close to 0.643.

Table 5: Processes of macro and industry-level inflation

Dep.Var:	π_t	π_t^i
π_{t-1}	0.643***	
	(0.161)	
π_{t-1}^i		0.455***
		(0.047)
Constant	-0.001	0.0176**
	(0.003)	(0.007)
Year fixed effects	No	Yes
Industry fixed effects	No	Yes
N	14	320
R^2	0.427	0.506

Notes: We regress the macro-level inflation rate (and the industry-specific inflation rate) on its (one-period) lagged term using annual data for 2004-2017. We include both the year and industry fixed effects into the AR(1) regression of the industry-specific inflation rate. Robust standard errors are reported in the parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

price changes. Specifically, firm ω's sales in period $t+1$ equals

$$R_{\omega,t+1} = p_{\omega,t+1} q_{\omega,t+1}.$$

In our model, the firm chooses and therefore knows its output level $q_{\omega,t+1}$ when forecasting its sales. Thus, we have

$$\log(q_{\omega,t+1}) - \log(q_{\omega,t}) \approx \frac{q_{\omega,t+1} - q_{\omega,t}}{q_{\omega,t}} \approx \frac{\mathbb{E}_{\omega,t}[R_{\omega,t+1}] - R_{\omega,t}}{R_{\omega,t}} - \frac{\mathbb{E}_{\omega,t}[p_{\omega,t+1}] - p_{\omega,t}}{p_{\omega,t}},$$

where "≈" means "approximately equal." As both the expected sales growth and the expected (average) change of output prices are reported in the data, we are able to construct the (percentage and logarithm of) realized output growth from period t to period $t+1$. Next, the logarithm of realized sales in period $t+1$ can be stated as:

$$\log R_{\omega,t+1} = \frac{\sigma-1}{\sigma}\log(q_t(\omega)) + \frac{1}{\sigma}\log(C_t) + \frac{\delta}{\sigma}\log(P_t) + \frac{\sigma-\delta}{\sigma}\log(P_{i,t}) + \frac{1}{\sigma}a_t(\omega).$$

Note that C_t, P_t and $P_{i,t}$ are year or industry-year specific. Therefore, we can regress $\log(R_{\omega,t+1})$ on $\log(R_{\omega,t})$ and control for the industry-year fixed effects and the change in output, $\log(q_{\omega,t+1}) - \log(q_{\omega,t})$. This regression yields the estimates of the persistence as well as the standard deviation of innovations for the process of $a_t(\omega)/\sigma$. As different firms likely have different long-run average demand shifters (i.e., different intercepts for the AR(1) process of the firm's demand shifter), we calculate the difference between the realized sales and its over-time mean for a given firm. Then, we use the demeaned sales to run the AR(1) regression.

Table 6 represents the estimation results. We have tried to include different sets of fixed effects, and the estimated persistence is robustly around 0.76, which is higher than the persistence of the two inflation processes. This persistence value at the annual level translates into a quarterly persistence of 0.94, which is close to the value used in the literature.[25] As the theory predicts, a one percent increase in the output from period t to $t+1$ (i.e., $\Delta q_{\omega,t-1,t} = 1$) results in a roughly $0.44\% - 0.52\%$ increase in sales. Finally, the estimated standard deviation of innovations to the demand shifter is between 0.081 to 0.104, which is much higher than the volatility of macro- and industry-level innovations. For future use, we choose the specification using the industry-year, size-year and region-year fixed effects as our main specification for the estimation. Under this specification, ρ_a is estimated to be 0.765, and the estimated standard deviation of innovations to the process of $a_t(\omega)/\sigma$ is 0.096. Table 10 in Online Appendix 6.1 reports the regression results using the alternative sample, which are similar to the regression results presented here.[26]

4.2 Serial Correlations

Next, we present the results of the serial correlation regressions for the three target variables in Table 7. We use the residual FE of the industry-specific inflation rate to run the AR(1) regression, as we want to exclude the component of the FE (of the industry-specific inflation rate) that comes from the mis-forecast of the macro inflation. Similarly, we use the residual

[25] Bloom et al. (2018) used 0.95 as the quarterly-level persistence of the demand/productivity shock.
[26] The estimated persistence is 0.686, and the standard deviation of innovations is 0.089.

Table 6: Processes of the demand shifter

Dep.Var:	$\log(R)^{demeaned}_{\omega,t}$			
$\log(R)^{demeaned}_{\omega,t-1}$	0.752***	0.764***	0.765***	0.754***
	(0.019)	(0.018)	(0.019)	(0.025)
$\Delta q_{\omega,t-1,t}$	0.440***	0.431***	0.430***	0.518***
	(0.029)	(0.028)	(0.030)	(0.033)
Constant	0.002	0.002	0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.001)
industry fixed effects	Yes	No	No	No
region fixed effects	Yes	Yes	No	No
size-year fixed effects	Yes	Yes	Yes	Yes
industry-year fixed effects	No	Yes	Yes	Yes
region-year fixed effects	No	No	Yes	Yes
firm fixed effects	No	No	No	Yes
N	4249	4231	4118	3880
R^2	0.624	0.659	0.679	0.773

Notes: We regress the logarithm of demeaned sales in period t on its one-period lag and the percentage change in output (i.e., quantity) produced from period $t-1$ to period t. Standard errors are clustered at the firm level and reported in parentheses. Top and bottom one percent of the logarithm of firm sales are trimmed. $\log(R)^{demeaned}_{\omega,t}$ is the logarithm of firms i's sales in period t, while $\Delta q_{\omega,t-1,t}$ is the percentage change output produced from period $t-1$ to period t (i.e., $\Delta q_{\omega,t-1,t} = 1$ means 1%). Note that each firm belongs to one of the four size-based bins in the data. The regression controls for various fixed effects such as the industry-year, size-year, and region-year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

FE of sales to run the serial correlation regression, as we want to exclude the component of the FE of sales that comes from the mis-forecast of the macro-level and industry-level variables.[27] We observe that the serial correlations of FEs of all three variables are positively significant, substantiating the existence of information rigidity.[28] Table 11 in Online Appendix 6.1 reports regression results using the alternative sample.

We performed several robustness checks to confirm the estimated serial correlation coefficients presented in Table 7. Specifically, we tried various sets of fixed effects to estimate the serial correlation regressions. The estimated serial correlation is around 0.34 for the FE of macro inflation and around 0.2 for the FE of industry-specific inflation. For the serial correlation of FE of sales, the estimated coefficient is around 0.14 (percentage sales FE) and 0.12 (logarithm of sales FE). One potential concern for the regression of FE of the industry-specific inflation rate is that there are several industries that have only a few observations of

[27]For the FE of the industry-specific inflation rate, size-bin-year fixed effects are teased out from the original FE. For the FE of sales, size-bin-industry-year fixed effects are teased out from the original FE.

[28]As the standard errors of the estimated AR(1) coefficients are small (0.01-0.02), the correlation of FEs should also be statistically different.

industry-specific inflation forecasts each year (see Table 4). Thus, the law of large numbers might not hold when we average out the firm-specific noise terms in a given year. In Table 12 of Online Appendix 6.2, we exclude industries where the number of (industry-specific) inflation expectations is too small (e.g., below 150 or 225 over 15 years) and re-run the serial correlation regression of FEs of the industry-specific inflation rate. The estimated coefficients are very similar to the one reported in Table 7.

Table 7: Serial correlation of forecast errors

Dep.Var:	$\mathbb{FE}^{\pi}_{\omega,t}$	$\mathbb{FE}^{\pi^i}_{\omega,t}$	$\mathbb{FE}^{pct,sales}_{\omega,t}$	$\mathbb{FE}^{log,sales}_{\omega,t-1}$
$\mathbb{FE}^{\pi}_{\omega,t-1}$	0.349***			
	(0.012)			
$\mathbb{FE}^{\pi^i}_{\omega,t-1}$		0.197***		
		(0.021)		
$\mathbb{FE}^{res\ pct,sales}_{\omega,t-1}$			0.143***	
			(0.021)	
$\mathbb{FE}^{res\ log,sales}_{\omega,t-1}$				0.119***
				(0.026)
$log(sales)_{\omega,t-1}$	-0.011	0.033	0.002	-0.000
	(0.008)	(0.025)	(0.001)	(0.001)
Constant	0.071	-0.701	-0.062*	0.008
	(0.207)	(0.613)	(0.034)	(0.033)
year fixed effects	No	Yes	No	No
industry fixed effects	Yes	Yes	No	No
region fixed effects	Yes	Yes	Yes	Yes
industry-year fixed effects	No	No	Yes	Yes
N	4689	4061	3828	3844
R^2	0.154	0.246	0.253	0.137

Notes: Standard errors are clustered at the firm level and reported in parentheses. $\mathbb{FE}^{\pi}_{\omega,t}$ is the forecast error of the macro inflation rate. Forecast errors of the industry-specific inflation rate are residual forecast errors. That is, size-bin-year fixed effects are teased out from the original forecast errors. As a result, $\mathbb{FE}^{\pi^i}_{\omega,t}$ is the residual forecast error of the industry-specific inflation rate. Forecast errors of sales are residual forecast errors. That is, size-bin-industry-year fixed effects are teased out from the original forecast errors. $\mathbb{FE}^{pct,sales}_{\omega,t}$ is the residual forecast error of firm sales in percentage term. $\mathbb{FE}^{pct,sales}_{\omega,t}$ is the residual logarithm of the forecast error of firm sales. The top and bottom 1% of the FEs are trimmed. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As the serial correlation coefficient equals the persistence of the dynamic process multiplied by one minus the Kalman gain, we derive the Kalman gains for the three target variables as follows (based on the estimated dynamics processes):

$$G_{\pi} = 0.457,\ G_{\pi^i} = 0.567,\ \text{and}\ G_a = 0.844.$$

Using equation (8) that states $G_x = 1 - e^{-2\kappa_x}$, we infer the corresponding values of the channel capacity as follows:

$$\kappa_{\pi}^{data} = 0.305,\ \kappa_{\pi^i}^{data} = 0.419, \text{ and } \kappa_{a}^{data} = 0.930. \tag{24}$$

Therefore, the firm allocates most attention to forecasting the demand shifter and least attention to forecasting the (macro) inflation rate. This is consistent with our estimates of the three processes, as the process of the demand shifter is more volatile than the two inflation processes.[29]

4.3 Backing out the Information Cost

In the final step of our empirical exercise, we calibrate the marginal cost of acquiring and processing information, λ, and the two elasticity parameters, δ and σ, by matching the three channel capacities obtained from the data in equation (24). First, from equations (12) and (15), it is clear that we have an over-identification problem. Specifically, as both the left- and right-hand-sides of equation (12) are homogeneous of degree one with respect to the three parameters (λ, δ, σ), only the ratios of δ/λ and σ/λ matter for the optimal attention allocation (and thus can be calibrated).[30] As a result, we can only calibrate two parameters in order to match the three values of the channel capacity. We follow the literature (e.g., Bernard et al. (2003)) by assuming $\sigma = 4$ (the elasticity of substitution between varieties)[31] and calibrate the other two parameters (δ and λ) to minimize the sum

[29] The Kalman gains and the channel capacities above are derived from estimated coefficients (i.e., the AR(1) coefficients of the three processes and those of the three FEs). As the standard error of the estimated AR(1) coefficient of the macro inflation rate is large (0.161), it is hard to conclude that the attention allocated to the macro inflation rate is *statistically* different from the one allocated to the industry-specific inflation rate. However, as G_{π} and G_{π^i} are much smaller than G_a and the standard error of the estimated AR(1) coefficients are small for industry and firm-level variables, we can conclude that the attention allocated to the macro (industry-specific) inflation rate is *statistically* different from the one allocated to the firm-specific demand.

[30] In other words, if a set of values of (σ, δ, λ) leads to perfectly matched moments of the channel capacities, scaling the three parameters up or down by the same proportion leads to the same outcome.

[31] Alternatively, we can also normalize one sensitivity parameter (e.g., w_a) to one which is equivalent to choosing a value for σ.

of the squared percentage difference between the model-implied channel capacity and its empirical counterpart (across three target variables):

$$\min_{\delta,\lambda} \quad \left(\frac{\kappa_{\pi}^{model} - \kappa_{\pi}^{data}}{\kappa_{\pi}^{data}}\right)^2 + \left(\frac{\kappa_{\pi^i}^{model} - \kappa_{\pi^i}^{data}}{\kappa_{\pi^i}^{data}}\right)^2 + \left(\frac{\kappa_{a}^{model} - \kappa_{a}^{data}}{\kappa_{a}^{data}}\right)^2, \tag{25}$$

where (κ_x^{model} and κ_x^{data}) ($x \in \{\pi, \pi^i, a\}$) are the channel capacities implied by the model and estimated from the data, respectively.

We use the function of `fmincon` in Matlab to search for values of (δ, λ) that minimize the objective function in equation (25), which leads to the calibrated values as

$$\delta = 2.484, \quad \lambda = 2.595 * \cdot 10^{-4}.$$

The corresponding moments implied from the model are

$$\kappa_{\pi}^{model} = 0.313,\ \kappa_{\pi^i}^{model} = 0.426, \text{ and } \kappa_{a}^{model} = 0.831, \tag{26}$$

which are close to their empirical counterparts in equation (24). Moreover, the standard deviations of FEs implied by the model (non-targeted moments) are

$$\Psi_{\pi}^{theory} = 1.03\%,\ \Psi_{\pi^i}^{theory} = 2.89\%,\text{and } \Psi_{a}^{theory} = 10.2\%,$$

which are very close to their empirical counterparts:

$$\Psi_{\pi}^{data} = 1.0\%,\ \Psi_{\pi^i}^{data} = 2.83\%, \text{ and } \Psi_{a}^{data} = 9.0\%.$$

Finally, the calibrated sensitivity parameters of the three target variables are

$$w_{\pi} = 6.170,\ w_{\pi^i} = 2.298, \text{ and } w_a = 1.510. \tag{27}$$

The interesting finding from the calibration is that the calibrated importance parameter is much bigger for the macro inflation rate than for the industry-specific inflation rate and the firm demand. There are two explanations for this. First, firms in our sample are the largest firms in Japan (i.e., public firms) and thus probably care about the macro condition much more than an average Japanese firm. Therefore, forecasting macro inflation well is probably crucial for them. In addition, there is another possibility that the marginal cost of acquiring and processing information (which is a subjective mental cost in RI models) is possibly variable-dependent and thus heterogeneous across target variables. In particular, as firms that in our merged dataset are very large firm, they can have more resources to acquire macro information at a lower cost. This implies that the marginal cost of acquiring and processing information *might* be lower for the macro variable than for the other two target variables for these large firms. If this is true, then the model-implied importance parameter of macro inflation will become smaller. We leave a further exploration of the above two hypotheses to future research.

4.4 Gain from Removing Information Rigidities

In this subsection, we implement a simple back-of-the-envelope calculation of the payoff gains when we remove information rigidities (i.e., setting the information cost to zero). Based on equation (14) and the variance of various FEs implied by our calibrated model, the overall gain from removing information frictions is

$$\begin{aligned} Gain &= Gain_{\pi} + Gain_{\pi^i} + Gain_a \\ &= 6.17 * (1.03\%)^2 + 2.30 * (2.89\%)^2 + 1.51 * (10.2\%)^2 \\ &= 0.065\% + 0.192\% + 1.571\% = 1.828\%. \end{aligned}$$

We refrain from emphasizing the magnitudes of the gains, as our model is a stylized and partial-equilibrium model. Rather, we want to emphasize the contribution made by removing the information rigidity associated with each variable to the overall gain. The contribution depends on two factors: the importance of forecasting the target variable (correctly) and the variance of the FE of the target variable. We find that removing information frictions concerning forecasting macro inflation only increases the firm's payoff slightly (less than 4% of the overall gain), as the variance of its FE is extremely small. Interestingly, this is true despite the fact that the importance of forecasting macro inflation is much higher than the other two variables. In fact, the payoff gain from *only* removing information frictions associated with forecasting macro inflation is less than 0.1%, which is consistent with the finding of a small welfare gain from the literature (e.g., Luo (2008) and Maćkowiak and Wiederholt (2015)). To the contrary, removing information frictions associated with forecasting industry-specific inflation and firm-specific demand accounts for most of the overall payoff gain, as the variance of the FE of these two variables is much larger than that of the macro inflation rate. In summary, we find that the variance of the FE determines the level of the payoff gain from removing information frictions (for a given target variable) in our context.

5 Conclusion

In this study, we utilize a novel Japanese firm-level panel dataset that contains quantitative forecasts of the macro inflation rate, the industry-specific inflation rate, and firm sales to infer the corresponding degrees of information rigidities at the macro-, industry- and firm-levels. We find that the degree of information rigidity concerning forecasting the macro target is higher than the one associated with forecasting industry inflation and firm's demand. This is consistent with the predictions of the RI model with elastic attention proposed in Sims (2010) and supports the argument in favor of the existence of state-dependent information rigidities.

Then, we use our model and data moments to calibrate the marginal cost of acquiring and processing information and the importance parameters of forecasting the three variables by matching the three channel capacities obtained from the data. It is shown that the calibrated importance parameter is by far bigger for the macro inflation rate than for the industry-specific inflation rate and the firm demand. In spite of this finding, we show that removing information rigidities associated with forecasting the industry-specific inflation rate and firm-specific demand would increase the firm's payoff substantially more, compared with removing the information rigidity associated with forecasting the macro inflation. This is true, as the variance of FE is much smaller for the macro inflation rate than the other two target variables. The finding of the heterogeneous importance parameters calls for further research on heterogeneous incentives of acquiring information across firms and heterogeneous costs of processing information.

References

Afrouzi, Hassan and Choongryul Yang, "Dynamic rational inattention and the Phillips curve," *Available at SSRN 3770462*, 2021.

Andrade, Philippe and Hervé Le Bihan, "Inattentive professional forecasters," *Journal of Monetary Economics*, November 2013, *60* (8), 967–982.

_ **, Olivier Coibion, Erwan Gautier, and Yuriy Gorodnichenko**, "No firm is an island? How industry conditions shape firms' aggregate expectations," Technical Report, National Bureau of Economic Research 2020.

Baker, Scott R, Tucker S McElroy, and Xuguang S Sheng, "Expectation Formation Following Large, Unexpected Shocks," *Review of Economics and Statistics*, 2020, *102* (2), 287–303.

Barrero, Jose Maria, "The micro and macro of managerial beliefs," *Available at SSRN 3196761*, 2020.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in International Trade," *American Economic Review*, 2003, *93* (4), 1268–1290.

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry, "Really uncertain business cycles," *Econometrica*, May 2018, *86* (3), 1031–1065.

Boivin, Jean, Marc P Giannoni, and Ilian Mihov, "Sticky prices and monetary policy: Evidence from disaggregated US data," *American Economic Review*, 2009, *99* (1), 350–84.

Chen, Cheng, Tatsuro Senga, Chang Sun, and Hongyong Zhang, "Uncertainty, imperfect information, and expectation formation over the firms' life cycle," Technical Report, RIETI Discussion Paper 18-E-010 2020.

Coibion, Olivier and Yuriy Gorodnichenko, "What can survey forecasts tell us about information rigidities?," *Journal of Political Economy*, 2012, *120* (1), 116–159.

_ **and** _ , "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 2015, *105* (8), 2644–78.

_ **,** _ **, and Saten Kumar**, "How do firms form their expectations? New survey evidence," *American Economic Review*, 2018, *108* (9), 2671–2713.

Hottman, Colin J, Stephen J Redding, and David E Weinstein, "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.

Kahneman, Daniel, *Attention and effort*, Vol. 1063, Citeseer, 1973.

Lucas, Robert E, "Expectations and the neutrality of money," *Journal of Economic Theory*, 1972, *4* (2), 103–124.

_ , "Some international evidence on output-inflation tradeoffs," *American Economic Review*, 1973, *63* (3), 326–334.

Luo, Yulei, "Consumption dynamics under information processing constraints," *Review of Economic dynamics*, 2008, *11* (2), 366–385.

_ **and Eric R Young**, "Signal extraction and rational inattention," *Economic Inquiry*, 2014, *52* (2), 811–829.

_ **and** _ , "Induced uncertainty, market price of risk, and the dynamics of consumption and wealth," *Journal of Economic Theory*, 2016, *163*, 1–41.

_ **, Jun Nie, Gaowang Wang, and Eric R Young**, "Rational inattention and the dynamics of consumption and wealth in general equilibrium," *Journal of Economic Theory*, 2017, *172*, 55–87.

Ma, Yueran, Tiziano Ropele, David Sraer, and David Thesmar, "A quantitative analysis of distortions in managerial forecasts," Technical Report, National Bureau of Economic Research 2020.

Mackowiak, Bartosz and Mirko Wiederholt, "Optimal sticky prices under rational inattention," *American Economic Review*, 2009, *99* (3), 769–803.

Maćkowiak, Bartosz and Mirko Wiederholt, "Business cycle dynamics under rational inattention," *Review of Economic Studies*, October 2015, *82* (4), 1502–1532.

Mankiw, N. Gregory and Ricardo Reis, "Sticky information versus sticky prices: A proposal to replace the New Keynesian phillips curve," *Quarterly Journal of Economics*, November 2002, *117* (4), 1295–1328.

Mankiw, N Gregory, Ricardo Reis, and Justin Wolfers, "Disagreement about inflation expectations," *NBER macroeconomics annual*, 2003, *18*, 209–248.

Meyer, Brent, Nicholas B Parker, Xuguang Simon Sheng et al., "Unit Cost Expectations and Uncertainty: Firms' Perspectives on Inflation," Technical Report 2021.

Miao, Jianjun, Jieran Wu, and Eric R Young, "Multivariate rational inattention," *Econometrica*, 2022, *90* (2), 907–945.

Muth, John F., "Optimal properties of exponentially weighted forecasts," *Journal of the American Statistical Association*, 1960, *55* (290), 299–306.

Muth, John F, "Rational expectations and the theory of price movements," *Econometrica*, 1961, pp. 315–335.

Paciello, Luigi and Mirko Wiederholt, "Exogenous information, endogenous information, and optimal monetary policy," *Review of Economic Studies*, 2014, *81* (1), 356–388.

Ryngaert, Jane, "What do (and don't) forecasters know about U.S. Inflation," Working Paper 2017.

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

Sims, Christopher A, "Rational inattention and monetary economics," in "Handbook of Monetary Economics," Vol. 3, Elsevier, 2010, pp. 155–181.

Tanaka, Mari, Nicholas Bloom, Joel M David, and Maiko Koga, "Firm performance and macro forecast accuracy," *Journal of Monetary Economics*, 2019.

6 Online Appendix (Not for Publication)

6.1 Empirical Results using an Alternative Dataset

In this subsection, we present empirical results using the alternative dataset. Based on equations (16), (17), (19), and the estimates presented in Tables 5, 10, 11, we derive the Kalman gains for the three target variables as follows:

$$G_\pi = 0.443,\ G_{\pi^i} = 0.374, \text{ and } G_a = 0.827.$$

The implied channel capacities are

$$\kappa_\pi^{data} = 0.293,\ \kappa_{\pi^i}^{data} = 0.234, \text{ and } \kappa_a^{data} = 0.877. \tag{28}$$

Therefore, the firm devotes most of its attention to forecasting the process of its demand shifter and least attention to forecasting the macro and industry-specific inflation rates.

Following the same procedure as in the main text, we calibrate our model by setting $\sigma = 4$ and minimizing the objective function defined in equation (25). The calibrated parameter values are

$$\delta = 2.628,\ \lambda = 3.507 \cdot 10^{-4}.$$

The corresponding moments implied from the model are

$$\kappa_\pi^{model} = 0.2546,\ \kappa_{\pi^i}^{model} = 0.2240, \text{ and } \kappa_a^{model} = 1.2193, \tag{29}$$

which are not far away from their empirical counterparts in equation (28). Moreover, the standard deviations of FEs implied by the model (non-targeted moments) are

$$\Psi_\pi^{theory} = 1.05\%,\ \Psi_{\pi^i}^{theory} = 2.96\%, \text{ and } \Psi_a^{theory} = 9.1\%,$$

which are very close to their empirical counterparts:

$$\Psi_{\pi}^{data} = 1.01\%,\ \Psi_{\pi^i}^{data} = 2.48\%,\ \text{and}\ \Psi_{a}^{data} = 11.0\%.$$

Finally, the calibrated sensitivity parameters of the three target variables are

$$w_{\pi} = 6.906,\ w_{\pi^i} = 1.882,\ \text{and}\ w_a = 3.352. \tag{30}$$

It is still true that the calibrated importance (or sensitivity) parameter is by far bigger for the macro inflation rate than for the industry-specific inflation rate and the firm demand. In addition, the calibrated δ barely changes from the main text. Finally, the calibrated marginal cost of acquiring and processing information is higher than in the main text. This makes sense, as firm (managers) forecast their sales earlier now (i.e., January instead April).

The gain from eliminating information frictions can be decomposed as follows:

$$\begin{aligned} Gain &= Gain_{\pi} + Gain_{\pi^i} + Gain_a \\ &= 6.906 \cdot (1.05\%)^2 + 1.882 \cdot (2.96\%)^2 + 3.352 \cdot (9.1\%)^2 \\ &= 0.076\% + 0.147\% + 2.776\% \approx 3.00\%, \end{aligned}$$

where the most gain comes from eliminating information frictions related to forecasting the firm-specific demand. This is true, as the standard deviation of forecast errors is by far the largest for firm sales.

6.2 Regression Results that Exclude Industries with Few Observations

In this subsection, we rerun the serial correlation regression of the FE of industry-specific inflation by excluding industries that have few observations. Regression results are reported

Table 8: Summary statistics of the inflation rates (sales forecasts made in Jan.)

	Obs.	mean	std. dev.	median
realized macro-level inflation rate	9405	−0.35%	1.17%	−0.70%
forecasted macro-level inflation rate	8165	−0.07%	0.77%	0.00%
realized industry-specific inflation rate	9405	0.15%	2.76%	0.19%
forecasted industry-specific inflation rate	7109	−0.02%	0.74%	0.00%

Notes: Realized macro-level and industry-specific inflation rates (23 industries) are obtained from the website of the Economic and Social Research Institute (ESRI) within the Cabinet Office and refer to the fiscal year (April to March). To exclude outliers, we trim the top and bottom one percent of observations of the forecasts. Time span: 2004-2017 (fiscal years).

Table 9: Summary statistics of forecast errors (sales forecasts made in Jan.)

	Obs.	mean	std. dev.	median
forecast error of macro-level inflation rate	7414	−0.26%	1.01%	−0.40%
forecast error of industry-specific inflation rate	6497	0.32%	2.48%	0.33%
(percentage) forecast error of sales	2615	−0.30%	10.43%	−0.02%
(logarithm) forecast error of sales	2614	−0.01	0.11	−0.00
residual forecast error of industry-specific inflation rate	6498	−0.02%	2.10%	−0.075%

Notes: Realized macro-level and industry-specific inflation rates (23 industries) are obtained from the website of the Economic and Social Research Institute (ESRI) within the Cabinet Office and refer to the fiscal year (April to March). The forecast error is defined as the difference between the realized value from the forecasted value. To exclude outliers, we trim the top and bottom one percent of observations of the FEs. Time span: 2004-2017 (fiscal years).

Table 10: Processes of the demand shifter (sales forecasts made in Jan.)

Dep.Var:	$\log(R)^{demeaned}_{\omega,t}$			
$\log(R)^{demeaned}_{\omega,t-1}$	0.657***	0.689***	0.686***	0.576***
	(0.031)	(0.031)	(0.034)	(0.050)
$\Delta q_{\omega,t-1,t}$	0.330***	0.320***	0.307***	0.367***
	(0.043)	(0.042)	(0.047)	(0.064)
Constant	0.006***	0.007***	0.006**	0.005***
	(0.002)	(0.002)	(0.002)	(0.001)
industry fixed effects	Yes	No	No	No
region fixed effects	Yes	Yes	No	No
size-year fixed effects	Yes	Yes	Yes	Yes
industry-year fixed effects	No	Yes	Yes	Yes
region-year fixed effects	No	No	Yes	Yes
firm fixed effects	No	No	No	Yes
N	2009	1973	1857	1594
R^2	0.527	0.601	0.633	0.752

Notes: We regress the logarithm of demeaned sales in period t on its one-period lag and the percentage change in output (i.e., quantity) produced from period $t-1$ to period t. Standard errors are clustered at the firm level and reported in parentheses. Top and bottom one percent of the logarithm of firm sales are trimmed. $\log(R)^{demeaned}_{\omega,t}$ is the logarithm of firms i's sales in period t, while $\Delta q_{\omega,t-1,t}$ is the percentage change output produced from period $t-1$ to period t (i.e., $\Delta q_{\omega,t-1,t} = 1$ means 1%). Note that each firm belongs to one of the four size-based bins in the data. The regression controls for various fixed effects such as the industry-year, size-year, and region-year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Serial correlation of FEs (sales forecasts made in Jan.)

Dep.Var:	$\mathbb{FE}^{\pi}_{\omega,t}$	$\mathbb{FE}^{\pi^i}_{\omega,t}$	$\mathbb{FE}^{pct,sales}_{\omega,t}$	$\mathbb{FE}^{log,sales}_{\omega,t-1}$
$\mathbb{FE}^{\pi}_{\omega,t-1}$	0.358***			
	(0.013)			
$\mathbb{FE}^{\pi^i}_{\omega,t-1}$		0.285***		
		(0.023)		
$\mathbb{FE}^{res\ pct,sales}_{\omega,t-1}$			0.110***	
			(0.040)	
$\mathbb{FE}^{res\ log,sales}_{\omega,t-1}$				0.119***
				(0.044)
$log(sales)_{\omega,t-1}$	-0.012	-0.006	0.003	0.002
	(0.010)	(0.023)	(0.003)	(0.003)
Constant	0.121	0.165	-0.068	-0.043
	(0.207)	(0.565)	(0.66)	(0.064)
year fixed effects	No	Yes	No	No
industry fixed effects	Yes	Yes	No	No
region fixed effects	Yes	Yes	Yes	Yes
industry-year fixed effects	No	No	Yes	Yes
N	3963	3471	1297	1306
R^2	0.162	0.243	0.149	0.141

Notes: Standard errors are clustered at the firm level and reported in parentheses. $\mathbb{FE}^{\pi}_{\omega,t}$i s the forecast error of the macro inflation rate. Forecast errors of the industry-specific inflation rate are residual forecast errors. That is, size-bin-year fixed effects are teased out from the original forecast errors. As a result, $\mathbb{FE}^{\pi^i}_{\omega,t}$ is the residual forecast error of the industry-specific inflation rate. Forecast errors of sales are residual forecast errors. That is, size-bin-industry-year fixed effects are teased out from the original forecast errors. $\mathbb{FE}^{pct,sales}_{\omega,t}$ is the residual forecast error of firm sales in percentage term. $\mathbb{FE}^{pct,sales}_{\omega,t}$ is the residual logarithm of forecast error of firm sales. The top and bottom 1% of the FEs are trimmed. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in Table 12.

Table 12: Serial correlation of FEs of industry-specific inflation rate

Dep.Var:	$\mathbb{FE}^{\pi^i}_{\omega,t}$			
$\mathbb{FE}^{\pi^i}_{\omega,t-1}$	0.186***	0.190***	0.185***	0.190***
	(0.022)	(0.023)	(0.022)	(0.023)
$log(sales)_{\omega,t-1}$	0.028	0.028	0.020	0.019
	(0.026)	(0.027)	(0.028)	(0.029)
Constant	-0.603	-0.601	-0.408	-0.374
	(0.635)	(0.661)	(0.683)	(0.718)
year fixed effects	Yes	No	Yes	No
industry fixed effects	Yes	Yes	Yes	Yes
region fixed effects	Yes	No	Yes	No
region-year fixed effects	No	Yes	No	Yes
Sample (obs. of forecasts)	≥ 150	≥ 150	≥ 225	≥ 225
N	3815	3696	3510	3381
R^2	0.241	0.292	0.241	0.292

Notes: Standard errors are clustered at the firm level and reported in parentheses. Forecast errors of the industry-specific inflation rate are residual forecast errors. That is, size-bin-year fixed effects are teased out from the original forecast errors. As a result, $\mathbb{FE}^{\pi^i}_{\omega,t}$ is the residual forecast error of the industry-specific inflation rate. In the first two columns, we exclude industries with observations less than 150. In the last two columns, we exclude industries with observations less than 225. The top and bottom 1% of FEs are trimmed. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6.3 Theoretical Appendix

In this theoretical appendix, we prove that the loss function due to information rigidities is minimized when the firm minimizes the variance of the forecasting/filtering error of the output. Note that the firm's profit is given by

$$\Pi_t(\omega) = q_t(\omega)\left[\left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma} q_t(\omega)^{-1/\sigma} - w_t,\right],$$

where we have used the result that $p_t(\omega) = \left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma} q_t(\omega)^{-1/\sigma}$. Under full information, the optimal output be written as:

$$q_t^{full}(\omega) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \frac{C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}}{w_t^{\sigma}}.$$

Taking log on both sides yields:

$$\log q_t^{full}(\omega) = \sigma \log\left(\frac{\sigma-1}{\sigma}\right) + \log C_t + \delta \log P_t + (\sigma-\delta)\log P_{i,t} + a_t(\omega) - \sigma\log((w_t)\,. \quad (31)$$

Thus, the profit under full information is

$$\Pi_t^{full}(\omega) = \frac{1}{\sigma}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)} w_t^{1-\sigma}.$$

Now, we use Taylor expansion (up to the second-order) to approximate the profit function under information rigidities. Specifically, we have

$$\begin{aligned}\Pi_t(\omega) &= q_t(\omega)\left[\left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma} q_t(\omega)^{-1/\sigma} - w_t\right] \\ &\approx \Pi_t^{full}(\omega) - \frac{1}{2\sigma}\frac{w_t^{\sigma+1}}{C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}}\left(\frac{\sigma-1}{\sigma}\right)^{-\sigma}\left(q_t(\omega) - q_t^{full}(\omega)\right)^2.\end{aligned}$$

Firm revenue under full information can be written as

$$\begin{aligned}R_t^{full}(\omega) &= \left(q_t^{full}(\omega)\right)^{1-1/\sigma}\left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma} \\ &= \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)} w_t^{1-\sigma} \\ &= \sigma \Pi_t^{full}(\omega).\end{aligned}$$

The normalized loss function can thus be written as:

$$\begin{aligned}\frac{\Pi_t(\omega) - \Pi^{full}(\omega)}{\Pi_t^{full}(\omega)} &= -\frac{\sigma-1}{2\sigma}\left(\frac{q_t(\omega) - q_t^{full}(\omega)}{q_t^{full}(\omega)}\right)^2 \\ &= -\frac{\sigma-1}{2\sigma}\left(\log q_t(\omega) - \log q_t^{full}(\omega)\right)^2.\end{aligned}$$

Under imperfect information on the target variables, the optimal choice of output can be written as:

$$\log q_t(\omega) = \sigma \log \left(\frac{\sigma - 1}{\sigma}\right) + \sigma \log \mathbb{E}_{\omega,t-1} \left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma} - \sigma \log (w_t) . \tag{32}$$

Therefore, the variance of the output deviation from the full-information scenario can be expressed as

$$\mathbb{E}_{\omega,t-1} \left(\log q_t(\omega) - \log q_t^{full}(\omega)\right)^2 = \sigma^2 \mathbb{E}_{\omega,t-1} \left[\log \left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma} - \log \left(\mathbb{E}_{\omega,t-1} \left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma}\right)\right]^2$$

which is the variance of the forecasting error of $\left(C_t P_t^{\delta} P_{i,t}^{\sigma-\delta} e^{a_t(\omega)}\right)^{1/\sigma}$. Using equations (31) and (32), we can easily obtain the expression for the expected loss of the output change due to information rigidities.