INSTITUTO DE ECONOMÍA

DOCUMENTO de TRABAJO

530

2019

Rational Inattention-driven dispersion with volatility shocks

Javier Turén.

www.economia.puc.cl • ISSN (edición impresa) 0716-7334 • ISSN (edición electrónica) 0717-7593

Rational Inattention-driven dispersion with volatility shocks*

Javier Turen †

October 2019

Abstract

This paper studies price-setting decisions under Rational Inattention. Prices are set by tracking an unobserved target whose distribution is also *unknown*. The distribution of the target can change over time depending on persistent and unanticipated volatility shocks that hit the economy. Information acquisition is dynamic and fully flexible since, given information acquired in the past, business owners choose the amount of information they collect as well as how they want to learn about both the outcome and its distribution. We show that by allowing for imperfect information to be the *unique* source of rigidity, the model can simultaneously reconcile several stylized facts in the microeconomic evidence on price setting, both at the cross-sectional and time series levels. Dynamic imperfect information endogenously generates persistence in beliefs, which is crucial in replicating the dynamic empirical behavior of prices.

Keywords: *Rational Inattention, Dynamic Information, Price Dispersion, Frequency*

JEL Classification: E31, E32, D82, D83.

*I am extremely grateful to Raffaella Giacomini and Vasiliki Skreta for invaluable guidance and advices. I am also grateful for the useful suggestions and comments of Antonio Cabrales, Morten Ravn, Filip Matejka, Mirko Wiederholt, Franck Portier, Ralph Luetticke, Nicolas Figueroa, Edouard Schaal, Carlos Carvalho, Roger E.A. Farmer, Ernesto Pasten, Juan Pablo Torres-Martinez, Nezih Guner, Davide Melcangi, Pau Roldan, Silvia Sarpietro and Carlo Galli. All errors and omissions are mine.

†Pontificia Universidad Católica de Chile, Instituto de Economía, Contact: jturen@uc.cl

1 Introduction

What are the dynamic implications of imperfect information for price setting? What are the effects of time-varying volatility shocks on price setting? To set prices, owners must first acquire information about unpredictable components of their industries such as elasticities, their current demand, or the state of the economy. In reality, these price decisions are made with only partial information about both the realization of shocks and the stochastic process that generates them. This is relevant as the distribution of shocks is not only unknown but also likely to change over time, reflecting unanticipated periods of lower or higher uncertainty such as recessions. Despite the limited information, a rational price setter would adjust their behavior in response to *perceived* changes in their environment, and economic aggregates reflect those adjustments.

This paper argues that the aggregate effects documented using micro evidence on prices, *both* at the cross-sectional and time series level, can be rationalized as an implication of the dynamic process of collecting imperfect information about unobserved shocks while trying to uncover the distribution that generates them. The results contribute to the literature on the implications of imperfect information on aggregate conditions, started by the seminal contributions of Phelps (1967) and Lucas (1972), by mapping the role of dynamic imperfect information on aggregate price stability.

We propose a model of endogenous attention with costly entropy reduction to study how firms set prices when the distribution of shocks is time-varying. The model follows the literature on "Rational-Inattention" (henceforth, RI) Sims (2003), and allows for a dynamic and fully flexible information scheme. While past acquired information is relevant, we do not impose further assumptions on the *amount* of information being acquired or how owners *choose* to acquire it, i.e., there are no parametric assumptions on the distribution of signals. Firms collect information to update their beliefs about the realization of an aggregate fundamental, along with the distribution that generated it. As the predictability of the outcome depends on the persistent parameters that govern the distribution, the incentives to acquire information respond to owners' idiosyncratic beliefs, creating a dynamic learning problem.

The theory of RI has proven consistent with the empirical behavior of firms while providing a mechanism able to reconcile both individual and aggregate decisions. Using novel data on firms' expectations in New Zealand, Coibion, Gorodnichenko and Kumar (2018) argues that firms' behavior towards acquiring information is in line with the predictions of RI. Matějka (2015) stressed how RI is consistent with discrete pricing (a feature that resembles price stickiness) while generating the negative hazard rate for price changes as suggested by the data. Maćkowiak and Wiederholt (2009) shows how the sluggish reaction of prices to aggregate shocks is consistent with price setters deciding how to optimally split their limited attention between idiosyncratic or aggregate conditions. This paper supports the relevance of studying the dynamic implications of imperfect information by showing how this friction *alone* can simultaneously explain several features of the data at the cross-sectional and time series levels.

The model endogenously generates a positive response of price-change dispersion to volatility shocks along with a positive comovement between the dispersion and frequency of price changes.[1] These two features are in line with recent empirical evidence on price setting at the micro-level. Bachmann, Born, Elstner and Grimme (2019), Drenik and Perez (2018) and Klepacz (2017) document the existence of a positive correlation between volatility shocks and price-change dispersion. Moreover, the existence of a positive correlation between price-change dispersion (intensive margin) and the frequency of price changes (extensive margin) was shown by Vavra (2013). As owners are active learners, the results are also consistent with the presence of time-varying attention as documented by Coibion and Gorodnichenko (2015).

Solving a model in which information acquisition is dynamic and fully flexible imposes several methodological challenges. These challenges arise precisely because of its flexible structure. Acquired information has an effect on both pricing decisions and posterior beliefs about the next period's distribution. To allow for a dynamic setting, a common assumption in the RI literature is to assume a Gaussian distribution for the shock process. Typically this distribution is known with certainty. This assumption combined with a quadratic loss function leads to a closed form for the optimal signal structure, given by the outcome realization plus normally distributed noise as in Woodford (2003) and Maćkowiak, Matějka and Wiederholt (2018). In this paper, however, the optimal structure of signals depends on how firms choose to attach different probabilities to each possible shocks distribution. As information acquired in the past guides their decisions, the challenge is how to characterize the effects of flexible current information on posterior beliefs. We circumvent this problem by building on the solution proposed by Steiner, Stewart and Matějka (2017). Furthermore, we provide an algorithm tailored to solve this dynamic learning model.

The model is simulated and calibrated to replicate several stylized facts of price changes. To the best of our knowledge, the extent by which a fully flexible dynamic RI model can match the time series features of the data along with the evidence at the intensive margin, has not been previously addressed in the literature. To match moments, we assume a parametric distribution for information costs. The magnitude of cost dispersion across firms is meaningful as it is roughly half of the average cost. Since the cost of information is one of the critical parameters in the RI literature, the results are informative as they shed light on the degree of dispersion of this rigidity across firms.

Imperfect information about the persistent parameters that govern the distribution endogenously generates persistence in beliefs. While the economy can stochastically evolve across different volatility states, with different distributions for the shock process, costly information

[1] We will make the distinction between "dispersion" and "volatility." In this context, dispersion refers to the spread (typically measured as the standard deviation) of endogenous variables for the cross section of firms. Meanwhile, volatility refers to the spread of exogenous shocks.

affects the probability of even noticing a state change. A higher cost of information (relative to the assumed cost distribution), lead firms to optimally design information and pricing strategies that do not waste further attention on recognizing any possible distribution change. Hence, their pricing decisions are based on an aggregate state, which is *perceived* as being absorbing. However, access to cheaper information allows firms to recognize a distribution change via their optimal strategies. Although their reaction to the new state can also be sluggish, after they start attaching more probability to the new possible distribution, they start modifying their information and pricing decisions accordingly. Therefore as the economy exogenously moves across different states, firms with heterogeneous beliefs coexist. This is the main mechanism that allows the model to replicate some of the key features of the data after aggregating all pricing decisions.

While imperfect information is the only friction, the model is still consistent with the fact that individual prices stay constant for some time (extensive margin). As argued by Matějka (2015) and Jung, Kim, Matejka, Sims et al. (2019), a Rationally Inattentive agent chooses to price discretely when the processes for the fundamentals are not Gaussian. Uncertainty about the correct distribution is modeled by assuming a mixture of normal distributions for the optimal price. Because of this assumption, agents will not change their prices in every period, creating price stickiness. The simulated duration is, however, shorter compared to the data. Adding further frictions, such as price rigidities within the described dynamic learning structure, emerges as a natural extension of this paper. The combination of menu-costs with heterogeneous persistent beliefs would presumably amplify the documented effects on price dispersion as the economy moves across different volatility states.

The ability of the model to replicate the aforementioned evidence at the time series level crucially depends on the simultaneous presence of a dynamic, flexible information framework, with time-invariant heterogeneous costs. The model then nests two previously studied settings in the RI literature. With full information about the shock distribution, the model becomes static with a Gaussian unobserved target-price. This setting resembles the one presented by Woodford (2003) and Maćkowiak and Wiederholt (2009). A dynamic model with homogeneous information costs was also analyzed by Matějka (2015). We also propose an alternative version where all firms collect information of the same optimal price. We show how each assumption on its own is not enough to simultaneously replicate the dynamic relationships suggested by the data.

The paper contributes to the price-setting literature with information frictions. Alvarez, Lippi and Paciello (2011) solves a price-setting problem with observation and menu costs. The authors show how these two costs complement each other, delivering different implications for the timing of price reviews. Gorodnichenko (2008) solves a model with information frictions and menu costs. Moscarini (2004) introduces a pricing problem with limited information, where agents are restricted to receive new information infrequently creating inertia in their behavior.

Woodford (2009) introduces a setting with menu-costs, where the decisions to conduct a price review is made under RI. Vavra (2013) studies the dynamic behavior of price setting and argues how a menu-cost model, with time-variant idiosyncratic shocks, can match the dynamic features of prices. While all these papers rely on the crucial role of price rigidities, this paper aims to highlight the role of information rigidities as a key driver behind aggregate decisions.[2] Baley and Blanco (2018) studies dynamic pricing with menu-costs and information rigidities. In their set up, the timing of volatility shocks is known with certainty, which is precisely the primary assumption this paper aims to relax.

RI models have proven useful when rationalizing the empirical behavior of micro prices along with their aggregate implications. Maćkowiak and Wiederholt (2009) proposes a pricing model with endogenous attention to explain the sluggish response of prices to aggregate shocks. Matějka (2015) alternatively introduce a model that does not rely on quadratic objectives nor Gaussian distributions, as in Maćkowiak and Wiederholt (2009), which endogenously generates price discreteness. Afrouzi (2018) solves a dynamic general equilibrium model with inattentive price setters, Gaussian signals, and strategic complementarities between them. Paciello and Wiederholt (2013) shows how under costly information, monetary policy can reduce inefficient price dispersion by affecting the response of profit-maximizing prices to unobserved markup shocks. Finally, Stevens (2019) presents a price-setting model with constrained information which can capture the heterogeneous patterns of adjustments observed in the data, along with the sluggish response of prices to shocks. This paper contributes to this literature by studying the unexplored ability of these models to match the aggregate implications of the two price margins while being consistent with the evidence at the micro-level.

The rest of the paper is structured as follows. In Section 2, we introduce the model set up and discuss the dynamic costly information setting. We then fully derive and characterize the solution to the problem. Section 3 presents the algorithm used to replicate both cross-sectional and time-series moments from the data. The main results of the paper are discussed in Section 4, where we lay out both individual and aggregate implications under persistent volatility shocks. Section 5 introduces some alternative specification for the model. Finally, Section 6 concludes.

[2] The results can also contribute to the discussion about optimal policy within a price setting framework. Paciello and Wiederholt (2013) shows how replacing a price rigidity with an imperfect information mechanism is not innocuous for policy counterfactuals. While an optimal policy argument is beyond the scope of this paper, the results certainly lead the discussion towards that direction.

2 The dynamic learning pricing model

2.1 Set up

The setting is a partial equilibrium model where time is discrete $t \geq 0$ and there are a fixed number of firms $i = 1, \ldots, N$. Firm owners choose prices p_{it} from a finite set Ω_p to maximize the present discounted value of profits. Each firm can adjust its price costlessly in every period so p_{it} is set to maximize current profits $\widehat{\Pi}(p_{it}, \widehat{p}_{it})$. Following Caplin and Leahy (1997) and Alvarez et al. (2011), the profit function is set equal to:

$$\widehat{\Pi}(p_{it}, \widehat{p}_{it}) = \gamma(p_{it} - \widehat{p}_{it})^2 \tag{1}$$

The objective function (1) can be interpreted as a second-order approximation of a more general profit function around its non-stochastic steady state. The details behind the approximation are presented in Appendix 7.1. The parameter γ represents the curvature of the demand function and $\widehat{p}_{it}$ is labelled as the idiosyncratic "price-target". Given the approximation, $\widehat{p}_{it}$ is a function of firms' marginal costs. In the model, owners do not have complete information about cost conditions as they cannot fully track the shocks affecting their production due to their own limitations in processing information.[3]

Imperfect information about the current distribution of $\widehat{p}_{it}$ is modeled in the following way. There are two independent shocks drawn in each period t from finite sets, $\sigma_t \in \Omega_\sigma$ and $\epsilon_{it} \in \Omega_\epsilon$. The price-target is assumed equal to $\widehat{p}_{it} = \sigma_t \epsilon_{it}$. Underlying the shocks evolution is a probability distribution induced by a Markov Chain on Ω_σ and a discretized Gaussian on Ω_ϵ, with mean zero and unit variance. Thus, while the former shock is persistent, the latter is i.i.d. The stochastic process of both shocks is common information across firms.[4] We assume $\Omega_\sigma := \{\sigma_L, \sigma_H\} \subseteq \mathbb{R}_+$, with $\sigma_H = \phi\sigma_L$, $\phi > 1$. The transition probabilities of switching from the σ_L to the σ_H state, and viceversa, are labelled as τ_{LH} and τ_{HL} respectively.

The two components of the target-price are aimed at capturing idiosyncratic and aggregate uncertainty across firms. Since neither of the two shocks is fully observed, firms are not only uncertain about the realization of $\widehat{p}_{it}$, they also do not know if the price was drawn from

[3] Bachmann and Moscarini (2011) argues how different cost variables (such as input price elasticities or costs structures) are hard to estimate by firms. Think about owners who want to maximize profits but have multiple demands on their time such as reading reports about the firm's inventory levels, projecting future sales, testing and developing new products, collecting information about clients' reactions to historical prices, among others. Information is imperfect in this case, as owners cannot possibly precisely remember all the information when setting a price.

[4] Based on the second order approximation, the target-price is equivalent to $log(P^*_{it})$, where P^*_{it} is a constant mark-up over time-varying idiosyncratic marginal costs, see Appendix 7.1. Hence, negative values of $\widehat{p}_{it}$ are consistent with $P^*_{it} \in (0, 1)$. Moreover, the assumed stationarity of the target-price can be understood as deviations from a (known) deterministic trend.

$N(0,\sigma_L^2)$ or $N(0,\sigma_H^2)$. Therefore we will now refer interchangeably to persistent volatility states and different distributions for the target-price.

2.2 Information Acquisition

To optimally set prices firms collect information about $\widehat{p}_{it}$ to minimize (1). As the predictability of $\widehat{p}_{it}$ is "state-dependent", owner's beliefs about its current distribution will discipline their efforts to collect information. The learning process is dynamic in the sense that previously acquired information is still relevant for current decisions due to the persistence of volatility states.

Firms enter each period with prior beliefs $g_{it}(\widehat{p}_{it}) = m_{it}(\sigma_t)h(\epsilon) \in \Delta(\Omega_{\widehat{p}})$ where $\Omega_{\widehat{p}} := \Omega_\sigma \times \Omega_\epsilon$. Hence, $\Delta(\Omega_{\widehat{p}})$ is the set of all probability distributions on $\Omega_{\widehat{p}}$. In the definition, $m_{it}(\sigma_t)$ and $h(\epsilon)$ are the prior probability measures of σ_t and ϵ_{it} respectively. Since the probability of $\epsilon \in \Omega_\epsilon$ is i.i.d. and its stochastic process is known, its prior probability is constant across firms.

Owners acquire information about $\widehat{p}_{it}$ by choosing a signal $s_{it} \in \Omega_s$, where $|\Omega_p| \leq |\Omega_s|$. Firms are rationally inattentive since through costly information they aim to reduce the entropy of their beliefs, Sims (2003). Entropy about $\widehat{p}_{it}$ is defined as $\mathcal{H}(\widehat{p}_{it}|\mathcal{S}_i^{t-1}) \equiv E[-log(\widehat{p}_{it})|\mathcal{S}_i^{t-1}]$, where $\mathcal{S}_i^{t-1} = \{s_{it-1}, s_{it-2}, \ldots, s_{i0}\}$. Then $\mathcal{S}_i^{t-1}$ is the information set generated by the history of signals from firm i up to $t-1$. Prior uncertainty about $\widehat{p}_{it}$ is then $\mathcal{H}(\widehat{p}_{it}|\mathcal{S}_i^{t-1}) = -\sum_\sigma \sum_\epsilon g_{it}(\widehat{p}_{it}|\mathcal{S}_i^{t-1}) log(g_{it}(\widehat{p}_{it}|\mathcal{S}_i^{t-1}))$, where sums are taken across all possible values of σ and ϵ in their sets.

In line with RI models, the reduction in uncertainty is quantified by Shannon (1948)'s measure of mutual information flow:

$$\mathcal{I}(\widehat{p}_{it}, s_{it}|\mathcal{S}_i^{t-1}) \equiv \mathcal{H}(\widehat{p}_{it}|\mathcal{S}_i^{t-1}) - E_{s_{it}}[\mathcal{H}(\widehat{p}_{it}|s_{it})|\mathcal{S}_i^{t-1}] \tag{2}$$

Information flow (2) is the difference between prior and posterior uncertainty about $\widehat{p}_{it}$, conditioned on lagged information.[5] Due to the Markov structure, all the relevant historical information is summarized by the lagged value of the signal, $\mathcal{S}_i^{t-1} = \{s_{it-1}\}$. Thus, as long as owners observe neither current nor lagged $\widehat{p}_{it}$ it is possible to assume perfect information about further historical outcomes without compromising the model's implications.

During each period t, owners choose an "information strategy" $f_{it}(s_{it}, \widehat{p}_{it}|s_{it-1}) \in \Delta_{g_{it}}(\Omega_s \times \Omega_{\widehat{p}})$ and a "pricing strategy" $p_{it} : \Delta(\Omega_{\widehat{p}}) \rightarrow \Omega_p$. Information acquisition is then summarized by the joint probability distribution of signals and optimal prices where $\Delta_{g_{it}}(\Omega_s \times \Omega_{\widehat{p}})$ is the set of all probability distributions on $\Omega_s \times \Omega_{\widehat{p}}$, consistent with prior beliefs $g_{it}(\widehat{p}_{it})$. After the

[5] As described, the entropy formula relies on logarithms which depending on the base, changes the units by which we measure information. If the log is base two, then the information is measured in bits, while if it is e it is measured in nats.

price-target is drawn, the choice of $f_{it}(s_{it}, \widehat{p}_{it}|s_{it-1})$ reflects the type of acquired signal based on the information simplification process chosen by each owner.

The expression for the mutual information in (2) can be written as a function of $f_{it}(s_{it}, \widehat{p}_{it}|s_{it-1})$.

Proposition 1 ***: Mutual Information Equivalence***

Shannon's mutual information (2) is equal to:

$$\mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1}) = \sum_{s}\sum_{\sigma}\sum_{\epsilon} f(s_{it}, \widehat{p}_{it}|s_{it-1}) log\left(\frac{f(s_{it}, \widehat{p}_{it}|s_{t-1})}{g(\widehat{p}_{it}|s_{t-1})f(s_{it}|s_{it-1})}\right) \tag{3}$$

Proof in Appendix 7.2.

By setting an information strategy $f(s_t, \widehat{p}_{it}|s_{t-1})$, owners are choosing the total amount of information $\mathcal{I}(\widehat{p}_{it}, s_t|s_{t-1})$ to be acquired during each period.

2.3 The problem in two stages

Let us discuss the timing of the model. Within each period owners face two decisions: given prior beliefs $g_{it}(\widehat{p}_{it}|s_{it-1})$ they choose $f_{it}(s_{it}, \widehat{p}_{it}|s_{it-1})$ and then, endowed with this new information, they set prices p^*_{it}. Owners are Bayesian as by combining posterior beliefs about σ_t with τ_{LH} and τ_{HL}, they form prior beliefs for the next period $g_{it+1}(\widehat{p}_{it+1}|s_{it}) = m_{it+1}(\sigma_{t+1}|s_{it})h(\epsilon)$.

The pricing strategy describes how owners react to the received signal s_{it} by mapping posterior beliefs $f(\widehat{p}_{it}|s_{it}, s_{it-1}) \in \Delta(\Omega_{\widehat{p}})$ to optimal prices $p^*_{it}(s_{it}|s_{it-1})$.

$$p^*_{it}(s_{it}|s_{it-1}) = \arg\max_{p_{it}} \sum_{\sigma}\sum_{\epsilon} \widehat{\Pi}(p_{it}, \widehat{p}_{it}) f_{it}(\widehat{p}_{it}|s_{it}, s_{it-1}) \tag{4}$$

At the information acquisition stage, owners face a trade-off. Signals with higher precision allow them to observe $\widehat{p}_{it}$ with less noise, where the precision is determined by the channel's capacity (3). While owners can constantly modify the capacity, the cost of each additional unit of information is given by $\lambda_i > 0$. The cost directly affects the profit function, and it is assumed to differ across firms.

Information is fully flexible as firms set the precision of their signals by choosing $f(\widehat{p}_{it}, s_{it}|s_{it-1})$ without adding further parametric assumption about its particular shape. The chosen form for the joint probability distribution determines total acquired information for each moment of time. Moreover, as states are unobserved, each information strategy ultimately depend on owner's *perceived* prior distribution for $\widehat{p}_{it}$:

$$\widehat{p}_{it} \sim m_{it}(\sigma_L|s_{it-1})N(0, \sigma^2_L) + (1 - m_{it}(\sigma_L|s_{it-1}))N(0, \sigma^2_H)$$

Where $m_{it}(\sigma_L|s_{it-1})$ is the prior probability attached to the low volatility state of firm i, given information acquired in the past.

At the first stage, given the policy function $p^*_{it}(s_{it}|s_{it-1})$, $\mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1}) \geq 0$ and $g_{it}(\widehat{p}_{it}|s_{it-1})$, firms choose the conditional distribution $f(s_{it}|\widehat{p}_{it}, s_{it-1})$ to maximize expected profits relative to the cost of information:

$$f(s_{it}|\widehat{p}_{it}, s_{it-1}) = \arg \max_{\widehat{f}(.)\in\Delta_g(\Omega_s)} \sum_s \sum_\sigma \sum_\epsilon \widehat{\Pi}(p^*_{it}, \widehat{p}_{it})\widehat{f}(s_{it}|\widehat{p}_{it}, s_{it-1})g(\widehat{p}_{it}|s_{it-1}) - \lambda_i \mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1})$$

The information strategy shapes the posterior distribution of signals, which is equivalent to choosing $f(\widehat{p}_{it}, s_{it}|s_{it-1})$.

As the only purpose of costly information is to inform pricing decisions, through signals the firm is *implicitly* and optimally choosing its optimal price by determining $f(\widehat{p}_{it}|s_{it}, s_{it-1})$. Therefore it is enough to solve for the optimal distribution of prices conditional on the realization of the target-price. Matejka and McKay (2014) and Matějka (2015) formally show this result for static RI problems, while Steiner et al. (2017) prove it within a dynamic setting with flexible information. Intuitively, each signal $s_{it} \in \Omega_s$ will be associated with just one price $p_{it} \in \Omega_p$. If two signals lead to the same price, and since entropy is a concave function, the firm could ended up setting the same price with a lower information cost.[6]

2.4 The dynamic RI problem

Let us now formally introduce the dynamic information acquisition problem. While the price-setting decision is static, the unobserved and persistent distribution of $\widehat{p}_{it}$ implies a correlation between consecutive periods. As the precision by which owners try to uncover the underlying state of the economy is subject to their choice, prior beliefs $m_{it}(\sigma_{jt}|p_{it-1}), j = L, H$ become the state variable of the problem.

During each period t, given $g_{it}(\widehat{p}_{it}|p_{it-1}) \in \Delta(\Omega_{\widehat{p}})$ and information costs $\lambda_i > 0$, owners choose $f_{it}(p_{it}, \widehat{p}_{it}|p_{it-1}) \in \Delta_g(\Omega_p \times \Omega_{\widehat{p}})$ to solve the dynamic problem:

$$V(m_{it}(\sigma_L|p_{it-1})) = \max_{f_{it}(p,\widehat{p}|p_{it-1})} \sum_\sigma \sum_\epsilon \sum_p [\widehat{\Pi}(p_{it}, \widehat{p}_{it}) + \beta V(m_{it+1}(\sigma_L|p_{it}))] f_{it}(p_{it}, \widehat{p}_{it}|p_{it-1})$$
$$- \lambda_i \mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1}) \tag{5}$$

[6]Moreover, since information is costly and $f(\widehat{p}_{it}, s_{it}|s_{it-1})$ is endogenous, necessarily $\mathcal{I}(\widehat{p}_{it}, p^*_{it}|s_{it-1}) \leq \mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1})$. The linearity of the cost function is relevant under a dynamic setting as it prevents the firm from stock unused information for future periods, Steiner et al. (2017).

Subject to:

$$\mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1}) \quad = \quad f_{it}(p, \widehat{p}|p_{it-1})log\left(\frac{f_{it}(p_{it}, \widehat{p}_{it}|p_{it-1})}{g_{it}(\widehat{p}_{it}|p_{it-1})f_{it}(p_{it}|p_{it-1})}\right) \tag{6}$$

$$g_{it}(\widehat{p}_{it}|p_{it-1}) = m_{it}(\sigma_t|p_{it-1})h(\epsilon) \quad = \quad \sum_p f_{it}(p, \widehat{p}_{it}|p_{it-1}) \tag{7}$$

$$m_{it+1}(\sigma_L|p_{it}) \quad = \quad \mathcal{T}_{t+1}(f_{it}(\sigma_L|p_{it})) \tag{8}$$

$$0 \quad \leq \quad f_{it}(p, \widehat{p}|p_{it-1}) \leq 1 \tag{9}$$

Owners maximize the expected value of $\widehat{\Pi}(p_t, \widehat{p}_{it})$ with respect to the perceived probability distribution of p_{it} and $\widehat{p}_{it}$ relative to the total information cost. The cost λ_i forces the firm to form a probabilistic conjecture of its optimal price given both the unobserved persistent and i.i.d. shocks. Since the space of prices and shocks is finite, the strategy space is compact. Therefore, from the continuity of the objective function, the RI problem has a solution.

The state variable in the value function (5) corresponds to the prior probability of the low volatility state. Equation (7) forces the chosen joint probability distribution to be consistent with owners' prior beliefs. Without this constraint, owners could "forget" relevant information acquired in the past. Equation (8) characterizes the belief updating process. In this equation, $\mathcal{T}_{t+1}$ represents the law of motion of σ_L based on the Markov switching probabilities, while its argument is the posterior probability of the distribution of $\widehat{p}_{it}$ having low volatility at time t.

The fully-flexible information scheme imposes a challenge on how to solve (5) as the shape of $f_{it}(p_{it}, \widehat{p}_{it}|p_{t-1})$ and its implications on $\mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1})$, has a non-linear effect on continuation values $V(m_{it+1}(\sigma_L|p_{it}))$. To tackle this issue, I build on the result proposed by Steiner et al. (2017).[7] The following system of non-linear equations characterizes the solution of (5) subject to equations (6) to (9).

[7]This paper argues that a dynamic RI problem consistent with (5) is equivalent to a control problem without uncertainty about $\widehat{p}_{it}$. Because of this equivalence, firm's continuation value are then a function of the *history* of prices and shocks, so the decision about the shape of the joint probability distribution does not affect $V(m_{it+1}(\sigma_L|p_{it}))$.

Proposition 2 ***: Solution of the dynamic RI problem***

$$m_{it}(\sigma_L|p_{it-1}) = (1-\tau_{LH})f_{it-1}(\sigma_L|p_{it-1}) + \tau_{HL}(1-f_{it-1}(\sigma_L|p_{it-1})) \tag{10}$$

$$f_{it}(p_{it}|\widehat{p}_{it}, p_{it-1}) = \frac{exp\left[\left(\Pi(p_{it},\widehat{p}_{it}) + \beta V(m_{it+1}(\sigma_L|p_{it}))\right)/\lambda_i\right] f_{it}(p_{it}|p_{it-1})}{\sum_{p'} \; exp\left[\left[\Pi(p'_{it},\widehat{p}_{it}) \; + \beta V(m_{it+1}(\sigma_L|p_{it}))\right)/\lambda_i\right] f_{it}(p'_{it}|p_{it-1})} \tag{11}$$

$$V(m_{it}(\sigma_L|p_{it})) = \lambda_i E\left[\sum_p exp\left[\left(\Pi(p_{it},\widehat{p}_{it}) + \beta V(m_{it+1}(\sigma_L|p_{it}))\right)/\lambda_i\right] f_{it}(p_{it}|p_{it-1})\right] \tag{12}$$

Proof in Appendix 7.3.

For any $\lambda_i > 0$, (10), (11) and (12) summarize the main equations that solves the problem. Equation (10) is the prior probability of being in the low volatility state as a function of the Markov transition probabilities and lagged acquired information. The expression then corresponds to the functional form of $\mathcal{T}_t$ in equation (8). The prior probability $m_{it}(\sigma_H|p_{it-1})$ is simply the complement of equation (10). These two probabilities are embedded into $g_{it}(\widehat{p}_{it}|p_{it-1})$ to force prior beliefs to be consistent with the joint probability distribution as stated in (7).

The form of the conditional probability $f_{it}(p_t|\widehat{p}_{it}, p_{it-1})$, i.e. the information strategy, is characterized in (11). The probability resembles the dynamic Logit formula except for the term $f_{it}(p_{it}|p_{it-1})$, which multiplies the cost-benefit ratio of choosing the price p_{it}. As $f_{it}(p_t|p_{it-1})$ is independent of realized shocks, it is interpreted as owner's "predisposition" to chose each $p_{it} \in \Omega_p$ without additional current information. Following Steiner et al. (2017)'s characterization, we interpret firms' predisposition as prices that are chosen with high probability (on average) across outcomes and states, i.e. $f(p_t|p_{t-1}) = E_{\widehat{p}_{it}}[f(p|\widehat{p}_{it}, p_{t-1})]$. The posterior probability $f_{it}(p_t|\widehat{p}_{it}, p_{it-1})$ is a function of λ_i as its magnitude determines the amount of information to process and, with this, the weight attached to prior probabilities. Pricing decisions are drawn from (11) reflecting the noisiness in signals, whilst being consistent with owners' idiosyncratic state-dependent beliefs. Equation (12) shows the expression for the continuation value. See Appendix 7.3 for the specific derivation of these last two expressions.

Due to imperfect information about both the outcome and its time-varying distribution, there is no specific closed form for the posterior probability $f_{it}(p_t|\widehat{p}_{it}, p_{it-1})$.[8] Moreover, as information cost non-linearly affects both the posterior probability and continuation values, it

[8]Starting from the same model but assuming that the underlying distribution of $\widehat{p}_{it}$ is known with certainty (i.e. a static framework), there would be a closed form expression for the posterior uncertainty. With a quadratic objective and Gaussian distributions, the model boils down to a Bayesian updating set up, where the posterior distribution of prices is equal to a weighted sum between prior beliefs and signals. Under RI, the weight attached to signals becomes the choice variable of the problem.

is difficult to anticipate how different values of λ_i would affect the information strategies. The model is then solved numerically.

3 Numerical solutions

3.1 The algorithm

Before numerically solving the model, we need to introduce further assumptions about the simplex of each variable. The computational intensity of RI models severely restricts this decision, Tutino (2013). Let $|\Omega_\epsilon| = 13$ and $|\Omega_p| = 25$ be the number of possible values that the idiosyncratic shock ϵ_{it} and prices p_{it}, can take respectively.[9] The different values for ϵ_{it} come from a linearly equally-spaced grid ranging from $-2\sigma_H$ to $2\sigma_H$. Since $g_{it}(\widehat{p}_{it}) = m_{it}(\sigma)h(\epsilon)$, the state variable is defined as the probability of being in the low state $m_{it}(\sigma_L) \in \Delta(\Omega_\sigma)$, where $\Delta(\Omega_\sigma)$ is the belief simplex. The dimension of the belief simplex is $|\Delta(\Omega_\sigma)| = 25$, where each point reflects distinct (equally spaced) values for the marginal probability of the economy being in the low volatility state.

The algorithm to solve the dynamic RI problem is as follows:

1. Fix a value for the idiosyncratic information acquisition cost, e.g. λ_1.

2. Given λ_1 and the belief simplex, compute prior beliefs $g(\widehat{p}_{it}) = m(\sigma_t)h(\epsilon_{it})$.

3. With $g(\widehat{p}_{it})$, the model is solved by Value Function Iteration.

   3.1. Starting with a guess for the vector $V(m_{t+1}(\sigma_L))$, we first solve the static RI problem. The algorithm computes $f(p_{it}, \widehat{p}_{it}|p_{it-1}) \in \Delta(\Omega_p \times \Omega_\sigma \times \Omega_\epsilon)$ which is the solution for the system of nonlinear equations (7), (11) and $f(p_t|p_{t-1}) = E_{p_t}[f(p|\eta_{it}, p_{it-1})]$.

   3.2. Given $f(p_{it}, \widehat{p}_{it}|p_{it-1})$, the conditional probability $f(\sigma|p_{it}, p_{it-1}) = \sum_\epsilon f(\sigma, \epsilon|p_{it}, p_{it-1})$ is computed for each $p_{it} \in \Omega_p$. Through (10), posterior beliefs become the prior beliefs for next period, which are used to update $V(m_{t+1}(\sigma_L))$.

   3.3. Relying on the definition for $V(m_{it}(\sigma_L|p_t))$ in (12), the algorithm iterates the value function until convergence when, within each iteration, it re-estimates $f(p_{it}, \widehat{p}_{it}|p_{it-1})$.

4. Repeat point 3 for all possible values in $\Delta(\Omega_\sigma)$, i.e. setting different priors $g(\widehat{p}_{it})$.

5. Repeat 2, 3, and 4 for all possible values for λ_i.

The setting of the model and the decision on the shape of the joint probability distribution resembles a filtering problem. The numerical discrepancies between filtering with discrete variables relative to continuous outcomes are not significant and depend on the nature of the approximation, Farmer (2016) and Farmer and Toda (2017).

[9] Hence, there are 650 possible combinations of results from the three random variables, $f(\sigma, \epsilon, p) = 2\times 13\times 25$.

3.2 Calibration

The parameters in the model are: the discount rate β, switching probabilities τ_{LH} and τ_{HL}, the price elasticity of demand η (which determines the curvature of demand γ), the volatility level in the two states $\sigma_L, \sigma_H = \phi\sigma_L$, and the cost of acquiring information $\{\lambda_i\}_{i=1}^N$. Each period is a month, so we set the discount factor equal to $\beta = 0.999$. The transition probabilities are assumed equal to $\tau_{LH} = 0.009$ and $\tau_{LH} = 0.0355$. These monthly transition probabilities imply a quarterly probability of 2.9% of switching from the low to the high volatility state and a 89% probability of remaining in the high volatility state. These numbers are roughly in line with Bloom, Floetotto, Jaimovich, Saporta-Eksten and Terry (2018)'s estimates for the transition across uncertainty regimes in the U.S. Following Nakamura and Steinsson (2008), the price elasticity of demand is assumed $\theta = 4$ (implying a 33% markup). Hence, $\gamma = -10$ as shown in Appendix 7.1.

The remaining parameters are calibrated to match different stylized facts of individual price changes, taken from microeconomic data.[10] In particular, all the stylized facts are taken from Vavra (2013). Using the algorithm described in 3.1, we simulate an economy with $N = 7,500$ firms and $T = 5,500$ periods. In the simulations, the economy evolves naturally across states and idiosyncratic shocks, where we rule out the first 500 periods. To set the heterogeneity, we assume there are 15 distinct values for λ, which are randomly and uniformly assigned across firms, i.e. $15 \times 500 = 7,500$. Without further evidence on the cost distribution, we assume $\lambda_i \sim N(\overline{\lambda}, \sigma_\lambda^2)$ with the distribution truncated at zero. Given $\overline{\lambda}$ and σ_λ, the different values of λ are given by the 15 equidistant percentiles (from 2.5 to 97.5) of this distribution. All these four parameters $\{\sigma_L, \phi, \overline{\lambda}, \sigma_\lambda^2\}$ are calibrated to replicate the data.

3.3 Matching moments

Table 1 show the moments chosen from the data and their simulated counterparts from the model. *Frequency* stands for the frequency of price reviews (fraction of prices that change) per month across firms. The kurtosis from the distribution of price changes is labeled as $Kurtosis(|\Delta p|)$. $E(|\Delta p|)$ is the average magnitude of a price change (in percentage points), and $E(Dispersion)$ is the average price-change dispersion across time. As these last three moments are computed conditioned on a price change occurring, i.e. $|\Delta p_{i,t}| \neq 0$, they aim to reflect the intensive margin of price changes. *Frequency*, on the other hand, captures the extensive margin of price adjustments. Besides these targeted moments, we assess the model's ability to replicate the data by focusing on additional (non-targeted) moments. *Fraction small* is the proportion of small price changes, where a small change $|\Delta p_{i,t}| < 0.5E(|\Delta p|)$. *Fraction*

[10] In line with this approach, Woodford (2009) also collects empirical facts from the literature to assess the ability of his model to replicate documented features.

up is the fraction of price changes that are increased, and $Corr(Dis, Freq)$ is the correlation between the price-change dispersion and the frequency of price changes.

The dynamic RI model can match these features of the data, except for the frequency of price changes. While prices remain constant in the model nearly half of the time, this is still not enough to fully replicate the empirical evidence. However, this result is still relevant as the model's ability to partially replicate the degree of micro price stickiness is attained assuming imperfect information as the unique friction.[11] Matějka (2015) showed that RI is consistent with price-setting evidence at the extensive margin. However, the intensive margin channel was muted in his analysis. The average absolute size of price changes was also calibrated by Maćkowiak and Wiederholt (2009), based on a RI model where prices change constantly. The results in Table 1 suggest that a dynamic version of a RI model with flexible information is consistent with the evidence at these two margins simultaneously. The model also matches the empirical comovement between price-change dispersion and frequency over time. In Section 4, we provide further intuition on the mechanism by which the model can rationalize this correlation.

Table 1: Matched Moments and Alternative Specifications

Targeted moments	Data	Baseline
$Frequency$	0.11	0.53
$Kurtosis(\|\Delta p\|)$	6.40	6.33
$E(\|\Delta p\|)$	0.077	0.079
$E(Dispersion)$	0.075	0.083
Non-Targeted moments		
$Fraction\ small$	0.33	0.088
$Fraction\ up$	0.65	0.499
$Corr(Dis, Freq)$	0.506	0.534

Notes: all moments are taken from Vavra (2013). Frequency is the fraction of prices that change per month, $Kurtosis(|\Delta p|)$ is the kurtosis coefficient of the distribution of absolute price changes, $E(|\Delta p|)$ is the average magnitude of non-zero price changes, and $E(Dispersion)$ is the average standard deviation of price changes over time conditioned on a price change occurring. In addition, $Fraction\ small$ is the percentage of small price changes, fraction up is the proportion of prices that change upwards and $Corr(Dis, Freq)$ is the time series correlation between the dispersion and the frequency of price changes.

Table 2 displays the calibrated parameters. The volatility of the price-target in the high state is double the volatility in the low state. The rise in volatility is in line with estimated uncertainty parameters during episodes of economic distress, Bloom et al. (2018). The magnitude of the standard deviation for the attention cost $\sigma_\lambda \approx 0.09$ is considerable as it is more than half of the average cost. While the calibrations rely on a parametric assumption about the distribution

[11] Targeting the frequency of price change is important not only to assess the possibility of replicating the positive correlation between frequency and price-change dispersion over time, but also as a relevant moment to pin down the average costs of information $\bar{\lambda}$ and its dispersion σ_λ.

of costs, they are informative as they provide a quantitative assessment about the potential dispersion of information rigidities across firms.

Table 2: Calibrated Parameters

Parameter	Value	Description
β	0.99	Discount Rate
γ	-6	Curvature of demand function
τ_{LH}	0.009	Monthly transition probability: low/high state
τ_{HL}	0.0355	Monthly transition probability: high/high state
σ_L	0.0742	Volatility in low state
ϕ	2.05	Increase in volatility in high state
$\overline{\lambda}$	0.17	Mean distribution information cost
σ_λ	0.09	Stdv distribution information cost

3.4 Information and pricing strategies

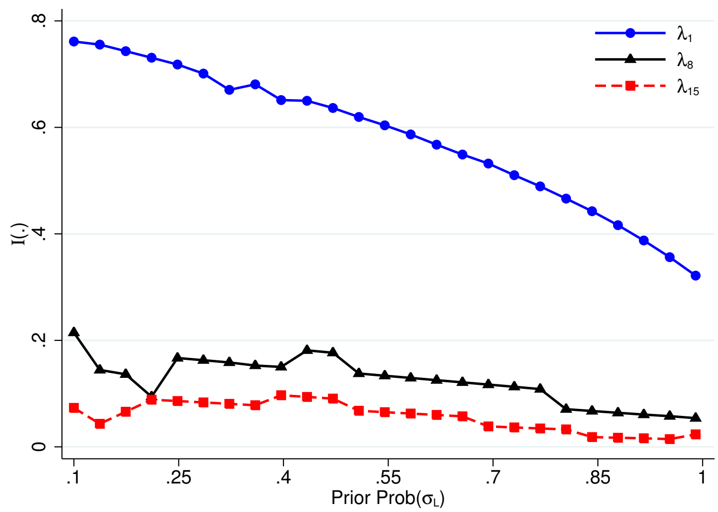
Given the calibrated parameters, let us describe the different implications of total acquired information on price setting as a function of their different cost levels.

Initially owners choose their information strategies, which will inform their pricing decisions. Uncertainty about the current distribution of $\widehat{p}_{it}$ implies that the decision made about total acquired information will depend on prior beliefs about each state. Due to the cost heterogeneity, we focus on three type of firms, specifically those facing low, medium, and high information costs. In particular, we focus on firms with $\lambda_1 < \lambda_8 < \lambda_{15}$ taken from the cost distribution. Figure 1 shows $\mathcal{I}(p_{it}, \widehat{p}_{it}|p_{it-1})$, as a function of distinct prior probabilities assigned to the low volatility state for these three type of firms. As the perceived predictability of the unobserved target increases, for owners it is now optimal to acquire less information. This behavior is consistent for all three types of firms. The speed by which total information increases, i.e., the learning rate, is however higher for the low-cost group. The cost heterogeneity creates non-trivial implications for firms' learning decisions, reflected in the relative distance between groups. While firms facing the lower costs choose to acquire more information, higher values of λ's affect firms' behavior to the extreme of making medium-cost firms behave similarly as the higher-cost group.

To minimize conditional variance (equation (4)) firms design their optimal information and pricing strategies conditioned on the costs they face. Information costs not only disciplines the search quality, they also force owners to form a probabilistic conjecture about the likelihood of extreme realization of the target, which affects their pricing decisions.

Let us focus on the simulated information strategies $f_{it}(p_{it}, \sigma_t, \epsilon_{it}|p_{it-1})$ for the same three cost levels λ_1, λ_8 and λ_{15}. For expository reasons, I compute strategies when price setters believe there is a 75%, 50% and 25% probability of being in the low volatility state. As the joint

Figure 1: Total Information



Notes: the figure presents total acquired information with respect to prior probabilities attached to the economy being in the low volatility state. The relationship is shown for three different information-costs values, $\lambda_1 < \lambda_8 < \lambda_{15}$.

probability distribution depends on three random variables, I calculate the "predisposition" $f_{it}(p_{it}|p_{it-1}) = \sum_{\sigma}\sum_{\epsilon} f_{it}(p_{it}, \sigma_t, \epsilon_t|p_{it-1})$, where the prior beliefs about σ_t are embedded in p_{it-1}. This is presented in the left panel of Figure 2. After acquiring new information about the realization of $\widehat{p}_{it}$, firms update the distribution to form the posterior $f_{it}(p_{it}|\widehat{p}_{it}, p_{it-1})$ from which the optimal price is finally drawn. This is shown in the right panel of the figure, assuming a given constant realization for $\widehat{p}_{it}$ drawn from the *low volatility* distribution. The outcome is represented by the vertical solid line.

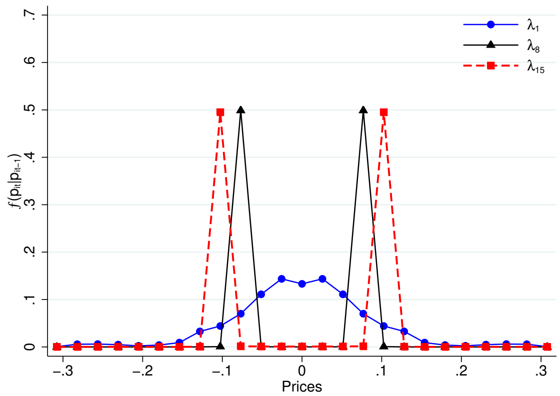
Costly information makes owners prone to choose using a smaller set of prices. Independently of prior beliefs, the predisposition $f_{it}(p_{it}|p_{it-1})$ degenerates into only two possible prices for λ_8 and λ_{15}. Following equation (11), the decision is to attach zero probability to choosing a price different from these two. The chosen magnitude of the two possible prices is, however, relevant as owners are trying to reduce their probability of making large mistakes on average to the utmost, while being consistent with their beliefs. A perceived higher probability of extreme realizations of $\widehat{p}_{it}$ is reflected in the decision of choosing between more extreme values. For λ_8, the owner slightly modifies her pricing strategy when the probability of being in the high volatility state is high enough (75%). Although she still allocates more than 60% probability to the two prices, she is also now choosing prices closer to the mean. The higher probability of being in σ_H pushes the owner to acquire further information (Figure 1), which is used to disentangle whether the realization of $\widehat{p}_{it}$ is close to the average or not. The optimal strategy is not to waste scarce attention in trying to uncover an extreme realization for the price which is, even if beliefs are correct, an event with very low probability.

Given the predispositions, the posterior probability for λ_8 and λ_{15} degenerates into the same two possible values. The relatively higher amount of information acquired by the middle-cost firm allows it to consistently set a price closer to the true outcome (grey vertical line) on average. This is clear by looking at the two middle panels of Figure 2. While the two firms start with the same predisposition, the frequency by which the λ_8 firm sets a price closer to the true value is higher due to its lower information cost. However, none of these firms would ever set $p_{it} = \widehat{p}_{it}$ as the posterior probability is zero at this point. When the uncertainty about the correct distribution is maximized, i.e., $Prob(\sigma_L) = 0.5$, both the magnitude of the errors and the price-change dispersion is amplified relative to the other cases. This is also a direct implication of the belief-driven endogenous attention. This feature of the posterior distribution is what allows the model to generate some degree of price stickiness.

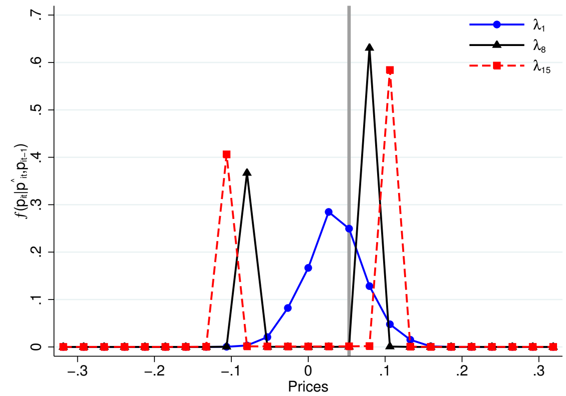
Cheaper information allows the owner to more precisely distinguish both the realization of the target-price and its underlying distribution. At the 75% probability the predisposition for λ_1 resembles a normal distribution with almost zero probability assigned to extreme realization of $\widehat{p}_{it}$. Instead of concentrating all the probability mass in a couple of prices, the firm allocates the prior probability in this way not to rule out any potential price ex-ante. The shape of the predisposition allows them to react immediately as new (highly precise) information arrives, distributing the posterior probability close to the true realization. Around 25% of time the λ_1 firm correctly set $p_{it} = \widehat{p}_{it}$ for all three cases. When the probability of σ_L decreases and firms collect more information, the distribution changes to attach more probability to extreme realizations, in exchange with reducing the probability of setting prices closer to the mean. Even if the firm wrongly attaches a high probability to the high volatility distribution, it is still able to closely track the optimal price.

Despite the different signal quality, any firm can have misperceived beliefs about the correct distribution. For example, if the economy currently switches to the low volatility state after being in the high state for several periods, even a highly informed owner (e.g., λ_1) may continue setting prices thinking that the economy is still in the less predictable state. It would be enough that the current realization for $\widehat{p}_{it}$ is close to the mean to prevent them from noticing any differences. The rate by which firms uncover a potential new state, which is linked to their information and pricing strategies, is what brings most of the interesting insights to the problem. In Appendix 7.4, we present further evidence on the importance of information gathering for firms despite their different cost levels.

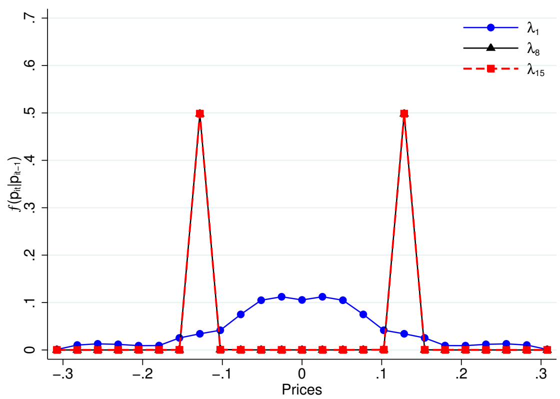
Figure 2: Firm's predisposition $f(p_{it}|p_{it-1})$ and conditional probability $f(p_{it}|\widehat{p}_{it}, p_{it-1})$



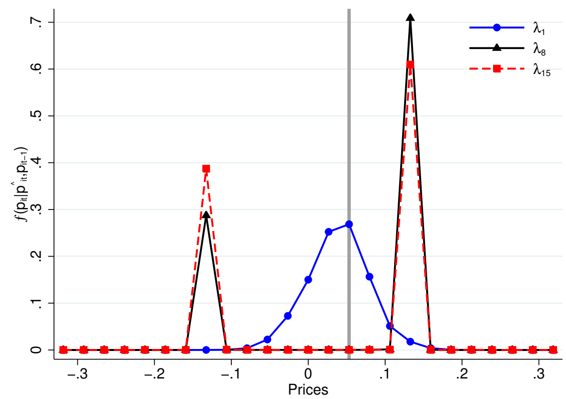
(a) Prior Probability $|\sigma_L = 75\%$



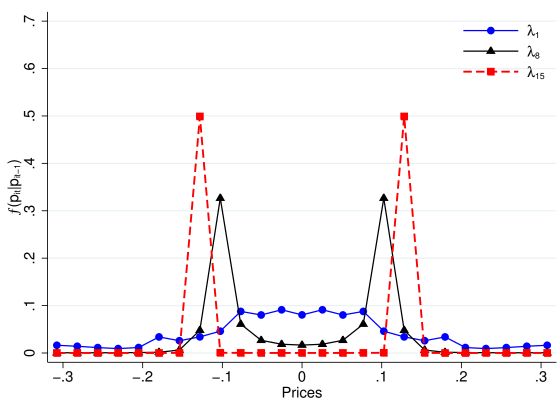
(b) Posterior Probability $|\sigma_L = 75\%$



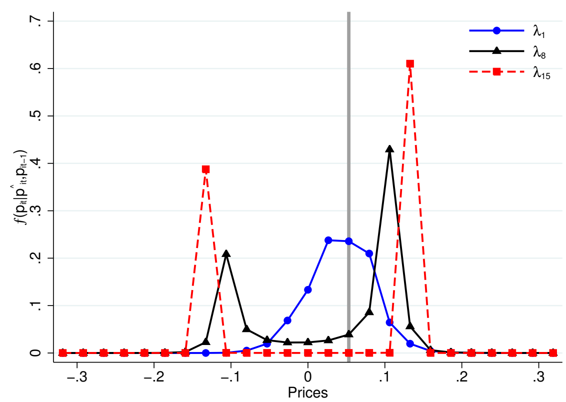
(c) Prior Probability $|\sigma_L = 50\%$



(d) Posterior Probability $|\sigma_L = 50\%$



(e) Prior Probability $|\sigma_L = 25\%$



(f) Posterior Probability $|\sigma_L = 25\%$

4 Delayed learning dynamics

In this section, we show the transition dynamics of the model by simulating an exogenous change of state. Initially, firms know with certainty the correct state at $T = 0$, after which they start updating their beliefs about the current state. We simulate two different paths, where the economy remains in the low(high) volatility state for 500 periods, and then it switches to the high(low) state at time T. Keeping the assumed transition of states constant, we simulate 100 economies with 1,500 firms where, as in the calibrations, we allocate the 15 different information costs uniformly across them. The results are then averages of the variables across economies at each point in time.

4.1 Firm Level Evolution

Initially we present the results at the firm level by focusing on the same three groups of firms λ_1, λ_8, and λ_{15}. Figure 3 shows the evolution of total acquired information, posterior beliefs about the high volatility state $f_{it}(\sigma_H|p_t, p_{it-1})$, and the average loss function $\widehat{\Pi}(p_{it}, \widehat{p}_{it}) - \lambda_i \mathcal{I}(\widehat{p}_{it}, p_{it}|p_{it-1})$ for the three types of firms after the state change (vertical dotted line). The left column shows the implications of switching from low to high while the right column presents the opposite.

When information is cheap, owners can more easily notice a state change and respond to the lower predictability of the target by acquiring more information. However, even when the cost is low, the impossibility to immediately notice the new distribution causes a sluggish reaction in the rate by which firms increase their information gathering. Imperfect information about persistent states endogenously generates persistence in beliefs in this setting. As volatility is expected to rise during economic recessions, the predictions are consistent with the presence of countercyclical attention, in line with the empirical results of Coibion and Gorodnichenko (2015).[12] Meanwhile, firms facing a higher cost of information does not react at all after the change of state. Although a non-absorbing Markov Chain governs the transitions, states are *perceived* as being absorbing for this group of firms due to their chosen information strategies.

The reasons behind the two different responses (sluggishness and perceived absorbing states) can be understood by studying the evolution of the posterior probability. While the firm with the lowest cost can notice the new distribution, it still needs approximately five months to start attaching more probability to being in σ_H (left-middle figure). Even though the cost is relatively small, it still prevents the firm from being entirely sure that the economy is in the high volatility state, as posterior beliefs do not go beyond 75%. In comparison the same firm needs only three months to realize the economy is back to the more predictable state (right-middle

[12]Based on the behavior of professional forecasters, the authors argue that the degree of information rigidities (a proxy for the total level of inattention) decreased during episodes of higher volatility in the U.S. They interpret this as an increase in the amount of collected information.

figure). As firms choose to collect more information when they start noticing the economy is in the high volatility state, it propitiates a faster reaction after a new state change. The model then generates asymmetric learning rates as a consequence of costly dynamic information.

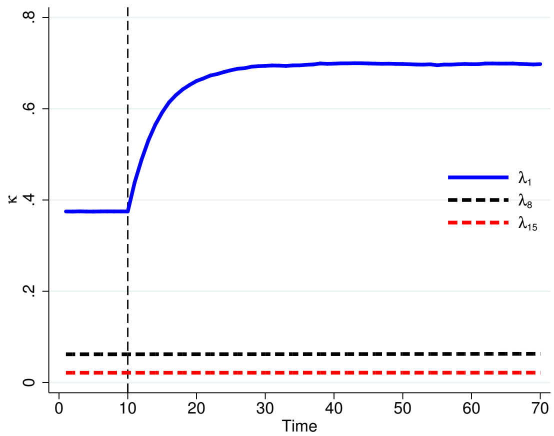
The reason why posterior beliefs for the other firms do not change is explained by the way owners designed their information strategies. As previously discussed (Figure 2), these firms endogenously choose not to waste attention on noticing any extreme realizations of the target-price. This decision completely rules out the possibility of noticing that the economy is in the high volatility state, delivering unresponsive pricing decisions. Certainty about the true initial state (at T = 0) followed by a prolonged episode of high volatility, is enough to leave this type of firms in an entirely uncertain situation, attaching an equal probability to each state (middle-right panel). Firms ended up choosing between two prices that are farther from the mean, relative to the case where they start in the low volatility state. This leads to a situation where their pricing decisions are irrelevant when discriminating between the two possible distributions.

Active learning within this dynamic setting generates disagreement about the correct underlying distribution for $\widehat{p}_{it}$, which finally affects both pricing and posterior information decisions. The empirical evidence also supports the model prediction about time-varying heterogeneous beliefs across firms, Kumar, Afrouzi, Coibion and Gorodnichenko (2015).[13]

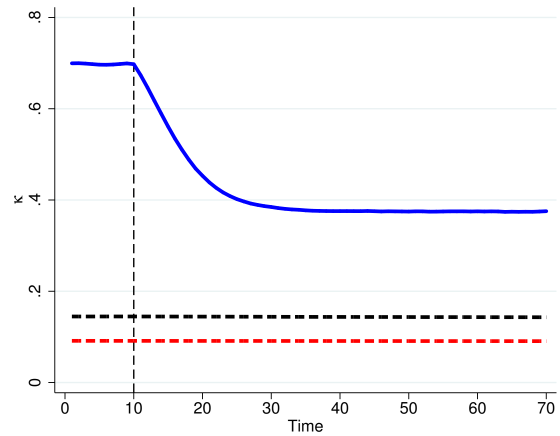
From the last row of Figure 3, we notice that although neither λ_8 nor λ_{15} firms notice any state change, the cost still matters as it leads to more precise signals which ultimately affect their losses.

[13] Although the paper documents the presence of time-varying beliefs about the inflation rate, we see this as a valid proxy for the beliefs about an aggregate price index such as $\widehat{p}_{it}$.

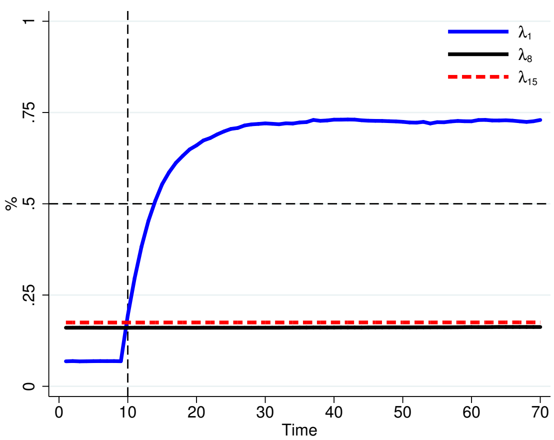
Figure 3: Firm's predisposition $f(p_{it}|p_{it-1})$ and conditional probability $f(p_{it}|\widehat{p}_{it}, p_{it-1})$



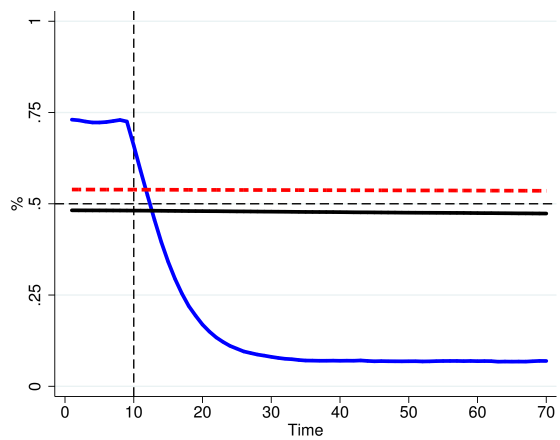
(a) Attention - Low → High



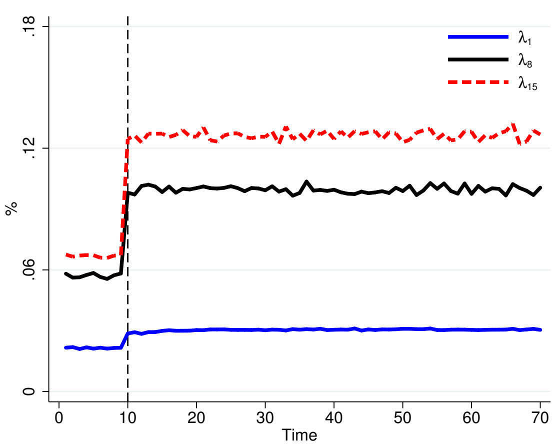
(b) Attention - High → Low



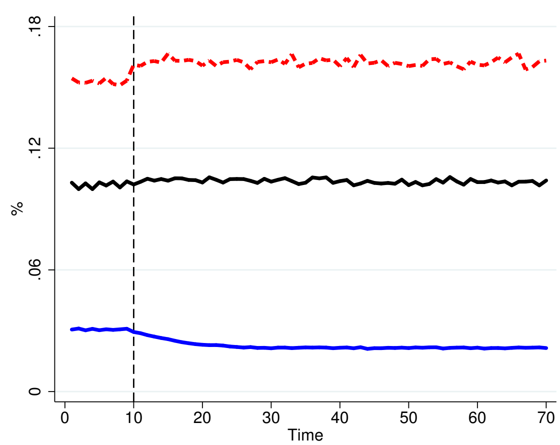
(c) Posterior $f(\sigma_H|p_{it}, p_{it-1})$ - Low → High



(d) Posterior $f(\sigma_H|p_{it}, p_{it-1})$ - High → Low



(e) Loss - Low → High



(f) Loss - Low → High

Notes: the vertical dotted black lines represent a change in the aggregate state. In the left panels, the economy moves from the low to the high volatility state. The right panels show the opposite transition. The two top figures present total acquired information $\mathcal{I}(\widehat{p}_{it}, p_{it}|p_{it-1})$. The middle figures show the evolution of the posterior probability of the economy being in the high volatility state while the bottom figures show the evolution of average loss.

4.2 Aggregate Evolution

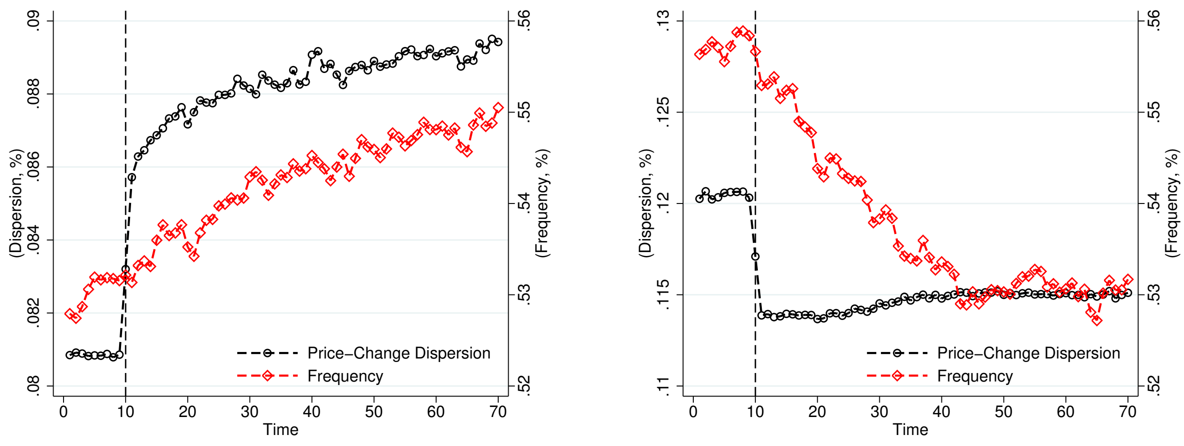
Let us now aggregate pricing decisions across firms. Figure 4 presents the evolution of the price-change dispersion (given by the standard deviation of Δp_{it}) and, in the secondary axis, the evolution of the frequency of price changes. The dynamic for both margins are computed following the same two simulated transitions. As the economy transits across states, both the price-change dispersion and the frequency of price changes commove positively, consistent with the data. Since recessions are episodes where aggregate volatility is expected to rise, the model can also rationalize the documented presence of counter-cyclical price change dispersion. While around half of the time prices are not adjusted in the model, we get an average increase of 4.1% during the high volatility state for the extensive margin, and a reduction of 5.2% for the low one.

While the response is sluggish and monotonically increasing when the economy enters into the less predictable state, the opposite transition brings a more immediate and dampened response for both series. This is an implication of the asymmetric efforts to collect information through the transition, which lead owners to disagree about the current and common distribution of the price-target. While a growing proportion of firms are noticing any new state, the rest will keep setting prices as if the state has remained constant, affecting the overall price-change dispersion. More informed firms also adjust their prices more or less frequently depending on the perceived predictability of the state. The reaction when the economy is back to the more predictable state is then faster as owners were already receiving more precise signals.

With full information, the volatility of $\widehat{p}_{it}$ is $\sigma_L \approx 0.07$ and $\sigma_H \approx 0.15$ (Table 2). As most owners set prices while perceiving each state as being absorbent, the rise of price dispersion is abated by the presence of Rational Inattentive owners during the low-high transition. Thus costly information contributes to price stability in this case. However, in the second case, high information costs distort relative prices preventing dispersion from decreasing. In this case, actual dispersion is 65% greater relative to the full information scenario. This is relevant due to the empirical evidence suggesting that the source of countercyclical price dispersion is mostly driven by agents responsiveness rather than higher volatility of exogenous shocks, Berger and Vavra (2019).

Through the results, we can directly map the implications of the proposed information-driven mechanism for price stability. This can have significant consequences for the policy design. In particular, the scope by which policies can effectively reduce price instabilities can be very different, depending on agents' beliefs about the overall state of the economy. In this case, by initially noticing that the high volatility state is the current state (although its unconditional probability is significantly lower), it is enough to force a situation where prices are persistently less stable compared to the actual state of the economy.

Figure 4: Aggregate Evolution



(a) Transition: Low → High

(b) Transition: High → Low

Notes: the figure presents the time series evolution of price-change dispersion and the frequency of price changes (secondary axis). The dotted vertical lines show the moment when the economy switches to the low or high volatility state respectively.

5 Alternative specifications

The three main ingredients of the model are dynamic information, heterogeneous information costs, and idiosyncratic shocks to $\widehat{p}_{it}$. In this section, we discuss the model's ability to replicate the aforementioned stylized facts after shutting down each of these channels in turn.

5.1 A Static version

The model has a static counterpart version which shares the same structure for $\widehat{p}_{it} = \sigma_t \epsilon_{it}$, but with full information about the current distribution. In this version σ_t is always known with certainty, meaning that firms acquire costly information to only track the realizations of $\widehat{p}_{it}$. The problem becomes a standard static RI problem with a quadratic objective and Gaussian signals. As discussed, in this context the optimal signal takes the parametric form $s_{it} = \widehat{p}_{it} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. Again, the precision of the signal σ_ε^{-2} is chosen by each firm given their information costs λ_i. Given the structure of the signal s_{it}, we expect full price flexibility in this case. With all firms changing their prices in every period, it is simply not possible for the model to replicate the correlation between the two price margins documented in the data.

5.2 Homogeneous information costs

Alternatively, we can study the role that heterogeneous costs play in the dynamic learning model. We will then solve an alternative version of the baseline model, with the only difference that now all firms are ex-ante identical in terms of their costs λ, i.e., $\sigma_\lambda = 0$.

5.3 Common price-target

The final alternative specification rules out the presence of idiosyncratic shocks, i.e. $\widehat{p}_{it} = \widehat{p}_t = \sigma_t \epsilon_t$. While keeping heterogeneity in costs, now all firms collect information about both the distribution and the realization of the *same* target-price. This assumption creates additional incentives for owners to anticipate what other firms are doing. This feature severely complicates the model as now owners should form higher-order beliefs within a setting where information is fully flexible. To rule out this possibility, and for keeping tractability in this case, we assume that owners operate in segmented markets, Gorodnichenko (2008). All the remaining assumptions are the same.

5.4 Implications for price change dispersion

To provide a fair comparison, we re-calibrate the parameters of these alternative specifications (except for the static version) to match the same targeted moments as in the original model. The results are presented in Table 3.

Table 3: Matched Moments and Alternative Specifications

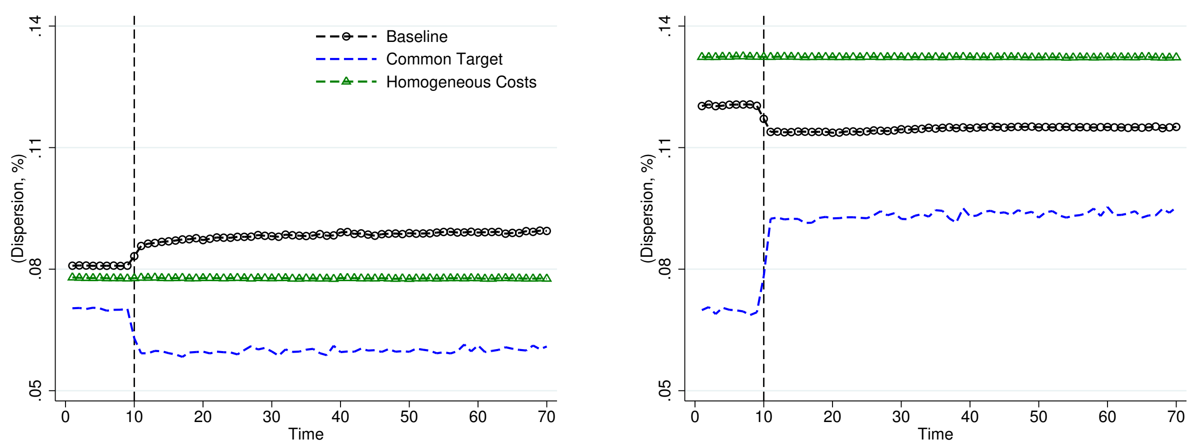
Targeted moments	Data	Baseline	$\lambda_i = \overline{\lambda}$	$\widehat{p}_{it} = \widehat{p}_t$
$Frequency$	0.11	0.53	0.504	0.53
$Kurtosis(\lvert\Delta p\rvert)$	6.40	6.33	5.903	6.23
$E(\lvert\Delta p\rvert)$	0.077	0.079	0.076	0.079
$E(Dispersion)$	0.075	0.083	0.077	0.070
Non-Targeted moments				
$Fraction\ small$	0.33	0.088	0.012	0.092
$Fraction\ up$	0.65	0.499	0.499	0.499
$Corr(Dis, Freq)$	0.506	0.534	-0.0278	-0.638

Overall, the alternative models are all successful in matching non-trivial moments from microeconomic data on price setting. Again, we do this without adding any further rigidities to this decision. However, *all* of these alternative versions fail to replicate the main two features of the time series of price changes: the positive correlation between price-change dispersion and frequency along with the positive reaction of dispersion after a volatility shock.

Relative to the correlation (last row of Table 3), the impossibility of generating this result was not obvious ex-ante. Firms with homogeneous information costs may still set different prices since they are drawn from posterior beliefs. Concerning the version with common $\widehat{p}_t$, the fact that firms still rely on different information costs leads to different information strategies. However, neither of these two cases can replicate this feature of the data.

Regarding the effect of a volatility shock, Figure 5 shows the evolution of price-change dispersion over the same two transitions discussed in 4.2. Except for the baseline case, the reaction of the price-change dispersion is not consistent with the expected effects of a volatility shock. Even after the recalibration, each state is perceived as absorbing for the homogeneous costs version. Hence, there is no reaction after either a positive or a negative volatility shock. When the price-target is common across firms, the dispersion reaction goes in the opposite direction relative to what is suggested by the data. Owners facing cheaper information reacts to the lower (greater) predictability of the target by acquiring more (less) information, with this changing their price updating patterns. As the remaining firms do not notice the state change, they do not modify their efforts to track the common optimal price. Hence, the overall price dispersion decreases after a positive shock and increases after a negative one. For completeness, the evolution of the extensive margin of price changes for the same three versions of the model is presented in Appendix 7.5.

Figure 5: Price-Change Dispersion - Model comparison



(a) Transition: Low → High

(b) Transition: High → Low

Notes: the figure presents the evolution of price-change dispersion when the economy moves from the low(high) volatility state to the high (low) state. The evolution is computed for the baseline model along with the two alternative specifications: homogeneous costs and the common price-target.

6 Conclusions

This paper addresses price-setting decisions under dynamic imperfect information acquisition. Rational Inattentive price setters collect information about an unobserved target-price before setting prices. Costly information serves two purposes: helping to determine the realization of the variable along with shedding light about the distribution that generated the target. The unobserved distribution is time varying to allow for different states of the economy where aggregate volatility can rise. Information is dynamic and fully flexible as owners choose the amount of information to acquire as well as how they want to learn about outcomes. This mechanism generates persistence in beliefs which is crucial to match distinct features of the data.

While imperfect information is enough to match the dynamic features of the data and can generate price stickiness, it is still not enough to fully match the fact that prices remain constant for several periods. A natural extension of this paper is to move closer to a setting that combines price rigidities with dynamic attention, in the line of Woodford (2009) or Stevens (2019).

As the model's solution does not depend on any specific objective function or on a particular parametric distribution for the unobserved shocks, it can be naturally extended to alternative settings beyond price-setting decisions. Concerning dynamic learning, it is relevant to more deeply explore the consequences of endogenous asymmetric learning rates over different states of the economy. While Van Nieuwerburgh and Veldkamp (2006) studied asymmetric responses due to imperfect information, there is no additional evidence in the context of costly entropy reduction, where heterogeneous learning rates arise endogenously due to agents' private efforts.

The main motivation behind this paper was to assess the time-varying implications of costly information and their relationship with overall price stability. As an unintended consequence of this friction, and after sufficiently long periods where the economy is in a less predictable state, prices could end up being much more volatile relative to a case where such rigidity is muted. The results can be relevant as they show the importance of communicational policies for monetary authorities, along with the timing of the announcements, to provide less informed agents with the expected duration of growth or crisis episodes.

References

Afrouzi, Hassan, "Strategic Inattention, Inflation Dynamics and the Non-Neutrality of Money," 2018.

Alvarez, Fernando E and Francesco Lippi, "A note on Price Adjustment with Menu Cost for Multi-product Firms," *Manuscript*, 2010.

_ , _ , and Luigi Paciello, "Optimal price setting with observation and menu costs," *The Quarterly Journal of Economics*, 2011, *126* (4), 1909–1960.

Bachmann, Rüdiger and Giuseppe Moscarini, "Business cycles and endogenous uncertainty," in "2011 Meeting Papers," Vol. 36 Society for Economic Dynamics 2011.

_ , Benjamin Born, Steffen Elstner, and Christian Grimme, "Time-varying business volatility and the price setting of firms," *Journal of Monetary Economics*, 2019, *101*, 82–99.

Baley, Isaac and Julio A Blanco, "Menu costs, uncertainty cycles, and the propagation of nominal shocks," *American Economic Journal: Macroeconomics*, 2018, *forthcoming.*

Berger, David and Joseph Vavra, "Shocks vs. Responsiveness: What Drives Time-Varying Dispersion?," *Journal of Political Economy*, 2019, *forthcoming.*

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

Caplin, Andrew and John Leahy, "Aggregation and optimization with state-dependent pricing," *Econometrica*, 1997, pp. 601–625.

Coibion, Olivier and Yuriy Gorodnichenko, "Information rigidity and the expectations formation process: A simple framework and new facts," *The American Economic Review*, 2015, *105* (8), 2644–2678.

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

Drenik, Andres and Diego Perez, "Price setting under uncertainty about inflation," *Working Paper*, 2018.

Farmer, Leland E, "The Discretization Filter: A Simple Way to Estimate Nonlinear State Space Models," *SSRN Working Papers*, 2016.

_ and Alexis Akira Toda, "Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments," *Quantitative Economics*, 2017, *8* (2), 651–683.

Gorodnichenko, Yuriy, "Endogenous information, menu costs and inflation persistence," Technical Report, National Bureau of Economic Research 2008.

Jung, Junehyuk, Jeong-Ho Kim, Filip Matejka, Christopher A Sims et al., "Discrete Actions in Information-constrained Decision Problems," *Review of Economic Studies*, 2019, *forthcoming.*

Klepacz, Matthew, "Price setting and volatility: Evidence from oil price volatility shocks," *Working Paper*, 2017.

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko, "Inflation targeting does not anchor inflation expectations: Evidence from firms in New Zealand," Technical Report, National Bureau of Economic Research 2015.

Lucas, Robert E, "Expectations and the Neutrality of Money," *Journal of economic theory*, 1972, *4* (2), 103–124.

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

_ , Filip Matějka, and Mirko Wiederholt, "Dynamic rational inattention: Analytical results," *Journal of Economic Theory*, 2018, *176*, 650–692.

Matějka, Filip, "Rationally inattentive seller: Sales and discrete pricing," *The Review of Economic Studies*, 2015, p. rdv049.

Matejka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *The American Economic Review*, 2014, *105* (1), 272–298.

Moscarini, Giuseppe, "Limited information capacity as a source of inertia," *Journal of Economic Dynamics and control*, 2004, *28* (10), 2003–2035.

Nakamura, Emi and Jón Steinsson, "Five facts about prices: A reevaluation of menu cost models," *The Quarterly Journal of Economics*, 2008, *123* (4), 1415–1464.

Nieuwerburgh, Stijn Van and Laura Veldkamp, "Learning asymmetries in real business cycles," *Journal of monetary Economics*, 2006, *53* (4), 753–772.

Paciello, Luigi and Mirko Wiederholt, "Exogenous Information, Endogenous Information, and Optimal Monetary Policy," *Review of Economic Studies*, 2013, *81* (1), 356–388.

Phelps, Edmund S, "Phillips curves, expectations of inflation and optimal unemployment over time," *Economica*, 1967, pp. 254–281.

Shannon, CE, "A mathematical theory of communication," *The Bell System Technical Journal*, 1948, *27* (3), 379–423.

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

Steiner, Jakub, Colin Stewart, and Filip Matějka, "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 2017, *85* (2), 521–553.

Stevens, Luminita, "Coarse pricing policies," *Review of Economic Studies*, 2019, *forthcoming.*

Tutino, Antonella, "Rationally inattentive consumption choices," *Review of Economic Dynamics*, 2013, *16* (3), 421–439.

Vavra, Joseph, "Inflation dynamics and time-varying volatility: New evidence and an ss interpretation," *The Quarterly Journal of Economics*, 2013, *129* (1), 215–258.

Woodford, Michael, "Imperfect common knowledge and the effects of monetary policy," *In Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps, ed. Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford*, 2003.

_ , "Information-constrained state-dependent pricing," *Journal of Monetary Economics*, 2009, *56*, S100–S124.

7 Appendix

7.1 Appendix A: profit function approximation

The derivation follows closely Alvarez and Lippi (2010). All firms share the same profit function $\Pi(P_t, Y_t, C_t) = Y_t P_t^{-\eta}(P_t - C_t)$. Where $\eta > 1$ represents the constant price elasticity, Y_t is the intercept of the demand (i.e. it's a demand shifter) and C_t is the marginal cost at time t. I assume that Y_t and C_t are perfectly correlated, i.e. when costs are high demand is also high. In order to approximate the objective function as (1), I compute a second order approximation of $\Pi(P_t, Y_t, C_t)$ around its frictionless price. In the RI context, the frictionless price is the optimal price under full information P_t^*.

The second order approximation of $\Pi(P_t, Y_t, C_t)$

$$\Pi(P_t, Y_t, C_t) \approx \Pi(P_t^*, Y_t, C_t) + \left.\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t}\right|_{P_t = P_t^*} (P_t - P_t^*) + \frac{1}{2}\left.\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2}\right|_{P_t = P_t^*} (P_t - P_t^*)^2$$

Which can be written:

$$\begin{aligned}\frac{\Pi(P_t, Y_t, C_t)}{\Pi(P_t^*, Y_t, C_t)} = 1 + &\frac{1}{\Pi(P_t^*, Y_t, C_t)}\left.\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t}\right|_{P_t = P_t^*} P_t^* \frac{(P_t - P_t^*)}{P_t^*} \\ &+ \frac{1}{2}\frac{1}{\Pi(P_t^*, Y_t, C_t)}\left.\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2}\right|_{P_t = P_t^*} (P_t^*)^2 \left(\frac{P_t - P_t^*}{P_t^*}\right)^2\end{aligned}$$

Taking the first and second order conditions:

$$\begin{aligned}\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t} &= Y_t P_t^{-\eta}\left[-\eta\left(\frac{P_t - C_t}{P_t}\right) + 1\right] \\ \frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2} &= -Y_t P_t^{-\eta-1}\eta\left[-\eta\left(\frac{P_t - C_t}{P_t}\right) + 1\right] - Y_t \eta P_t^{-\eta-2} C_t\end{aligned}$$

From the first order conditions, the optimal price is simply a constant mark-up over marginal cost $P_t = \frac{\eta}{\eta-1} C_t$. Evaluating the first and second order conditions at the optimal price:

$$\begin{aligned}\left.\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t}\right|_{P_t^*} &= 0 \\ \left.\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2}\right|_{P_t^*} &= -\eta Y_t C_t \left(\frac{1}{P_t^*}\right)^2 \left(\frac{\eta}{\eta - 1} C_t\right)^{-\eta}\end{aligned}$$

The maximized value of the profits:

$$\Pi(P_t^*, Y_t, C_t) = Y_t \left(\frac{\eta}{\eta - 1}\right)^{-\eta} C_t^{1-\eta} \left(\frac{1}{\eta - 1}\right)$$

Therefore, the term:

$$\frac{1}{2}\frac{1}{\Pi(P_t^*, Y_t, C_t)}\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2}\bigg|_{P_t}(P_t^*)^2 = \frac{-\eta Y_t C_t \left(\frac{\eta}{\eta-1}C_t\right)^{-\eta}}{Y_t \left(\frac{\eta}{\eta-1}\right)^{-\eta} C_t^{1-\eta}\left(\frac{1}{\eta-1}\right)} = -\eta(\eta - 1)$$

Finally, the second order approximation:

$$\frac{\Pi(P_t, Y_t, C_t) - \ \Pi(P_t^*, Y_t, C_t)}{\Pi(P_t^*, Y_t, C_t)} = -\frac{1}{2}\eta(\eta - 1)\left(\frac{P_t - P_t^*}{P_t^*}\right)^2 + o\left(\frac{P_t - P_t^*}{P_t^*}\right)$$

Where I can finally define $\gamma \equiv -\frac{1}{2}\eta(\eta-1)$, $\widehat{\Pi}(p_{it}, \widehat{p}_{it}) = log(\Pi(P_t, Y_t, C_t)) - log(\Pi(P_t^*, Y_t, C_t))$, $p_t = log(P_t)$ and $\widehat{p}_{it} = log(P_t^*)$ as stated in equation (1).

7.2 Appendix B: equivalence of mutual information

Information entropy is a measure about the uncertainty of a random a variable. Consider a random variable X with finite support Ω_x, which is distributed according to $f \in \Delta(\Omega_x)$. The entropy of X, is defined by:

$$\mathcal{H}(X) = -\sum_{x \in \Omega_s} f(x) log f(x)$$

With the convention that $0 log 0 = 0$. In RI, the acquired amount of information is measured by entropy reduction. Given the signal s_t, entropy reduction is measured by mutual information, which in the context of this dynamic model is:

$$\mathcal{I}(\widehat{p}_{it}, s_t | s_{t-1}) = \mathcal{H}(\widehat{p}_{it} | s_{t-1}) - E_{s_t}[\mathcal{H}(\widehat{p}_{it} | s_t) | s_{t-1}]$$

Given the entropy, the target-price $\widehat{p}_{it} = \sigma_t \epsilon_{it} \in \Omega_{\widehat{p}}$, and the definition for mutual information we can prove:

$$
\begin{aligned}
\mathcal{I}(\widehat{p}_t, s_t|s_{t-1}) &= \mathcal{H}(\widehat{p}_t|s_{t-1}) - E_{s_t}[\mathcal{H}(\widehat{p}_t|s_t)|s_{t-1}] \\
&= \sum_s f(s_t|s_{t-1}) \left[\sum_\sigma \sum_\epsilon f(\widehat{p}_t|s_t, s_{t-1}) log(f(\widehat{p}_t|s_t, s_{t-1}))\right] \\
&- \sum_\sigma \sum_\epsilon g(\widehat{p}_t|s_{t-1}) log(g(\widehat{p}_t|s_{t-1})) \\
&= \sum_s \sum_\sigma \sum_\epsilon f(s_t, \widehat{p}_t|s_{t-1}) log(f(\widehat{p}_t|s_t, s_{t-1})) - \sum_\sigma \sum_\epsilon \left[\sum_s f(s_t, \widehat{p}_t|s_{t-1})\right] log(g(\widehat{p}_t|s_{t-1})) \\
&= \sum_s \sum_\sigma \sum_\epsilon f(s_t, \widehat{p}_t|s_{t-1}) log\left(\frac{f(\widehat{p}_t|s_t, s_{t-1})}{g(\widehat{p}_t|s_{t-1})}\right) \\
&= \sum_s \sum_\sigma \sum_\epsilon f(s_t, \widehat{p}_t|s_{t-1}) log\left(\frac{f(s_t, \widehat{p}_t|s_{t-1})}{g(\widehat{p}_t|s_{t-1}) f(s_t|s_{t-1})}\right)
\end{aligned}
$$

Using the notation $\sum_x = \sum_{x \in \Omega_x}$.

From the second to the third line of the equivalence we rely on the fact that the prior distribution (marginal) is characterized as the sum of the joint probability distribution $f(s_t, \widehat{p}_t|s_{t-1})$ across all potential signals. The final expression is then what is shown in equation (3).

7.3 Appendix C: solution of the dynamic RI problem

In this section, I show how to derive the solution for the dynamic RI problem formally introduced in Section 2.4. Given prior beliefs $g(\widehat{p}_t|p_{t-1})$, firms choose the conditional probability distribution of prices $f_t(p_t|\widehat{p}_t)$ (equivalent of choosing $f(p, \widehat{p}_{it})$) in each point of the simplex $\Omega_p \times \Omega_\sigma \times \Omega_\epsilon$. To simplify notation, I will omit the lagged price conditioning and focus on a representative firm $\lambda_i = \lambda$.

Since the prior belief about the volatility distribution $m_t(\sigma_L)$ is the state variable of the problem, we can write the Bellman equation:

$$
V(m_t(\sigma_L)) = \max_{f_t(p_t|\widehat{p}_t)} \sum_\sigma \sum_\epsilon \sum_p [\widehat{\Pi}(p_t, \widehat{p}_t) + \beta V(m_{t+1}(\sigma_L))] f_t(p_t|\widehat{p}_t) g_t(\widehat{p}_t) \;\; - \;\; \lambda \mathcal{I}(\widehat{p}_t, p_t)
$$

Where:

$$
\mathcal{I}(\widehat{p}_t, p_t) = f_t(p_t, \widehat{p}_{it}) log\left(\frac{f_t(p_t, \widehat{p}_t)}{g_t(\widehat{p}_t) f_t(p_t)}\right) = f_t(p_t|\widehat{p}_t) g_t(\widehat{p}_t)[log(f_t(p_t|\widehat{p}_t)) - log(f_t(p_t))]
$$

The function is also maximized subject to the constraint on the prior (7). The first order condition of $V(m_t(\sigma_L))$ with respect to $f_t(p_t|\widehat{p}_{it})$:

$$\begin{aligned}g_t(\widehat{p}_t)\left[\widehat{\Pi}(p_t,\widehat{p}_{it})+\beta V(m_{t+1}(\sigma_L))+\beta\left[\frac{\partial V(m_{t+1}(\sigma_L))}{\partial m_{t+1}(\sigma_L)}\times\frac{\partial m_{t+1}(\sigma_L)}{\partial f_t(p_t|\widehat{p}_t)}\right]\right]\\ -\lambda g_t(\widehat{p}_t)[log(f_t(p_t|\widehat{p}_t))+1-log(f_t(p_t))-1]-g_t(\widehat{p}_t)\mu(\widehat{p}_t)=0\end{aligned}\tag{13}$$

The last term on the left hand side of equation (13), $\mu(\widehat{p}_{it})$, corresponds to the Lagrange multiplier of the constraint attached to the prior, equation (7).

Embedded in equation (13) is the effect of the current information strategy on posterior beliefs, $\frac{\partial m_{t+1}(\sigma_L)}{\partial f_t(p_t|\widehat{p}_t)}$. As discussed, posterior beliefs will later become the prior for $t+1$, $g_{t+1}=m_{t+1}(\sigma)h(\epsilon)$. The known i.i.d. structure of the idiosyncratic shocks ϵ_t implies that the chosen information strategy is not going to affect beliefs about this marginal distribution. Moreover, as stressed by Steiner et al. (2017), we can treat the effects of current information on future beliefs about the persistent state σ_t as fixed. The authors shows that a dynamic RI problem such as the one presented in this paper, is equivalent to a control problem without uncertainty about persistent states.[14] Therefore $\frac{\partial m_{t+1}(\sigma_L)}{\partial f_t(p_t|\widehat{p}_t)}=0$ and given $g_t(\widehat{p}_t)\geq 0$ and $\lambda>0$, equation (13) becomes:

$$\begin{aligned}\frac{\Pi(p_t,\widehat{p}_t)\;+\beta V(m_{t+1}(\sigma_L))-\mu(\widehat{p}_t)}{\lambda} &= \;log\left(\frac{f(p_t|\widehat{p}_t)}{f_t(p)}\right)\\ exp\left(\frac{\Pi(p_t,\widehat{p}_t)\;+\beta V(m_{t+1}(\sigma_L))}{\lambda}\right)exp\left(\frac{-\mu(\widehat{p}_t)}{\lambda}\right) &= \frac{f(p_t|\widehat{p}_t)}{f_t(p)}\\ \Rightarrow f(p_t|\widehat{p}_t)= \;exp\left(\frac{\Pi(p_t,\widehat{p}_t)\;+\beta V(m_{t+1}(\sigma_L))}{\lambda}\right)&f_t(p_t)\phi(\widehat{p}_t)\end{aligned}$$

Where:

$$\phi(\widehat{p}_t)\equiv exp\left(\frac{-\mu(\widehat{p}_t)}{\lambda}\right)\tag{14}$$

[14]The intuition behind the result is the following. In the control problem, while firms have full information about the current and past history of shocks, they face a trade-off of optimizing their flow utility $\widehat{\Pi}(p_t,\widehat{p}_t)$ against a control cost given by: $E_{f(p_t|\widehat{p}_{it})}[log(f(p_t|\widehat{p}_t))-log(q(p_t|\widehat{p}_t)|z^t]$. The variable z^t stands for the entire history of past shocks and prices. The cost is determined by the deviation of the final action with respect to some default action $q(p_t|\widehat{p}_{it})$. By relying on properties of the entropy, the paper shows an equivalence between a control and a dynamic Rational Inattention problem. Thus the inattention problem is solved by initially solving the control problem with observable states, characterizing the optimal conditional probability for each default rule $f(p_t|\widehat{p}_t)$, and then choosing q. As states are observable in the control problem, the solution ignores the effects of information acquisition on future beliefs (i.e., treat them as fixed) when solving the dynamic RI problem.

Finally, due to the restriction on the prior:

$$
\begin{aligned}
g_t(\widehat{p}_t) &= \sum_{p'} f_t(p_t'|\widehat{p}_t) g(\widehat{p}_t) \\
&= \sum_{p'} exp\left(\frac{\Pi(p_t', \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) f_t(p_t')\phi(\widehat{p}_t) g(\widehat{p}_t) \\
\Rightarrow \phi(\widehat{p}_t) &= \frac{1}{\sum_{p'} \ exp\left(\frac{\Pi(p_t', \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) f_t(p_t')}
\end{aligned}
$$

Combining this expression with (14), and adding the conditioning on lagged prices, we get the expression for the optimal posterior distribution of prices given the unobserved target, (11):

$$
f_t(p_t|\widehat{p}_t, p_{t-1}) = \frac{exp\left[\left(\Pi(p_t, \widehat{p}_t) + \beta V(m_{t+1}(\sigma_L|p_t))\right)/\lambda\right] f_t(p_t|p_{t-1})}{\sum_{p'} \ exp\left[\left[\Pi(p_t', \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L|p_t))\right)/\lambda\right] f_t(p_t'|p_{t-1})}
$$

The expression for the value function is then simply given by plugging this expression into (5):

$$
\begin{aligned}
V(m_t(\sigma_L)) &= \lambda \sum_{\sigma}\sum_{\epsilon}\sum_{p} f(p_t, \widehat{p}_t)\left(\sum_{p} exp\left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) f(p_t)\right) \\
&= \lambda E\left[\sum_{p} \ exp\left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L s))}{\lambda}\right) f(p_t)\right]
\end{aligned}
$$

7.4 Appendix D: information bounds

The solution of the model, and in particular its parameters, calls for validation in the sense that the overall process of actively seeking costly information must always be desirable for firms if they are going to do so. We can then compare the outcomes under RI relative to two extreme cases: Full information and no information. Under full information (FI) the cost of acquiring information is $\lambda_i = 0$ for all firms, while under no information (NI), the cost $\lambda_i \to \infty$. In the former case, firms perfectly track the optimal price p_{it}^*, whereas in the latter the absence of information lead firms to rationally set their prices equal to the unconditional mean of the target, $p_{it}^*(NI) = E[\sigma_t\epsilon_{it}] = E[\sigma_t]E[\epsilon_{it}] = 0$.

These extreme cases introduce two normative bounds for the solution of the RI model. Based on objective (1), the static profit loss under FI is $\widehat{\pi}_t^{FI} = 0$, while $\widehat{\pi}_t^{NI} = \gamma\sigma_j^2$, where $j = L, H$ depending on the realization of the state. In the case of RI, $\widehat{\pi}_t^{RI} = \gamma(p_{it}^* - \widehat{p}_{it})^2$ which varies according to the stochastic choice of p^*_{it}. In this case, the agent decides to use part of his mental capacity to uncover the realization and the distribution of the price-target. Acquiring costly information is always desirable if it loss is bounded within these two extreme cases.

$$0 = \widehat{\pi}_t^{FI} < \widehat{\pi}_t^{RI} < \widehat{\pi} s_t^{NI} = \gamma \sigma_j^2 \tag{15}$$

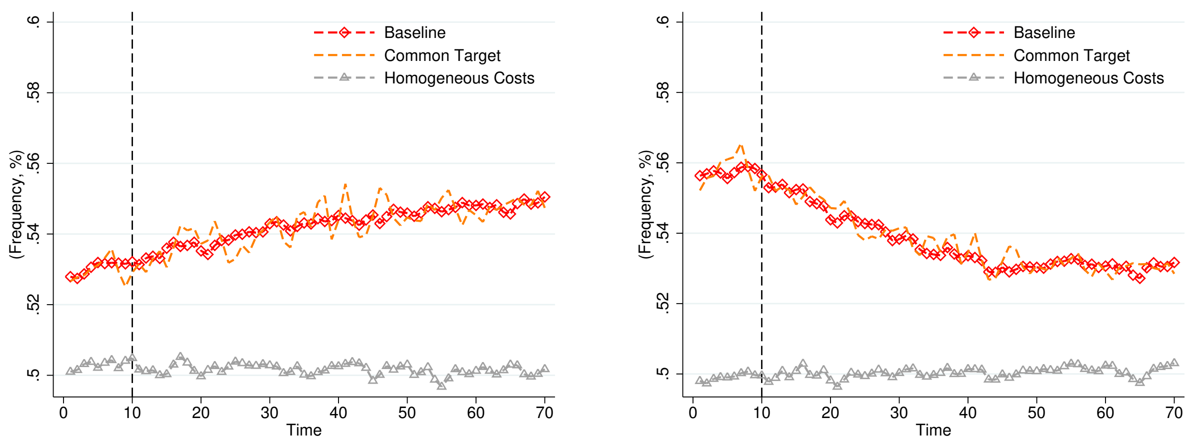
The following table shows the average loss under RI for all 15 different firm types and also the ratio between the loss under RI relative to NI. The agent is always better off by acquiring costly information if this ratio is less than one. This is confirmed by the results, which holds independently of being in either state. However, the gains are notably higher when the economy is in the more volatile state.

Table 4: Information Bounds

	Low Volatility		High Volatility	
Profit Loss	RI	RI/NI	RI	RI/NI
λ_1	0.012	0.187	0.014	0.061
λ_2	0.024	0.373	0.062	0.271
λ_3	0.032	0.495	0.067	0.293
λ_4	0.037	0.573	0.072	0.313
λ_5	0.041	0.633	0.075	0.329
λ_6	0.045	0.694	0.078	0.342
λ_7	0.046	0.716	0.083	0.362
λ_8	0.047	0.735	0.088	0.383
λ_9	0.050	0.777	0.090	0.395
λ_{10}	0.052	0.812	0.093	0.406
λ_{11}	0.059	0.919	0.095	0.416
λ_{12}	0.058	0.905	0.100	0.436
λ_{13}	0.059	0.919	0.100	0.436
λ_{14}	0.059	0.923	0.110	0.481
λ_{15}	0.059	0.925	0.118	0.517

7.5 Appendix E: Evolution of alternative models

Figure 6: Price-Change Frequency - Model comparison



(a) Transition: Low → High

(b) Transition: High → Low

Notes: The figure presents the time series evolution of price-change dispersion and the frequency of price changes (secondary axis). The dotted lines show the time frame when the economy is in a recession.