Rational Inattention-driven dispersion with volatility shocks*

Javier Turen
Pontificia Universidad Católica de Chile †

October 11, 2018

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 hits the economy. Information acquisition is dynamic and fully flexible since, given information acquired in the past, owners choose the amount of information to collect as well as how they want to learn about both the outcome and its distribution. I show that allowing for imperfect information as the *unique* source of rigidity, the model is able to simultaneously reconcile several stylized facts from the microeconomic evidence on price-setting, both at the cross-sectional and time series levels. The model is consistent with countercyclical price dispersion and with the presence of a positive correlation between dispersion and frequency of price changes, two features supported by the data. Dynamic imperfect information endogenously generates persistence in beliefs, which is crucial to replicate the dynamic empirical behavior of prices, without further assumptions about price-rigidities.

Keywords: *Active learning, 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 useful suggestions and comments to Antonio Cabrales, Morten Ravn, Filip Matejka, Mirko Wiederholt, Franck Portier, Ralph Luetticke, Nicolas Figueroa, Roger E.A. Farmer, Ernesto Pasten, Juan Pablo Torres-Martinez, Nezih Guner, Davide Melcangi, Silvia Sarpietro and Carlo Galli. All errors and omissions are mine.

†jturen@uc.cl

1 Introduction

Firms constantly and actively collect information to guide their decisions. Before setting prices, owners must first acquire information about unpredictable components of their industries, such as: elasticities or the state of the economy, which are hardly fully observed. In reality, this task is presumably harder as the features that generated these shocks may also be unknown. This is relevant since the distribution of shocks is likely to change over time, reflecting unanticipated periods of lower or higher uncertainty, such as recessions. Within this natural framework, the implications of imperfect information about shocks *and* their distributions on firms decisions, are certainly not clear. I show in this paper that this channel is quantitatively relevant as it is able to *simultaneously* reconcile several stylized facts from the micro price-setting literature, both at the cross-sectional and time-series level.

I 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), where I allow for a dynamic and fully-flexible information scheme. While past acquired information is relevant, I do not impose further assumptions on the *amount* of information to acquire or how owners *choose* to acquire it, i.e. there are no parametric assumptions on the distribution of signals. To set prices optimally, firms constantly 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 volatility shocks that affects the economy, the incentives to acquire information change as a function of owners beliefs about the current distribution, making the learning problem dynamic.

After calibrating the main parameters of the model, I show how dynamic imperfect information is consistent with countercyclical price dispersion along with generating a positive comovement between price-change dispersion and the frequency of price changes. The simultaneous presence of these two features is consistent with the data, as documented by Vavra (2013) using microeconomic evidence from price-setting.[1] In addition, my model is also consistent with the simultaneous presence of small and large prices changes at the cross-section level, as stressed by Klenow and Kryvtsov (2008). Finally as owners are active learners in the model, its implications are also in line with the presence of counter-cyclical attention, as documented by Coibion and Gorodnichenko (2015).

By combining price-rigidities with time-varying idiosyncratic shocks, Vavra (2013) shows how the aforementioned set of stylized facts is consistent with models of state-dependent pricing. I argue that even if we abstract from any of these two assumptions, and if we study

[1]Throughout the paper I will constantly make the distinction between "dispersion" and "volatility". In this context, dispersion refers to the spread (typically measured as the inter-quantile range or the standard deviation) of endogenous variables with respect to the cross-section of firms. On the other hand, volatility refers to the spread of exogenous shocks.

instead price-setting decisions under imperfect information it is still possible to rationalize the same behavior. Besides the theoretical exercise, there are important reasons why it is relevant to provide an information-based explanation that departs from price-rigidities. Microeconomic evidence suggests that managerial costs related to information processing represent a significantly larger proportion of total expenditures, relative to cost associated with price-changes, Zbaracki, Ritson, Levy, Dutta and Bergen (2004). In addition, my model illustrates a possible channel through which imperfect information can affect price stability, whilst being consistent with the data. Replacing a price-rigidity with an imperfect information mechanism is not innocuous for policy counterfactuals. Paciello and Wiederholt (2013) highlights the different implications for optimal policy within a price-setting framework, after allowing for active learning under RI. While in the paper I do not provide an optimal policy argument, the results certainly lead the discussion towards this direction.

Solving a model where information acquisition is dynamic and fully-flexible within a non-stationary framework, impose several methodological challenges. The difficulties arise precisely because of its flexible structure. In the model, acquired information has an effect on both pricing decisions and posterior beliefs about next's period distribution. To allow for a dynamic setting, a common assumption in the RI literature is to allow for a Gaussian distribution for the shock process, which is known with certainty. This assumption combined with a quadratic loss function, leads to a close form for the optimal signal structure, given by the true outcome realization plus normally distributed noise, Woodford (2003) and Maćkowiak, Matějka and Wiederholt (2018a). Uncertainty about the distribution of shocks, leads the optimal structure of signals to depend on the relative likelihood assigned to the economy being in each unobserved state. This impose a challenge when it comes to characterize the effects of current information on posterior beliefs. I tackle this problem by relying on the solution proposed by Steiner, Stewart and Matějka (2017). My model builds on this result, by allowing for different values of information costs and for state-dependent distributions. Then, I provide an algorithm to numerically solve this dynamic model with flexible information.

The model is simulated and calibrated to replicate documented properties of individual price changes in microeconomic datasets. To the best of my knowledge, the extend to which a fully-flexible dynamic RI model is able to match the data, has not been addressed by the literature before. This result strengthen the importance of incorporating limited information related constraints to understand economic outcomes and agent's behavior, as discussed by Mackowiak, Matejka and Wiederholt (2018b) and Gabaix (2017).

Since the model aims to reconcile the dynamic features of price dispersion, in the simulations the high volatility state is interpreted as a proxy for economic recessions in the data.[2] To match moments, I assume a parametric distribution for information costs. While the estima-

[2]This feature is in line with papers discussing how recessions are moments of a significant increase in uncertainty, which is captured by a rise in the volatility of both aggregate and idiosyncratic shocks, Bloom (2009) and Jurado, Ludvigson and Ng (2015).

tion depends on this assumption, I show that the magnitude of cost dispersion across firms is meaningful as it represents almost half of the average cost. Being the cost of information one of the key parameters in the RI literature, the results are informative as they shed light on how dispersed this rigidity can be. The ability of the model to match the large proportion of small price changes at the cross-section level, is an implication of having active learners with dispersed information costs.[3]

Imperfect information about a persistent distribution endogenously generates persistence in owners beliefs. Based on simulated data, I quantify the degree of beliefs persistence. I show that when the economy enters into a high volatility state, a firm with the lowest cost of information (with respect to all possible discretized values) needs 5 months to recognize the change, while a firm facing the highest costs need 9 periods. I quantify the role of beliefs persistence in price-change dispersion by decomposing its variance. I argue that at the onset of a state change, the proportion of the dispersion explained by owner's different beliefs about the shock distribution increase by 25%.

The ability of the model to replicate the aforementioned positive correlation between dispersion and frequency of price changes crucially depends on the concurrent 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). Whilst, a dynamic model with homogeneous information costs has also been studied by Matějka (2015). After re-calibrating the parameters under these two alternative specifications, I show how each assumption on its own is not enough to replicate the dynamic relationships suggested by data.

The baseline model presents some shortcomings. Firstly, it cannot fully replicate the price-stickiness observed in the data, where individual prices stays constant for some time. Although prices does not change regularly in the model, its simulated duration is shorter relative to the data.[4] Embedding 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. Secondly, while all firms track the same target-price, they cannot infer what others are doing i.e. there is no strategic complementarity in the model.

[3] In the literature of price-rigidities, matching this feature is not trivial. A common assumption within these models is to allow for stochastic menu-costs, Dotsey, King and Wolman (1999). As discussed in section 2.3, the source of the rigidity in my learning model is modeled as a fixed-effect across firms.

[4] The impossibility of matching this feature was not obvious ex-ante. As argued by Matějka (2015) and Jung, Kim, Matejka, Sims et al. (2015), a Rationally Inattentive agent chooses to price discretely when the processes for the fundamentals are not fully Gaussian. This would resembles price-stickiness without assuming further price-adjustment costs. I conjecture that the simulated price-stickiness is not enough to match the data, as the model still relies on a quadratic loss function independently that the shock is distributed according to a mixture of normals.

As discussed by Hellwig and Veldkamp (2009) and Yang (2015), complementarity in actions brings incentives to know what others know. As signal's precision are a non-parametric function of idiosyncratic costs and persistent beliefs, the task of inferring the time-varying posterior beliefs of others and from this, their distinct pricing decisions is clearly not trivial.

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 paper shows how these two costs complements 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 fresh information only at irregular moments of time, creating inertia in their behavior. Woodford (2009) introduces a setting with menu-costs, where the decision to conduct a price review is made under RI. In all these papers the results are driven by the crucial role of price-rigidities, while in this paper all the implications are a pure consequence of information rigidities. In addition, this literature has not look at the dynamic behavior of price-setting while trying to match the stylized facts, while in this paper these two tasks are performed simultaneously. In a recent paper, 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 main assumption this paper aims to relax.

RI models have proven useful to rationalize the empirical behavior of prices along with its aggregate implications. Maćkowiak and Wiederholt (2009) proposes a pricing model with endogenous attention from firms 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. Finally, 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. This paper contributes to this literature by studying the unexplored consequences of volatility shocks on price dispersion, whilst calibrating a fully dynamic imperfect information model able to match empirical facts from the data.

The rest of the paper is structured as follows. In Section 2, I introduce the model set-up and discuss the dynamic costly information setting. I then fully derive and characterize the solution of the problem. Section 3 presents the algorithm to numerically solve the model and then I show how the model is able to replicate both cross-sectional and time-series moments from data. The main results of the paper are presented in Section 4, where I 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 dynamic learning pricing model

2.1 Set up

The model is a partial equilibrium model with discrete time $t = 0, 1, \dots$ and a fixed number of firms $i = 1, \dots, N$. 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 profit $\widehat{\Pi}(p_{it}, \widehat{p}_t)$. Following Caplin and Leahy (1997) and Alvarez et al. (2011), the profit function is set equal to:

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

The objective (1) can be thought 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 $\widehat{p}_t$ is labelled as the "price-target". Due to the approximation, $\widehat{p}_t$ is a function of marginal costs which, I assume, are not perfectly observed by firms. Owners does not have complete information about cost conditions, as they cannot fully observe all the aggregate shocks affecting their production due to their own limitations to process information.[5]

There are two independent shocks drawn in each period t from finite sets, $\sigma_t \in \Omega_\sigma$ and $\epsilon_t \in \Omega_\epsilon$. The price-target is set equal to $\widehat{p}_t = \sigma_t \epsilon_t$. Underlying the evolution of the shocks, there 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 the two shocks is common information across firms.[6]

Following Gorodnichenko (2008), I assume firms operate in segmented markets so owners does not track prices set by others. In addition, I assume there are two different values for the persistent shock $\Omega_\sigma := \{\sigma_L, \sigma_H\} \subseteq \mathbb{R}_+$, 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. Thus, the realization of $\widehat{p}_t$ is determined by a discretized Gaussian distribution where its persistent volatility can be either low (L) or high (H).

[5] Bachmann and Moscarini (2011) also assumes an unobserved cost structure for firms. They argue how different cost variables (such as, input price elasticities or costs structures) are hard to estimate by firms. Think about owners that wants to maximize profits but needs to split their time in 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 remember all the information precisely, before setting a price.

[6] The definition of the target price as $\widehat{p}_t = \sigma_t \epsilon_t$, may lead to assume that the model allows for negative prices. 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 marginal costs, see Appendix 7.1. Hence, negative values of $\widehat{p}_t$ are consistent with $P^*_{it} \in (0, 1)$.

2.2 Information Acquisition

Firms start each period with prior beliefs $g_{it}(\widehat{p}_t) = m_{it}(\sigma_t)h(\epsilon) \in \Delta(\Omega_{\widehat{p}})$ 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_t respectively. Since the probability of $\epsilon \in \Omega_\epsilon$ is i.i.d. and its stochastic process is known, its prior probability is constant.

To set prices, owners acquire costly information about $\widehat{p}_t$ by choosing a signal $s_{it} \in \Omega_s$, where $|\Omega_p| \leq |\Omega_s|$. Firm are rationally inattentive since through costly information they aim to reduce the entropy of their beliefs, Sims (2003). Entropy (uncertainty) about $\widehat{p}_t$ is defined as $\mathcal{H}(\widehat{p}_t|\mathcal{S}_i^{t-1}) \equiv E[-log(\widehat{p}_t)|\mathcal{S}_i^{t-1}]$, 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 of firm i up to $t-1$. Given the discretization, the prior uncertainty about $\widehat{p}_t$ is $\mathcal{H}(\widehat{p}_t|\mathcal{S}_i^{t-1}) = -\sum_\sigma \sum_\epsilon g_{it}(\widehat{p}_t|\mathcal{S}_i^{t-1})log(g_{it}(\widehat{p}_t|\mathcal{S}_i^{t-1}))$. In the definition, sums are taken across all possible realizations of σ and ϵ in their sets.

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

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

Information flow (2) is defined as the difference between prior and posterior uncertainty (after observing s_{it}) about $\widehat{p}_t$, conditioning on lagged information.[7] 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 $\widehat{p}_t$ 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" $f_{it}(s_{it}, \widehat{p}_t|s_{it-1}) \in \Delta_{g_{it}}(\Omega_s \times \Omega_{\widehat{p}})$ 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 $g_{it}(\widehat{p}_t)$. After the price-target is realized, the choice of $f_{it}(s_{it}, \widehat{p}_t|s_{it-1})$ reflects the type of acquired signal based on the mental simplification process decided by the owner.

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

Proposition 1 : *Mutual Information Equivalence*

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

[7]As described, the formula of the entropy 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.

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

Proof in Appendix 7.2.

Thus, by setting an information strategy $f(s_t, \widehat{p}_t|s_{t-1})$ and given prior beliefs, owners are choosing the total amount of information $\mathcal{I}(\widehat{p}_t, s_t|s_{t-1})$ to acquire during period t.

2.3 The problem in two stages

Let us discuss the timing of the model. Within each period and after the realization of the two shocks, owners face two decisions: given prior beliefs $g_{it}(\widehat{p}_t|s_{it-1})$, they choose $f_{it}(s_{it}, \widehat{p}_t|s_{it-1})$ and then endowed with 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 $g_{it+1}(\widehat{p}_{t+1}|s_t) = m_{it+1}(\sigma_{t+1}|s_t)h(\epsilon)$.

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

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

At the information decision stage, owners face a trade-off. Signals with higher precision allows them to observe $\widehat{p}_t$ 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 extra unit of information is given by $\lambda_i > 0$. This cost affects directly the profit function and it is assumed different across firms.

Information is fully-flexible in the model as firms set the precision of their signals by choosing the shape of $f(\widehat{p}_t, s_{it}|s_{it-1})$, which determines total acquired information. With no further parametric assumptions about the shape of the joint probability distribution, owners privately decide how they want to learn about an outcome drawn from a mixture of normal distributions. As states are unobserved, the information strategies ultimately depend on owners idiosyncratic *perceived* prior distribution for $\widehat{p}_t$:

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

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

At the first stage, given the policy function $p^*_{it}(s_{it}|s_{it-1})$ and $\kappa_{it} \geq 0$ defined as (3), firms set their information strategies by choosing the conditional distribution $f(s_{it}|\widehat{p}_t, s_{it-1})$ given the time-invariant linear cost λ_i:

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

Owners decide the precision of signals based on prior beliefs about the true distribution of $\widehat{p}_t$, to maximize their expected profits relative to the cost of information.

The information strategy then determines the posterior distribution of signals which, given $g(\widehat{p}_t|s_{t-1})$, is then equivalent to set $f(\widehat{p}_t, s_{it}|s_{it-1})$. 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 setting $f(\widehat{p}_t|s_{it}, s_{it-1})$. 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 shows this for static RI problems, while Steiner et al. (2017) prove it within a dynamic setting with flexible information. Intuitively, if there are two signals which delivers the same price and since the entropy is a concave function, the firm could ended up setting the same price while lowering its information costs.[8] As prices $p_{it} \in \Omega_p$ are then associated with at most one signal $s_{it} \in \Omega_s$, the information strategy can be equivalently written as the joint probability distribution between optimal and target prices.

2.4 The Dynamic RI Problem

Let us formally introduce the dynamic information acquisition problem. While the price-setting decision is static, the unobserved and persistent distribution of $\widehat{p}_t$ implies a correlation between consecutive periods. As the precision by which owners tries to uncover the underlying state of the economy is subject to their choice, prior beliefs about the probability of each persistent state $m_{it}(\sigma_{jt}|p_{it-1})$ where $j = L, H$, becomes the state variable of the problem. With only two states, the state variable is a scalar which is convenient for writing the recursive problem.

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

$$V(m_{it}(\sigma_L|p_{it-1})) = \max_{f_{it}(p,\widehat{p}|p_{t-1})} \sum_\sigma \sum_\epsilon \sum_p [\widehat{\Pi}(p_{it}, \widehat{p}_t) + \beta V(m_{it+1}(\sigma_L|p_{it}))] f_{it}(p, \widehat{p}|p_{it-1}) - \lambda_i\kappa_{it} \quad (4)$$

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

Subject to:

$$\kappa_{it} = f_{it}(p,\widehat{p}|p_{it-1})log\left(\frac{f_{it}(p,\widehat{p}|p_{it-1})}{g_{it}(\widehat{p}|p_{it-1})f_{it}(p|p_{it-1})}\right) \tag{5}$$

$$g_{it}(\widehat{p}_t|p_{it-1}) = m_{it}(\sigma_t|p_{it-1})h(\epsilon) = \sum_p f_{it}(p,\widehat{p}_t|p_{it-1}) \tag{6}$$

$$m_{it+1}(\sigma_L|p_{it}) = \mathcal{T}_{t+1}(f_{it}(\sigma_L|p_{it})) \tag{7}$$

$$0 \leq f_{it}(p,\widehat{p}|p_{it-1}) \leq 1 \tag{8}$$

Owners aim to maximize the expected value of $\widehat{\Pi}(p_t,\widehat{p}_t)$ with respect to the perceived probability distribution of p_t and $\widehat{p}_t$, relative to the total cost of information. The cost λ_i forces the firm to form a probabilistic conjecture of its optimal price, given 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 (4) corresponds to the prior probability of being in the low volatility state. The solution of the problem depends on several restrictions. Equation (5) corresponds to the aforementioned expression for total acquired information. Equation (6) forces the chosen joint probability distribution being consistent with owner's prior beliefs. Without this constraint, owners could "forget" relevant information acquired in the past. Equation (7) characterizes the beliefs 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 attached to the current distribution of $\widehat{p}_t$ having low volatility. Finally, equation (8) ensures that the joint probability distribution is defined correctly.

2.5 Solving the model

A fully-flexible information scheme impose a challenge on how to characterize the solution of (4) as the shape of $f_{it}(p_{it},\widehat{p}_t|p_{t-1})$ and its implications on κ_{it}, has a non-linear effect on continuation values $V(m_{it+1}(\sigma_L|p_{it}))$. To tackle this issue, I rely on the solution proposed by Steiner et al. (2017). This paper argues that a dynamic RI problem consistent with (4) is equivalent to a control problem without uncertainty about $\widehat{p}_t$. 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 $V(m_{it+1}(\sigma_L|p_{it}))$. Therefore, the solution for the dynamic problem boils down to solving a collection of static problems with time-varying priors.

The solution of (4) subject to (5) - (8), is characterized by the following system of non-linear equations.

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

$$m_{it}(\sigma_L|p_{it-1}) = (1-\tau_{LH})f_{it-1}(\sigma_L|p_{it-1}) + \tau_{HL}(1-f_{it-1}(\sigma_L|p_{it-1})) \tag{9}$$

$$f_{it}(p_t|\widehat{p}_t, p_{it-1}) = \frac{exp\left[\left(\Pi(p_{it},\widehat{p}_t) + \beta V(m_{it+1}(\sigma_L|p_{it}))\right)/\lambda_i\right] f_{it}(p_t|p_{it-1})}{\sum_{p'} exp\left[\left[\Pi(p'_{it},\widehat{p}_t) + \beta V(m_{it+1}(\sigma_L|p_{it}))\right)/\lambda_i\right] f_{it}(p'_t|p_{it-1})} \tag{10}$$

$$V(m_{it}(\sigma_L|p_{it})) = \lambda_i E\left[\sum_p exp\left[\left(\Pi(p_{it},\widehat{p}_t) + \beta V(m_{it+1}(\sigma_L|p_{it}))\right)/\lambda_i\right] f_{it}(p_t|p_{it-1})\right] \tag{11}$$

Proof in Appendix 7.3.

For any given value of $\lambda_i > 0$, (9), (10) and (11) summarizes the main equations that solve the problem. Equation (9) 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 (7). The prior probability for the high volatility state $m_{it}(\sigma_H|p_{it-1})$ is then the complement of (9). These two probabilities are embedded into $g_{it}(\widehat{p}_t|p_{it-1})$ to force prior beliefs being consistent with the joint probability distribution, as stated in (6).

The conditional probability of p_t given unobserved shocks and lagged prices $f_{it}(p_t|\widehat{p}_t, p_{it-1})$, i.e. the information strategy, is characterized in (10). The probability resembles the Dynamic Logit formula, except for the term $f_{it}(p_t|p_{it-1})$ which multiplies the benefit-cost ratio of setting price p_t. As $f_{it}(p_t|p_{it-1})$ is independent of realized shocks it is interpreted as owner's "predisposition" to chose $p_t \in \Omega_p$ without additional current information. Specifically, Steiner et al. (2017) characterized firm's predisposition as prices that are chosen with high probability on average across outcomes and states, i.e. $f(p_t|p_{t-1}) = E_{\widehat{p}_t}[f(p|\widehat{p}_t, p_{t-1})]$. The posterior choice probability $f_{it}(p_t|\widehat{p}_t, p_{it-1})$ is then 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 (10) reflecting the noisy signals that owners receive, whilst being consistent with their own state-dependent beliefs. Finally, equation (11) shows the expression for the continuation value of the firm. I will refer to 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 close-form for the posterior probability $f_{it}(p_t|\widehat{p}_t, p_{it-1})$.[9] Moreover as information cost affects non-linearly both the posterior probability and continuation values, it is difficult to anticipate how the different values of λ_i would affect the information strategies. Without any known form of posterior uncertainty, the model is solved numerically.

3 Numerical Solutions

3.1 The algorithm

In this section, I provide an iterative algorithm to solve the dynamic information problem (9), (10) and (11, where its main parameters are calibrated.

Before solving the model numerically, I rely on further assumptions about the number of points on the simplex for each variable. The computational intensity of RI models severely restricts this decision, Tutino (2013). Let $|\Omega_\epsilon| = 11$ and $|\Omega_p| = 21$ be the number of possible values that the idiosyncratic shock ϵ_t and prices p_{it}, can assume respectively. The different values that $\widehat{p}_t$ can take, comes from a linearly equally spaced grid ranging from $-2\sigma_H$ to $2\sigma_H$. Based on the definition for $\widehat{p}_t = \sigma_t \epsilon_t$, the two unobserved states for σ_t and the assumption for $|\Omega_\epsilon|$, $|\Omega_{g(\widehat{p})}| = 21$.[10] Since $g_{it}(\widehat{p}_t) = m_{it}(\sigma)h(\epsilon)$, the state variable in the discretized problem 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 assumed $|\Delta(\Omega_\sigma)| = 21$, where each point reflects distinct (equally spaced) values for the marginal probability of being in the low state, in the (0,1) interval.

With the assumed discretization, I now describe the algorithm to solve the dynamic RI problem.

1. The procedure starts by fixing a value for the idiosyncratic cost, e.g. λ_1.

2. Given λ_1 and starting from the first value in $\Delta(\Omega_\sigma)$, we set prior beliefs $g(\widehat{p}_t) = m(\sigma_t)h(\epsilon)$, with dimension 2×11.

3. Fixing the value for $g(\widehat{p}_t)$, the model is solved by Value Function Iteration.

[9] Starting from the same model but where the underlying distribution of $\widehat{p}_t$ is known with certainty (i.e. a static framework), there would be a close 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.

[10] The number reflects that the average for the target $E(\widehat{p}) = 0$, is the same under the two possible distributions.

3.1. Starting with a guess for the vector $V(m_{t+1}(\sigma_L))$, the algorithm solves the static problem by computing $f(p_t, \widehat{p}_t|p_{t-1}) \in \Delta(\Omega_p \times \Omega_\sigma \times \Omega_\epsilon)$ which solves the system of nonlinear equations consistent with (6), (10) and $f(p_t|p_{t-1}) = E_{\widehat{p}_t}[f(p|\widehat{p}_t, p_{t-1})]$.[11]

3.2. With the chosen $f(p_t, \widehat{p}_t|p_{t-1})$, the marginal $f(\sigma|p_t, p_{t-1}) = \sum_\epsilon f(\sigma, \epsilon|p_t, p_{t-1})$ is calculated for each $p_t \in \Omega_p$. Through (9), posterior beliefs are updated to form prior beliefs for the next period, which are used to update $V(m_{t+1}(\sigma_L))$.

3.3. Using the expression for $V(m_{it}(\sigma_L|p_t))$ in (11), the algorithm iterates the Value Function until convergence where within each iteration it re-estimates $f(p_t, \widehat{p}_t|p_{t-1})$.

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

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

The price-tracking setting of the model and the decision on the shape of the joint probability distribution, resembles a filtering problem. The discretized setting for prices and shocks could immediately raise some concerns on its consequences for filtering. Departing from a continuos setting is not a numerical issue depending on the accuracy of the discrete approximation.[12] Evidence on discrete filtering support the previous statement. The numerical discrepancies between filtering with discrete relative to continuos outcomes are not significant, and depends on the nature of the discrete approximation, Farmer (2016) and Farmer and Toda (2017).

3.2 Calibration

The set of unknown parameters in the model are: the discount rate β, the two Markov Switching probabilities τ_{LH}, τ_{HL}, the price elasticity of demand η (which defines the curvature of demand γ), the low and high volatility states $\sigma_L, \sigma_H = \phi\sigma_L$ and the different information costs assigned to firms $\{\lambda_i\}_{i=1}^N$. Each period is assumed to be a month, so I set the discount factor equal to $\beta = 0.999$. The two transition probabilities are choose to be consistent with the literature of uncertainty shocks, $\tau_{LH} = 0.01$ and $\tau_{LH} = 0.036$. These monthly transition probabilities implies a quarterly probability of moving from the low to the high volatility state of 2.9% and a probability of remaining in the high volatility state of 89%. These numbers are roughly in line with Bloom, Floetotto, Jaimovich, Saporta-Eksten and Terry (2014) estimates for the U.S. Finally, the price elasticity of demand is set at $\theta = 5$ (implying a 25% markup). This is also consistent with existing models of price rigidities, Burstein and Hellwig (2006). Hence, $\gamma = -\frac{1}{2}\theta(\theta - 1) = -10$ as derived in appendix 7.1.

The remaining four parameters are calibrated to replicate different stylized facts on individual price changes reported from microeconomic data sets. In line with this approach, Woodford

[11] Although the static solution of the model is extremely computationally intensive, I gained a lot of efficiency by iterating directly over the FOC condition, as suggested by Lewis (2009).

[12] Tauchen (1986) discussed optimal ways to discretize a stationary continuous process

(2009) also collects empirical facts from several studies to assess the ability of his model to replicate documented features. Particularly, I take the stylized facts from the results in Klenow and Kryvtsov (2008) and Vavra (2013). These two papers rely on the Bureau of Labor Statistics monthly micro data which is used to construct the CPI in the US. Klenow and Kryvtsov (2008) reported the presence of both small and large price adjustments. They argued that almost half of the time a price is changed, the magnitude of the adjustment is less than 5%.[13] In addition, Vavra (2013) provided evidence on countercyclical price-change dispersion by showing that the standard deviation of price changes (from the cross section of firms) increases by approximately 25% during episodes of high volatility (NBER recessions). He also shows that the frequency of price changes also increases during recessions, leading to a positive comovement between frequency and price change dispersion over time.

The model-implied moments are generate by simulating an economy with $N = 7,500$ firms and $T = 5,500$ periods, using the algorithm described in 3.1. In the simulations the economy is allowed to evolve naturally across states and shocks, and I rule out the first 500 periods. To set the heterogeneity, I 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 about the cost distribution, I assume $\lambda_i \sim N(\overline{\lambda}, \sigma_\lambda^2)$ where the distribution is truncated at zero. Given $\overline{\lambda}$ and σ_λ, the different values of λ are set by the 15 equidistant percentiles, from 2.5 to 97.5 of this distribution. The remaining four parameters of the baseline model $\{\sigma_L, \phi, \overline{\lambda}, \sigma_\lambda^2\}$ are finally calibrated to replicate the data.

3.3 Matching Moments

The first two columns of Table 1 shows the moments chosen from the data and its simulated counterparts using the baseline model. The first two targeted moments correspond to evidence at the cross-sectional level, while the remaining two are evidence at the time series level. All these measures are calculated conditioning on a price change occurring. In the table, $Prob(|\Delta p|) < 5\%$ is the proportion of prices changes that are smaller than 5%. $Kurtosis(|\Delta p|)$ is the coefficient of kurtosis from the distribution of prices changes. $Stdv(Dispersion)$ and $Stdv(Frequency)$ stands for the relative standard deviation of dispersion and frequency of price changes respectively. As shown below, while the model reproduces some degree of price-stickiness, the predicted frequency of price changes is higher relative to the data. I then targeted the relative dispersions to have comparable (standardized) variability-related measures, allowing me to assess the extend by which the model replicates the dynamic evolution of price changes.

According to the results, the model is able to reproduce these four features of the data simultaneously. I believe this is one of the first studies to match different moments from data

[13] In addition, Midrigan (2011) shows how the distribution of price changes Δp resembles a normal distribution centered at zero, with a standard deviation of 8.2%.

through a purely dynamic RI model, i.e. without imposing additional frictions. The model also fairly replicates additional non-targeted moments. While the simulated $Median|\Delta p|$ is only half of what is observed in the data, the model is close to match the proportion of price changes that are less than 2.5% in absolute value. One of the main features of the baseline specification is that its able to endogenously generate the positive correlation between dispersion and the frequency of price changes, without further assumptions. The simulated correlation $Corr(Dis, Freq)$ is approximately 0.2, while in the data is roughly 0.28. In section 4, I simulated a transition between states to stress its pricing dynamics and to show the mechanism by which the model generates this positive correlation.

Table 1: Matched Moments and Alternative Specifications

Targeted Moments	Data	Baseline	Static	Homogeneous Costs
$Prob(\|\Delta p\|) < 5\%$	0.443	0.434	0.401	0.497
$Kurtosis(\|\Delta p\|)$	6.403	6.098	6.433	6.374
$Stdv(Dispersion)$	0.354	0.323	0.341	0.296
$Stdv(Frequency)$	0.120	0.096	0.173	0.091
Non-Targeted Moments				
$Median\|\Delta p\|$	0.097	0.050	0.039	0.048
$Prob(\|\Delta p\|) < 2.5\%$	0.254	0.193	0.258	0.209
$Corr(Dis, Freq)$	0.276	0.199	-0.009	-0.075

Notes: The $Prob(|\Delta p|) < 5\%$, $Prob(|\Delta p|) < 2.5\%$ and the $Median|\Delta p|$ comes from Tables III and IV in Klenow and Kryvtsov (2008). These first two moments represent the proportion of firms that produced small adjustment on their prices, 5% and 2.5%, respectively. $Median|\Delta p|$ corresponds to the median of the absolute price growth. The remaining moments comes from Table I and IV in Vavra (2013). $Kurtosis(|\Delta p|)$ represents the Kurtosis of the distribution of absolute price change, $Stdv(Dispersion)$ and $Stdv(Frequency)$ stands for the Relative Standard Deviation (coefficient of variation) for dispersion and frequency of prices changes and finally, $Corr(Dis, Freq)$ is the time series correlation between dispersion and the frequency of price changes.

The calibrated parameters are shown in Table 2. According to the results, the volatility of the price target in the high state increases by 74% with respect to the low state $\sigma_L = 0.091$. The rise in volatility is in line with estimated uncertainty parameters during episodes of economic distress, Bloom et al. (2014). The average cost of attention is estimated at 0.039, which accounts for a 4% of the average revenues of a firm. The magnitude of the standard deviation for the attention cost $\sigma_\lambda = 0.02$ is considerable as it represents almost half of the average cost. Being one of the key parameters in the theory of RI, the results are relevant as they shed light on the degree of dispersion of information rigidities across firms. While the calibrations rely on a parametric assumption about costs, they are informative to provide a quantitative assessment on the potential spread of information-related frictions.

Table 2: Calibrated Parameters

Parameter	Value	Description
β	0.99	Discount Rate
γ	-10	Curvature of demand function
τ_{LH}	0.01	Monthly transition probability: low/high state
τ_{HL}	0.036	Monthly transition probability: high/high state
σ_L	0.091	Volatility in low state
ϕ	1.74	Increase in volatility in high state
$\overline{\lambda}$	0.039	Mean distribution information cost
σ_λ	0.02	Stdv distribution information cost

3.4 Quantitative exploration of the model

Based on calibrated parameters, let us describe the different implications of information decisions on price-setting as a function of information costs.

3.4.1 Simulated Information Strategies

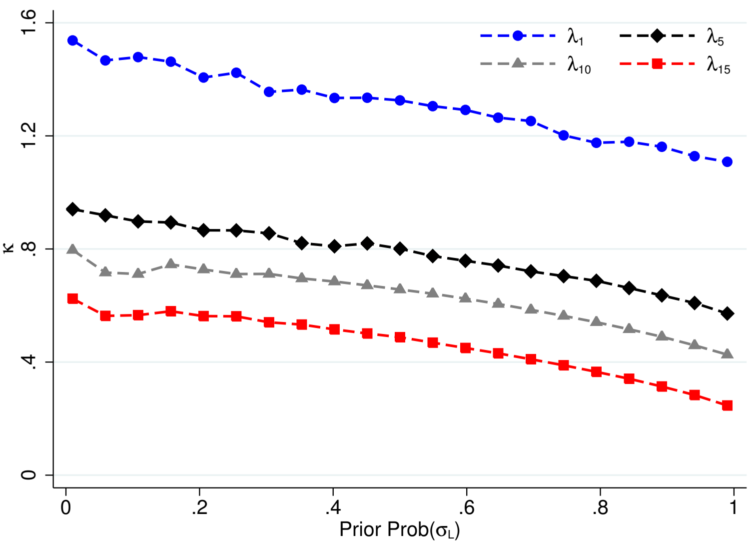
Before setting prices, owners set their information strategies. Uncertainty about the distribution of $\widehat{p}_t$ leads information decisions to depend on prior beliefs about each state. Figure 1 shows total acquired information κ_{it}, as a function of different prior probabilities assigned to the economy being in the low volatility state. The relationship is plotted for four different values of the information costs, where $\lambda_1 < \lambda_5 < \lambda_{10} < \lambda_{15}$.[14] As the perceived predictability of the unobserved-target increases, owners find it optimal to acquire less information. Hence the flow of information, which determines the learning rate, is not only disciplined by the cost. It is also affected by owners subjective beliefs about the underlying distribution that is generating the price-target.

Let us provide further intuition on how owners learn by simulating conditional probabilities. For expository reasons, I present the simulated information strategies for λ_1 and λ_{15} and for different prior probabilities. This is shown in Figure 2. Aa the joint probability distribution depends on three random variables, i.e. $f_{it}(p_{it}, \sigma_t, \epsilon_t | p_{it-1})$, I present compute the "predispositions" $f_{it}(p_{it}|p_{it-1}) = \sum_\sigma \sum_\epsilon f_{it}(p_{it}, \sigma_t, \epsilon_t | p_{it-1})$. Embedded in p_{it-1} are the prior beliefs about σ_t.

Imperfect information and uncertainty about the correct distribution of the price-target, make owners more prone to choose within a smaller set of prices. The left panel of Figure 2 shows the information strategies for λ_{15}, where the price-setter attaches high (80%), medium (50%) and low probability (20%) of being in the low volatility state. As information is very imprecise, instead of being predisposed to set prices closer to the mean the firm attaches more

[14] The information costs are extracted from the 15 normally distributed values for λ. Their labels corresponds to the 1st, 5th, 10th and 15th values.

Figure 1: Total Information

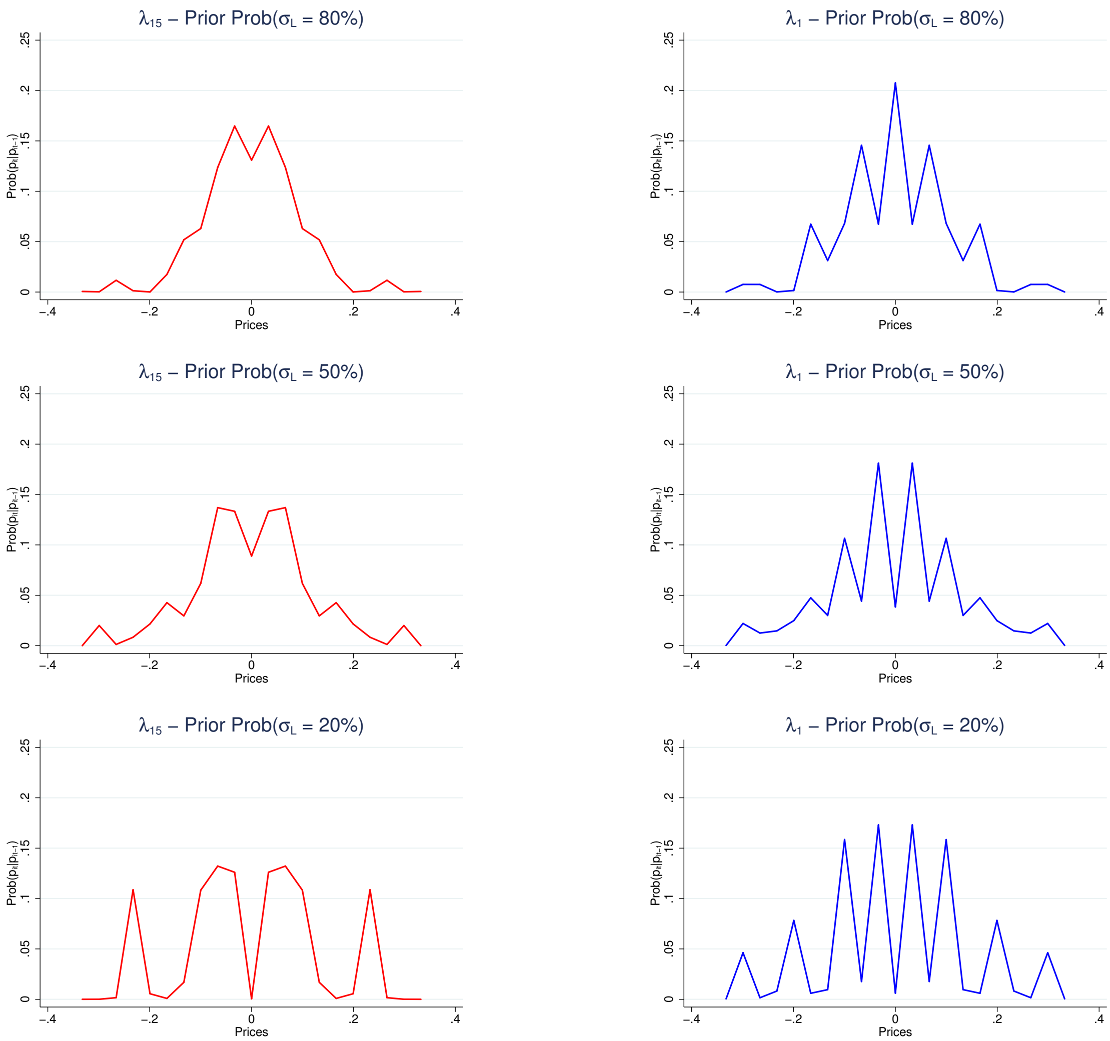


Notes: The figure presents total acquired information with respect to prior probabilities attached to the economy being in the low volatility state. The relationship is shown across different values of the information costs.

probability to prices away from it. Prices that were assigned with a larger probability ex-ante, i.e. before new information, are weighted more heavily when setting posterior beliefs. Hence, this subset of prices are more likely to be finally chosen as owners ultimately set prices by randomizing over the posterior distribution. This way, the owner is confident that at least she is reducing the probability of making large mistakes on average, to the utmost. This is consistent with her objective to minimize conditional variance. As the prior probability assigned to the low volatility state decreases, the owner modifies her information strategy by moving more probability mass towards extreme realizations, which are now perceived more likely. When the prior probability is low (lowest-left panel), the distribution suggests that the owner spends all its information efforts in noticing the sign of the outcome along with uncovering any potential extreme realization.

With additional information, owners will distinguish more precisely the relative position of the target and will distribute the probability accordingly. The right panel of Figure 2 shows the same three cases for the lowest information cost. As owners observe more precise signals, they are less inclined to set any specific price beforehand. They allocate their ex-ante probability more evenly across all potential realization of the target, in order to attach a significant amount of weight to new information. Under high probability of being in the low volatility state (top-right panel) the shape resembles a Normal distribution. When new information suggests that the economy may be in the less predictable state, owners again redistribute prior probability from the average towards extreme realizations.

Figure 2: Evolution of firm's predisposition



Notes: The two panels show the firms predisposition evolution $f_{it}(p_t|p_{it-1})$ for different prior beliefs about the low volatility distribution. The evolution on the left (blue lines) is due to the low information cost firm, while the evolution on the right (red lines) corresponds to the high information cost firm.

Information costs not only disciplines the quality of the search, they also force owners to form a probabilistic conjecture about the likelihood of extreme realization of the target, which lastly affects their pricing decisions over time.

3.4.2 Implications of (heterogeneous) imperfect information

In Table 3, I present the consequences of different information costs with respect to profit loss $\widehat{\pi}_{it}$ and price changes Δp_{it}. As expected, access to lower information costs allow owners to track the realizations more closely leading to lower profit losses. This feature is always true independently of the state of the economy however during recessions, the differences are amplified.

As a consequence of the normal distribution assumed for the target price, the average for Δp_{it} is zero. However, there is a clear negative relationship between the cost of attention and the dispersion of price changes. This is precisely what allows the model to replicate the proportion of small and large price adjustments observed in the data. The impossibility of tracking $\widehat{p}_t$ closely, lead high cost firms being more predisposed to set prices around the average (as shown in figure 2). Thus although prices can change at any time, the magnitude of the change will be bounded due to price-setters own decisions on how they choose to learn. The magnitude of the adjustments increases as the outcome is observe with higher precision, which is the case for lower values of λ. The results are interesting as they provide an alternative explanation to this feature of the data, driven by cognitive limitations to collect information.

Table 3: Implications of Imperfect Information

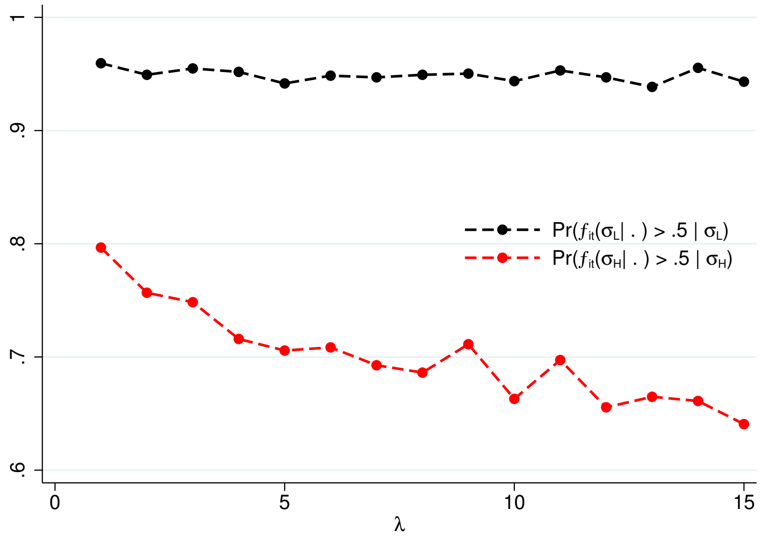
	All		Low Volatility		High Volatility	
Profit Loss	Mean	Stdv	Mean	Stdv	Mean	Stdv
λ_1	0.0027	0.0014	0.0030	0.0011	0.0015	0.0015
λ_5	0.0146	0.0043	0.0142	0.0020	0.0161	0.0084
λ_{10}	0.0226	0.0065	0.0219	0.0052	0.0250	0.0095
λ_{15}	0.0381	0.0176	0.0361	0.0163	0.0451	0.0201
Δp_{it}						
λ_1	0.00002	0.0743	0.0002	0.0593	-0.0005	0.1134
λ_5	0.00002	0.0658	0.0001	0.0524	-0.0004	0.1004
λ_{10}	0.00002	0.0606	0.0001	0.0467	-0.0004	0.0955
λ_{15}	0.00002	0.0508	0.0001	0.0368	-0.0004	0.0840

Notes: In the table, the values of the costs are $\lambda_1 = 0.0064$, $\lambda_5 = 0.0298$, $\lambda_{10} = 0.0463$ and $\lambda_{15} = 0.0776$. The values are computed as the average of the four different categories across firms and time.

3.4.3 Belief-driven pricing decisions

How frequent pricing decisions are made under "correct" beliefs about the current distribution? Figure 3 provides a measure to assess the accuracy of beliefs. In the figure, the black and the red line represents the $Prob(f_{it}(\sigma_L|p_{it}, p_{it-1}) > \frac{1}{2}|\sigma_t = \sigma_L)$ and $Prob(f_{it}(\sigma_H|p_{it}, p_{it-1}) > \frac{1}{2}|\sigma_t = \sigma_H)$ respectively, i.e. the probability that posterior beliefs about the *correct* distribution are greater than a half. Conditional probabilities are calculated for all distinct values of λ.

Figure 3: Accuracy of Beliefs



Notes: The figure presents the probability of setting prices based on the correct distribution for $\widehat{p}$. The black and red lines represents $Prob(f_{it}(\sigma_L|p_{it}, p_{it-1}) > \frac{1}{2}|\sigma_t = \sigma_L)$ and $Prob(f_{it}(\sigma_H|p_{it}, p_{it-1}) > \frac{1}{2}|\sigma_t = \sigma_H)$ respectively. This is a measure on how likely is that owners set prices attaching higher probability to the true distribution.

The results suggest that when the economy is in the low volatility state, around 95% of times owners set prices based on correct beliefs about the underlying distribution. This feature is independent of the magnitude of information costs and its explained by the higher persistence of expansions relative to recessions.[15] This is not longer true during high volatility episodes as the cost of information harms the precision of beliefs. In this case the accuracy starts at 80% for the lowest value of λ, and then starts falling monotonically as the cost increases reaching 65% when λ_{15}. Thus when the economy enters into a high uncertainty state (e.g. a recession), a significant proportion of firms would keep making pricing decisions *as if* the state is still the low volatility one. This time-varying discrepancy of beliefs is then crucial to explain the sources of price dispersion, as discuss in the next section.

Additional quantitative implications of the model are discussed in the appendix. Owners always face an outside option: set the optimal price at the unconditional mean of $\widehat{p}_t$ (which is

[15] According to the transition probabilities τ_{LH}, τ_{HL}, the unconditional probability of the low state is 78%.

independent of the distribution) without paying for further information. In section 7.4 of the appendix, I show that firms always prefer to acquire information independently of the value of λ. Moreover, I introduce a measure to assess the robustness of estimated costs by bounding their values based on the two polar cases: Full information and no information.

4 Delayed Learning Dynamics

In this section, I show the transition dynamics of the model by simulating an exogenous change of state in the economy. I assume the economy remains in the low volatility state for several months (300 periods) and then, it enters into a recession at time T which last for 28 months (the average duration of a crisis given the calibrated parameters). After this, the economy moves back to the low volatility state. Keeping the assumed transition of states constant, I simulate a 1,000 economies with 150 firms each where, as in the calibrations, I allocate the 15 different information costs uniformly across them.[16] Finally, I average the different variables across economies at each point in time.

4.1 Firm Level Evolution

Initially, I present results for two different firms at the two extremes of the cost distribution λ_1 and λ_{15}. The evolution of total acquired information κ_{it}, posterior beliefs about the high state $f_{it}(\sigma_H|p_t, \widehat{p}_t, p_{it-1})$ and the absolute magnitude of price changes $|\Delta p_{it}|$ for both costs are shown in Figure 4. In the figure, the two dotted vertical lines represents the high volatility state.

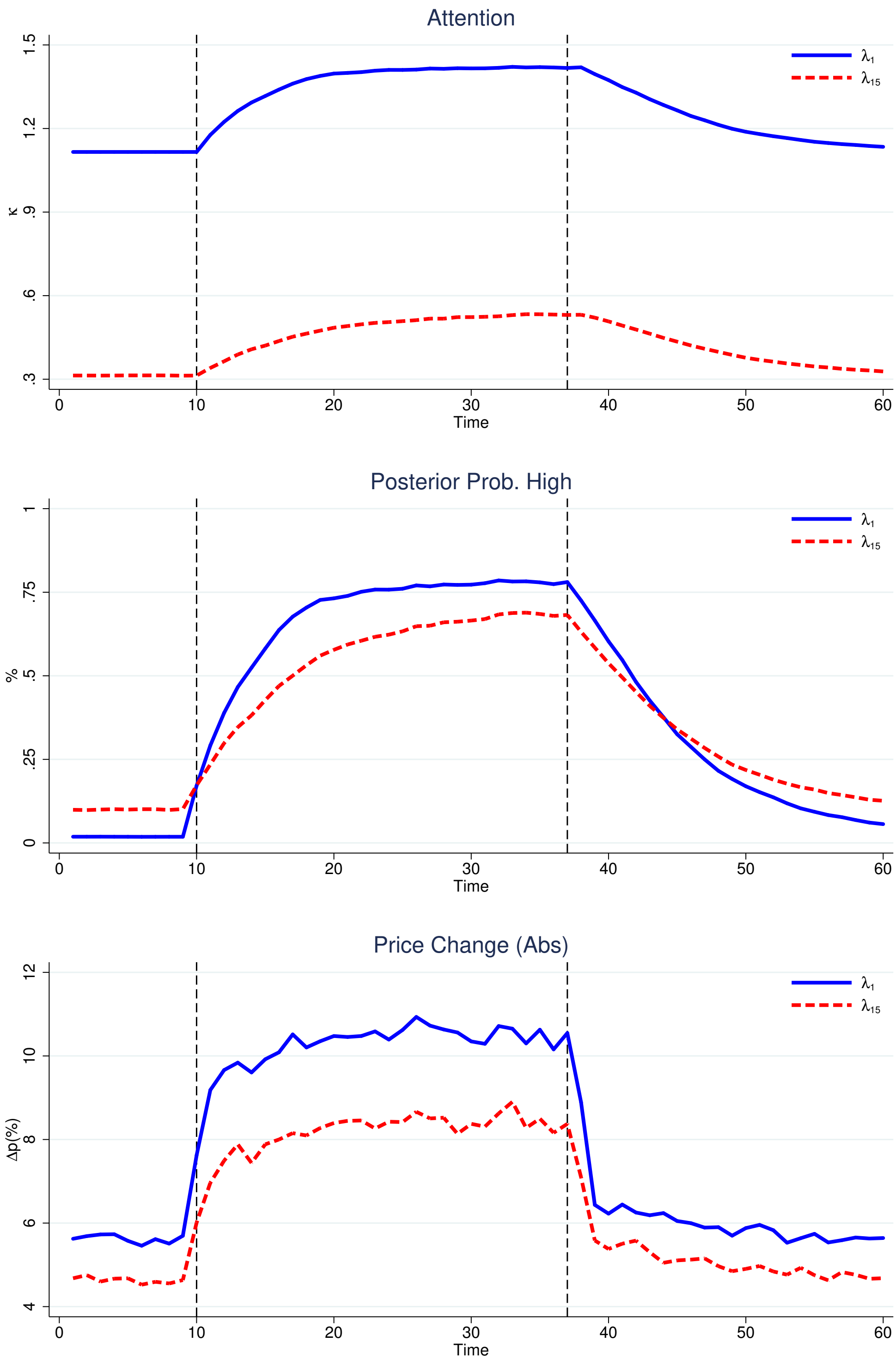
According to the time series evolution of total information and due to active learning, owners endogenously acquire more information under the less predictable state. The impossibility to notice immediately the new distribution cause a sluggish reaction in the rate by which firms increase their acquired information.[17] As economic recessions are episodes of a significant rise in uncertainty, the predictions of the model are consistent with the presence of countercyclical attention, in line with the empirical results in Coibion and Gorodnichenko (2015). Based on the behavior of professional forecasters, the authors argue that the degree of information rigidities (a proxy for the total level of inattention) went down during episodes of higher volatility in the U.S. This is interpreted as an increase in the amount of collected information.

The middle panel of Figure 4 shows the evolution of posterior beliefs about the probability of being in the high volatility state. Imperfect information about persistent states endogenously generates persistence in beliefs. Costly information prevents owners to notice a change in

[16] Generating 1,000 replicas of the state transitions brings significant computational challenges. To partially reduce the length of computing times, I took a smaller set of firms, relative to what was assumed for the calibration.

[17] The RI model is solved based on natural logs. To measure the total amount of information in "bits", I scaled the total amount of information by $\frac{1}{log_2(exp(1))}$.

Figure 4: Time varying evolution: Firm Level



Notes: In all the figures, the vertical dotted black lines represent the high volatility episodes. The top figure presents the evolution of total acquired information κ_{it} for the firms with low information cost (solid blue line) and for the high information cost one (red dashed line). The middle figure shows the evolution of the posterior probability of the economy being in the high volatility state while the bottom figure shows the evolution of the absolute value of price changes Δp_{it}.

the distribution immediately. Particularly, the rate by which the beliefs updating process is delayed is disciplined by the magnitude of information costs. While the low cost owner (λ_1) starts attaching higher probability to the high state after the fifth month, the high cost owner (λ_{15}) does it at the ninth month. Hence, the model generates disagreement about the true underlying distribution for $\widehat{p}_t$ during the transition, which finally affects both pricing and posterior information decisions. In addition, the model's prediction about time-varying heterogeneous beliefs across firms is also supported by the data, Kumar, Afrouzi, Coibion and Gorodnichenko (2015).[18]

The effects on $|\Delta p_{it}|$ are shown in the bottom panel of Figure 4. While the magnitude of price changes is expected to increase under the high volatility state, firms does not adjust immediately after a change of state. The figure shows a clear delay in the rate by which firms amend their pricing decisions during the first months of the high volatility state. This is the direct consequence of heterogeneous and persistent beliefs about the true distribution of the price-target. The bigger gap between the two firms during high volatility episodes, indicates the presence of higher price-change dispersion which is consistent with the empirical evidence.

The transition dynamics are different depending on the state of the economy. When a high volatility episode is over, owners need around four to five months to notice the change independently of their information costs. This contrast their behavior at the onset of the recession. The asymmetric reaction is explained by the different information strategies. Owners keep collecting additional information even after the less predictable state is over, due to their impossibility to notice immediately the new distribution. While this amount of information is not optimal from a cost-effectiveness analysis, higher information efforts allows them to notice the new state relatively faster than before causing asymmetric learning rates over the cycle.

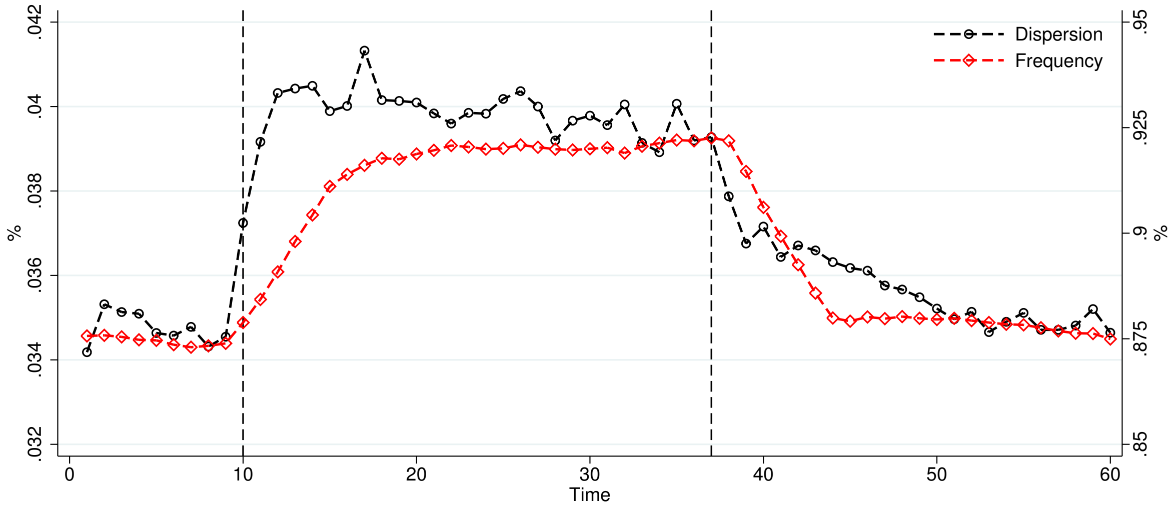
4.2 Aggregate Evolution

To study the overall implications of costly information over time, I aggregate information and pricing decisions across all firms. Figure 8 presents the time series evolution of price-change dispersion (measured by the inter quantile range) and in the secondary axis, the evolution of the frequency of price changes. Although there are no specific costs to adjust prices in the model, the presence of dynamic imperfect information is enough to simultaneously generate countercyclical price dispersion, with an average increase of 16% during the high volatility state, along with a positive correlation between dispersion and frequency of price changes. Price dispersion is time-varying due to owners private efforts to collect noisy information, which prevents them to recognize promptly a new distribution. In addition, price dispersion provides a measure of the dynamic inefficiencies caused by the information rigidity. Since all firms track the same $\widehat{p}_t$, dispersion is zero under full-information. This is relevant, as it revisits the

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

implications of imperfect information on the distortion of relative prices, through a dynamic learning mechanism.

Figure 5: Aggregate Evolution



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

While the possibility to replicate some nontrivial stylized facts is always desirable to validate the insights of the model, the richness of the exercise rely on the alternative assumptions that accomplish this. The existing literature rationalized this dynamic behavior as a consequence of price-rigidities combined with firms facing time-varying idiosyncratic shocks, Vavra (2013). None of these assumptions are need in this dynamic information model. Understanding the sources of price distortions through an information-driven rather than a price-rigidity mechanism is important for the design of policies. In particular, the scope by which policies can effectively reduce price instabilities are very different, depending if the source of the distortion comes from information rather than price-setting frictions. This line of reasoning does not imply that price-rigidities are not important. Ideally we will like to move closer to a setting that combines price-stickiness with dynamic attention.[19] As the main motivation of the paper is to stress the implications of dynamic imperfect information on pricing decisions, I leave this possibility open for future work.

Regarding the presence of time-varying idiosyncratic shocks, the baseline model intentionally rules out this possibility. However, I can incorporate this feature within my setting by assuming $\widehat{p}_{it} = \sigma_t \epsilon_{it}$. Allowing for $\widehat{p}_{it}$ would certainly make the set-up less controversial, as now the model incorporates the presence of idiosyncratic shocks at the firm level along with ruling out the

[19]As noticed in Figure ??, while the baseline model has proven effective to match the dynamics it is still not able to match the level of price-stickiness.

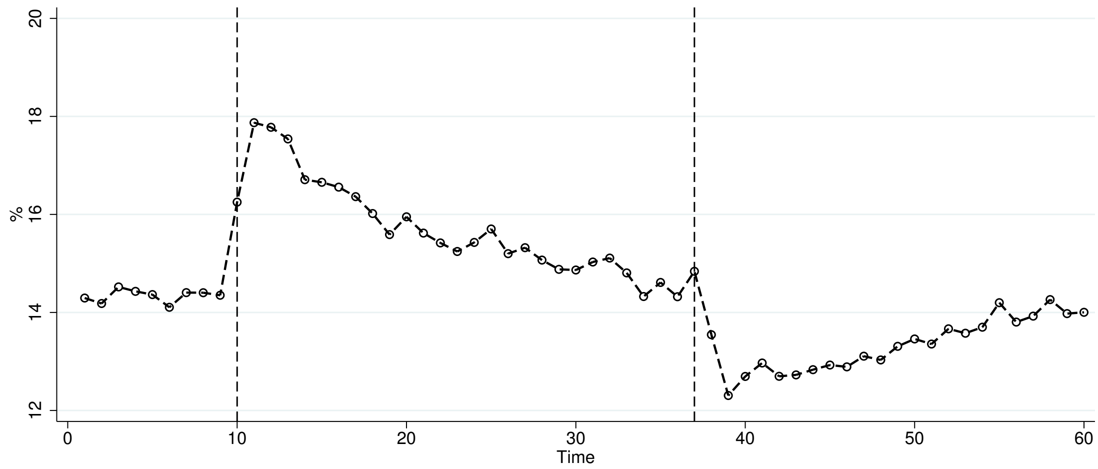
possibility of learning from others. I conjecture that the additional level of dispersion imposed by ϵ_{it}, would amplify price-dispersion making the task of replicating the data less challenging. I depart from this assumption mostly to stress the implications of imperfect information to the utmost. Under $\widehat{p}_{it}$ price-change dispersion can now be labelled as efficient as firms constantly track different prices. This would obscure the mapping of imperfect information on deviations from optimal values along with its implications during highly unpredictable episodes, when information is valued the most.

4.3 What drives the increase in dispersion?

Although the model replicates the empirical evolution of price-change dispersion its dynamic features call for further explanation. As the price-target becomes less predictable during high volatility states, it is not clear the extend by which the effects on price dispersion are due to exogenous shocks relative to owners endogenous reaction to the unobserved new distribution. 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 (2017).

To shed light on this feature, I decompose the total variance of price changes conditioning on idiosyncratic information costs, $Var(\Delta p_{it}) = E[Var(\Delta p_{it})|\lambda] + Var[E(\Delta p_{it})|\lambda)]$. Since costs are assigned randomly, I can group firms based on this time-invariant feature. The first element on the right hand side of $Var(\Delta p_{it})$ captures the price change dispersion within firms sharing the same information costs, while the second element computes the dispersion between firms facing different costs. The proportion of total variance explained by heterogeneous costs, the between effect, is shown in Figure 6. To characterize its time varying evolution, I replicate the same transition dynamics as in the previous section.

Figure 6: Variance Proportion - Between Effect



Notes: The figure presents the evolution of the proportion of total variance explained by the between effect. This effect is captured by the time-invariant heterogeneous information costs that owners face in the model. The transition dynamics across states is the same as in the previous sections, where the vertical dotted lines shows the recession period.

During the low volatility state the between effect accounts for around 14% of total dispersion. At the onset of the more uncertain state, and although the target becomes exogenously less predictable, the proportion of dispersion cause by endogenous responses increase by a non-negligible 25%. The initial rise in price dispersion is then mostly driven by the heterogeneous and persistent beliefs about the current distribution. Since beliefs guide pricing decisions, price dispersion is endogenously amplified. The between effect starts to decrease monotonically as owners recognize the new distribution. This behavior is then in line with the empirical evidence, as the detrimental effects of a recession are initially attributed to a rise in responsiveness rather than higher volatility of shocks.

5 Alternative specifications

The model has two distinct features that makes it appealing to understand the sources behind price-change dispersion: dynamic information and the presence of heterogeneous information costs. To gain intuition about the role each of these assumptions plays in the results, in this section I stress the main consequences of abstracting from each of these two channels.

5.1 A Static problem

The baseline model is dynamic due to firms impossibility to observe the persistent distribution. Alternatively, I can solve a simpler version of the model where I keep the structure for the unobserved target-price $\widehat{p}_t = \sigma_t \epsilon_t$, but where owners know with certainty the current state of the economy σ_t. In this setting, firms acquire costly information to track the i.i.d. shock ϵ_t. The problem becomes a standard static RI problem with a quadratic objective and Gaussian signals. To make the results comparable with the discretized baseline model, I solve the model by maximizing (1) relative to costly information (5), subject to (6) and (8). The solution of the static problem is:

$$f_{it}(p_t|\widehat{p}_t) = \frac{exp\left[\left(\Pi(p_t, \widehat{p}_t)\right)/\lambda_i\right] f_{it}(p_t)}{\sum_{p'} exp\left[\left[\Pi(p'_t, \widehat{p}_t)\right)/\lambda_i\right] f_{it}(p'_t)} \tag{12}$$

Where the optimal price p^*_{it} is drawn from equation (12). As in the baseline model, $f_{it}(p_t)$ stands for firm's predisposition to set price p_t but now pricing decisions are independent of lagged prices.

5.2 Homogeneous Information costs

Alternatively, I study the implications of allowing for homogeneous information costs. This model shares the same dynamic set-up as in the baseline specification, but without dispersion in information costs λ, i.e. $\sigma_\lambda = 0$.

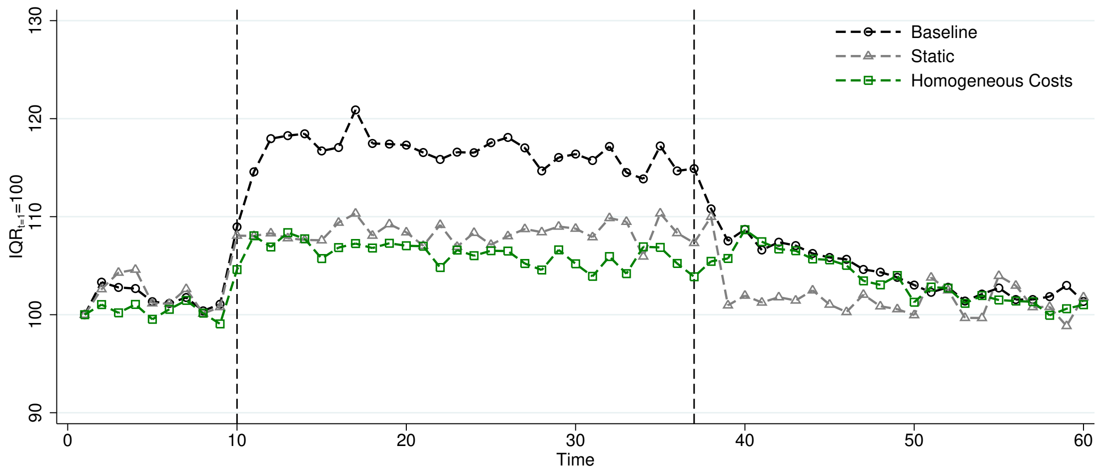
5.3 Implications for price change dispersion

To provide a correct comparison, I re-calibrate the parameters of these alternative specifications to match the same targeted moments as in the baseline model. The results are presented in the last two columns in Table 1.

In the static set-up, both the Kurtosis and the relative standard deviation of price change dispersion are closer to the data relative to the baseline case. Higher relative precision to replicate Kurtosis is also true under the homogeneous costs specification. Besides the relative precision to match targeted moments these results supports a broader issue: despite being computational intensive, models allowing for costly acquisition of information are proven useful to match non-trivial moments from microeconomic data on price-setting. However *neither* the static nor the homogeneous cost model is able to generate the positive correlation between dispersion and frequency of price changes. The failure to generate this result was not obvious ex-ante. Firms with homogeneous information costs can still set different prices since they are drawn from their (equal) posterior beliefs. Moreover these two alternative settings also generate higher information acquisition during high volatility episodes, leading to time-varying frequencies of price updating. However the interplay between a dynamic information setting and heterogeneous costs, is what allows the baseline model to replicate the dynamic features of data. The presence of persistent beliefs which generates sluggish pricing reactions combined with heterogeneous beliefs about the unobserved distribution, is what endogenously generates the positive comovement between price dispersion and frequency.

Figure 7 shows the evolution of price dispersion for the three alternative models. To make the evolutions comparable, I recalculated the simulated transition using the original parameters as in Table 2, and I normalize the first values of the dispersion to 100. The amplification effect caused by the two assumptions in the baseline model is now more marked. When the economy enters into the high volatility state, price dispersion increases by 8% and 7% approximately in the static and homogeneous costs cases, while the increase in the baseline set-up accounts for 16% approximately. In appendix 7.5, I present a broader comparison of the three models by looking at the (unscaled) evolution of price dispersion, frequency of price changes and the percentage of firms updating their total information.

Figure 7: Price Dispersion - Model Comparison



Notes: The figure presents the evolution of price change dispersion for the baseline model (black dotted line), static model (gray line) and homogeneous costs (green line). In the figure, the first observation of price dispersion is normalized to a 100.

While the static setting is not particularly meaningful to discipline the dynamics of price dispersion, it provides interesting insights on the effects of endowing firms with additional costless information. Certainty about the current distribution makes the problem of collecting information easier as now owners focused all their information efforts on uncovering the realized target-price. This leads to a reduction in the inefficient price dispersion of 10% relative to the baseline case. The static case resembles a setting where firms receive costless and fully-precise signals about the actual state of the economy at the beginning of each period. As previously discussed, these type of results can have broader implications as they support the design of policies aiming to manipulate agents expectations, by providing them with additional accurate information.

6 Conclusions

This paper studied price-setting decisions under dynamic costly information. In line with existing models, owners collect information about an unobserved target-price before setting prices. In the model, besides the outcome, firms does not perfectly observe the persistent distribution that generated the target. Information is dynamic and fully-flexible as owners choose total information to acquire as well as how they want to learn about the outcome. After calibrating the parameters, I argued that the model is able to rationalize several stylized facts from the micro price-setting literature, where imperfect information is the unique rigidity. The model generates persistence in beliefs which are crucial to match the data. The paper stressed the importance of incorporating costly information as an relevant friction faced by price-setters. While imperfect information is enough to match the dynamic features of the data, this rigidity can be complemented with common assumptions about state-dependent pricing. The possibility of matching the data relies heavily on the combination of a dynamic setting with time-invariant heterogenous information cost. By abstracting from any of these two channels, the model is not capable to simultaneously replicate the dynamic features of the data.

Although the paper revolves around price-setting decisions, the model is general and tractable enough to be extended to alternative settings where the intension is to study agents decisions within a dynamic and fully-flexible learning scheme. This is because the solution does not depend on any specific objective function or in a particular parametric distribution for the unobserved shocks. With respect to dynamic learning, it can be interesting to explore further some of the results in future projects. For instance, I think is important to explore deeply the consequences of endogenous asymmetric learning rates over different states of the economy. While asymmetric responses due to imperfect information have been studied before, Van Nieuwerburgh and Veldkamp (2006), there is no additional evidence in the context of costly entropy reduction where different learning rates arise as a consequence of agents private efforts.

The main motivation behind this paper was to assess the time-varying implications of costly information for the inefficient allocation of prices. Understanding the sources of inefficient price dispersion through a purely information set-up it is then crucial, as it dynamic patterns are expected to react to different economic environments or communicational policies.

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.

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?," Technical Report, National Bureau of Economic Research 2017.

Bloom, Nicholas, "The impact of uncertainty shocks," *econometrica*, 2009, *77* (3), 623–685.

_ , Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J Terry, "Really uncertain business cycles," Technical Report, National Bureau of Economic Research 2014.

Burstein, Ariel T. and Christian Hellwig, "Prices and Market Share in a Menu Cost Model," *Working Paper UCLA*, 2006.

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.

Dotsey, Michael, Robert G King, and Alexander L Wolman, "State-dependent pricing and the general equilibrium dynamics of money and output," *The Quarterly Journal of Economics*, 1999, *114* (2), 655–690.

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.

Gabaix, Xavier, "Behavioral inattention," Technical Report, National Bureau of Economic Research 2017.

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

Hellwig, Christian and Laura Veldkamp, "Knowing what others know: Coordination motives in information acquisition," *The Review of Economic Studies*, 2009, *76* (1), 223–251.

Jung, Junehyuk, Jeong-Ho Kim, Filip Matejka, Christopher A Sims et al., "Discrete Actions in Information-constrained Decision Problems," Technical Report, working paper 2015.

Jurado, Kyle, Sydney C Ludvigson, and Serena Ng, "Measuring uncertainty," *American Economic Review*, 2015, *105* (3), 1177–1216.

Klenow, Peter J and Oleksiy Kryvtsov, "State-dependent or time-dependent pricing: Does it matter for recent US inflation?," *The Quarterly Journal of Economics*, 2008, *123* (3), 863–904.

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.

Lewis, Kurt F, "The Two-Period Rational Inattention Model: Accelerations and Analyses," *Computational Economics*, 2009, *33* (1), 79–97.

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.

Mackowiak, Bartosz, Filip Matejka, and Mirko Wiederholt, "Rational Inattention: A Disciplined Behavioral Model," *Working Paper*, 2018.

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.

Midrigan, Virgiliu, "Menu costs, multiproduct firms, and aggregate fluctuations," *Econometrica*, 2011, *79* (4), 1139–1180.

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

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.

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.

Tauchen, George, "Finite state markov-chain approximations to univariate and vector autoregressions," *Economics letters*, 1986, *20* (2), 177–181.

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.

Yang, Ming, "Coordination with flexible information acquisition," *Journal of Economic Theory*, 2015, *158*, 721–738.

Zbaracki, Mark J, Mark Ritson, Daniel Levy, Shantanu Dutta, and Mark Bergen, "Managerial and customer costs of price adjustment: direct evidence from industrial markets," *Review of Economics and statistics*, 2004, *86* (2), 514–533.

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. its 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 context of Rational Inattention, the frictionless price is then 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 markup 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 $\gamma \equiv -\frac{1}{2}\eta(\eta - 1)$, $\widehat{\Pi}(p_{it}, \widehat{p}_t) = log(\Pi(P_t, Y_t, C_t)) - log(\Pi(P_t^*, Y_t, C_t))$, $p_t = log(P_t)$ and $\widehat{p}_t = log(P_t^*)$ 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 Rational Inattention, the acquired amount of information is measured by Entropy reduction. Given a collected signal s_t, entropy reduction is measured by mutual information, which in the context of this dynamic model is:

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

Given the definition of entropy and the mutual information, and relaying on the notation $\sum_x = \sum_{x \in \Omega_x}$, it is possible to prove:

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

Particularly, from the second to the third line of the equivalence I relied on the fact that the prior distribution (marginal) is characterized as the sum of the joint probability distribution $f(s, \widehat{p}|s_{t-1})$ across all potential values of the signal. The final expression for the mutual information, then coincides with what was presented 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 Rational Problem introduced in section 2.4. Given prior beliefs $g(\widehat{p}|p_{t-1}))$, firms choose the conditional probability distribution of prices $f_t(p|\widehat{p}_t)$ (equivalent to choose $f(p,\widehat{p}_t)$) 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 $\lambda_i = \lambda$. The Bellman representation of the model:

$$V(g_t(\widehat{p})) = \max_{f_t(p|\widehat{p}_t)} \sum_{\sigma} \sum_{\epsilon} \sum_{p} [\widehat{\Pi}(p_t, \widehat{p}_t) + \beta V(g_{t+1}(\widehat{p}))] f_t(p|\widehat{p}_t) g_t(\widehat{p}) - \lambda \kappa_t$$

Where:

$$\kappa_t = f_t(p, \widehat{p}_t) log \left(\frac{f_t(p, \widehat{p}_t)}{g_t(\widehat{p}_t) f_t(p)} \right) = f_t(p|\widehat{p}_t) g_t(\widehat{p}) [log(f_t(p|\widehat{p}_t)) - log(f_t(p))]$$

The function is also maximize subject to the constraint on the prior (6). The first order condition of $V(g_t(\widehat{p}))$ with respect to $f_t(p|\widehat{p}_t)$:

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

Where:

$$\frac{\partial g_{t+1}(\widehat{p})}{\partial f_t(p|\widehat{p}_t)} = h(\epsilon) \frac{\partial m_t(\sigma)}{\partial f_t(p|\widehat{p}_t)} \tag{14}$$

The last term on the left hand side of equation (13) $\mu(\widehat{p}_t)$, corresponds to the Lagrange multiplier of the constraint attached to the prior, equation (6).

Equation (14) represents the effect of current information strategy on posterior beliefs. Prior beliefs about the shock ϵ_t are independent of acquired information due to their i.i.d. structure. As stressed by Steiner et al. (2017), I can treat the effects of information on future beliefs as fixed. This is due to the equivalence between this dynamic Rational Inattention problems and a control problem without uncertainty about states.[20]

Since $\partial m_t(\sigma)/\partial f_t(p|\widehat{p}_t) = 0$, $g_t(\widehat{p}) \geq 0$ and $\lambda > 0$, equation (13) then becomes:

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

Where I defined:

$$\phi(\widehat{p}_t) \equiv exp\left(\frac{-\mu(\widehat{p}_t)}{\lambda}\right) \tag{15}$$

By the restriction on the prior:

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

Combining this expression with (15), and adding the conditioning on lagged prices, we get the expression for the optimal posterior distribution of prices given the unobserved target, (10):

[20]The intuition behind the result is the following: In the control problem, while firms have full information about current and past history of shocks, they face a trade off: optimizing her flow utility $\widehat{\Pi}(p_t, \widehat{p}_t)$ against a control cost given by: $E_{f(p_t|\widehat{p}_t)}[log(f(p_t|\widehat{p}_t)) - log(q(p_t|\widehat{p}_t)|z^t]$. 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 $q(p_t|\widehat{p}_t)$. By relying on properties about the entropy, the paper shows an equivalence between a control and 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 optimizing q. Since states are observable in the control problem, the solution ignores the effects of information acquisition on future beliefs (i.e. treat them as a fixed) when solving the dynamic Inattention problem.

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

The expression for the value function, is then simply given by plugging this expression (4):

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

7.4 Appendix D: Information Bounds

Information frictions introduce by the RI model prevents firms to use all the available information. Nevertheless the solution of the model, and particular its parameters, needs to be validated in the sense that the overall process of actively seeking information must be attractive for firms given their idiosyncratic costs. A useful exercise is then to compare the outcomes under RI with respect to its 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 with No Information (NI), the cost firms $\lambda_i \to \infty$. In the former case, firms perfectly track the optimal price $p^*_{it}(FI) = \widehat{p}_t$, whereas in the latter the absence of information lead firms to rationally set their optimal prices equal to the unconditional mean of the target, $p^*_{it}(NI) = E[\widehat{p}_t] = 0$. While under neither of the two cases there is room for cross-sectional disagreement on price changes, still the comparison is useful as a way to validate the chosen parameters.

These cases introduce two normative bounds for the solution of the RI model which are relevant due to the calibrated dispersion of idiosyncratic costs. Based on firm's objective (1), the static profit loss under FI is $\widetilde{\widetilde{\pi}}_t^{FI} = 0$, while $\widetilde{\widetilde{\pi}}_t^{NI} = \gamma\sigma_j^2$, where $j = L, H$ depending on the realization of the state. In the case of RI $\widetilde{\widetilde{\pi}}_t^{RI} = \gamma(p^*_{it} - \widehat{p}_t)^2 + \lambda_i \kappa^*_{it}$ which varies according to the stochastic choice of p^*_{it} and hence κ^*_{it}.[21] Intuitively the net loss under RI must be within these two extreme cases.

$$0 = \widetilde{\widetilde{\pi}}_t^{FI} < \widetilde{\widetilde{\pi}}_t^{RI} < \widetilde{\widetilde{\pi}}_t^{NI} = \gamma\sigma_j^2 \tag{16}$$

The difference between FI and RI is interpreted as the loss due to the information friction while the difference with respect to NI is then the net gain from actively collecting information. Table 4 shows that across the different values of λ and states, the net loss is always within the bounds. According to the parameters, the total variance and the variance under the low and high states is 0.121, 0.085 and 0.251, respectively. Them, given the value of γ I compute the relative loss under RI over the loss with NI. As expected in all cases, the ratio is less than one suggesting that firms are always willing to collect costly information independently of states.

[21] In terms of notation, I introduce "∼" to refer to differentiate the net profit loss function, i.e. the static loss after incorporating the information costs, from the gross profit loss.

Table 4: Information Bounds

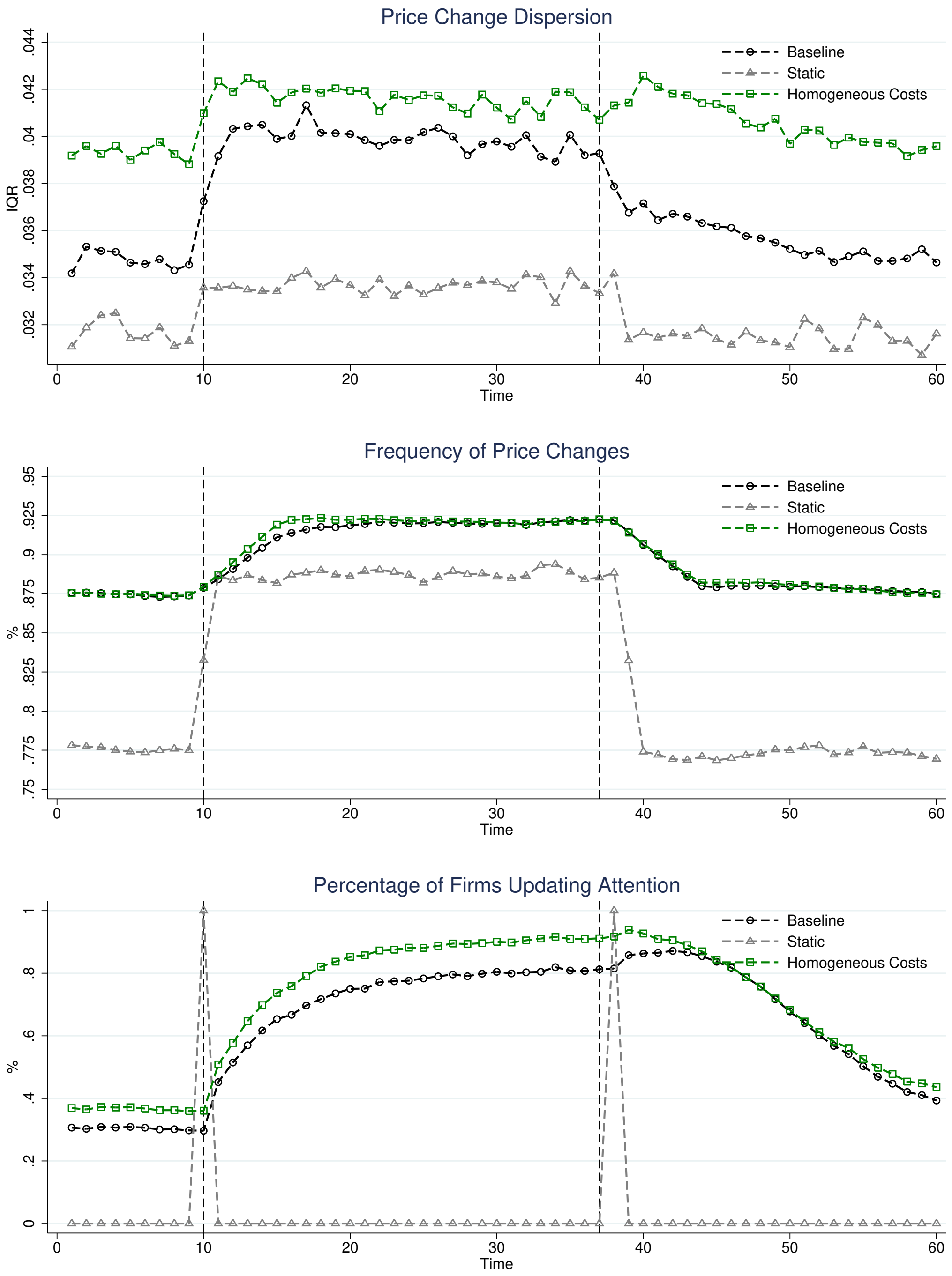
	All		Low Volatility		High Volatility	
Net Profit Loss	RI	RI/NI	RI	RI/NI	RI	RI/NI
λ_1	0.010	0.086	0.010	0.122	0.010	0.041
λ_2	0.022	0.181	0.022	0.258	0.022	0.087
λ_3	0.028	0.230	0.028	0.328	0.028	0.111
λ_4	0.032	0.263	0.032	0.374	0.032	0.127
λ_5	0.035	0.290	0.035	0.412	0.035	0.140
λ_6	0.038	0.313	0.038	0.446	0.038	0.152
λ_7	0.040	0.335	0.040	0.476	0.041	0.162
λ_8	0.043	0.357	0.043	0.509	0.043	0.173
λ_9	0.046	0.377	0.046	0.537	0.046	0.183
λ_{10}	0.048	0.393	0.048	0.560	0.048	0.190
λ_{11}	0.050	0.413	0.050	0.588	0.050	0.200
λ_{12}	0.053	0.436	0.053	0.620	0.053	0.212
λ_{13}	0.056	0.460	0.056	0.655	0.056	0.223
λ_{14}	0.060	0.495	0.060	0.706	0.060	0.239
λ_{15}	0.067	0.550	0.066	0.783	0.067	0.266

7.5 Appendix E: Evolution of alternative models

In this section, I describe the evolution of the three alternative models for the evolution of price dispersion, the frequency of price changes and the percentage of firms updating their attention. According to the upper figure, even under homogeneous costs there is persistent dispersion of prices. This is because, despite sharing the same cost, the optimal price is set by drawing from posterior beliefs $f_{it}(p_t|\widehat{p}_t, p_{it-1})$, according to equation (10). Interestingly, under both the baseline scenario and homogeneous costs the evidence supports the presence of asymmetric reactions with respect to a change of state, which is a feature of the dynamic setting.

By looking at the combine evolution of dispersion and the frequency of price changes, it may seems that these two alternative specifications are able to capture the positive correlation suggested by data. However, in terms of their levels, the magnitude by which dispersion rise does not seems particularly meaningful to actually generate the positive correlation. Finally, the lower panel shows the main implication of the static setting. Under this scenario, firms noticed the change of state with full precision, which leads all of them to adjust their total attention immediately. This is the main difference with a dynamic setting. Imperfect information about the states, makes the attention reaction sluggish, where the rate by which firms update their attention is disciplined by their own information costs.

Figure 8: Aggregate Evolution



Notes: In all the figures, the vertical dotted black lines represent the high volatility episode. The top figure presents the aggregate inter-quantile range evolution of price changes. The middle figure shows the evolution of the frequency of price changes while the bottom figure shows the percentage of firms updating their information capacity, κ_{it}. Each figure presents three cases. The black line represents the baseline model, the grey line corresponds to the static setting and the green line accounts for the homogeneous costs specification.