# Rational Inattention-driven dispersion over the business cycle*

Javier Turén
University College London †

November 13, 2017

**Job Market Paper**

Please click here for the most recent version

**Abstract**

This paper develops a model in which firms acquire costly information to make pricing decisions. Prices are set by tracking an unobserved target whose volatility depends on a persistent state of the economy. Firms are Rationally Inattentive since they face different information processing costs when learning the target. By embedding heterogeneous time-invariant information costs in this persistent volatility setting, I show that the model *endogenously* generates countercyclical dispersion in price changes, as documented by recent empirical findings. Costly information generates a delay in the rate at which firms' recognize any change of state, leading to different pricing decisions through the transition. Endogenous information and heterogeneous costs *alone* are enough to replicate the empirical time-varying evolution of the dispersion of price changes, as well as the positive co-movement between the dispersion and frequency of price changes.

Keywords: *Rational Inattention, Price Dispersion, Frequency, Endogenous Information.*
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, Davide Melcangi and Silvia Sarpietro. All errors and omissions are mine.

†javier.roman.12@ucl.ac.uk

# 1 Introduction

Seminar work by Lucas (1973), was the first to stress how imperfect information distorts relative prices, by setting them apart from optimal values. The consequences of price-setting distortions are well known, ranging from inefficient production to compromising its signaling role. This paper revisits this classical motivation by asking: What are the aggregate implications for price distortions when firms endogenously acquire imperfect information, over the business cycle? I show that a model embedding costly information within a dynamic price-setting framework generates inefficient and time-varying price dispersion, while *simultaneously* explaining both micro and aggregate empirical facts on prices.

In this setting, firms are Rationally Inattentive since they collect costly information about an aggregate fundamental, which depends on the state of the economy, that reduces the entropy of their beliefs. The first contribution of the paper is to show how endogenous information delays the rate by which firms recognize a change of state, which *amplifies* price distortions over the cycle. By allowing for imperfect information as the *only* source of rigidity in the model, I discuss how the interplay between dynamic information setting and heterogeneous time-invariant information costs are enough to replicate the countercyclical price change dispersion observed in the data, as documented by Vavra (2013). The second main contribution of the paper, is to show how through a purely dynamic Rational Inattention model, I am able to replicate several empirical facts from the price-setting literature which, to the best of my knowledge, have not been documented before. Given the recent empirical evidence on the presence of limited attention across firms, Kumar, Afrouzi, Coibion and Gorodnichenko (2015) and Bartoš, Bauer, Chytilová and Matějka (2016), the results strengthens the implications of allowing for imperfect information as a relevant source of rigidity.

Inefficient price dispersion arise as a large group of ex-ante identical firms set different prices with respect to their optimal common value, as a consequence of their own idiosyncratic information costs. As price dispersion brings an aggregate inefficient allocation of production, it is then crucial to understand its main drivers along with its dynamic properties.[1] In order to do so, I introduce a dynamic information acquisition model in the original spirit of Sims (2003). To set prices optimally, firms need to learn about the realization of an aggregate "target" price, which is drawn from an unknown distribution. Specifically, the volatility of the distribution depends on the persistent state of the economy, which is unobserved by firms.[2] Firms set their information strategies by choosing the total amount of information to collect, but information is costly. Learning occurs by observing noisy signals, where signal's precisions are a function

[1] Alternatively, price dispersion among a group of firms with heterogeneous production costs and costless information, can be labeled as "efficient", since their discrepancies are not a result of any type of rigidity.

[2] 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, in this setting: the unobserved target price.

of total acquired information. Rational Inattention models quantify total information based on Shannon (1948)'s uncertainty reduction. Through costly information, firms are not only uncovering the realization of the target price, they are also reducing their uncertainty about the latent state of the economy, which increase their knowledge about the true distribution of the unobserved target.

Learning about a persistent state makes the problem of acquiring information dynamic: total information observed today has non-negligible consequences for tomorrow. Since firms attach positive weights to their different signals, the sole presence of heterogenous costs is enough to generate inefficient price dispersion. However, by embedding the dynamic information structure into this setting, I show that the model *endogenously* generates an amplification in price dispersion as the economy moves from a low to a high volatility state. As stressed below, this countercyclical behavior is fully consistent with data. The intuition behind this result is the following: Imperfectly acquired information delays the rate by which firms learns about any change of state. The amount of time firms need to notice a change is completely disciplined by their idiosyncratic information costs. After characterizing the dispersion of information costs, I compare the evolution of beliefs of two firms taken from the lower and the upper end of the cost distribution. I show that, when the economy moves from the low to the high volatility state, the firm with the lowest cost is able to recognize the change after 5 periods, while the firm with the highest cost needs 9 periods to notice the change. Hence, price dispersion increase throughout the transition. This is because firms set prices based on misspecified beliefs, which are an implication of their own information choices.

Imperfect information as the only source of rigidity is not only important to stress the implications for price distortions, it also makes the model tractable. Although Rational Inattention models are computationally intensive to solve, a tightly parametrized model as the one presented in this paper, allows me to explore its quantitative implications by matching several stylized facts from the literature on the micro evidence of price-settings. By matching moments, I am able to pin down relevant parameters in order to characterize a parametric distribution for information costs. This feature has not been addressed by the literature on Rational Inattention before, with the exception of Woodford (2009).[3] Based on estimated parameters, I show that the model matches some non-trivial cross-sectional moments of the distribution of price changes, such as the large proportion of small changes, as stressed by Klenow and Kryvtsov (2008). Moreover, the model also replicates recent evidence related to the time-varying evolution of price changes. Vavra (2013) documents that price change dispersion is countercyclical and also reports the existence of a positive co-movement between the frequency of price changes and price change dispersion. The implications of my endogenous information model are also

[3]However, while Woodford (2009) aims to match similar moments as in this paper, he introduced a model with Rational Inattention *and* price rigidities in the form of menu-costs. As mentioned, I alternatively matched moments by relaying on a purely dynamic Rational Inattention setting. Moreover, he estimates an average value for the information cost, while I also characterize its dispersion across firms.

completely in line with these two facts. Finally, the results are also consistent with empirical evidence on state-dependent attention. In the model, price-setters endogenously increase their attention (total information) as the economy transits from the low to the high volatility state. Recently, Coibion and Gorodnichenko (2015) provided empirical evidence supporting this behavior. The authors show how the total level of estimated inattention decreases during recessions while increases during normal times.

A final methodological contribution of the paper is related to the specific solution of the model. Particularly, in order to derive and solve the model, I use state-of-the-art techniques which have not fully developed up until recently. The main difficulty in solving models with imperfect information about persistent variables is precisely its dynamic structure. Acquired information has an effect on both pricing decisions and on firms' posterior beliefs. Steiner, Stewart and Matějka (2017) recently tackled this problem by arguing that it is possible to disregard the effects of current information on continuation values after showing the equivalence between a dynamic Rational Inattention problem and a control problem with observables states. My model builds on Steiner et al. (2017) by allowing for unknown values of information costs and by extending the unobserved process to a mixture of two Gaussian distributions.

The ability of the model to generate the aforementioned positive correlation between dispersion and the frequency of price changes crucially relies on the concurrent presence of a dynamic framework with time-invariant heterogeneous costs. The model then nests two previously studied settings in the literature. By shutting down imperfect information about the state the model becomes static with a Gaussian unobserved target price. This setting resembles the one presented by Maćkowiak and Wiederholt (2009) and Woodford (2003). While a dynamic model with homogeneous costs has also been studied by Matějka (2015), I show how neither alternative version of the model is able to replicate the dynamic relationship between the dispersion and frequency of price changes.

The paper contributes to the optimal pricing literature with information frictions. Alvarez, Lippi and Paciello (2011) characterize a price acquisition problem with observation and menu costs. The paper shows how the 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 Rational Inattention. In all off these papers their main results are driven by the crucial role of price rigidities, while in this paper all the implications arise as a consequence of a purely information rigidity setting.

Rational Inattention 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 (2017) solves a dynamic general equilibrium model with inattentive price-setters 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. By introducing the business cycle dimension, this paper complements these results by showing how endogenous attention is able to reconcile newer empirical evidence on time-varying behavior of prices.

The rest of the paper is structured as follows. Section 2 discusses the main empirical facts of both price setting and the existing evidence of limited attention across firms. In Section 3 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 4 presents the algorithm to numerically solve the model and then I discuss 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 5, where I lay out both individual and aggregate implications over the business cycle. Finally, section 6 concludes.

# 2 Empirical Facts

## 2.1 Price Setting Behaviour

Empirical evidence on prices from microdata is significant with several well known stylized facts. Below, I enumerate the most relevant ones for the purposes of this research. Afterwards in section 4.2, I will show how the model is able to rationalized these empirical features.

**Price Change dispersion over the business cycle**. As already mentioned, Vavra (2013) was the first to document the interesting time varying behavior of the dispersion of price changes and the frequency of updating over the business cycle. In particular, he argues that the standard deviation of price changes (across the cross section of firms) increases by $\approx 25\%$ during episodes of high volatility (NBER recessions). He also showed how the frequency of price changes, also increases during recessions leading to a positive co-movement between this series and the dispersion of prices changes. The evidence is robust at industries and aggregate levels. These two new combined were proven relevant, since they contradicted all existing models on price rigidities.

**Small and large changes in prices**. Klenow and Kryvtsov (2008) find evidence of both small and large price adjustments: Almost half of the price are changed by less than 5%. Moreover, according to Midrigan (2011), the distribution of price changes $\Delta p$ resembles a normal distribution centered at zero, with a standard deviation of 8.2%.

## 2.2 Imperfect Information across firms

In the model, there is a crucial role for heterogeneous information costs across firms. While in principle, this assumption may account for the presence of large and small firms with different budgets to collect fresh information, more workers exclusively dedicated to study market trends or firms that because of the scale, have the possibility of testing and developing new products. Besides these intuitive reasons, specific empirical evidence on the presence of inattention across firms, is still relatively thin. Below, I discuss two recent papers documenting this fact.

**Inattention across firms**. By relying on a unique survey of firms from New Zealand, Coibion, Gorodnichenko and Kumar (2015) documents the presence of significant heterogeneity in attentiveness across firms. Particularly, while a significant proportion of firms are well informed about recent values for the inflation rate, some other firms reports values that are very far from actual lagged values or the Central Bank historical target. The authors stressed, how the wide dispersion of beliefs is then in line with models of imperfect information.

**Inattention within firms**. Bartoš et al. (2016) presents evidence on endogenous allocation of costly attention, within firms across two different countries. The empirical results are consistent with a model where different acquired knowledge is an implication of the agent's own efforts to collect information. Although the paper focused on attention discrimination during the hiring process of potential employees, it still shades some light on the actual presence of inattention within firms.

# 3 Baseline Model

## 3.1 Set up

The baseline set up assumes an economy with discrete time $t = 0, 1, \ldots$ and a fixed number of firms, $i = 1, \ldots, N$. Firms choose prices $P_{it}$ to maximize profits. Let $\Pi(P_{it}, Y_t, C_t) = Y_t P_{it}^{-\eta}(P_{it} - C_t)$ be the profit function, where $\eta > 1$ represents the constant elasticity of demand, $Y_t$ is a demand shifter for the offered product and $C_t$ is the marginal cost which is common across firms. Following Caplin and Leahy (1997) and Alvarez et al. (2011), this general profit function can be written as a second order approximation of the log of $\Pi(P_{it}, Y_t, C_t)$ around its optimal price. The approximation is derived in Appendix 7.1.

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

Where $p_{it} \in \Omega_p \subseteq \mathbb{R}_+$ is the log of $P_{it}$ (the chosen price), $\widehat{p}_t$ is the log of the aggregate price target $P_t^*$ (optimal price). The target $\widehat{p}_t$ is equivalent to the log of a constant markup over the time varying marginal costs. Due to information frictions, the optimal target price $\widehat{p}_t$

is unobserved by firms. The parameter $\gamma = -\frac{1}{2}\eta(\eta-1)$ represents the second order loss per period related to setting a price different from the optimum.[4] As $\widehat{\Pi}(p_{it}, \widehat{p}_t)$ does not depend on the prices set by other firms, the set up assumes that firms operates in segmented markets.[5]

The target evolves responds to time varying shocks, unobserved by firms. Since $\widehat{p}_t$ is a function of marginal costs, the assumption resembles unobserved aggregate shocks to the productivity of the firm.[6] The price target is assume equal to $\widehat{p}_t = \sigma_t \epsilon_t$. It is composed by two payoff-relevant shocks: a persistent shock $\sigma_t$ and an i.i.d. shock $\epsilon_t$. The persistent shock takes two possible values: $\sigma_t \in \{\sigma_L, \sigma_H\} \subseteq \mathbb{R}_+$, where $\sigma_H = \phi\sigma_L$, $\phi > 1$. The probability of the two states is governed by a Markov Chain with transition probabilities $\tau_{LH}$ and $\tau_{HL}$, where $\tau_{LH} \neq \tau_{HL}$. The second shock also takes values within a finite set, $\epsilon_t \in \Omega_\epsilon \subseteq \mathbb{R}$. The probability of each realization of $\epsilon_t$ is determined by a discrete approximation of a continuous i.i.d. process $\epsilon_t \sim N(0,1)$. The two shocks are independent of each other and the stochastic properties of both of them are common knowledge across firms.[7]

The structure of $\widehat{p}_t = \sigma_t \epsilon_t$, is motivated by the stylized facts. The finite set for $\Omega_\epsilon$ allows the target (and optimal prices) to vary across different realizations. The other component $\sigma_t$, incorporates business cycle dynamics by allowing the economy to move stochastically between two persistent states, one with low volatility and one with high volatility. In the model, firms not observe the current nor the lagged realization of $\widehat{p}_t$ and also they do not observe the pricing decisions of other firms. As explained below, imperfect information about the target lead firms to minimize their *expected* loss. If the loss is observed, and given the set prices, firms could back-out the true realization of the target with full precision.

## 3.2 Information Acquisition

Rational Inattention originally started with Sims (2003) idea of linking economic decisions to the process of acquiring relevant information. According to this theory, people's actions are based on their own capabilities to acquire and digest information, which are determined by their mental capacities (human brain). Sims (2003) characterized this limited capacity as a "channel" that only transmit information units at a finite rate. In the model, firms acquire information about $\widehat{p}_t$ to set $p_{it}$, so the two variables share a joint probability distribution. Ideally, a measure of the amount of collected information, must then be a function of this joint distribution. One

[4]Hence, $\gamma$ is a measure of the curvature of the demand function.

[5]Afrouzi (2017) discuss a setting with Rational Inattentive price-setters firms and strategic complementarities between them.

[6]Bachmann and Moscarini (2011) also allows for an unobserved cost structure for firms. They argued how different cost variables (such as, price elasticities or costs stuctures) are hard to estimate by firms.

[7]The definition of the target price as $\widehat{p}_t = \sigma_t \epsilon_t$ and $\epsilon_t \sim N(0,1)$ may lead to assume that the model allows for negative prices. Given the second order approximation, the target price is defined as $log(p^*_{it}) \equiv \widehat{p}_t$, with $p^*_{it}$ the optimal price. Hence, negative values of $\widehat{p}_t$ are consistent with the optimal price being between zero and one.

of the main appealing features of Rational Inattention is that it proposes such a measure, based on Shannon (1948)'s measure of mutual information.

Before information is collected, firms enter every period $t$ with their own prior beliefs about $\widehat{p}_t$. Let $m_{it}(\sigma)$, $h_{it}(\epsilon)$ and $g_{it}(\widehat{p}) \equiv m_{it}(\sigma)h_{it}(\epsilon)$ be the prior probability measures of $\sigma_t$, $\epsilon_t$ and $\widehat{p}_t$ respectively. Since probability of each value of $\epsilon \in \Omega_\epsilon$ is i.i.d., and its stochastic process is common information across firms, its prior probability is constant, $h_{it}(\epsilon) = h(\epsilon)$.

According to information theory, the entropy (uncertainty) about $\widehat{p}_t$ is defined as $\mathcal{H}(\widehat{p}_t|\mathcal{S}_i^{t-1}) = E[-log(\widehat{p}_t)|\mathcal{S}_i^{t-1}]$, where $\mathcal{S}_i^{t-1}$ collects all the lagged history of signals about $\widehat{p}$ of firm $i$, up to period $t-1$. The history of signals are relevant due to the persistent component of the process. Due to the finite values of the random shocks, the entropy of $\widehat{p}_t = \sigma_t\epsilon_t$ is:

$$\mathcal{H}(\widehat{p}_t|\mathcal{S}_i^{t-1}) = -\sum_{\sigma}\sum_{\epsilon} g_{it}(\widehat{p}|\mathcal{S}_i^{t-1})log(g_{it}(\widehat{p}|\mathcal{S}_i^{t-1})) \tag{2}$$

In terms of notation, $\sigma$ and $\epsilon$ subscripts in the expression, means that the sums are taken with respect to all possible realizations of the shocks in their sets. At each period, firms choose their *information strategies* in order to set prices $p^*_{it}$. Let $\Omega_s$ be the signal space, where $|\Omega_p| \leq |\Omega_s|$. Starting from prior beliefs, firms set their information strategies by choosing any signal $s_{it} \in \Omega_s$ about the realized $\widehat{p}_t$. Hence, firms chose what information to process along with its quality. High precision signals not only allow to track the unobserved target closely, but also provides relevant information about the current state of the economy, which ultimately affects posterior beliefs about the next period state. Prior beliefs for the next period $m_{it+1}(\sigma)$, depend on transition probabilities and current posterior beliefs about the two states, hence firm's priors $g_{it}(\widehat{p})$ becomes a state variable in the problem, making the problem of acquiring information dynamic.

Through signals $s_{it}$, firms reduce their uncertainty about the realized $\widehat{p}_t$. The reduction in uncertainty is given by Shannon's measure of mutual information flow:

$$\mathcal{I}(\widehat{p}_t, s_{it}|s_{it-1}) = \mathcal{H}(\widehat{p}_t|s_{it-1}) - E_{s_{it}}[\mathcal{H}(\widehat{p}_t|s_{it})|s_{it-1}] \tag{3}$$

The information flow $\mathcal{I}(\widehat{p}_t, s_{it}|s_{it-1})$ is defined as the difference between prior (2) and posterior uncertainty about $\widehat{p}_t$, after acquiring the signal $s_{it}$.[8] Due to the Markovian structure of the persistent states, all the relevant past information is summarized by the lagged value of the signal, $\mathcal{S}_i^{t-1} = s_{it-1}$. Shannon's mutual information (3) can be equivalently written as a

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

function of the joint probabability distribution of signals and shocks, $f_t(s, \widehat{p}|s_{t-1})$, conditional on lagged information.

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

*The mutual information (3) is equal to:*

$$\mathcal{I}(\widehat{p}_t, s_t|s_{t-1}) = \sum_s \sum_\sigma \sum_\epsilon f(s, \widehat{p}|s_{t-1}) log\left(\frac{f(s, \widehat{p}|s_{t-1})}{g(\widehat{p}|s_{t-1})f(s|s_{t-1})}\right) \tag{4}$$

*Proof in Appendix 7.2.*

According to proposition 1, the amount by which uncertainty is reduced is then a function of the prior $g(\widehat{p}|s_{t-1})$ and the joint probability distribution between signals and the price target, $f(s, \widehat{p}|s_{t-1})$. Since uncertainty is reduced by acquiring signals, $\mathcal{I}(\widehat{p}_t, s_t|s_{t-1})$ is then interpreted as an information channel with a given "capacity". Hence, by entering each period with a specific prior, the total flow of information that will pass through the channel is determined by this joint probability distribution.

## 3.3 The problem in two stages

The timming of the model is the following. Within each period, firms face two different decisions: Firstly, given prior beliefs $g(\widehat{p}_t|s_{t-1})$, they acquire information by choosing signals about the current unobserved realization of $\widehat{p}_t$ and second, based on collected information they set prices $p^*_{it}$. Firms are Bayesian since from their posteriors beliefs about the current state of the economy, they update their priors for next period $g(\widehat{p}_{t+1}|s_{t-1})$, based on the transition probabilities. A useful result that allows to solve this two stage model in just one stage, is the *equivalence* between signals and actions. More precisely, for each signal $s_{it}$ there is a single price $p_{it}$ that solves the problem. Matejka and McKay (2014) and Matějka (2015) formally shows this for static rational inattention problems while Steiner et al. (2017) prove it within a dynamic setting. In this section, I will explain the main intuition behind this equivalence and then I formally solve the model in one stage.

In the second stage, firms set it optimal price $p^*_{it}(s_{it}|s_{it-1})$, given some realization of the signal $s_{it}$. The signal gives the firm a belief about the realized target $\widehat{p}_t$, which is used to maximize the expected profit. More precisely, given $s_{it}$, the firm's policy function at $t$ is set by:

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

In the first stage, firms faces an information trade-off. While more *precise* signals allows them to observe $\widehat{p}_t$ with less noise, the precision of information is determined by its channel's capacity (3). To endogenously increase their capacity, firms must be willing to pay an idiosyncratic lineal cost $\lambda_i$, per unit of information. The unique source of ex-ante heterogeneity in the model is due to the time-invariant idiosyncratic $\lambda_i$, which affects directly the profit function of the firms.[9] Given prior beliefs $g_{it}(\widehat{p}|s_{it-1})$, firms set the precision of their signals by *choosing* the posterior probability distribution $f(s_{it}|\widehat{p}_t, s_{it-1})$. The particular shape of $f(s_{it}|\widehat{p}_t, s_{it-1})$ is completely decided by firms. A common result in Rational Inattention, Sims (2006) and Maćkowiak, Matějka and Wiederholt (2017), suggests that by assuming a quadratic objective function and Gaussian processes for the unobserved variables, the distribution of optimal signals is also Gaussian. Hence, the information strategy boils to choosing the variance of normally distributed signals. In this model, although the objective function is still quadratic, the unobserved target is drawn from a *mixture* of two normal distributions, so the optimality of Gaussian signals does not longer hold. Heterogeneous costs of attention, leads to different beliefs about the current state of the economy, which also depends on previously collected information. Thus, in this setting, the precision of signals does not have *any* ex-ante specific parametric shape.

Firms chooses signals about $\widehat{p}_t$, by characterizing the posterior distribution $f(s_{it}|\widehat{p}_t, s_{it-1})$. Given the prior $g(\widehat{p}_t|s_{it-1})$, and due to Bayes Law, choosing $f(s_{it}|\widehat{p}_t, s_{it-1})$ is equivalent to choose $f(\widehat{p}_t, s_{it}|s_{it-1})$. By Proposition 1, there is a mapping between the joint probability distribution and the amount of collected information, $\kappa_{it} = \mathcal{I}(\widehat{p}_t, s_{it}|s_{it-1})$. Therefore in the first stage, *given* the policy function $p^*_{it}(s_{it}|s_{it-1})$, firms choose the amount of acquired information by setting $f(s_{it}|\widehat{p}_t, s_{it-1})$ conditional on paying the lineal cost $\lambda_i$.

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

The signal distribution, determines the posterior beliefs about the unobserved target from which the new optimal price is finally drawn. Since information is costly and the joint distribution is endogenous, $\mathcal{I}(\widehat{p}_t, p^*_{it}|s_{it-1}) \leq \mathcal{I}(\widehat{p}_t, s_{it}|s_{it-1})$. Chosen signals are always consistent with the optimal price, therefore by choosing the posterior distribution of the signal, the firm is *simultaneously* and optimally choosing its optimal price. Because of the concavity of the mutual information and the linearity of the cost, each signal $s_{it} \in \Omega_s$ is linked to only one price $p_{it} \in \Omega_p$. By contradiction, if there are two signals which delivers the same price, due to the concavity of the entropy the firm would ended up with the same price, while lowering the infor-

[9] In Rational Inattention models, the cost is interpreted as the disutility resulting from a non-optimal action due to information rigidities. While throughout the paper I constantly refer to "firms", I can rationalize the presence of costs affecting profits by thinking on the firm's "owners" or "main sellers". While clearly, they want to maximize the profit of the firm they can face several information costs such as: reading reports about the firm's inventory levels, testing and developing new products, collect information from the sales division about stocks, among other.

mation cost. The linearity of the cost function prevents the firm to stock unused information for future periods. The equivalence between signals and prices allows to solve the model in just one stage.

## 3.4 The Dynamic Rational Inattention Problem

At each time period $t$, given prior beliefs $g_{it}(\widehat{p}|p_{it-1})$, firms maximize their expected profits by choosing the amount of information to collect. Their information strategies $f_{it}(p,\widehat{p}|p_{it-1}) = f_{it}(p|\widehat{p},p_{it-1})g_{it}(\widehat{p}|p_{it-1})$ are a solution to the objective function:

$$\max_{f_{it}(p,\widehat{p}|p_{t-1})} \sum_{t=0}^{\infty} \beta^t \sum_{p} \sum_{\sigma} \sum_{\epsilon} \widehat{\Pi}(p_t,\widehat{p}_t) f_{it}(p,\widehat{p}|p_{t-1}) - \lambda_i \kappa_{it} \tag{5}$$

Subject to:

$$\kappa_{it} = f_{it}(p,\widehat{p}|p_{it-1}) log\left(\frac{f_{it}(p,\widehat{p}|p_{it-1})}{g_{it}(\widehat{p}|p_{it-1}) f_{it}(p|p_{it-1})}\right) \tag{6}$$

$$g_{it}(\widehat{p}) = m_{it}(\sigma)h(\epsilon) = \sum_{p} f_{it}(p,\widehat{p}|p_{it-1}) \tag{7}$$

$$m_{it+1}(\sigma) = T_{t+1}(\sigma_L|\sigma_t,p_{it}) f_{it}(\sigma_t|p_{it}) \tag{8}$$

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

Where $p_{it} \in \Omega_p$, $\sigma_t \in \{\sigma_L,\sigma_H\}$ and $\epsilon_t \in \Omega_\epsilon$. The first term of the objective function is the expected value of the profit $\widehat{\Pi}(p_t,\widehat{p}_t)$ and the second is the lineal idiosyncratic cost $\lambda_i$ on information capacity $\kappa_{it}$. The problem is then to maximize the expected value of $\widehat{\Pi}$ with respect to the *perceived* probability distribution of $p_t$ and $\widehat{p}_t$ relative to the cost of acquiring information. The cost $\lambda_i$, forces the firm to form a probabilistic conjecture of the optimal price given the unobserved shocks. Since the space of prices and shocks is finite, the strategy space is compact. Therefore, from the continuity of the objective function, the Rational Inattention problem has a solution.

The solution of the maximization problem depends on several restrictions. Equation (6) corresponds to the aforementioned endogenous quantity of acquired information at each time. Equation (7) restricts the joint probability distribution to be consistent with the firm's prior beliefs. Without constraint (7), the firm could then"forget" the acquired information previously collected. Posterior beliefs about the economy being in the low state in the next period, are expressed in equation (8). Since there are two persistent states, I define the posterior probability as a function of $\sigma_L$ and then characterize $\sigma_H$ as the residual. In this equation, $T_{t+1}(\sigma_L|\sigma_t, p_{it})$ represents the law of motion of $\sigma_L$, based on the Markov switching probabilities, the current unobserved state and the chosen price. Finally, equation (9) ensures that the joint probability distribution is defined correctly.

## 3.5 Solving the model

The decision of acquiring information in this setting is dynamic due to the correlation between consecutive states. Prior beliefs $g_{it}(\widehat{p})$ becomes a state variable in the problem, since through the acquired information firms form their posterior beliefs about the current state, which then becomes the new prior for next period. Thus, the maximization problem (5) is written recursively based on its Bellman representation.

$$V(g_{it}(\widehat{p})) = \max_{f_{it}(p,\widehat{p}|p_{t-1})} \sum_{\sigma}\sum_{\epsilon}\sum_{p}[\widehat{\Pi}(p_{it}, \widehat{p}_t) + \beta V(g_{it+1}(\widehat{p}))] f_{it}(p, \widehat{p}|p_{it-1}) - \lambda_i \kappa_{it} \tag{10}$$

Subject to equations (6) to (9). To solve the dynamic attention problem (10) it is necessary to characterize the effects of current acquired information on firms posterior beliefs. However, Steiner et al. (2017) documents the equivalence between a dynamic rational inattention problem as (10) and a control problem without uncertainty about $\widehat{p}_t$. While deriving the solution for the attention problem it is then possible to omit the effects of current information on continuation values, where now $V(g_{it+1}(\widehat{p}))$ only depends on past information and the current persistent state. Therefore, solving the dynamic attention problem is then equivalent to solve a sequence of static problems. The solution of the dynamic Rational Inattention problem (10), its characterized by the following system of non lineal equations.

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

$$(1 - \tau_{12}) f_{it-1}(\sigma_L|p_{it-1}) + \tau_{21}(1 - f_{it-1}(\sigma_L|p_{it-1})) = \sum_{p} f_{it}(p, \widehat{p}|p_{it-1}) \tag{11}$$

$$f_{it}(p_t|\widehat{p}_t, p_{it-1}) = \frac{exp\left[\left(\Pi(p_t, \widehat{p}_t) + \beta V(g_{it+1}(\widehat{p}))\right)/\lambda_i\right] f_{it}(p_t|p_{it-1})}{\sum_{p'} exp\left[\left[\Pi(p'_t, \widehat{p}_t) + \beta V_{t+1}(g_{it+1}(\widehat{p}))\right)/\lambda_i\right] f_{it}(p'_t|p_{it-1})} \tag{12}$$

$$V(g_{it}(\widehat{p})) \;=\; \lambda_i E\left[\sum_{p} exp\left[\left(\Pi(p_t,\widehat{p}_t)+\beta V(g_{it+1}(\widehat{p}))\right)/\lambda_i\right] f_{it}(p_t|p_{it-1})\right] \tag{13}$$

*Proof in Appendix 7.3.*

For a given value of $\lambda_i$, the optimal joint distribution of prices and shocks should be consistent with equations (11), (12) and (13). Equation (11) is derived by combining constraints (8) and (7). The probability of choosing price $p_t$ given the unobserved shocks and lagged prices, $f_{it}(p_t|\widehat{p}_t, p_{it-1})$ is characterized in equation (12). In the numerator, the benefit-cost ratio of choosing price $p_t$ (the term inside squared brackets) is multiplied by firm $i$ "predisposition" to set that particular price. Because the function $f_{it}(p_t|p_{it-1})$ only depends on lagged pricing decisions (omitting the realization of $\widehat{p}$), it is interpreted as the firm own bias to set price $p_t \in \Omega_p$. Moreover, since the target $\widehat{p}_t$ is unobserved, the optimal price $p^*_{it}$ is finally drawn from the conditional density (12). Finally the expression for $V(g_{it}(\widehat{p}))$, equation (13), is derived after characterizing the conditional probability of prices.

Two crucial implications arise from the nonlinear system of equations. Firstly, there is no closed form solution for the attention problem. As derived in Steiner et al. (2017) (Theorem 1, page 13), firms "predisposition" are equal to the average optimal choice at each time, $f(p_t|p_{t-1}) = E_{\widehat{p}_t}[f(p|\widehat{p}_t, p_{t-1})]$. In a setting with no uncertainty about $\sigma_t$ (i.e. a static problem) and due to Gaussian shocks, Bayesian Updating leads to the posterior distribution of prices being equal to a lineal combination between prior beliefs and new information. The weights on each component are endogenously given by the signal precision. Hence, while uncertainty about the persistent state of the economy allows me to characterize the time-varying evolution of beliefs, the trade-off is that it makes the problem less tractable. Particularly, in order to solve the problem I then need to rely on numerical methods. Second, the solution of the problem highlights the implications of heterogeneous information costs. Through the definition of $f(p_t|p_{t-1})$, it is clear that $\lambda_i$ affects *both* the probability of choosing price $p_t$, $f_{it}(p_t|\widehat{p}_t, p_{it-1})$, and firm's "predisposition". However, since the joint probability distribution of prices and shocks is determined non-parametrically, there is no direct way to characterize ex-ante how the shape of the distribution its going to change depending on different values of the information cost.

To characterize the aggregate implications of heterogeneous attention, in this paper I complement the crucial insights in Steiner et al. (2017) in two different ways: I solve and derive the model allowing for different information costs $\lambda_i$ and also, I introduce a numerical algorithm to solve these type of models in practice.[10] The specific details of the numerical solutions of the model are discussed in the next section.

[10] With respect to the heterogeneous costs of information, in Steiner et al. (2017) the authors starts by normalizing the value of the cost before deriving the solution of the model. While allowing for an extra parameters could be interpreted as a minor contribution, it actually affects the form of the value function (13).

# 4 Numerical Solutions

## 4.1 Calibration

The set of unknown parameters of the model are: the discount rate $\beta$, the two Markov Switching probabilities $\tau_{LH}, \tau_{HL}$, the price elasticity of demand $\eta$ (which defines the curvature of the demand $\gamma$), the volatility under the low and high states, $\sigma_L, \sigma_H = \phi\sigma_L$ and the heterogeneous information costs assigned to each firm, $\{\lambda_i\}_{i=1}^{N}$. Due to the computational burden of solving the model and to estimate its key parameters, I start by calibrating a subset of them. Each period is assumed to be a month, so I set the discount factor equal to $\beta = 0.999$. I also fixed the two transition probabilities to be in line with the literature of uncertainty shocks, $\tau_{LH} = 0.01$ and $\tau_{LH} = 0.036$. The monthly transition probabilities implies a quarterly probability of changing from the low to the high state of 2.9% and a probability of staying in the high volatility state of 89%. These numbers are 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 in line 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 estimated to replicate several stylized moments of the microeconomic evidence on price setting, summarized by the results in Klenow and Kryvtsov (2008) and Vavra (2013). These two papers rely on the Bureau of Labor Statistics (BLS) monthly micro data which is used to construct the CPI in the US. Additionally, the papers collected information across the same three metropolitan areas. While the time span is different, 1988-2004 in Klenow and Kryvtsov (2008) with respect to 1988-2012 in Vavra (2013), when the statistics reported in the two papers coincides, there are no significant discrepancies between their reported values. I combine evidence from these two sources because there are some relevant moments, that are more useful to pin down certain parameters of the model, that were only reported in one of the papers.

To solve the model numerically, I need further assumptions on the number of points on the simplex of each variable. The fact that Rational Inattention models are very computationally intensive to solve, restricts the amount of points in the simplex, Tutino (2013). Let $|\Omega_\epsilon| = 11$ and $|\Omega_p| = 21$ be the number of possible values that the idiosyncratic shock $\epsilon_t$ and prices $p_{it}$, can assume respectively. The discretization of the unobserved target $\widehat{p}_t$ is a lineally equally spaced grid with 21 different values, starting from $-2\sigma_H$ to $2\sigma_H$. Since prior beliefs become the state variable of the problem, I define the belief simplex. Given $\widehat{p}_t = \sigma_t\epsilon_t$, the two unobserved states $\sigma_t$ and the assumption for $|\Omega_\epsilon|$, I set $|\Omega_{g(\widehat{p})}| = 21$ as the total number of points in the belief simplex. From the definition of $g(\widehat{p}) = m_{it}(\sigma)h(\epsilon)$, I define the state variable in the discretized problem as the probability of being in the low state.

After discretizing the relevant variables, I solve the dynamic Rational Inattention problem through the following algorithm. (i) The procedure starts by fixing a value for the idiosyncratic cost, e.g. $\lambda_1$. (ii) Given the cost and for each value in the belief simplex (prior beliefs), the algorithm computes the optimal probability distribution $f_{it}(p_t|\widehat{p}_t, p_{it-1})$ by solving the nonlinear system of equations consistent with (11) - (12).[11] (iii) Based on the chosen probability distribution for each point in the belief simplex, the prior belief for next period $g_{it+1}(\widehat{p})$ is computed which is used to set the value function $V(g(\widehat{p}))$ according to (13). (iv) Iterate the Value Function until convergence and finally (v) Repeat all the previous steps, for each possible value of the idiosyncratic costs.

The price-tracking setting of the model and the decision on the shape of the joint probability distribution (which determines the precisions of signals), resembles a filtering problem. The discretize 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. Tauchen (1986) discussed optimal ways to discretize a stationary continuous process. Evidence on discrete filtering support the previous statement. The numerical discrepancies between filtering with discrete relative than continuos outcomes, are not significant and relies on the nature of the discrete approximation, Farmer (2016) and Farmer and Toda (2017).

The model-implied moments are generated after simulating an economy with $N = 7,500$ firms and $T = 5,500$ periods. The first 500 periods are ruled out and I allow the economy to evolve naturally across states and shocks. To build the heterogeneity, I assume there are 15 potential different values for $\lambda$. Then, I randomly assigned the different costs uniformly across firms, i.e. $15 \times 500 = 7,500$. Without no further information about the cost distribution, I assume the different costs are drawn from a *truncated* normal distribution, $\lambda_i \sim N(\overline{\lambda}, \sigma_\lambda^2)$. The distribution is truncated at zero, for the costs being correctly defined. Given $\overline{\lambda}$ and $\sigma_\lambda^2$, the 15 different values of $\lambda$ are set by the equidistant percentiles, from 2.5 to 97.5, of this distribution. Starting each period from the firm specific prior beliefs, the sequence of optimal price $p_{it}^*$ is drawn from the optimal posterior distribution $f_{it}(p_t|\widehat{p}_t, p_{it-1})$. The simulated moments are then computed by exploiting the cross-section and time series dimension of optimal prices across firms. I target four moments, to estimate the remaining two parameters of the baseline model, $\sigma_L$ and $\phi$, along with the mean and variance of the assumed cost distribution.

## 4.2 Matching Moments

Table 1 reports the data and simulated moments while Table 2 summarize both the calibrated and estimated parameters. The first two targeted moments corresponds to evidence at the cross-sectional level, while the remaining two are related to time series evidence. It is important to mention that $Stdv(Dispersion)$ and $Stdv(Frequency)$ stands for the relative standard deviation of the dispersion and frequency of price changes respectively. The fact that prices

[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).

are fully flexible in the model, impose an additional challenge in matching both the level and the dynamics of these two series. By matching the relative standard deviations, I can focus on the main motivation of the paper: the dynamic evolution of price changes. As stressed in table 1, my baseline - tightly parametrized model - with costly acquisition of information is able to match different moments from the micro price setting literature. Particularly, the model jointly replicates the presence of small and large price adjustments along with the aggregate dynamic evolution of price changes which, as stated in section 2.1, are not easy to reconcile even for price rigidities models. Moreover, the model also fairly replicates other (non-targeted) relevant moments, such as the positive correlation between the dispersion and the frequency of price changes.

Table 1: Matched Moments

| Targeted Moments | Data | Model |
|---|---|---|
| $Prob(\|\Delta p\|) < 5\%$ | 0.443 | 0.433 |
| $Kurtosis(\|\Delta p\|)$ | 6.403 | 6.020 |
| $Stdv(Dispersion)$ | 0.354 | 0.316 |
| $Stdv(Frequency)$ | 0.120 | 0.097 |
| Non-Targeted Moments | | |
| $Median\|\Delta p\|$ | 0.097 | 0.051 |
| $Prob(\|\Delta p\|) < 2.5\%$ | 0.254 | 0.192 |
| $Corr(Dis, Freq)$ | 0.276 | 0.155 |

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 represents 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.

According to the estimated parameters in table 2, the volatility of the price target in the high state increases by 75% with respect to the low state $\sigma_L = 0.097$. The rise in volatility is in line with the estimation for macro uncertainty parameters in Bloom et al. (2014). The average cost of attention is estimated at 0.039, which represents on average a 4% of the total revenues of a firm. Importantly, the standard deviation for the attention cost $\sigma_\lambda = 0.02$ is non-meaningful representing almost half of the average cost. This suggests a significant dispersion across attention levels from firms, which is in line the findings in Coibion et al. (2015) and Kacperczyk, Van Nieuwerburgh and Veldkamp (2016). Particularly, the former paper also assumes heterogeneity in fund manager skills (skills as different attention costs), to rationalize the increase in dispersion during recessions. While linked to the implications in this paper, Kacperczyk et al. (2016) does not fully rely on a dynamic setting.

Table 2: Model Parameters

| Calibrated | Value | Description |
|---|---|---|
| $\beta$ | 0.99 | Discount Rate |
| $\gamma$ | 10 | Curvature of demand function |
| $\tau_{LH}$ | 0.01 | Monthly transition probability: low/high state |
| $\tau_{HL}$ | 0.036 | Monthly transition probability: high/high state |
| Estimated | | |
| $\sigma_L$ | 0.092 | Volatility in low state |
| $\phi$ | 1.75 | Increase in volatility in high state |
| $\overline{\lambda}$ | 0.039 | Mean distribution information cost |
| $\sigma_\lambda$ | 0.02 | Stdv distribution information cost |

## 4.3 Quantitative exploration of the model

With the estimated parameters, I document several stylized facts about price setting decisions and acquisition of information to provide further relevant intuition on the results. Table 3 shows the average and standard deviation of profit loss $\widehat{\pi}_{it}$, the price changes $\Delta p_{it}$, the average level of information capacity $\kappa_{it}$ and the information capacity growth $\Delta\kappa_{it}$, across four different values of the cost of information $\lambda$. The results in the first panel of Table 3 show how firms with lower attention costs, which are able to track the unobserved price closely, attained lower profit losses. With respect to price change decisions, the average price change of firms is zero. This is a consequence of the normal distribution assumed for the target price, which is centered at zero. With respect to the dispersion, there is a clear negative relationship between the cost of attention and the dispersion of price changes. These results provide further insights on why the model replicates the small and large price adjustments documented in the data. The impossibility of closely tracking the target price, lead firms with higher information costs to endogenously allocate their limited attention around the potential realizations of $\widehat{p}_t$ with the highest probability, which lead to small reactions of their optimal prices with respect to firms with lower information costs.

The last two panels of Table 3 provide evidence on the amount of information endogenously collected by firms. While it should be expected that higher information costs crowds-out total acquired information, it is important to assess the evolution of attention growth. On average, the growth of attention and the cost of information share a positive relationship. This behavior reflects one of the implications of allowing the economy to move stochastically across different states. Throughout the transitions, the changes of attention are relatively smoother for low cost firms while they are more abrupt for high cost firms. This is a consequence of the relative difference on the level of acquired information $\kappa_{it}$. By making small adjustment in $\kappa_{it}$, low cost firms can quickly recognize any change of states, while less attentive firms needs to produce significant changes in their acquired information to notice the same change. As stressed be-

Table 3: Stylized Facts

| | All | | Low Volatility | | High Volatility | |
|---|---|---|---|---|---|---|
| Profit Loss | Mean | Stdv | Mean | Stdv | Mean | Stdv |
| $\lambda_1$ | 0.0028 | 0.0014 | 0.0028 | 0.0014 | 0.0027 | 0.0014 |
| $\lambda_5$ | 0.0147 | 0.0043 | 0.0146 | 0.0042 | 0.0148 | 0.0048 |
| $\lambda_{10}$ | 0.0226 | 0.0064 | 0.0226 | 0.0063 | 0.0228 | 0.0066 |
| $\lambda_{15}$ | 0.0381 | 0.0170 | 0.0380 | 0.0169 | 0.0383 | 0.0173 |
| $\Delta p_{it}$ | | | | | | |
| $\lambda_1$ | 0.00002 | 0.0752 | 0.0002 | 0.0604 | -0.0005 | 0.1138 |
| $\lambda_5$ | 0.00002 | 0.0661 | 0.0001 | 0.0535 | -0.0004 | 0.0992 |
| $\lambda_{10}$ | 0.00002 | 0.0619 | 0.0001 | 0.0480 | -0.0004 | 0.0968 |
| $\lambda_{15}$ | 0.00002 | 0.0519 | 0.0001 | 0.0384 | -0.0004 | 0.0845 |
| $\kappa_{it}$ | | | | | | |
| $\lambda_1$ | 1.1883 | 0.1300 | 1.1872 | 0.1287 | 1.1924 | 0.1346 |
| $\lambda_5$ | 0.6826 | 0.0897 | 0.6824 | 0.0894 | 0.6836 | 0.0909 |
| $\lambda_{10}$ | 0.5391 | 0.0919 | 0.5388 | 0.0916 | 0.5400 | 0.0929 |
| $\lambda_{15}$ | 0.3662 | 0.0815 | 0.3658 | 0.0811 | 0.3676 | 0.0832 |
| $\Delta\kappa_{it}$ | | | | | | |
| $\lambda_1$ | 0.0004 | 0.0312 | -0.0020 | 0.0108 | 0.0093 | 0.0632 |
| $\lambda_5$ | 0.0008 | 0.0334 | -0.0020 | 0.0186 | 0.0107 | 0.0616 |
| $\lambda_{10}$ | 0.0013 | 0.0407 | -0.0021 | 0.0257 | 0.0136 | 0.0713 |
| $\lambda_{15}$ | 0.0025 | 0.0486 | -0.0021 | 0.0245 | 0.0196 | 0.0918 |

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.

low, the rate by which firms recognize any change of state is completely disciplined by their idiosyncratic costs.

Further quantitative implications of the model are discussed in the appendix. Firms always face an outside option: set the optimal price at the unconditional mean of $\widehat{p}_t$ and not acquire any extra information. In section 7.4 of the appendix, I show under each of the possible values for $\lambda$, firms are always better off by acquiring costly information relative to the no information case. Moreover, I introduce a natural way to assess the robustness of the estimated costs by bounding their values between the two extreme information cases: Full information and no Information.

# 5 Delayed Learning Dynamics

In this section, I assess the transition dynamics of the model by simulating an exogenous change of state in the economy. Initially, the economy stays in the low state for a sufficient number of months (300 periods) and then, switches to the high volatility state for 28 months. Afterwards, the economy goes back to the low state, where it stays until the end of the simulated period. The length of the recession is set to 28 months, since this is the average duration of the high volatility state according to the calibrated transition probabilities. Keeping the evolution of states constant, I simulate significantly large number of economies 1,000 with 150 firms each, where as in the calibrations, I allocate the 15 different attention costs uniformly across the total number of firms.[12] Finally, I average each variable across economies, for each point in time.

## 5.1 Firm Level Evolution

To gain intuition of what drives the dynamics of price adjustments, I start by solving the attention problem for two different firms. With the solution, I characterize their different "predispositions" $f_{it}(p_t|p_{it-1}) = \sum_\sigma \sum_\epsilon f_{it}(p, \widehat{p}|p_{it-1})$. The presence of idiosyncratic attention costs, cause different pricing reactions due to their impossibility to immediately notice a change of state. To show this, I present the evolution of $f_{it}(p_t|p_{it-1})$ before and after the first change of state. The information cost of each firm are set equal to $\lambda_1$ and $\lambda_{15}$, i.e. the first and the last value of the distribution of $\lambda$, and I labelled them as "low $\lambda$" and "high $\lambda$".

Figure 1 shows the predisposition of these two firms at times: $T-1$, $T$, $T+5$ and $T+10$, where $T$ represents the month where the economy changed from the low to the high volatility state. The evolution for the firm with low $\lambda$ is presented in the in the left panel, while the right panel shows the evolution for high $\lambda$. For expository reasons, I also plot the actual distribution from which the optimal price is drawn at each period (dashed line). The predisposition is computed

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

*after* the firm choose the shape of the joint probability distribution, so it is plotted at the end of each period. At $T-1$, the firm's pricing strategies looks very different. The low cost firm acquired very precise signals about $\widehat{p}_t$, therefore its chosen probability distribution overlaps the distribution of the unobserved target. The discrepancies between the two distributions, are then only attributed to the presence of the information cost.[13] The shape of the predisposition of the high cost firm, reflects its lower attention. This firm does not attach the highest probability mass at zero, it prefers to put more probability weight to prices around the mean. Because of its higher cost, the shape of the predisposition suggests that the firm design its pricing strategy to disentangle if the realization of the target was either negative or positive. While both firms aim to minimize a quadratic loss function, the rational response of facing a high information cost, is to choose a signal which is informative about the "sign" of the target, decreasing the probability of larger errors. Finally, and despite their different costs, both firms are certain that the economy is in the low volatility state since they attach zero probability to extreme realizations of the price target.

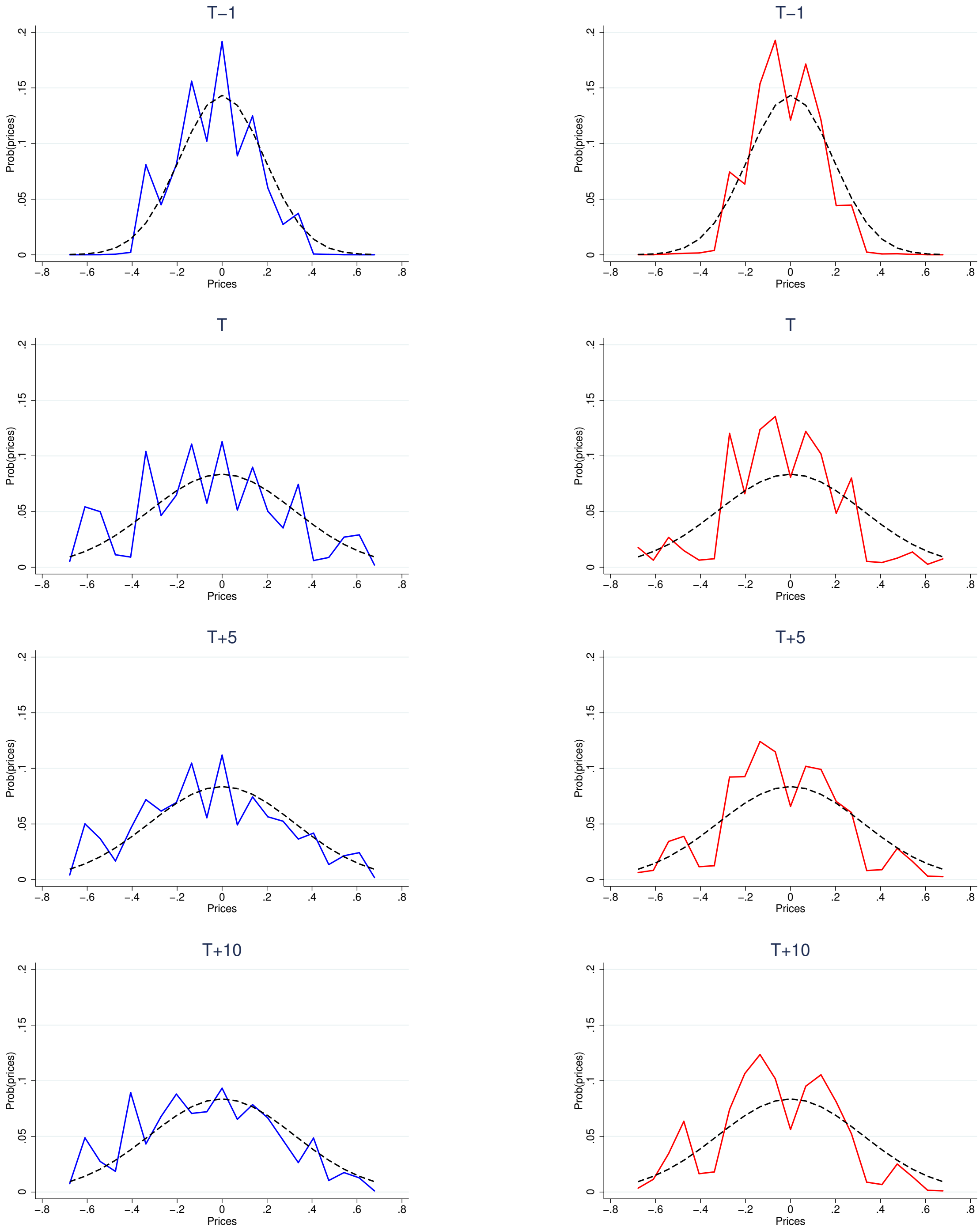
Interesting reactions appears after a the economy moves to the high volatility state at time $T$. The "low $\lambda$" firm immediately starts adjusting it price strategy, by redistributing probability weight towards more extreme realizations of the price target. Five to ten months after the change, this firm has already changed its predisposition completely where again its predisposition overlaps the true distribution of the price target. On the other hand, the reaction of the "high $\lambda$" firm is completely different. By the end of time $T$, the only noticeable reaction of the firm is that it moved more probability weight towards higher prices, but still it attaches almost zero probability to extreme outcomes. Intuitively, this firm keeps behaving as if the economy was still in the low volatility state, and mistakenly attributes the more extreme realization of the target to its own less precise signal. While after 10 months, the high cost firm certainly attached more weight to extreme realizations of the target, still there is a significant discrepancy with respect to the true distribution from which the optimal price is drawn.

The evolution of total amount of attention $\kappa_{it}$, the absolute magnitude of price changes $|\Delta p_{it}|$ and the posterior beliefs of being in the high state $f_{it}(\sigma_H | p_t, \widehat{p}_t, p_{it-1})$ for both the "low $\lambda$" and "high $\lambda$" firms are shown in figure 2. According to the time series evolution of total information, firms endogenously acquire more information under the less predictable state. However, the impossibility to notice immediately a change of state, cause a sluggish reaction in the rate by which firms increase their total attention.[14] While the magnitude of price changes is expected to

[13] Although the distributions may look asymmetric this is mostly due to random issues. The probabilities attached to each price (which determines the shape of the function) are computed by averaging them across the different simulated economies. Since the predispositions are also functions of lagged prices, which are drawn from the conditional density (12), the final shape of the distribution is then also affected by its own price randomness. The higher the number of simulated economies, the more symmetric the distribution should finally look.

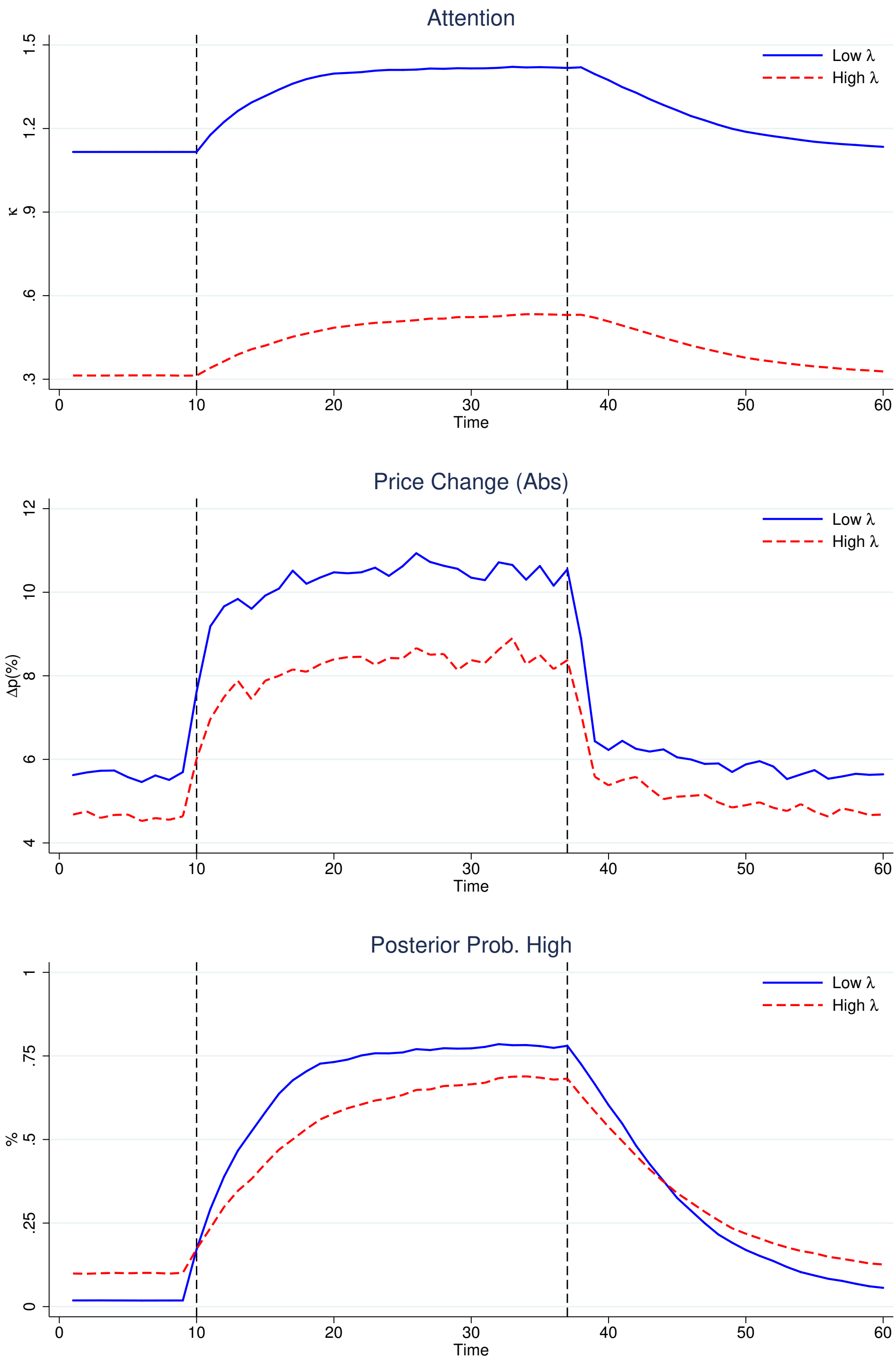
[14] The Rational Inattention model is solved based on natural logs. To measure the total amount of information in "bits", I scaled the total amount of information by $\frac{1}{log_2(exp(1))}$.

Figure 1: Evolution of firm's predisposition




Notes: The two panels show the firms predisposition evolution $f_{it}(p_t|p_{it-1})$ at different moments of time. In the figure, T stands for the month where the economy changed from the low to the high volatility state. 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.

Figure 2: Time varying evolution: Firm Level




Notes: In all the figures, the vertical dotted black lines represents the high volatility episodes. The top figure presents the evolution of total acquired information $\kappa_{it}$ 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 absolute value of price changes $\Delta p_{it}$ while the bottom figure shows the evolution of the posterior probability of the economy being in the high volatility state.

increase under the high volatility state, firms rationally take different pricing decisions after the onset of the recession. There is a noticeable delay in the rate by which firms amend their pricing strategy during the first months of the recession, reflecting their own incomplete information about being in the more unlikely state. Additionally, from the observed behavior, it is expected to observe an increase in the price change dispersion during recessions, as suggested by the bigger difference between the price strategies of the two firms. The bottom panel of 2 closes the intuition behind price setting behavior, by showing the evolution of firms posterior beliefs of being in the high volatility state. Costly information delayed the rate by which firms notice any change of state. The "low $\lambda$" firm attaches higher probability of being in the high state after 5 months into the recession, while the "high $\lambda$" firms takes 9 months. Hence, the information asymmetries, endogenously amplify the transition effects when the economy moves to a less predictable state, evidenced by the different pricing reactions.

It is interesting how the transition dynamics are very different when the economy goes back to the low volatility state. Particularly, after 4-5 months both firms already noticed the change. The asymmetric reaction is again due differences in information costs. Rational Inattentive firms are aware that their collected information is by definition noisy. While during a recession, the probability of extreme realizations of the target price increases, initially firms attributes these higher outcomes to their own limitations to observe information. This is suggested by the evolution of the posterior beliefs of being in the high state as presented in figure 2. Firms need several months to "convince" themselves that the economy actually moved to a state with lower probability of occurrence, which leads them to collect more costly information. Finally, and as a consequence of their higher level of information, when the economy moves back to the more persistent state, they are able to notice the change more quickly, which explains the reported asymmetry.

### 5.2 Aggregate Evolution

To explore the overall implications of costly information through the business cycle, I aggregate the information and pricing decisions across the cross-section of firms. The top and middle panels in figure 6, presents the time series evolution of price disagreement (measured by the inter quantile range) and the frequency of price changes.[15] In line with the empirical evidence on price setting, both series share a positive co-movement over the business cycle.

While the possibility to replicate some nontrivial stylized facts is always desirable to validate the insights of the model, the aggregate results supports a bigger point. The rise in both the dispersion and frequency of price changes, emerges as an endogenous response to the het-

[15]Following Vavra (2013) I measured price change dispersion by conditioning on non zero price changes. In a model with price rigidities this is important to focus on the dispersion across firms whose shocks pushed them outside their inaction regions. In an information model, this is also important since it allows to focus on dispersion among the "price updaters".

erogeneous cost of acquiring information within a dynamic Rational Inattention setting. The baseline set up of the model starts by intentionally ruling out any exogenous and time-varying source of dispersion in the model, to precisely stress the main implications behind endogenous attention. For instance, although the importance of idiosyncratic shocks at the firm level are undeniable, incorporating this feature into the model will certainty made the task of generating dispersion in price changes more straightforward, while obscuring the implications of costly information. Particularly, by assuming $\widehat{p}_{it} = \sigma_t \epsilon_{it}$ I can incorporate idiosyncratic costs within the model setting, but since all firms now tracks different targets, price change dispersion arises as an implication of this assumption.

The baseline model is also consistent with the empirical evidence on time varying evolution of attention. As stressed in the bottom panel of 6, the proportion of firms updating information increase monotonically with the change of state. After 28 months into the recession, around 80% of firms keeps updating their total information on a regular basis. The evidence is in line with the empirical results in Coibion and Gorodnichenko (2015). This evidence suggest 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.

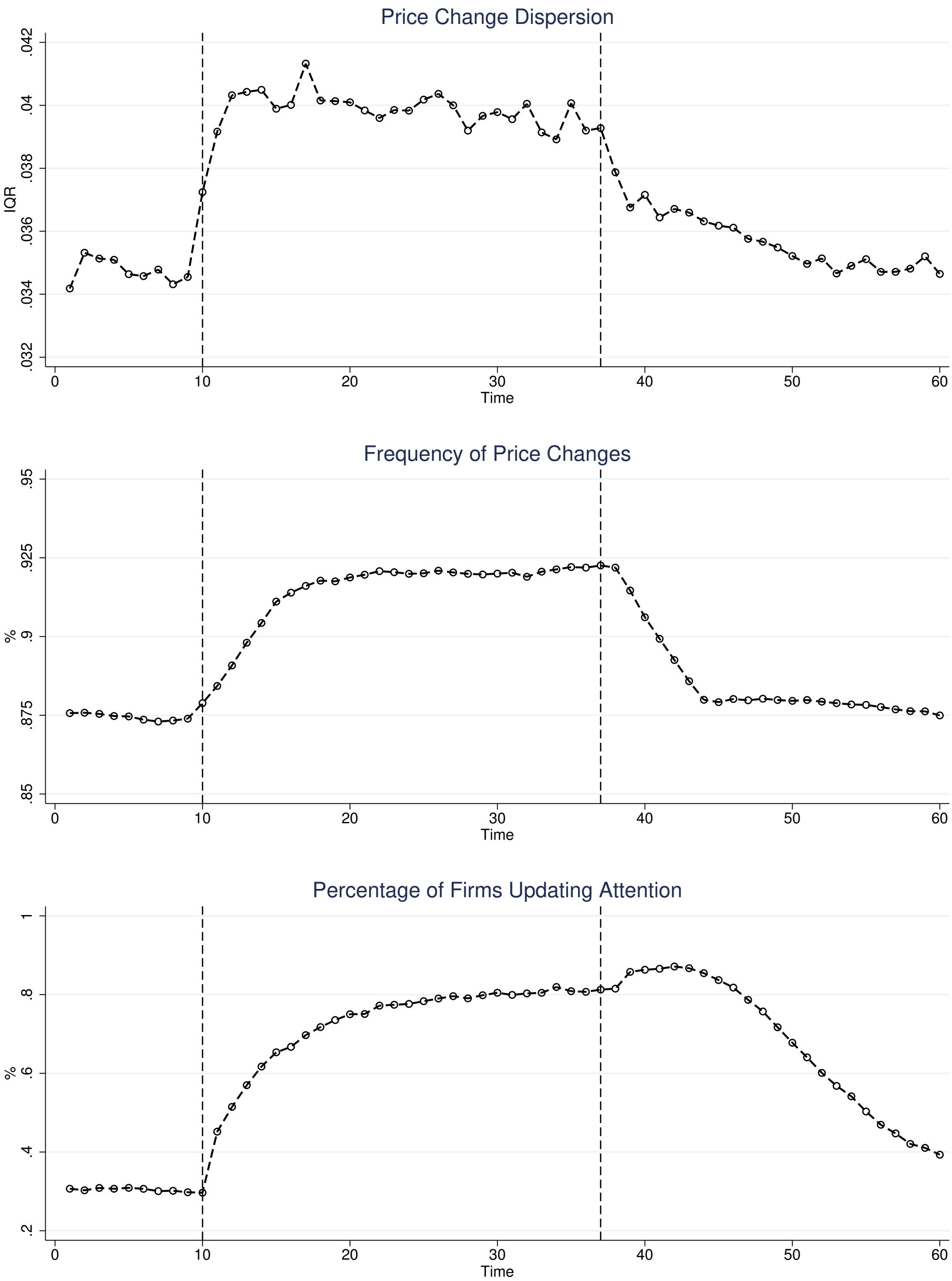
## 5.3 What drives the increase in dispersion?

Although the model replicates the documented evolution of price change dispersion, the main reasons driving its dynamic features are not obvious. Since during a recession, the economy becomes less predictable, it is not clear how to disentangle the effects of exogenous shocks from the firms own endogenous reactions to a change in state. To shed light on this feature, I decompose the total variance of price changes given firm's idiosyncratic information costs.

$$Var(\Delta p_{it}) = E[Var(\Delta p_{it})|\lambda] + Var[E(\Delta p_{it})|\lambda)] \tag{14}$$

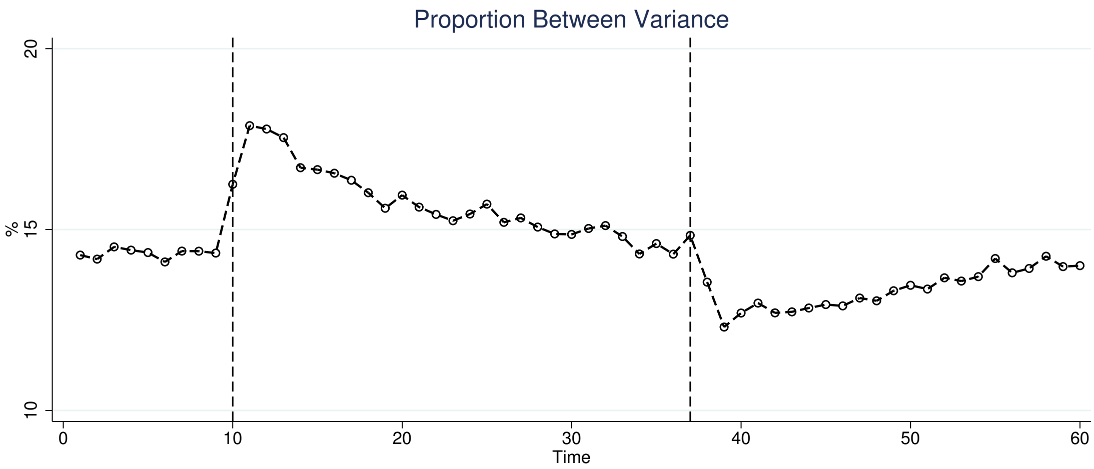
As information costs are randomly assigned, I can group firms given this time invariant feature. The first element on the right hand side of (14) captures the price change dispersion within firms sharing the same information cost while the second element computes the dispersion between firms with heterogeneous costs. Thus, the proportion of the overall variance explained by allowing for heterogeneous costs (between effect) is shown in figure 4. To characterize its time varying evolution, I replicate the same transition dynamics as in the previous section.

Figure 3: Aggregate Evolution




Notes: In all the figures, the vertical dotted black lines represents 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, $\kappa_{it}$.

Figure 4: Variance Proportion - Between Effect




As described in the figure, during the low volatility state the between effect accounts for almost 14% of the total dispersion. At the onset of the recession, and although the overall economy becomes less predictable, the proportion of the dispersion cause by the endogenous response increase to 18%, a non-negligible 25% increase. I interpret the increase in dispersion as a consequence of costly information, since as the economy enters into the recession (and firms already recognize the change of state) the proportion of the between effect starts decreasing. Hence, the evidence suggests that the sole presence of information frictions is enough to amplify the initial detrimental effects of a recession.

## 5.4 Implications of Heterogenous costs

The model have two distinct features that makes it appealing to understand the sources behind price change dispersion: its dynamic nature and the presence of heterogeneous information costs. As explained, dynamic acquisition of information and idiosyncratic costs, are responsible for the delay by which firms notice a change of state and also amplify the dispersion over the cycle. Therefore, in this section I stress the main implications and consequences after shutting down each of these two channels.

### 5.4.1 A Static problem

The baseline model is dynamic due to firms impossibility to observe the persistent state. Alternatively, I can solve a simpler version of the model where I assume that while the target price have the same structure as before, $\widehat{p}_t = \sigma_t \epsilon_t$, firms perfectly observe the current state of the economy $\sigma_t$. Thus, firms acquire costly information to track the i.i.d. shock $\epsilon_t$. The problem becomes a standard static Rational problem with a quadratic objective and Gaussian signals. To make the results comparable with the discretized baseline model, I solve the model by maximizing (5) subject to (6), (7) and (9).[16] The solution of the static problem is:

$$f_{it}(p_t|\widehat{p}_t) \quad = \quad \frac{exp\left[\left(\Pi(p_t,\widehat{p}_t)\right)/\lambda_i\right] f_{it}(p_t)}{\sum_{p'} exp\left[\left[\Pi(p'_t,\widehat{p}_t)\right)/\lambda_i\right] f_{it}(p'_t)} \tag{15}$$

Where the optimal price $p^*_{it}$ is drawn from equation (15). As in the baseline model, $f_{it}(p_t)$ stands for firm's predisposition to set price $p_t$.

### 5.4.2 Homogeneous Information costs

In this case, I asses the implications of allowing for heterogeneous information costs. The model share the same set up as in the baseline setting, but there is dispersion in $\lambda$, i.e. $\sigma_\lambda = 0$.

### 5.4.3 Implications for price change dispersion

Under these two cases, I re-estimate and recompute the target models. The results are presented in table 4. In the static set-up, both the Kurtosis and the relative standard deviation of price change dispersion are closer to the actual moments relative to the baseline case. Higher relative precision to replicate Kurtosis, is also true under the homogeneous costs scenario. Aside from the relative precision to match target moments, these results supports a broader issue: Despite its computational intensity, models allowing for costly acquisition of information, are proven successful in matching non-trivial moments from the price setting literature. Nevertheless, *neither* the static nor the homogeneous cost model can replicate the positive correlation between dispersion and frequency of price changes. The impossibility to generate this result was not obvious ex-ante. Firms with equal information costs still can set different prices, they are drawn from their (homogeneous) posterior beliefs. Moreover, these two alternative settings generates higher attention during high volatility episodes, so I certainly expected time-varying frequencies of price updating. According to the results, the interplay between sluggish pricing reactions (dynamic setting) and dispersed beliefs about the current state (heterogeneous costs) as the economy moves through the cycle, is what endogenously generates the positive co-movement between these two series.

In Figure 5, I compare the evolution of price dispersion between the three different models. To make the evolutions comparable, I re-estimated the simulated transition using the original values for the parameters as of Table 2 and I normalized the first values of the dispersion to 100. The amplification effect of under the baseline model is clear. At the beginning of the recession,

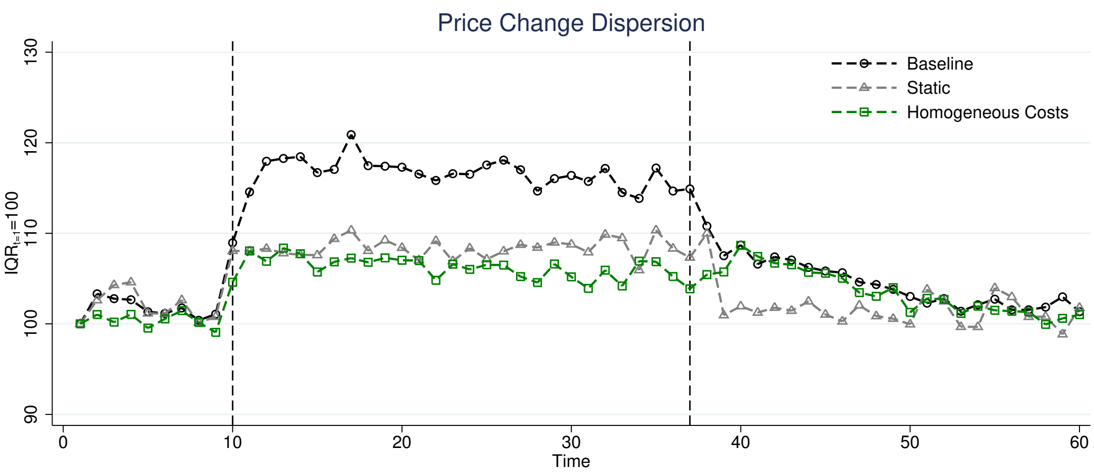
[16] Still, the general results apply since the unconditional distribution from which firms draw their optimal prices is Gaussian, as expected.

Table 4: Moments Alternative Specifications

| Targeted Moments | Data | Baseline | Static | Homogeneous Costs |
|---|---|---|---|---|
| $Prob(\|\Delta p\|) < 5\%$ | 0.443 | 0.433 | 0.401 | 0.497 |
| $Kurtosis(\|\Delta p\|)$ | 6.403 | 6.020 | 6.433 | 6.374 |
| $Stdv(Dispersion)$ | 0.354 | 0.316 | 0.341 | 0.296 |
| $Stdv(Frequency)$ | 0.120 | 0.097 | 0.173 | 0.091 |
| Non-Targeted Moments | | | | |
| $Median\|\Delta p\|$ | 0.097 | 0.051 | 0.039 | 0.048 |
| $Prob(\|\Delta p\|) < 2.5\%$ | 0.254 | 0.192 | 0.258 | 0.209 |
| $Corr(Dis, Freq)$ | 0.276 | 0.155 | -0.009 | -0.075 |

price dispersion increased by 8% and 7% approximately in the static and homogeneous costs respectively, while under the baseline set-up the increase is of 18% approximately. To stress its different implications and levels, 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 5: 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 it not particularly meaningful to discipline the dynamics of price dispersion, it do provides interesting insights on the effects of endowing firms with more (costless) information. As expected, moving towards a scenario with higher information leads to a reduction in price dispersion of 10% relative to the baseline setting. Intuitively, we could think about the static results as a setting where firms receive costless signals about the actual state of the economy, at the beginning of each period. These types of results, can certainty have broader implications for the design of policies aiming to manipulate agents expectations, by providing them with more accurate information.

# 6 Conclusions

This paper stressed how a dynamic model with costly acquisition of information, is enough to endogenously generate countercyclical inefficient price dispersion in line with recent data. By structurally estimating the model, I showed that endogenous attention generates a delay in the rate by which firms noticed any change of state, with interesting dynamics through the cycle. The presence of costly information amplifies the effects of exogenous shocks, where at the onset of a recession a significant increase in the inefficient price dispersion is caused by firms misspecified beliefs.

The dynamic dependence between price dispersion and the frequency of price changes predicted by the model relies heavily on the combination of a dynamic setting with time invariant heterogenous information cost. By shutting down, each of these two channels the model is incapable to rationalized this empirical feature of the data. Moreover, there are some very interesting results that demands for a further exploration in future projects. Particularly, there is an asymmetric response between the rate by which firms recognize a change from the low to the high state, compared to when the economy transits from the high to the low state. While asymmetric responses due to imperfect information have been studied before, Van Nieuwerburgh and Veldkamp (2006), there is no further evidence in the context of Rational Inattention.

The model presented in this paper is tractable enough to be extended to other settings. While I intentionally wanted to rule out any feedback effect between variables, the structure of the model can be fitted into a general equilibrium framework, as the one proposed in Woodford (2009). The main motivation for this paper was to assess the time-varying implications of endogenous information for the inefficient allocation of prices. The results are interesting since they are not only able to replicate newer dynamic features of price changes, it also supports the presence of imperfect information as a relevant constraint agents face when making different economic decisions.

# References


**Afrouzi, Hassan**, "Strategic Inattention, Inflation Dynamics and the Non-Neutrality of Money," 2017.

**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.

**Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka**, "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 2016, *106* (6), 1437–1475.

**Bloom, Nicholas, 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.

**_ , _ , and Saten Kumar**, "How do firms form their expectations? new survey evidence," Technical Report, National Bureau of Economic Research 2015.

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

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

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

**Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp**, "A rational theory of mutual funds' attention allocation," *Econometrica*, 2016, *84* (2), 571–626.

**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.

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

**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**, "The rational inattention filter," *Working Paper*, 2017.

**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.

— , “Rational inattention: Beyond the linear-quadratic case,” *The American economic review*, 2006, *96* (2), 158–163.

**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.

# 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, $Y_t$ represents 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, 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 friction price is given by 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 $\Omega_x$, 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 rely 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 (4).

### 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 3.4. 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 (7). The first order condition of $V(g_t(\widehat{p}))$ with respect to $f_t(p|\widehat{p}_t)$:

$$g_t(\widehat{p})\left[\widehat{\Pi}(p_t,\widehat{p}_t)+\beta V(g_{t+1}(\widehat{p}))+\beta\left[\frac{\partial V(g_{t+1}(\widehat{p}))}{\partial g_{t+1}(\widehat{p})}\times\frac{\partial g_{t+1}(\widehat{p})}{\partial f_t(p|\widehat{p}_t)}\right]\right]$$
$$-\lambda g_t(\widehat{p})[log(f_t(p|\widehat{p}_t))+1-log(f_t(p))-1]-g_t(\widehat{p})\mu(\widehat{p}_t)=0 \tag{16}$$

Where:

$$\frac{\partial g_{t+1}(\widehat{p})}{\partial f_t(p|\widehat{p}_t)}=h(\epsilon)\frac{\partial m_t(\sigma)}{\partial f_t(p|\widehat{p}_t)} \tag{17}$$

The last term on the left hand side of equation (16) $\mu(\widehat{p}_t)$, corresponds to the Lagrange multiplier of the constraint attached to the prior, equation (7).

Equation (17) represents the effect of current information strategy on posterior beliefs. Prior beliefs about the shock $\epsilon_t$ 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.[17]

Since $\partial m_t(\sigma)/\partial f_t(p|\widehat{p}_t)=0$, $g_t(\widehat{p})\geq 0$ and $\lambda>0$, equation (16) then becomes:

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

Where I defined:

$$\phi(\widehat{p}_t)\equiv exp\left(\frac{-\mu(\widehat{p}_t)}{\lambda}\right) \tag{18}$$

By the restriction on the prior:

[17] 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.

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

Combining this expression with (18), and adding the conditioning on lagged prices, we get the expression for the optimal posterior distribution of prices given the unobserved target, (12):

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

The expression for the value function, is then simply given by plugging this expression (10):

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

## 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 validation of 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 $\widetilde{\widehat{\pi}}_t^{FI} = 0$, while $\widetilde{\widehat{\pi}}_t^{NI} = \gamma\sigma_j^2$, where $j = L, H$ depending on the realization of the state. In the case of RI, $\widetilde{\widehat{\pi}}_t^{RI} = \gamma(p_{it}^* - \widehat{p}_t)^2 + \lambda_i\kappa_{it}^*$ which varies according to the stochastic choice of $p_{it}^*$ and hence $\kappa_{it}^*$.[18] Intuitively, the net loss under RI must be within these two extreme cases.

$$0 = \widetilde{\widehat{\pi}}_t^{FI} < \widetilde{\widehat{\pi}}_t^{RI} < \widetilde{\widehat{\pi}}_t^{NI} = \gamma\sigma_j^2 \tag{19}$$

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 for actively tracking information. Table 5 shows that across the different values of $\lambda$ 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 $\gamma$ 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 across states.

Table 5: Information Bounds

| | All | | Low Volatility | | High Volatility | |
|---|---|---|---|---|---|---|
| Net Profit Loss | RI | RI/NI | RI | RI/NI | RI | RI/NI |
| $\lambda_1$ | 0.010 | 0.086 | 0.010 | 0.122 | 0.010 | 0.041 |
| $\lambda_2$ | 0.022 | 0.181 | 0.022 | 0.258 | 0.022 | 0.087 |
| $\lambda_3$ | 0.028 | 0.230 | 0.028 | 0.328 | 0.028 | 0.111 |
| $\lambda_4$ | 0.032 | 0.263 | 0.032 | 0.374 | 0.032 | 0.127 |
| $\lambda_5$ | 0.035 | 0.290 | 0.035 | 0.412 | 0.035 | 0.140 |
| $\lambda_6$ | 0.038 | 0.313 | 0.038 | 0.446 | 0.038 | 0.152 |
| $\lambda_7$ | 0.040 | 0.335 | 0.040 | 0.476 | 0.041 | 0.162 |
| $\lambda_8$ | 0.043 | 0.357 | 0.043 | 0.509 | 0.043 | 0.173 |
| $\lambda_9$ | 0.046 | 0.377 | 0.046 | 0.537 | 0.046 | 0.183 |
| $\lambda_{10}$ | 0.048 | 0.393 | 0.048 | 0.560 | 0.048 | 0.190 |
| $\lambda_{11}$ | 0.050 | 0.413 | 0.050 | 0.588 | 0.050 | 0.200 |
| $\lambda_{12}$ | 0.053 | 0.436 | 0.053 | 0.620 | 0.053 | 0.212 |
| $\lambda_{13}$ | 0.056 | 0.460 | 0.056 | 0.655 | 0.056 | 0.223 |
| $\lambda_{14}$ | 0.060 | 0.495 | 0.060 | 0.706 | 0.060 | 0.239 |
| $\lambda_{15}$ | 0.067 | 0.550 | 0.066 | 0.783 | 0.067 | 0.266 |

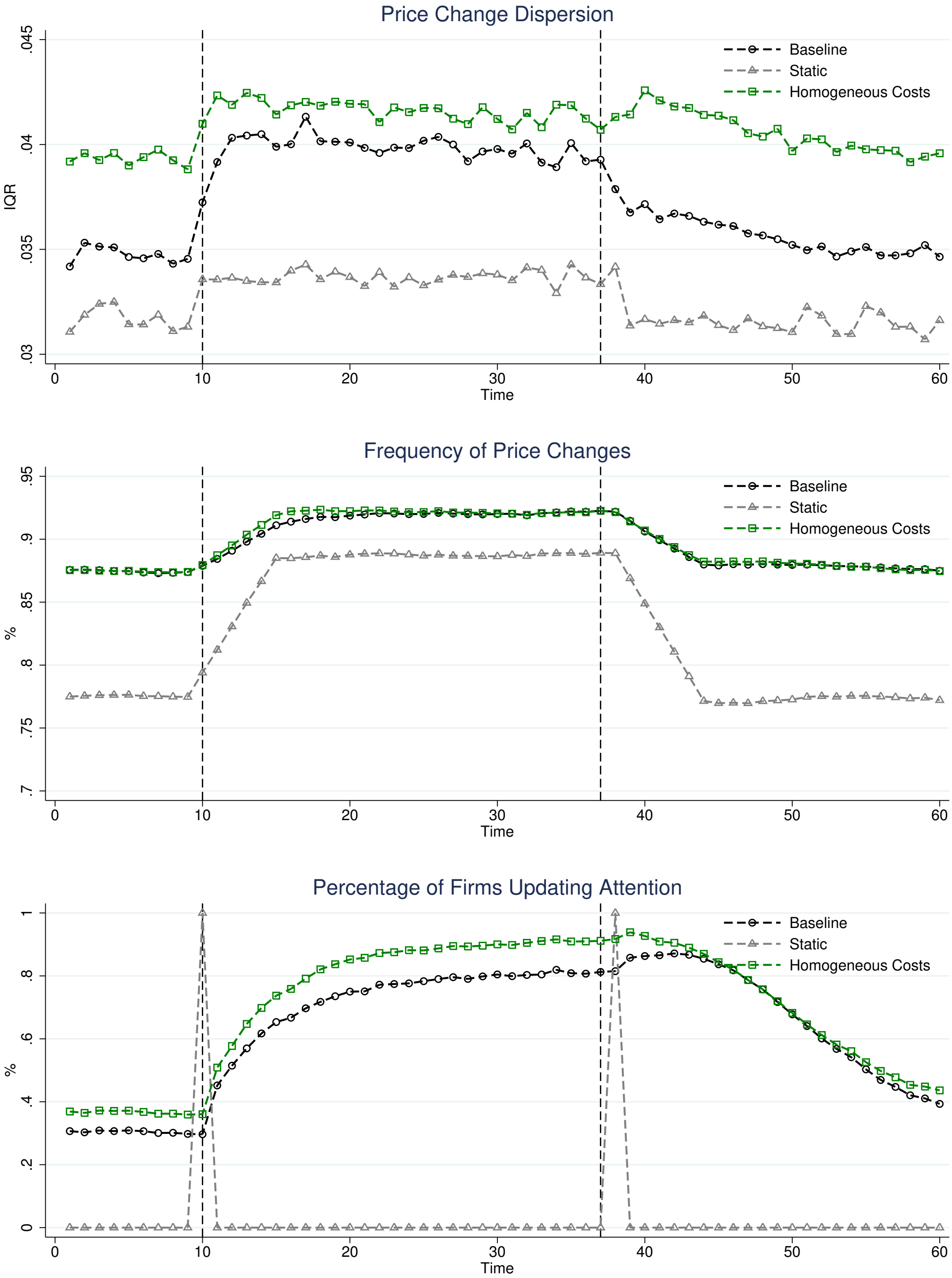
[18] In terms of notation, I introduce "$\sim$" to refer to differentiate the net profit loss function, i.e. the static loss after incorporating the information costs, from the gross profit loss.

## 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 (12). 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 6: Aggregate Evolution




Notes: In all the figures, the vertical dotted black lines represents 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, $\kappa_{it}$. 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.