Rationally Inattentive Seller: Sales and Discrete Pricing

Filip Matějka[1]
CERGE-EI[2]

Abstract

There are often intervals between price changes. This rigidity is usually modeled by an explicit adjustment cost. I show that such dynamics of prices can also represent the optimal actions of price-setters who have difficulty processing information about new shocks. This paper presents a model of a rationally inattentive seller. The model generates a wide spectrum of observed price series properties that sticky-price models cannot explain. The one information constraint implies that prices move back and forth between a few rigid values, sales are short-lasting, or that responses to persistent shocks are sluggish. This is the first pricing model that fully implements rational inattention with no simplifying assumptions on the functional forms of the processed signals.

Keywords: rational inattention, sticky prices, sales.

JEL: D83, E31, E52

[1]I am especially grateful to Per Krusell and Chris Sims for invaluable comments. I would also like to thank Nobuhiro Kiyotaki, Esteban Rossi-Hansberg, Ricardo Reis, Alisdair McKay, Mike Woodford, Ming Yang, Michal Bauer, Michal Kejak and Rick Stock for comments and numerous helpful discussions.

[2]A joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Academy of Sciences of the Czech Republic.

1 Introduction

Macroeconomists study nominal rigidities as sources of the real effects of monetary policy. Models that are built to assess these effects are typically based on explicit assumptions of price stickiness using Calvo-style adjustments or some form of menu cost. Bils and Klenow (2004), however, cast doubts on these assumptions by finding that individual prices do not stay fixed for long periods of time. When the models are calibrated to fit the observed frequency of price changes, then the implied real effects of nominal shocks are very small. Bils and Klenow (2004) thus motivated macroeconomists to focus on prices at the micro level, too.

An alternative line of the modeling of nominal rigidities is based on the assumption that agents cannot attend to all the available information about new shocks. This idea was proposed by Christopher Sims, formulated in a framework called "rational inattention"(Sims, 1998, 2003). I show in this paper that information frictions in the form of rational inattention can in fact generate nominal rigidities with several appealing properties that other models cannot account for. The most important results of this paper are about prices at the micro level.

I present a model where a rationally inattentive seller processes information about a volatile unit input cost and sets the price to maximize his profit. The question is how prices respond to cost shocks. The model generates the following results, which agree very well with the recent empirical literature: (A) Prices do not change all the time, but they do change frequently. (B) Prices tend to change back and forth between exactly the same values. (C) Most price changes are sales-like short-term movements. (D) Prices respond to persistent shocks with a delay.

These results are driven by how the seller processes information about the unit input cost. If the seller was able to observe the input cost, then his pricing strategy would be perfectly flexible. A profit-maximizing price would be set so that the mark-up over the marginal cost would equal the optimal mark-up, which is constant in the model. However, an inattentive agent does not observe the cost, and so might not achieve the optimal mark-up, either. Different assumptions about how knowledge is formed generate different

pricing strategies.

The virtue of this paper, and rational inattention, is its robustness. The model is not based on any explicit assumptions about how the seller allocates his attention. Rationally inattentive agents actively choose how to optimally allocate their limited attention; they do not just passively acquire imperfect information of a given form. The choice then determines the nature of posterior uncertainty. Simply put, such agents pay more attention to exactly the pieces of information that are more important to maximize their objectives. Different objectives can thus drive different allocations of attention and different dynamics of the resulting actions.

Since the seller cannot acquire perfect information, his pricing actions are imperfect and delayed. It turns out that misjudging the input cost when it is low is more costly to the seller, so he pays more attention to shocks leading to low costs, which then implies more flexible low prices and sales-like movements.

Perhaps most interestingly and surprisingly, the rationally inattentive seller chooses to price discretely, i.e. he sets up a price plan consisting of a few prices and charges only one of them even when the input cost is continuously distributed. This implies that prices are likely to stay fixed when cost shocks are small. Although there are no explicit adjustment costs and all functional forms in the model are continuous, the seller chooses to price discretely in order to economize on his information capacity. In other words, considering his cognitive limitations, the seller's optimal strategy is a discretized price plan. The less information the seller processes, the lower the number of different prices he chooses to charge. In the model in this paper, I show the existence of isolated price points analytically, under fairly restricting assumptions, and also observe it in a wide range of numerical examples.

This result provides further evidence that nominal rigidity can, in fact, be driven by forces that are quite different from the explicit adjustment costs in sticky-price models. This distinction might be particularly important when modeling rigidity under non-standard economic conditions. While, for instance, Calvo-style models fix the frequency of price adjustments, the form of nominal rigidity based on rational inattention emerges endogenously and is thus responsive to changes in economic conditions.

The rest of the paper is organized as follows. The following section is devoted to the related literature. Section 3 derives the basic model. Since the notion of attention allocation is still novel, I formulate the model in this section with iid input cost only, which makes the problem essentially a static one and its implications are easier to comprehend. Solutions to the model are studied both analytically and numerically in Section 4. Section 5 then discusses an extension of the baseline model with two input cost shocks differing in persistence, which allows studying the temporal effects of information frictions.

2 Related Literature

I build on Sims' research on rational inattention, which applies the findings of information theory (Shannon, 1948). Information theory is a celebrated concept that studies the limitations of physical communication channels.[3]

My paper is not the first model of pricing that uses rational inattention. Maćkowiak and Wiederholt (2009) were the first to do so. They show that in such a model nominal aggregates do respond to money shocks sluggishly. Their model is more complex than the one presented in this paper,[4] but in order to solve the model they simplify the problem of attention allocation slightly. Simply put, they assume that uncertainty about variables is always Gaussian[5] and that information about different shocks has to be processed separately.

My model differs from that in Maćkowiak and Wiederholt by not assuming any specific form of posterior uncertainty. In fact, this paper addresses Sims (2006), who claims that the nature, not just the quantity, of agents' uncertainty is subject to choice. The model of Maćkowiak and Wiederholt generates a very appealing result of price dynamics at the aggregate level. However, individual prices in their model change too often, in fact, they change all the time. The findings of my paper are thus important, because they show that

[3] See Cover and Thomas (2006) for a good review.

[4] They even solve a DSGE model in Maćkowiak and Wiederholt (2010).

[5] This approach is typical for other important contributions in the literature on rational inattention, including studies of the consumption-savings problem (Luo, 2008) or portfolio choices (Van Nieuwerburgh and Veldkamp, 2010; Mondria, 2010).

when rational inattention is implemented in its unconstrained version, it can also account for evidence at the micro level, unlike sticky-price models.

The discreteness result in this paper is very closely related to the findings in Matějka and Sims (2010), which is a technical study. In that paper, we explore the discreteness of actions under information constraints in a class of tracking problems.[6] Discreteness is a common feature of solutions to tracking problems. It emerges any time the tracked variable has bounded support, but it does not emerge universally. This class does not, however, include the seller's pricing decision problem.

Finally, this study relates to the empirical literature, which follows the findings of Bils and Klenow (2004). The implications of the model are particularly close to the empirical findings in Eichenbaum et al. (2008). The striking feature of their data set is that all prices, including sales, often switch back and forth between exactly the same values,[7] just like in the presented model. The most quoted price of each quarter is often its top price. They also infer that it is the rigidity of the "reference price", the quarter's most frequently quoted price, that is the useful statistic for assessing aggregate nominal rigidity. Eichenbaum et al. (2008) then propose a model with a price plan, which is a finite set of prices between which adjustment is costless, while changing the plan is costly. My model's findings provide motivation for such a setup and also explain why the price plan's rigidity is more closely related to aggregate rigidity than the rigidity of single prices.

[6] With the objective of satisfying $U(x, y) = \hat{U}(x - y)$.

[7] They use the weekly scanner data of a major U.S. retailer.

3 Model

In this section, I first briefly introduce the model's basic building blocks, then I formalize the notion of "the processing of information". After that I state the seller's optimization problem of choosing what information to process and what price to set, and finally I formulate the model's equilibrium in Definition 2.

A monopolistic seller incurs a random unit input cost μ and sets the price p that maximizes the expectation of his profit Π,

$$\Pi(\mu, p) = d(p)(p - \mu), \tag{1}$$

where $d(p)$ is a fixed demand curve the seller faces. The demand curve is a function of price only; it is decreasing and convex.

The non-standard feature of the simple model is that the seller cannot observe the realized unit input cost, but he first needs to process information about it. How the agent allocates his attention influences the nature of his posterior uncertainty about the variable. If the posterior knowledge is given by a perceived distribution with a pdf $k(\mu)$, then the seller chooses the price maximizing the following expectation of profit:

$$p[k] = \arg\max_{\hat{p}} \int \Pi(\mu, \hat{p}) k(\mu) d\mu. \tag{2}$$

Different strategies of attention allocation generate different sets of $k(\mu)$. In this model, we fix the amount of information the seller can process, but we allow him to choose exactly how to process it, which is represented by choosing what "shapes" of posterior knowledge $k(\mu)$ can be realized. This is the main distinction of this model from the earlier literature on rational inattention, where the posterior uncertainty was assumed to be Gaussian. The quantification of the amount of processed information is specified in Section 3.1. This quantification defines constraints on the collections of $\{k(\mu)\}$ that are achievable by the seller.

Naturally, the more information the seller processes the more precise knowledge about μ he acquires, which helps him set the price closer to its optimum. However, in the class of strategies that exhaust the seller's information capacity, he chooses the one that generates forms of posterior knowledge that are the most favorable with respect to his

objective $\Pi(\mu, p)$. For some intuition on the attention allocation before the main model is formulated, see an example in Appendix A.

Choosing what to pay attention to is in the rest of the section formalized as optimization over the signaling devices that are available to the agent. Such devices are described by the distributions of their signals s and the collections of posterior knowledge they generate conditional on these signals $\{f(\mu|s)\}$. Once some posterior knowledge is realized, then the seller sets the price to $p[f(\mu|s)]$ according to (2). The seller thus chooses how to process information, which is described by a joint distribution of s and μ, while considering what profit outcomes it leads to, through (1) and (2).

Time is discrete. The unit input cost is i.i.d., it is drawn from the same distribution in each period. When the cost is i.i.d., then no knowledge about the current input cost carries over to the next period.[8] Although the time series of prices is simulated, this model is essentially static: it is a repetition of the one-period model. The timing of events within each period is as follows.

1. The unit input cost μ is drawn.
2. The seller processes information about the realized μ.
3. The seller sets the price p to maximize the expectation of the profit given his knowledge about μ.
4. The amount given by $d(p)$ is sold.

3.1 Information processing

This part establishes what forms of uncertainty are achievable by the seller who processes a given amount of information. The rational inattention framework applies the results of information theory, which was introduced in Shannon (1948). Information theory provides an understanding of what information can be passed through channels that transmit blocks of symbols at a limited rate. Therefore, the theory describes what the rationally inattentive seller is able to find out about the unit input cost if he can inspect only a limited number

[8] I discuss an extension with a serially correlated unit input cost in Section 5.

of digits of various cost indicators in his store, or if he can ask his subordinates a limited number of questions with answers of yes or no, or if he is constrained to read a text report of a given limited length, etc.

The agent's information processing mechanism is determined by the channel and the coding he uses. In other words, the mechanism describes what signaling devices the agent pays attention to. Coding is the meaning assigned to each received symbol. For instance, if the agent asks questions with answers of yes or no, then the code specifies what question generates a particular symbol. The seller's mechanism of information processing can thus describe what specific questions the seller asks or what digits of what indicators he looks at.

Ultimately, the information processing mechanism controls to what the agent pays attention. Appendix A provides an example of a seller who can ask only a single question. The seller knows the unit input cost is drawn from a uniform distribution between \$1.0 and \$2.0. Asking "Is the cost below \$1.2?" pays more attention to low cost than asking "Is the cost below \$1.8?".

One extremely useful implication of information theory is that we do not need to specify which information channels the seller uses. It turns out that we can classify all of them by a single quantity: information capacity. Given this quantity we can state exactly what knowledge about the unit input cost is achievable and what is not. For such a classification, information theory uses the concept of entropy.

The entropy of the distribution of a random variable is the fundamental measure of uncertainty about the variable. Before the seller starts processing information about the realized unit input cost, he possesses some prior knowledge about it. We assume the prior knowledge coincides with the true distribution from which μ is drawn, let its pdf be $g(\mu)$. Entropy H of the distribution is the following quantity :

$$H\left[g(\mu)\right] = -\int g(\mu)\log g(\mu)d\mu. \tag{3}$$

The more concentrated the distribution is, the lower the entropy and the better knowledge the agent possesses. To further lower the distribution's entropy, i.e. to improve the knowledge, the seller needs to receive signals on the realized value of μ, i.e. he has to

process some information.

The information content carried by a signal is the expected reduction of entropy due to an observation of a signal. It is called the "mutual information" between the random variable and the signal. Letting s be a signal, the mutual information is:

$$I[\mu; s] \quad = \quad H\left[g(\mu)\right] - \mathbb{E}_s H\left[f(\mu|s)\right], \tag{4}$$

where $H\left[f(\mu|s)\right]$ is the entropy of the posterior distribution conditional on the observation of s.

Any information processing mechanism can be formally defined by the resulting $f(s|\mu)$; what signals are transmitted, and received, given the realized value of the random unit input cost is μ. $f(s|\mu)$ together with the prior $g(\mu)$ form the posterior knowledge $f(\mu|s)$ through Bayes law.

$$f(\mu|s) = \frac{f(s|\mu)g(\mu)}{\int f(s|\mu)g(\mu)d\mu}. \tag{5}$$

Using these pdf's[9] we can express the mutual information (4) in the following form as a functional of the joint distribution of μ and s.

$$I[f(\mu, s)] = \int f(\mu, s) \log\left(\frac{f(\mu, s)}{g(\mu)f(s)}\right) d\mu ds, \tag{6}$$

where the joint pdf $f(\mu, s) = f(s|\mu)g(\mu)$.

It is the coding theorem, one of the cornerstones of information theory, that states the sufficient and necessary conditions for what information can be passed through an information channel. The theorem says that any information channel can transmit any message with information content less than the channel's capacity; the agent only needs to use different codes, i.e. different sequences of questions. If the seller's information capacity per period is κ, then $f(s|\mu)$ is achievable with arbitrary precision by some mechanism of processing information if and only if

$$I[f(\mu, s)] \leq \kappa. \tag{7}$$

Summary: Choosing how to process information is equivalent to selecting the conditional distribution $f(s|\mu)$. $f(s|\mu)$ together with the prior knowledge about the unit input

[9] To be mathematically precise, we should use probability measures instead of pdf's. However, the only difference would be the exposition.

cost $g(\mu)$ form $f(\mu, s) = f(s|\mu)g(\mu)$, which in turn determines the posterior knowledge $f(\mu|s)$ conditional on any received signal s. The seller observes the signal s and infers μ. Typically, he would like to receive signals generating posteriors $f(\mu|s)$ that are Dirac delta functions: that would represent perfect knowledge about μ. However, the channel's information capacity limits how tight the signals can on average be. Using the coding theorem, we know that the average reduction of entropy of the seller's knowledge in one period cannot exceed the seller's information capacity κ.

3.2 Choosing how to process information and what price to set

Now we are almost ready to state the seller's problem. A rationally inattentive agent chooses: 1. how to process information through a channel of a limited information capacity, and 2. how to respond to the realized posterior knowledge.

1. There exists a mechanism of processing information for any $f(s|\mu)$ that the seller chooses to achieve, as long as it satisfies (7).

2. Once a particular signal on μ is realized, the agent chooses an optimal response, $p = \tilde{P}(s)$, maximizing the expected profit.

$$\tilde{P}(s) = \arg\max_{p} \int \Pi(\mu, p) f(\mu|s) d\mu, \tag{8}$$

   where the posterior knowledge $f(\mu|s)$ is given by Bayes law (5).

The two decisions, 1. and 2., are not independent. While deciding on the optimal mechanism of processing information, $f(s|\mu)$, the agent is aware of his policy function, $\tilde{P}(s)$. Choosing how to process information thus takes the following form:

$$f(s|\mu) = \arg\max_{\hat{f}(S|\mu)} E[\Pi] = \arg\max_{\hat{f}(s|\mu)} \int \Pi\Big(\mu, \tilde{P}(s)\Big) \hat{f}(s|\mu) g(\mu) d\mu ds, \tag{9}$$

subject to (7) and (8). Notice that $f(s|\mu)g(\mu)$ is the perceived probability density of a realization of $\Pi\Big(\mu, \tilde{P}(s)\Big)$.

Since the demand function $d(p)$ is a decreasing and convex function of price, then the profit function is single-peaked and (8) has a unique solution. Given a signal, there is

only one optimal price. We can therefore equivalently describe the information processing mechanism in terms of the resulting distribution of prices that the received signals lead to, instead of in terms of the distribution of signals. We thus substitute $f(s|\mu)$ by $f(p|\mu)$, where $p = \tilde{P}(s)$. The whole optimization problem can be formulated in terms of a joint distribution of μ and p.

The seller's rational inattention problem takes the following final form, which describes the simultaneous choices of the information processing mechanism and the pricing response to the posterior knowledge.

Definition 1. ***The decision strategy of a rationally inattentive seller.*** *Let $g(\mu)$ be the sellers's prior knowledge about the unit input cost, κ be the seller's information capacity and $\Pi(\mu, p)$ be the profit function. His decision strategy $f(\mu, p)$ is a solution to the following maximization problem:*

$$f(\mu, p) = \arg\max_{\hat{f}(\cdot,\cdot)} E[\Pi(\mu, p)] = \arg\max_{\hat{f}(\cdot,\cdot)} \int_{\mu} \int_{p} \Pi(\mu, p) \hat{f}(\mu, p) d\mu dp, \tag{10}$$

subject to

$$\int_{p} \hat{f}(\mu, p) dp = g(\mu) \qquad \forall \mu \tag{11}$$

$$\hat{f}(\mu, p) \geq 0, \qquad \forall \mu, p \tag{12}$$

$$I[\hat{f}(\mu, p)] \leq \kappa. \tag{13}$$

(11) requires consistency with prior knowledge, (12) states the non-negativity of a probability distribution, and (13) is the information constraint.

Definition 2. ***Equilibrium.*** *Let the unit input cost be i.i.d. In each period t, the unit input cost is drawn from a distribution whose pdf is $g(\mu)$. The equilibrium of the model is the joint distribution $f(\mu, p)$ and stochastic processes $\{\mu_t\}$ and $\{p_t\}$, such that:*

1. *$f(\mu, p)$ is a solution to* (10)-(13),
2. *unit input costs $\{\mu_t\}$ are drawn from $g(\mu)$,*
3. *prices $\{p_t\}$ are drawn from $f(p|\mu_t)$, and*
4. *amounts sold are $\{d(p_t)\}$.*

4 Solving the model

This section starts with a couple of analytical insights into the solution's properties. The optimal strategies of a rationally inattentive seller are dispersed about the perfect information strategy: the higher the seller's information capacity, the more concentrated the strategies are. These insights are based on the first order conditions for the optimal achievable form of posterior knowledge.

Then I show, under relatively strong assumptions, that the support of the resulting distribution of prices consists of isolated points. This feature of the solution generates rigidity of prices at the micro level. Finally, I express how profit losses from misjudging the realized μ vary with the level of μ, and I find that the seller faces higher losses when the unit input cost is low. The seller thus allocates more of his information capacity to lower input costs, which results in finer responses to such realizations of shocks. The second part of this section is devoted to numerical solutions, which allow us to study the quantitative properties of the seller's pricing strategies and the resulting time series of prices.

The seller chooses the optimal strategy, which is a solution to (10)-(13). The demand function takes the following form

$$d(p) = p^{-\theta}, \tag{14}$$

where θ is price elasticity. The unit input cost is continuously distributed, typically uniformly, over a bounded interval. The solution to the seller's problem (10)-(13) takes the form of a joint distribution $f(\mu, p)$, which summarizes both the decisions on how to process information and how to respond to signals.

The information constraint (13) is binding whenever the seller's information capacity is finite. Finite information capacity does not suffice for the seller to perfectly observe the unit input cost, since the cost is continuously distributed over an interval. If $\kappa = \infty$, (13) is not binding and the realized μ is known with certainty. Observing the cost is equivalent to acquiring posterior knowledge with all probability mass concentrated at the true value of μ. The conditional $f(p|\mu)$ is degenerate at the profit-maximizing optimal price $p = p_{opt}(\mu)$,

$$p_{opt}(\mu) = \frac{\theta}{\theta - 1}\mu. \tag{15}$$

On the other hand, if $\kappa < \infty$, then the seller acquires only imperfect posterior knowledge, upon which he makes the pricing decision.

First, we can derive the first order condition for the posterior distribution.

Proposition 1. *If $\kappa < \infty$ and if $f(p) > 0$, then the first order condition states:*

$$f(\mu|p) = h(\mu)e^{\Pi(\mu,p)/\lambda}, \tag{16}$$

where $h(\mu) = e^{-\nu(\mu)/\lambda}g(\mu)$ is independent of p, $\nu(\mu) \in L^{\infty}(\mathbb{R}^N)$ is the Lagrange multiplier on the consistency with the prior (11)*, and $\lambda \in \mathbb{R}$ is the multiplier on the information constraint* (13)*.*

Proof: Appendix B.

Multiplying by $f(p)$, the first order condition (16) also implies the following for the joint pdf of μ and p:

$$f(\mu, p) = h(\mu)f(p)e^{\Pi(\mu,p)/\lambda}. \tag{17}$$

It is not possible to solve for $h(\mu)$ and $f(p)$ in general. However, (17) is useful to grasp some understanding of the properties of the optimal strategy.[10] The seller chooses what pieces of information about cost μ to process based on: i) what he knew in advance, which is given by $g(\mu)$ and ii) the relative importance of various pieces of information given by the shape of the profit function $\Pi(\mu, p)$. Exploring (17), we find:

Corollary 1. *The seller's strategy $f(\mu, p)$ is concentrated in the regions of high profit $\Pi(\mu, p)$, the more so the lower λ is, which is when the information capacity κ is high. High κ allows the seller to form more precise posterior knowledge. When λ goes to zero, then $f(\mu, p)$ collapses on $p = p_{opt}(\mu)$, which is given by* (15)*.*

Moreover, since Proposition 1 implies that $h(\mu) > 0$ whenever $g(\mu) > 0$, then we can infer from (17) that the optimal strategies are not deterministic.

[10] Using this formula, we could infer that quadratic objectives together with Gaussian priors generate Gaussian posteriors as the optimal ones. The rational inattention problem simplifies significantly. Moreover, the variance of all posteriors equals $2^{-2\kappa}\sigma^2_{prior}$. See Sims (2003) and Maćkowiak and Wiederholt (2009).

Corollary 2. *If $\kappa < \infty$, then*

A) the posterior knowledge $f(\mu|p)$ has the same support as the prior $g(\mu)$,

B) the support of the conditional distribution of prices $f(p|\mu)$ is the same as the support of the overall distribution of prices $f(p)$.

This corollary states that to allocate his information capacity efficiently, the seller never acquires signals that rule out some values of input cost with certainty. As long as the information constraint is binding, all posterior distributions overlap completely and all prices in the unconditional support can be realized for all input costs, but with different conditional probabilities. This can also be seen by expressing a variation of the entropy of the posterior, $-\int f(\mu|p)\log f(\mu|p)d\mu$. It is infinite if $f(\mu|p) = 0$ on a set of positive measure, since the derivative of $d(-x\log x)/dx = \infty$ at $x = 0$. The marginal change of the entropy is infinite, and therefore the marginal value of the unit of such information completely ruling some values of μ is for a bounded profit function zero.

The next proposition states that even though the unit input cost is continuously distributed, the resulting distribution of prices can consist of isolated points only.

Proposition 2. *If the unit input cost is uniformly distributed in $[\mu_{min}, \mu_{min} + \Delta]$, where $\Delta < \mu_{min}/(1+\theta)$, then there exists $\kappa_0 > 0$ such that for all $\kappa < \kappa_0$ the distribution of prices is discrete.*

Proof: Appendix D.

The discreteness of prices is a striking feature that arises despite the fact that all functional forms appearing in the model are purely continuous. It turns out that the first order condition, (16), can hold in isolated points only. Under the assumptions above, I actually show that there exists κ that generates at most two price points. It is quite possible that a stronger claim regarding the discreteness existence could hold, such that the price points are isolated whenever the prior distribution is bounded. In fact, the discreteness emerges in all numerical solutions.

The discreteness result is related to the findings in the technical study Matějka and Sims (2010). That study proves the existence of discreteness under constraints on information

capacity in a wide class of tracking problems, where the objective has the following form: $U(x, y) = \tilde{U}(x - y)$. Matějka and Sims (2010) show that if the prior distribution of x is bounded, or has sufficiently thin tails, then the distribution of the action variable y is always discrete. This class of tracking problems does not, however, include the seller's pricing decision problem presented here.[11]

Discreteness occurs as the optimal response to the shape of prior distribution. The agent would like to acquire the posterior with a specific form of noise, given by the first order condition (16). The form of noise depends on the shape of the profit function. However, if the prior distribution is bounded, signals close to the bounds have a very different form of noise from those in the middle of the range of input costs. The agent knows the bounds a priori, so no noise carries over outside of the range of costs. Such signals with no noise in some regions are more costly. The agent therefore chooses signals leading to prices further away from the optimal limiting prices.

The first order condition (16) cannot hold on an interval of prices. It would be too costly to choose continuously changing posteriors and make them consistent with the prior.[12] Since entropy is a concave function of the underlying distribution, collapsing together two posterior distributions leading to nearby prices generates higher posterior entropy and thus lower information flow. The discreteness occurs when the gain from this economizing on information outweighs the loss from a further departure from the optimal prices. The seller saves on information capacity, which he can use elsewhere, by simplifying his decision problem through considering only a finite number of price points.

The final proposition of this section is concerned with the optimal allocation of attention across the levels of the unit input cost. The seller improves his knowledge in regions where the marginal profit from processing extra information is the highest and the seller's

[11] The discreteness of prices can also be proven in another setup, which is close to the seller's problem. It can be shown analytically that if the seller maximizes the expectation of the logarithm of the profit, $\log(\Pi(\mu, p))$, instead of pure $\Pi(\mu, p)$, then complete discreteness occurs even without any additional assumptions. The logarithm together with the exponentials in (16) turn the first order condition into an equation with polynomials, which makes the proof simpler.

[12] This discussion does not apply, for instance, to quadratic objectives together with normally distributed priors. The optimal posteriors given by (16) are Gaussian, and they sum up to the prior for the Gaussian $f(p)$.

responses to μ in these regions are therefore relatively more precise. Losses in profit due to a small departure of Δp from the optimal price are equal to $1/2\frac{d^2\Pi(\mu,p)}{dp^2}\Delta p^2$, where the price deviation Δp equals $\frac{dp_{opt}(\mu)}{d\mu}\epsilon$, if the seller misjudges the unit input cost only by a small amount ϵ.

Proposition 3. *The loss in profit due to misjudging the unit input cost μ by a small amount ϵ is equal to $L(\mu)\epsilon^2/2$, where the loss factor $L(\mu)$ is the following quantity:*

$$L(\mu) = -\left(\frac{dp_{opt}(\mu)}{d\mu}\right)^2 \frac{d^2\Pi(\mu,p)}{dp^2}\bigg|_{p=p_{opt}(\mu)}. \tag{18}$$

To maximize expected profit. The seller processes more information about the regions of μ, where $L(\mu)$ is relatively high.

Proof: Appendix C.

Corollary 3. *For the profit function $p^{-\theta}(p-\mu)$, the loss factor* (18) *takes the form:*

$$L(\mu) = \theta\left(\frac{\theta}{\theta-1}\right)^{-\theta}\mu^{-\theta-1}, \tag{19}$$

which is proportional to $\mu^{(-\theta-1)}$. Given the demand elasticity θ and the amount of error ϵ, profit losses are larger for lower μ. The seller chooses to pay more attention to low unit input costs.

When the unit input cost is low, then the seller has a chance to generate high profit by selling a high amount. Moreover, charging a low price is relatively riskier, since the seller could end up selling a high amount at a small or even a negative margin. The seller could thus lose relatively more of his potential profit by deviating from the optimal prices when the cost is low. In the following section, we discuss that this result drives the fact that the highest price (regular price) is more rigid than the low prices (sales). It is not efficient for the seller to pay much attention to small variations in the unit input cost when it is high.

The presented model is intentionally very simple in order to be as little stylized as possible. It is only the unit input cost that influences the changes of the optimal price. However, most of the paper's results would hold at least qualitatively even for various other shock specifications. The seller processes information to target the stochastic optimal price. Pricing would most likely be imperfect, discrete, and asymmetric even when the shock

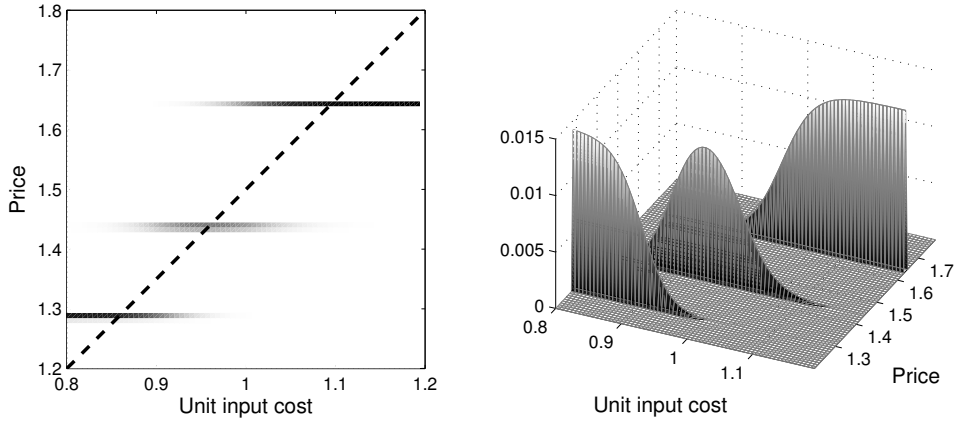

Figure 1: Joint distribution: $\theta = 3$, $\kappa = 1$ bit, uniformly distributed unit input cost

specification was selected differently or if, for instance, the optimal price varied due to the desire to discriminate among consumers.[13]

4.1 Numerical solutions

Complete solutions to (10)-(13) can be obtained numerically, since the problem is a maximization of a linear objective on a convex set.[14] The 2D domain of μ and p was discretized by introducing 70×70 grid points.

Figure 1 shows the optimal joint distribution $f(\mu, p)$ plotted in two different ways. It is a solution to a setup with $\theta = 3$, $\kappa = 1$ bit, and the unit input cost uniformly distributed over $(0.8, 1.2)$. Figure 2 a) presents the corresponding simulated time series of the resulting prices, with i.i.d. unit input cost drawn from the prior distribution in each period. Exhibit b) shows the time series for $\theta = 10$, $\kappa = 0.5$

Let us first inspect the basic properties of the solution presented in Figure 1. The dashed line in the graph on the left of the figure represents the optimal pricing strategy $p_{opt}(\mu) = \frac{3}{2}\mu$ arising under perfect information, (15).

[13] See an earlier version of this paper at http://iweb.cerge-ei.cz/pdf/wp/Wp408.pdf, with a section on demand shocks.

[14] I used a solver called LOQO, Vanderbei (1999), which is based on interior point methods. These methods are efficient for large scale convex optimization problems such as this one. However, standard descend-based methods would work, too.

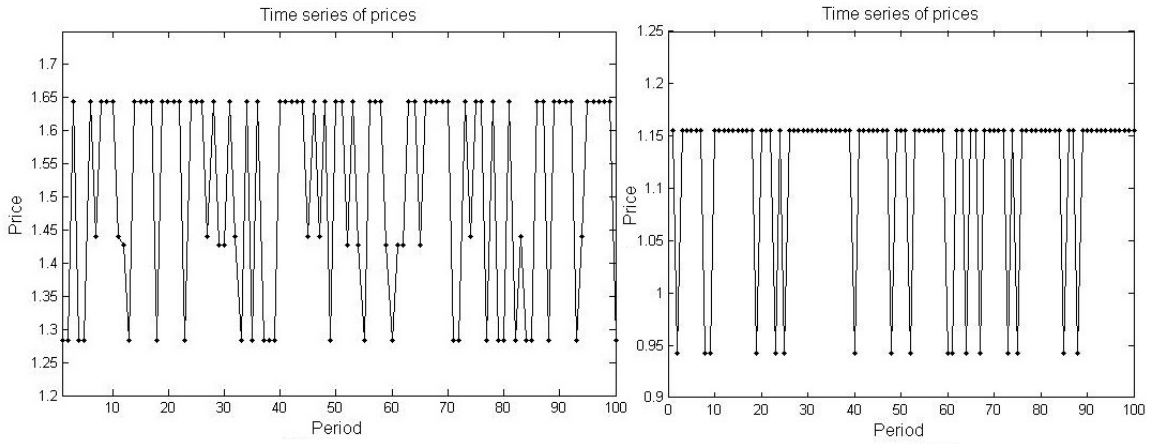

Figure 2: Simulated time series of prices: a) $\theta = 3$, $\kappa = 1$, b) $\theta = 10$, $\kappa = 0.5$

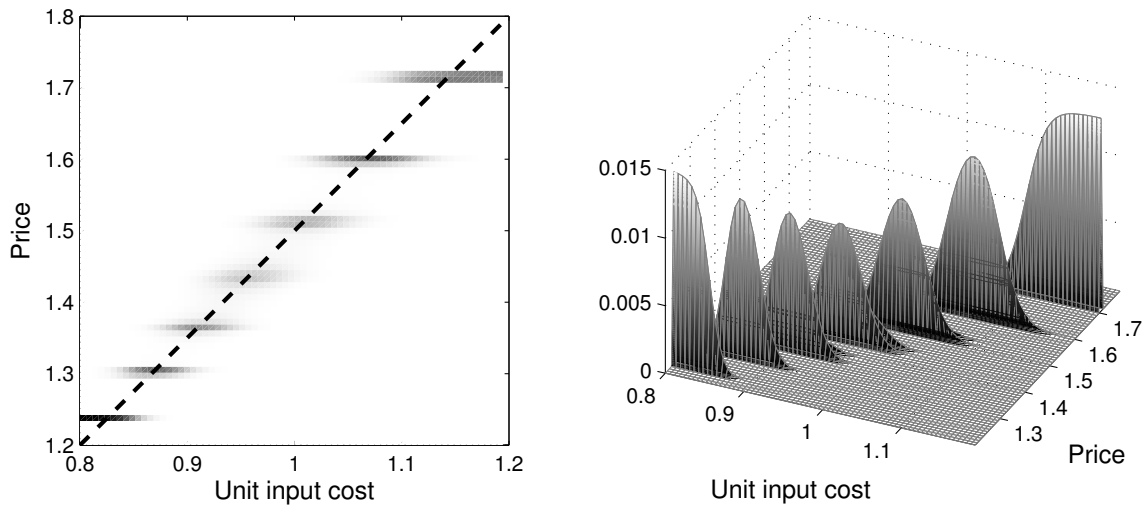

Figure 3: Joint distribution: $\theta = 3$, $\kappa = 2$ bits

The joint distribution is somewhat concentrated about the optimal pricing strategy, as Corollary 1 suggests. Figure 3 shows the solution for $\kappa = 2$; the higher the information capacity, the closer the pricing is to the optimal strategy, i.e. to the case when μ is observed by the seller.

The humped-shaped curves on the right of Figure 1 represent various realizations of posterior knowledge $f(\mu|p)$. The knowledge is imperfect. In fact, Corollary 2 states that these distributions overlap completely, which also implies that all prices that are realizable for one unit input cost can be set when another unit input cost is drawn.

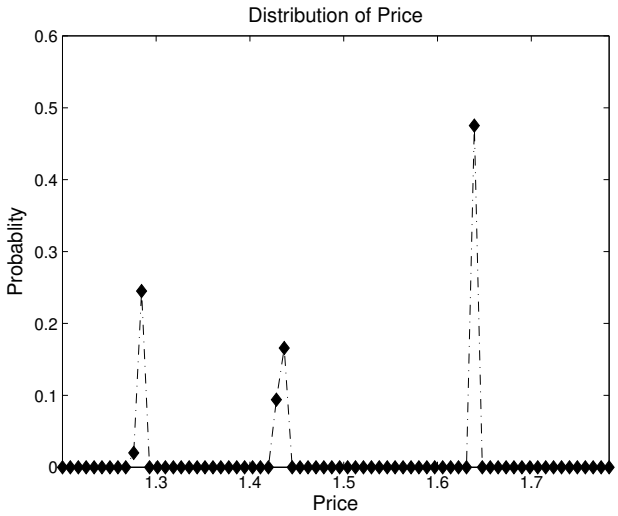

Figure 4: Distribution of prices, $\theta = 3$, $\kappa = 1$ bit

Rigid price values: What we can also see in Figure 1 is that the seller chooses to process information in such a way that only three different forms of posterior knowledge $f(\mu|p)$ are realized. Although the unit input cost is continuously distributed, the seller always charges a price in only three narrow regions. It is probably just a feature of the numerical solutions, which are not exact, that some of these regions consist of two grid-points and not just one. This is the pure discreteness that is claimed in Proposition 2. The distribution of prices is shown in Figure 4; it is the marginal of $f(\mu, p)$. For low input costs, the seller is most likely to realize the posterior knowledge on the left in the right exhibit of Figure 1. That input cost is most probably somewhere between 0.8 and 0.95. This posterior leads to a choice of $p = 1.28$. Higher input costs are likely to generate one of the two other signals, which lead to $p = 1.43$ and $p = 1.64$.

Rationally inattentive agents have complete freedom in acquiring signals they find useful as long as their information capacity is not exceeded. The seller can, for instance, choose to receive a signal with Gaussian noise, which would generate continuously distributed prices. Or he can observe the first digit of the unit input cost of a related product. The first digit of another product's unit input cost takes a finite number of values. Such an information processing mechanism would thus form a finite number of distinct posteriors, just like in Figure 1. The seller in our model endogenously decides to process information in a way that is more similar to this second example.

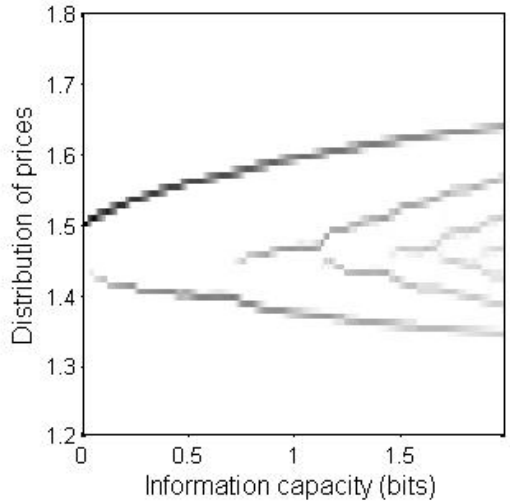

Figure 5: Bifurcation diagram of prices, $\theta = 3$

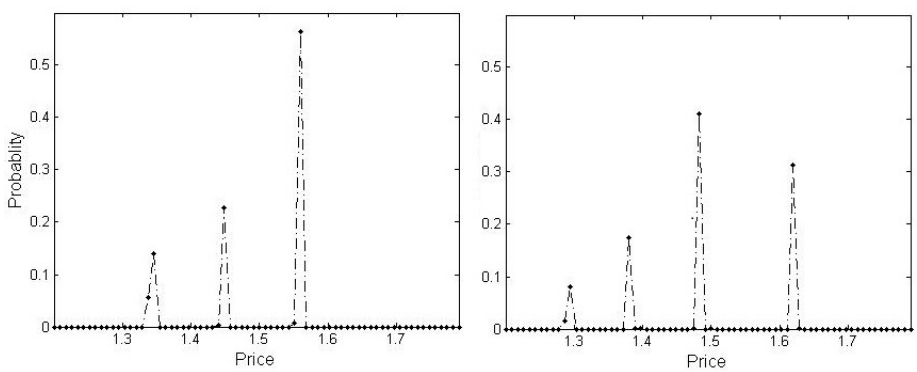

Figure 6: Distribution of prices for triangular distribution of costs, $\theta = 3$, a) $\kappa = 1$ bit, b) $\kappa = 0.5$ bit

The more information the seller processes, the finer the discretization. The bifurcation diagram in Figure 5 shows price distributions as a function of information capacity. When the capacity increases, new price points emerge. There is one price only when $\kappa = 0$, the second price emerges immediately as $\kappa > 0$, the third price at about $\kappa = 0.7$, etc.

Reference prices: We also find that the optimal pricing is asymmetric. These frequencies of the occurrence of different prices are not the same. The top price is usually quoted most often, while lower prices appear to be more flexible. Eichenbaum et al. (2008) call the most quoted price the "reference price". In Figure 4 we see that the top price is quoted 50% of the time, while the probability of the other two prices is about 25%. Corollary 3 states that

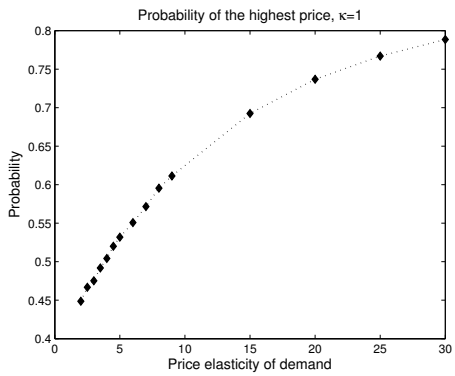

Figure 7: Asymmetry of pricing

the seller chooses to process the least amount of information in the region of high costs. This induces him to discretize high prices more coarsely. The highest price is charged to a wide range of different unit input costs and is thus realized relatively more often. Figure 6 shows that the distribution of prices is discrete and asymmetric even for non-uniform cost distributions. The presented result is a solution to a problem with a triangular distribution of unit input cost of the form: $g(\mu) \propto (0.2 - |\mu - 1|)$ for $\mu \in (0.8, 1.2)$. When $\kappa = 1$, then the top price is not the most frequent, but the asymmetry is still apparent.

For the uniform distribution of the input costs, the probability of the highest price as a function of θ is shown in Figure 7. Numerical solutions suggest that higher θ increases the asymmetry of pricing, and so does Corollary 3. We already discussed that the asymmetry is driven by the negative derivative of the loss factor $L(\mu)$, (19), with respect to μ. More specifically, the attention allocation depends on the relative magnitudes of $L(\mu)$ at different levels of μ. The ratio is $L(\mu_1)/L(\mu_2) = (\mu_1/\mu_2)^{-\theta-1}$. Higher θ increases the attention's sensitivity to μ.

Conjecture 4. *For the uniform distribution of unit input costs, the top price in the sample is the most probable to be realized in each period. The effect is stronger for a higher elasticity of demand θ.*

4.2 How the results agree with data

The model generates time series of prices that are appealing in various aspects. The whole distribution of prices forms a price plan. In each period, the seller charges one of the plan's

prices, but chooses not to select any values in between. There is no cost of switching from one price value within the plan to another. These findings correspond perfectly to the price-plan model that was proposed in the empirical study of Eichenbaum et al. (2008). In agreement with the data, prices in my model change frequently and switch back and forth between a few different values, and most price movements appear to be sales-like movements (Nakamura and Steinsson, 2008). The most quoted price is often the highest price.

The results of the presented model rest on the selection of parameters. The parameters are: a distribution of unit input cost $g(\mu)$, the price elasticity of demand θ, and the information capacity κ. Unfortunately, the cost data used in Eichenbaum et al. (2008) is proprietary. What the authors do report is the median standard deviation of log-unit input cost, which equals 0.12. I assume μ to be i.i.d. and uniformly distributed, and that each realization corresponds to a weekly unit input cost. The width of the uniform $g(\mu)$ is set to match the standard deviation of costs reported in Eichenbaum et al. (2008).

Regarding θ, estimates in the literature vary between 3 and 10.[15] Picking the information capacity κ is trickier. It is realistic that the price setters in large retail stores with more than 10,000 products do not follow all the details of all cost movements, but the proper level of κ is not obvious. In this paper, κ is the calibration parameter. Time series of prices generated by the model are similar to the price series in scanner data for information capacity that is less than 2 bits. Table 1 summarizes comparisons of the model's results with the findings in Eichenbaum et al. (2008). $\theta = 3$ generates a markup distribution much closer to the data than what is generated by $\theta = 10$, for which markups are too volatile. The standard deviation of prices does not depend on the parameters very much. Finally, higher θ implies a higher asymmetry of pricing, but the differences are not striking. Overall, $\{\theta = 3, \kappa = 1\}$ probably provides the best fit among the evaluated combinations of parameters.

Distribution of markups: If the input cost can take an infinite number of values, then the seller deviates from the optimal markup $1/(\theta - 1)$ with probability 1. When $\theta = 3$, standard deviations of log-markup are 0.16 for $\kappa = 1$ and 0.08 for $\kappa = 2$. The

[15] Kehoe and Midrigan (2007) study a heterogeneous menu cost model generating flexible sales and aggregate rigidity. They use $\theta = 3$.

	st.dev. of log (markup)	st.dev. of log (price)	profit loss	% at ref. price
Eichenbaum et al. (2008)	0.11	0.14	n.a.	60%
$\theta = 3, \kappa = 0.5$	0.27	0.09	1.8%	65%
$\theta = 3, \kappa = 1$	0.16	0.10	0.8%	50%
$\theta = 3, \kappa = 2$	0.08	0.11	0.2%	25%
$\theta = 10, \kappa = 0.5$	0.93	0.09	18%	79%
$\theta = 10, \kappa = 1$	0.7	0.10	8.2%	62%
$\theta = 10, \kappa = 2$	0.33	0.11	1.8%	38%

Table 1: Comparative statics, st.dev. of log(unit input cost) is 0.12

median standard deviation of log-markup in Eichenbaum et al. (2008) is 0.11. The more information that is processed, the less volatile the markups are, since prices are more tightly distributed around the optimal price. Figure 8 shows the standard deviation of markups and the resulting loss in profits as functions of the information capacity. The profit loss is the difference between expected profit of the optimal pricing strategy (15) and the expected profit of the informationally constrained strategy. We see the losses from imperfect information are fairly low. For $\theta = 3$ and 1 bit of information capacity, the seller loses only 0.8% of profit and already 2 bits of information are sufficient to recover 99.8% of expected profit under perfect information.

Reference prices: The model generates asymmetry in pricing, with the highest price typically being the most quoted, which is in agreement with the data.[16] Eichenbaum et al. (2008) report that prices stay at the reference price about 60% of the time; such a fraction is generated by the model for instance when $\{\theta = 9, \kappa = 1\}$ or $\{\theta = 3, \kappa = 0.5\}$. If $\{\theta = 3, \kappa = 1\}$, then the probability of the highest price is about 50%.

State dependence: Eichenbaum et al. (2008) measure the probability of changing a weekly price as a function of a percentage deviation from a hypothetical markup. Given a realized input cost, a hypothetical markup is a markup that would be realized if the price stayed constant. Their result together with the results of our model for $\theta = 3$ are

[16] Figure 7 shows the probability of the highest (and most quoted) price as a function of θ, for $\kappa = 1$.

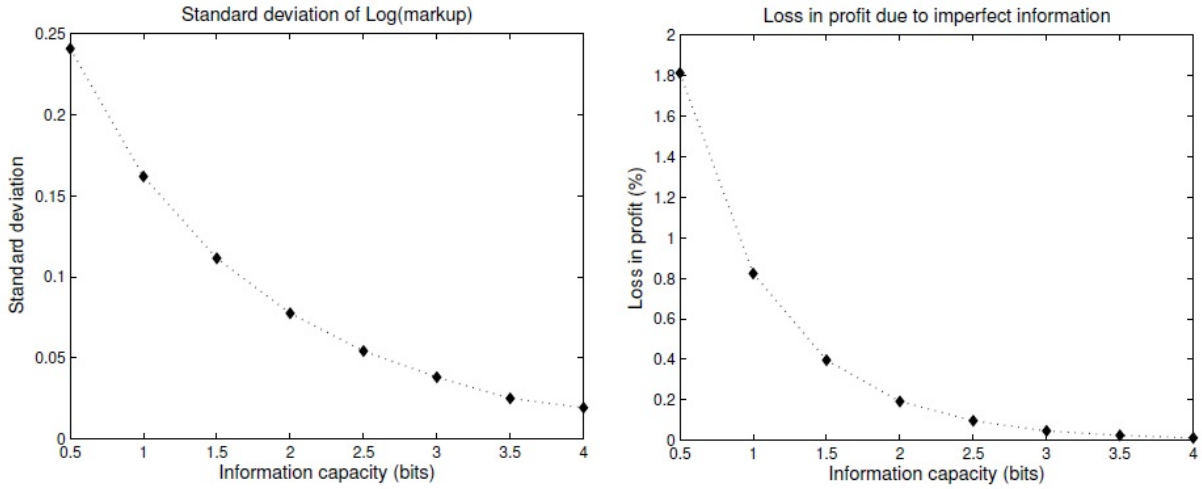

Figure 8: Implications of information capacity for sub-optimality of pricing, $\theta = 3$

presented in Figure 9. They are qualitatively similar. Pricing is highly state-dependent. Quantitative agreement is better for $\kappa = 2$ than for $\kappa = 1$.

Volatility of prices: Eichenbaum et al. (2008) find that the standard deviation of logs of prices is slightly higher than the deviation of logs of costs, 0.14 versus 0.12. In this model, it is the other way around, about 0.10 versus 0.12. This unfavorable difference is however smaller than for menu cost. For instance, marginal costs in the menu cost model of Golosov and Lucas (2007) are 40% more volatile than prices.

5 Extension: serially correlated unit input cost and delayed actions

This section discusses an extension of the original model (10)-(13) that makes the processing of information a dynamic problem. I introduce a serially-correlated component to the unit input cost, so knowledge acquired in one period forms the prior in the following one. Similarly as in Sims (2003) or Maćkowiak and Wiederholt (2009), I study responses to two distinct sources of shocks, which differ in their persistence. The other authors explore dynamic problems with quadratic objectives and autoregressive Gaussian processes for shocks, while I solve a case with a fully nonlinear objective. On the other hand, I choose

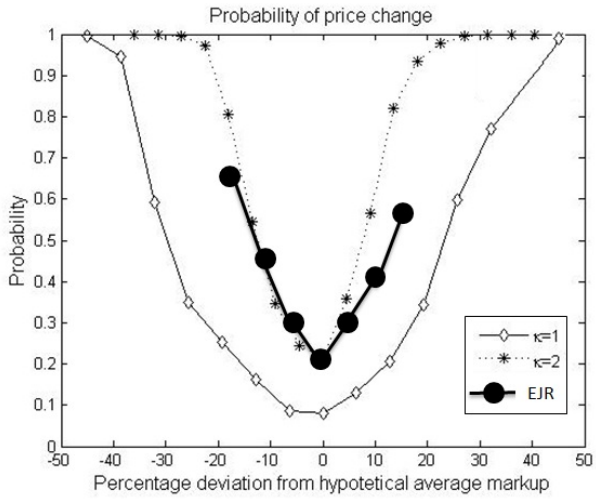

Figure 9: State dependence, $\theta = 3$

simpler stochastic properties of the shocks.

Solving fully-nonlinear dynamic problems under rational inattention is computationally extremely demanding. The seller's state variable is his current knowledge of the cost's components, which is an infinitely dimensional object if the components are continuously distributed.

Let the unit input cost be a product of the following two variables: a) the i.i.d. part μ, and b) the Markov variable A, which switches between two levels: A_L and A_H with the transition probability equal to τ. The advantage of this setup is that we can study the implications of the persistent A while being able to solve the non-quadratic problem. Knowledge about a binary variable is described by the probability of each of the two states. It is therefore just a scalar, which is very convenient. Let the seller's state variable x be the probability of $A = A_H$.

μ is supposed to represent the volatile part of the input cost specific to the seller, while A plays the role of a slowly moving aggregate variable, e.g. the price level. A is an index shifting the distribution of the nominal input cost $A\mu$. Due to the shocks' constrained structure, this section's results are useful for qualitative comparisons, but less so for estimation of any realistic dynamics of aggregate variables.

5.1 Formulation of the extended model

In each period t, the seller processes information about the two components of the unit input cost. The profit function takes the following form:

$$\Pi(A, \mu, p) = p^{-\theta}(p - A\mu). \tag{20}$$

The seller starts the period with a prior on A and μ given by

$$g_t(A, \mu) = g_{1,t}(A)g_2(\mu), \tag{21}$$

where g_2 is is the fixed pdf of the iid component, while the prior on A, $g_{1,t}$, can vary across periods. Shocks to the two components are independent of each other. The way the seller chooses to processes information determines the collection of potential posteriors, and then priors in the following period.

The seller's decision strategy is again described by a joint pdf $f_t(A, \mu, p)$. As in the iid case, the pdf needs to be consistent with the prior distribution and the seller cannot in one period process more information than κ.

$$\int f(A, \mu, p)dp = g_1(A)g_2(\mu) \tag{22}$$

$$I\left[f(A, \mu, p)\right] \leq \kappa. \tag{23}$$

Since A is persistent, the posterior knowledge $f_t(A|p_t)$ translates into the next period's prior. Considering that the probability of transition to the opposing state is τ, we get

$$g_{1,t+1}(A_H) = f_t(A_H|p_t)(1 - \tau) + \Big(1 - f_t(A_H|p_t)\Big)\tau. \tag{24}$$

Starting with a given prior $g_{1,0}$ at $t = 0$, the seller sequentially chooses strategies $\{f_t\}$ to maximize the discounted expected profits,

$$\sum_{t=0}^{\infty} \beta^t \int \Pi(A, \mu, p) f_t(A, \mu, p) dA d\mu dp, \tag{25}$$

subject to (21)-(24).

This sequential problem can be formulated recursively. Let V be the maximum attainable value of the objective above. It is a function of the prior in period zero, $g_{1,0}$. The value function $V(x)$ satisfies the following Bellman equation.

$$V(x_t) = \max_{f_t} \int \Big[\Pi(A, \mu, p) + \beta V(x_{t+1})\Big] f_t(A, \mu, p) dA d\mu dp, \tag{26}$$

subject to (21)-(24), where x_t stands for the prior probability in period t that A is in the high state, $x_t \equiv g_{1,t}(A_H)$.

Definition 3. ***Equilibrium of the extended model.*** *The equilibrium of the model is the value function $V(x)$ and stochastic processes $f_t(A, \mu, p)$, $\{\mu_t\}$, $\{A_t\}$, $\{x_t\}$ and $\{p_t\}$, such that:*

1. *$V(x)$ is a solution to* (26)-(23),
2. *initial values A_0 and $g_{1,0}(A_H)$ are given,*
3. *indexes $\{A_t\}$ are generated by a binary Markov process with a symmetric transition probability τ,*
4. *unit input costs $\{\mu_t\}$ are drawn from $g_2(\mu)$,*
5. *for a given x_t, $f_t(A, \mu, p)$ is a solution to the right hand side of* (26),
6. *prices $\{p_t\}$ are drawn from $f_t(p|A_t\mu_t)$,*
7. *$\{x_t\}$ are generated by the low of motion* (24),
8. *amounts sold are $\{d(p_t)\}$.*

In each period, the seller maximizes the right hand side of (26). Not only does the seller consider the current profit, but also the effects of the currently acquired knowledge on the future profits.

5.2 Numerical solutions

The value function, the fixed point of (26), can be found by iterations, while the right hand side of (26) is solved using the same techniques as in Section 4. Figure 10 shows the results of simulations over 120 periods for $\kappa = 1$, $\theta = 3$, μ uniformly distributed over $(0.8, 1.2)$, $A_L = 1$, $A_H = 1.1$, $\tau = 0.002$, and $\beta = 0.9992$. One period represents one week, which makes the annual discount factor equal to 0.96. $\tau = 0.002$ implies that the probability of changing the state (a 10% shock to the aggregate variable) at least once

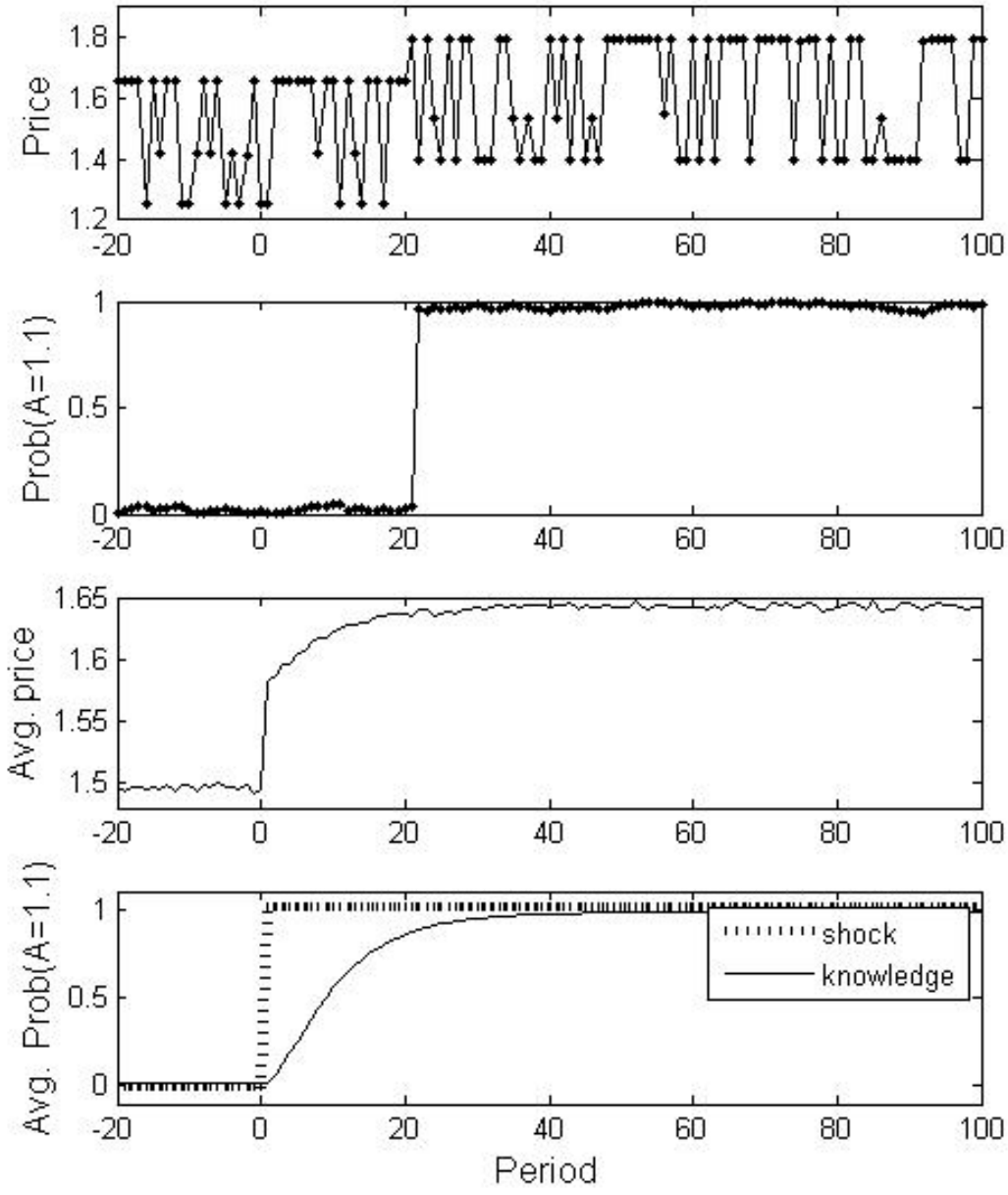

Figure 10: Two stochastic variables, abrupt transition, $t = 0.002$, $\kappa = 1$, (a) simulated prices series, (b) simulated knowledge about A, (c) average prices, (d) average knowledge.

during a year is about 10%. There is a simulated shock to A in period 1, when A switches from $A_L = 1.0$ to $A_H = 1.1$.

The top series in the figure presents one realization of price series $\{p_t\}$, and the second one shows the corresponding time-series of knowledge $\{g_t(A_H)\}$ in the same simulation. The price setter targets the optimal price $\frac{\theta}{\theta-1}A_t\mu_t$. Although $A_t\mu_t$ is distributed over a continuous range in every period, prices again exhibit the rigidity of values. Given prior knowledge about A, the distribution of prices forming a current price plan is discrete. However, when knowledge about A changes, the distribution of prices determining the seller's price plan changes too.

The abrupt adjustments of knowledge and price plans are typical for all simulations with the given parameters. What varies from one simulation to another is the period in which the seller finds out that A has probably switched to a new value. Once the seller starts realizing that A might have switched, then he endogenously allocates more of his information capacity specifically to A, and his knowledge shifts quickly. When A stays at the higher level, the agent is likely to acquire posterior knowledge that $A\mu$ is high several periods in a row, and the price plan's top price is even more likely to be realized. At first, $A\mu$ is attributed to high μ rather than to a transition of A, since the transition is far less likely. The seller starts adapting his knowledge about A and reallocating his attention only after a streak of high $A\mu$ is realized.[17]

On the other hand, Figure 11 shows that the adjustment of knowledge about the Markov variable A can in other cases be gradual. This arises when A is less stable than before and information capacity is lower, $\tau = 0.02$ and $\kappa = 0.5$. Knowledge starts adjusting almost immediately after the shock, but the rate of adjustment is much lower. Now, the seller devotes less of his information capacity specifically to A, but a shock to $A\mu$ is more likely to be attributed to A. Price plans consist of two points only. The plans are fairly rigid while A is stable, but become more flexible after the shock. The flexibility is only

[17] The described interdependence of shocks, knowledge and price responses rather resembles the implications of the signal extraction model in Lucas (1972), but some implications of rational inattention and signal extraction go in the opposite directions, see Sims (2003). In signal extraction, highly volatile variables are tracked with the least amount of error, while in rational inattention it is the stable variables that are easy to follow.

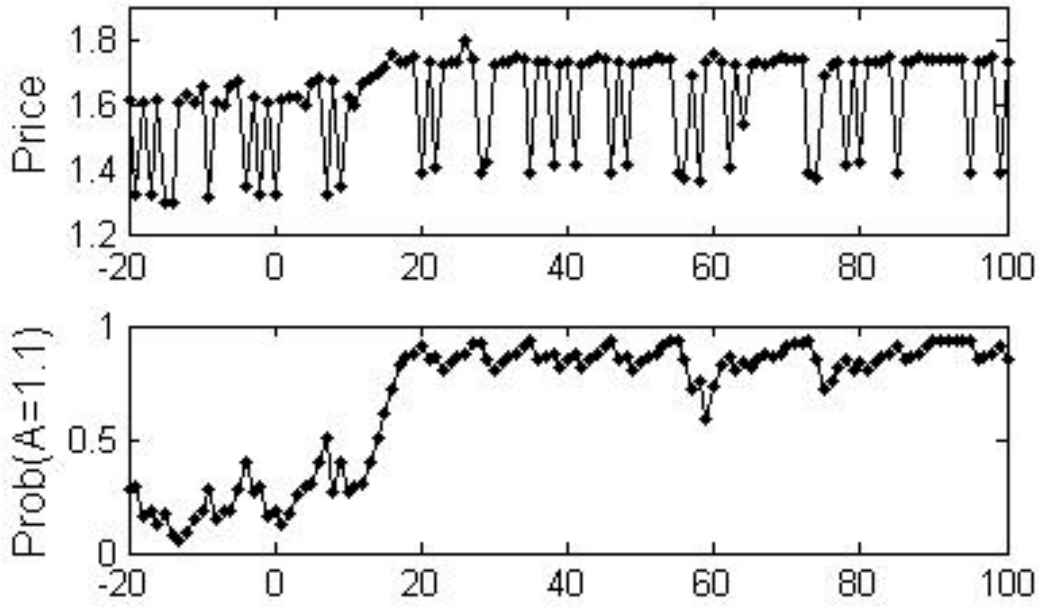

Figure 11: Gradual adjustment of knowledge, (a) simulated prices series, (b) simulated knowledge about A

in shifting the plan by an amount caused by the persistent aggregate shock, while no new price points arise within the plan. If the persistent shock was small, then the range of flexibility would be small too. Introducing a negligible menu cost in this model after all would suffice to stabilize these plans, while prices would still move almost freely among the plan's points. However, with no explicit adjustment cost and, for instance, purely deterministic A with an upward trend, the model would generate continuously shifting plans.

The bottom two series in Figure 10 are prices and knowledge averaged over 10,000 simulation runs. Their responses to the shock in period 1 are delayed. These series represent average realizations in the population of sellers where noise in information processing is not correlated across agents. Average price is the aggregate nominal level. The average knowledge about A shifts slowly, as more and more agents find out about the switch. On the other hand, the aggregate nominal level adjusts quite fast, although full adjustment is still delayed. The source of the quick initial change is the adjustment in the frequency of sales. We discuss this point below. This effect is not present in the model of Maćkowiak and Wiederholt who assume that information about different shocks needs to be processed

separately.

Joint signals: Since the rationally inattentive seller is allowed to process joint signals about several variables, then resulting actions and also knowledge about the joint characteristics typically adjust faster than knowledge about single slowly moving variables. After an aggregate shock to A, knowledge about A and thus the price plan do not change at first, but the frequency of sales together with the aggregate nominal level partly adjust.

The seller can pay attention directly to changes of $A\mu$, not necessarily separately to its components. Since μ is relatively more volatile, such changes are at first attributed to shocks to μ rather than to A. If a rationally inattentive agent tracks some joint characteristics of several variables, a dynamic path of knowledge about a single variable depends on how unexpected a shock to the variable is, while the path of price responses to the same shock depends on how unexpected is the resulting shock to the joint characteristics of interest.

On the other hand, if the agents are assumed not to be able to process signals that are informative about several variables, such as in Maćkowiak and Wiederholt (2009), then knowledge about a variable is always perfectly in line with the responses to the variable. Pricing responses to shocks to A would be as smoothed and as delayed as knowledge about A.

Endogenous infrequent updating of knowledge: The emerging jumpy adjustment of knowledge as in Figure 10 is often assumed in the literature on information frictions in the form of infrequent reviews of economic conditions.[18] Figure 12 presents solutions to the dynamic problem with μ being fixed at 1, the only stochastic variable left is A, $\kappa = 0.02$. The average price is fully smoothed and delayed, while single simulations display sudden changes of knowledge and price. The seller's knowledge does not change gradually. It stays constant for a while and then suddenly switches at one specific moment. This moment of transition is, however, not given deterministically.

Such a dynamics of knowledge resembles the one assumed in the sticky-information model introduced in Mankiw and Reis (2002). They postulate that agents' information updating

[18] See Mankiw and Reis (2002) and Alvarez et al. (2010), and Woodford (2008), who combines rational inattention with acquisition of perfect knowledge before a price change.

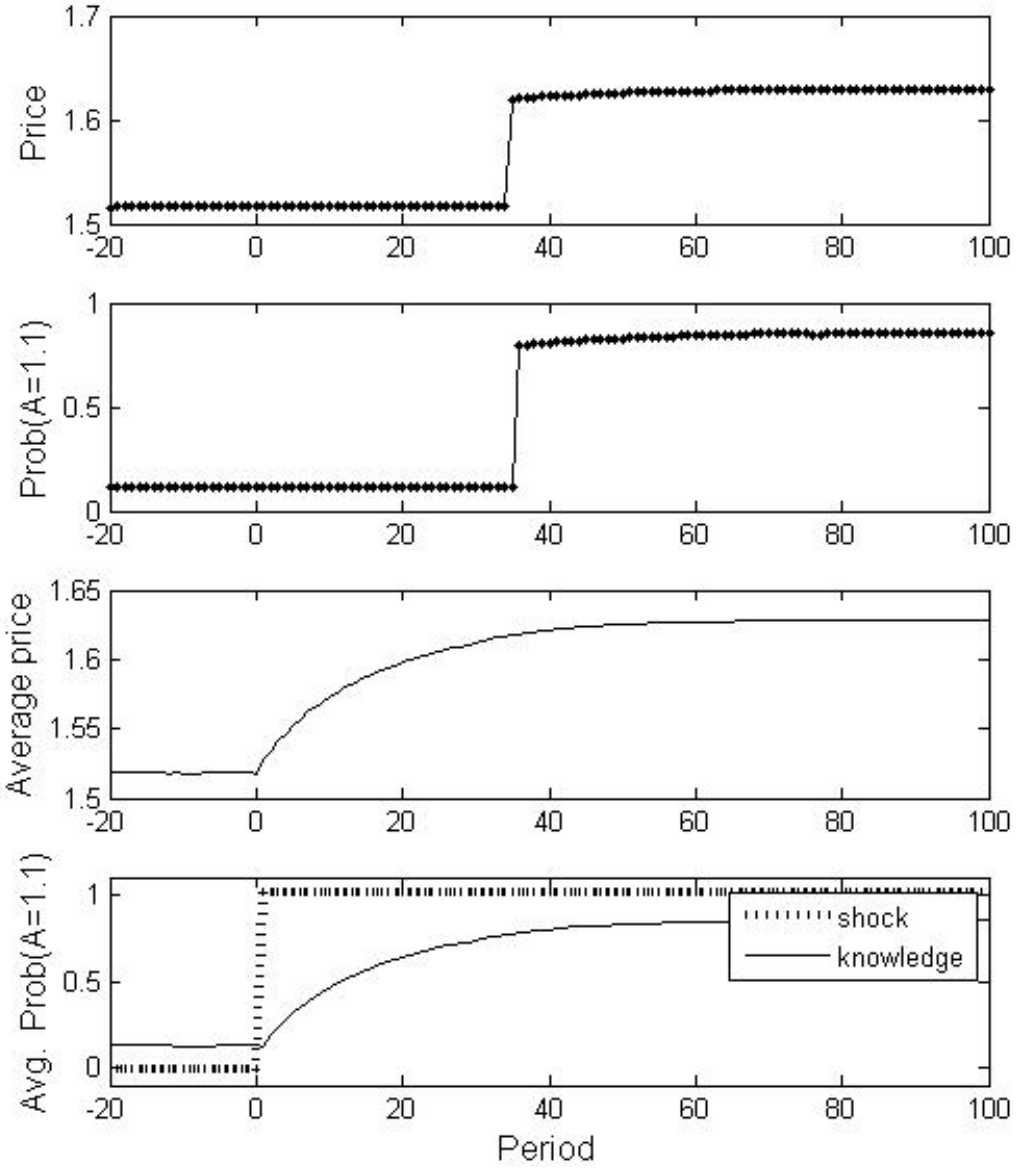

Figure 12: Responses to shocks to A, fixed μ

is staggered, and when agents update they acquire perfect information. In each period, a firm updates with a certain probability ν, i.e. only a fraction ν of firms update information in that period. Figure 12 shows that a similar form of updating can emerge under rational inattention, too. The observed dynamics is driven by the discussed discreteness in responses: agents sometimes prefer to receive a few different signals only rather than a complete spectrum of them. Receiving one signal instead of another then results in a sudden and significant change of knowledge.

5.3 Implications for the Modeling of Aggregate Nominal Rigidities

I discussed above that a rationally inattentive seller chooses to follow a "price plan", which is described by a distribution of prices that can be realized in a given period. If aggregate conditions are more persistent than idiosyncratic, then the idiosyncratic shocks drive price movements looking like sales and aggregate shocks shift the whole price plans. The price plan changes whenever the seller's knowledge about the state of persistent variables changes.

The source of aggregate nominal rigidity is that the seller's price plan shifts after an aggregate nominal shock with a delay. The delayed response is, however, partially offset by an adjustment of sales frequency. This is due to the difference between seller's perceived and the true distribution of the unit input costs, when the aggregate shock is misjudged. After a positive persistent shock, he at first mistakenly attributes the resulting deviation in input cost to a transient shock, which is more likely to occur. The seller thus does not alter his price plan, but he is less likely to charge its low prices. The reference price is fixed for a while, but the frequency of sales decreases. The seller does not know about the persistent shock, but his prices do partially reflect it. This effect is not present when he cannot process information about different shocks jointly.

Eichenbaum et al. (2008) observe in the data that while prices change very frequently, the value of the reference price stays quite rigid. The presented model generates similar pricing patterns. However, the reference price in our model does not have any special significance; it is just one of the characteristics of the whole generated distribution of prices. Eichenbaum et al. claim that it is the rigidity of the reference price, not single

prices, that is the useful statistics for macroeconomic analysis. Interestingly enough, the dynamics of the reference does in our model indeed mimic the seller's current knowledge about the state of the slowly moving index A. Current knowledge uniquely determines the selection of a price plan together with its reference price. However, knowledge about the persistent shock does not need to be perfectly in line with dynamics of responses to this shock, which we see in the bottom two series of Figure 10.

The degree of aggregate nominal rigidity depends on the price plan's adjustment as well as on the adjustment of sales frequency. Rational inattention thus does not provide support for a simple rule of removing sales from data for macroeconomic purposes. It does, however, relate to the heterogeneous menu cost model in Kehoe and Midrigan (2007) with low adjustment cost for sales prices and higher for regular prices, and even more to the sticky-plan model in Eichenbaum et al. (2008), where movements between the plan's prices is costless, only changing the plan is costly. These models also generate aggregate nominal rigidity despite the flexibility of sales.[19]

In rational inattention, it is more costly for the seller to pay attention purely to a persistent shock, since such an activity provides low marginal profit due to the shock's infrequent occurrence. In general, models based on menu cost that were to mimic rational inattention would need to assume a high cost of adjustments corresponding to shocks to the optimal price that are not driven by the most likely shocks.

However, we should not expect that any simple menu cost model can fit the properties of the pricing of rationally inattentive sellers in general. Adjustment frictions based rational inattention are not explicitly definable in the form of a constant menu cost, they endogenously depend on current knowledge, stochastic properties of shocks, past shocks realizations, etc. Naturally, it is the model presented in Maćkowiak and Wiederholt (2009) that is the proper starting point from rational inattention's point of view.

[19] Guimaraes and Sheedy (2008) is another model with flexible sales and monetary non-neutrality.

6 Conclusion

This paper shows that rational inattention can account for several empirical findings about nominal rigidities. Most importantly, I find that some stylized facts at the micro level that constitute puzzles for sticky-price models can be accounted for by information frictions in the form of rational inattention. Rational inattention can drive sluggish responses to new shocks. Moreover, the rationally inattentive seller chooses to follow a price plan with a finite number of distinct prices. Such a plan results in prices that do not always change, but they do change frequently switching between a few values. This strategy emerges as an optimal utilization of the agent's limited cognitive abilities. The price plan itself is typically asymmetric, with higher probabilities of realizing higher prices. This is due to potentially higher losses in the case when the seller misjudges the input cost when it is low. Such asymmetry leads to more rigid high prices, which closely resembles the "regular" or "reference" prices discussed in the empirical literature.

Information frictions can result in dynamics of prices that could in some ways be confused with the existence of explicit adjustment costs. It is, however, very important to distinguish what the true driver of nominal rigidity is. While, for instance, Calvo-style models are calibrated to a fixed frequency of price changes, rigidity based on rational inattention emerges endogenously and is thus responsive to changes in economic conditions.

References

Alvarez, F. E., Lippi, F., and Paciello, L. (2010). Optimal price setting with observation and menu costs. NBER Working Papers 15852, National Bureau of Economic Research, Inc.

Bils, M. and Klenow, P. J. (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy*, 112(5):947–985.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley, Hoboken, NJ.

Eichenbaum, M., Jaimovich, N., and Rebelo, S. (2008). Reference prices and nominal rigidities. NBER Working Papers 13829, National Bureau of Economic Research, Inc.

Golosov, M. and Lucas, R. E. (2007). Menu costs and phillips curves. *Journal of Political Economy*, 115:171–199.

Guimaraes, B. and Sheedy, K. D. (2008). Sales and monetary policy. CEP Discussion Papers dp0887, Centre for Economic Performance, LSE.

Kehoe, P. J. and Midrigan, V. (2007). Sales and the real effects of monetary policy. Technical report.

Lucas, R. J. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, 4(2):103–124.

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

Maćkowiak, B. and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, 99(3):769–803.

Maćkowiak, B. A. and Wiederholt, M. (2010). Business cycle dynamics under rational inattention. CEPR Discussion Papers 7691, C.E.P.R. Discussion Papers.

Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics*, 117(4):1295–1328.

Matějka, F. and Sims, C. A. (2010). Discrete actions in information-constrained tracking problems. Technical report, Princeton University.

Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory*, 145(5):1837 – 1864.

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

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27.

Sims, C. A. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy*, 49(1):317–356.

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

Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, 96(2):158–163.

Van Nieuwerburgh, S. and Veldkamp, L. (2010). Information acquisition and under-diversification. *Review of Economic Studies*, 77(2):779–805.

Vanderbei, R. J. (1999). Loqo: an interior point code for quadratic programming. *Optimization Methods and Software*, 11.

Woodford, M. (2008). Information-constrained state-dependent pricing. NBER Working Papers 14620, National Bureau of Economic Research, Inc.

A Simple example of attention allocation

This example has a flavor of what is further generalized in the main model of this paper. The example possesses an element of the selection of optimal attention allocation.

Let the profit-maximizing seller know that the input cost is uniformly distributed between 1 and 2, and let his mechanism of processing information be constrained to the observation of a binary indicator set up in his store. The indicator lights up whenever the unit input cost is above a certain threshold x. The seller can however set this level x himself. This particular assumption on the specific form of the information processing mechanism is not inherent to the true framework of rational inattention and it does not apply to the main model.

The question is: what is the optimal x? Given x, the seller acquires only two different forms of posterior knowledge: uniform distribution over $(1, x)$ or over $(x, 2)$. In the first case, his profit maximizing price is $\frac{\theta}{\theta-1}(1+x)/2$, while in the second it is $\frac{\theta}{\theta-1}(x+2)/2$. To find the optimal x, we maximize the ex ante expectation of profit before the binary signal is received. It is:

$$E[\Pi] = \int_1^x \left(\frac{\theta}{\theta-1}(1+x)/2\right)^{-\theta} \left(\frac{\theta}{\theta-1}(1+x)/2 - \mu\right) d\mu + \\ + \int_x^2 \left(\frac{\theta}{\theta-1}(x+2)/2\right)^{-\theta} \left(\frac{\theta}{\theta-1}(x+2)/2 - \mu\right) d\mu. \tag{27}$$

For $\theta = 2$, the optimal $x = \sqrt{2}$. It is less then the midpoint of the unit input cost at $\mu = 3/2$. Numerical solutions for other values of θ are shown in Figure 13. The seller chooses to pay more attention to lower input cost. If low cost is realized, then he attains relatively more precise knowledge: $\mu \in (1, x)$. When the input cost is low, the optimal price is also lower and the potential sales are higher, which also means higher potential losses from deviating from the optimal price. It turns out that the seller charges the higher price $\frac{\theta}{\theta-1}(x+2)/2$ with a probability higher than 0.5, because $(x-1) < (2-x)$.

Although the rational inattention setup does not restrict agents with such a choice of signals, this simplistic model illustrates the mechanism how an asymmetric objective generates asymmetric actions through the choice of attention allocation.

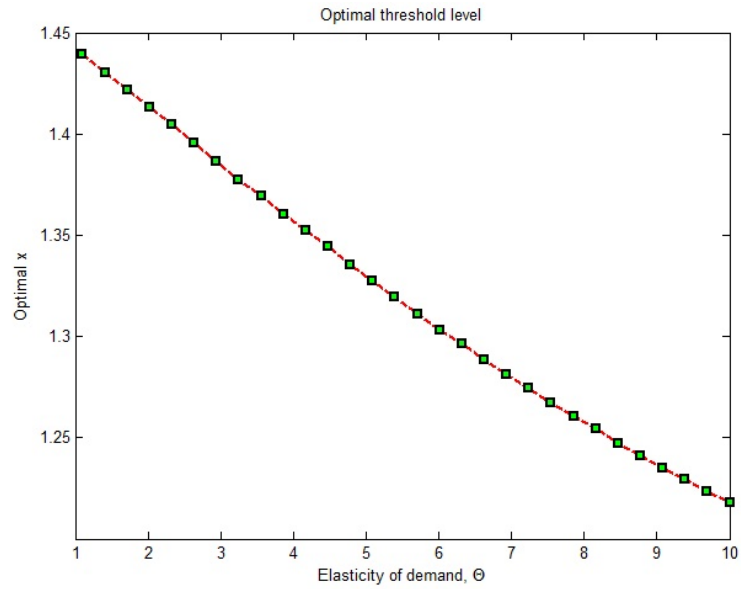

Figure 13: Attention allocation; Choosing optimal threshold level for a binary signal

B First order condition

The first order condition, equation (16), is derived here. The mutual information can be written as

$$I(\mu;p) = \int f(\mu,p)\log\frac{f(\mu|p)}{\int f(p_1)f(\mu|p_1)dp_1}d\mu dp. \tag{28}$$

The Lagrangian of (10)-(13) is:

$$\begin{aligned}\mathcal{L} = &\int \Pi(\mu,p)f(\mu,p)d\mu dp - \lambda\left[\int f(\mu,p)\log\frac{f(\mu|p)}{\int f(p_1)f(\mu|p_1)dp_1}d\mu dp - \kappa\right]\\ &- \int \nu(\mu)\left[\int f(\mu,p)dp - g(\mu)\right]d\mu,\end{aligned}$$

where $\lambda \in \mathbb{R}$ and $\nu \in L^\infty(\mathbb{R}^N)$ are Lagrange multipliers. The first order condition with respect to $f(\mu|p)$ is

$$f(p)\left(\Pi(\mu,p) - \lambda\log\frac{f(\mu|p)}{g(\mu)} - \nu(\mu)\right) = 0.$$

If $f(p) > 0$ and $\lambda > 0$ we obtain the first order condition (16):

$$f(\mu|p) = h(\mu)e^{\Pi(\mu,p)/\lambda},$$

where $h(\mu) = e^{-\nu(\mu)/\lambda}g(\mu)$.

C Approximate losses

Here I express how the profit losses depend on the level of the realized unit input cost. Let us assume that noise in the posterior knowledge is low, so that the certainty equivalence holds; we ignore the 3rd order effects. Let the true realized value of the input cost be μ^*. The seller acquires posterior knowledge with a mean $\mu' = \mu^* + \epsilon$, where ϵ is the small error. Due to the certainty equivalence, the agent chooses to charge a price p equal to $p_{opt}(\mu')$, (15).

Let $\Pi_{\mu*}(\mu') = \Pi(\mu*, p_{opt}(\mu'))$; its Taylor expansion about μ^* is

$$\begin{aligned}
\Pi_{\mu^*}(\mu') &= \Pi_{\mu^*}(\mu^*) + \frac{d\Pi_{\mu^*}(\mu')}{d\mu'}\epsilon + \frac{d^2\Pi_{\mu^*}(\mu')}{d\mu'^2}\epsilon^2/2 + \mathcal{O}(\epsilon^3) \\
&= \Pi(\mu^*, p_{opt}(\mu^*)) + \Big[\frac{d^2 p_{opt}(\mu')}{d\mu'^2}\frac{d\Pi(\mu^*,p)}{dp} \\
&\quad + \Big(\frac{dp_{opt}(\mu')}{d\mu'}\Big)^2 \frac{d^2\Pi(\mu^*,p)}{dp^2}\Big]\epsilon^2/2 + \mathcal{O}(\epsilon^3)\Big|_{\mu'=\mu^*, p=p_{opt}(\mu^*)} .
\end{aligned}$$

The linear term drops out since $\Pi_{\mu^*}(\mu')$ attains its maximum at $\mu' = \mu^*$. Similarly, $d\Pi(\mu^*, p)/dp = 0$ at the optimal response $p_{opt}(\mu^*)$. The change in profit due to the signal imperfection thus takes the form:

$$\begin{aligned}
\Delta\Pi &= \Pi_{\mu^*}(\mu') - \Pi_{\mu^*}(\mu^*) \\
&= \Big(\frac{dp_{opt}(\mu')}{d\mu'}\Big)^2 \frac{d^2\Pi(\mu^*,p)}{dp^2}\epsilon^2/2 + \mathcal{O}(\epsilon^3)\Big|_{\mu'=\mu^*, p=p_{opt}(\mu^*)} . \qquad (29)
\end{aligned}$$

The leading term is quadratic with a negative coefficient, profit is a concave function of the perceived μ' with the maximum at the true value $\mu*$. If the error ϵ is small, then the change in profit can be approximated by $-L(\mu^*)\epsilon^2/2$, where the approximative loss factor $L(\mu^*)$ is

$$L(\mu^*) = -\Big(\frac{dp_{opt}(\mu^*)}{d\mu^*}\Big)^2 \frac{d^2\Pi(\mu^*,p)}{dp^2}\Big|_{p=p_{opt}(\mu^*)} . \qquad (30)$$

The formula recognizes that the loss depends on the curvature of the profit function and also on how far away from the optimal price the realized price is. The more sensitively prices change with posteriors on the unit input cost, the further away can the realized p be from the optimal one.

Decreasing the error ϵ leads to profit gains that are proportional to $L(\mu)$. If the seller can decide in what regions of μ to pay more attention and decrease the noise, then he does so

for μ^*, where the loss factor $L(\mu^*)$ is higher.

D Discreteness

Proof of Proposition 2: If $f(p) > 0$, then the first order condition (16) holds. Since $f(\mu|p)$ is a pdf, it integrates to 1,

$$\int h(\mu) e^{p^{-\theta}(p-\mu)/\lambda} d\mu = 1. \tag{31}$$

If the distribution of prices is not discrete, then there exists a limit point p^* of points, where (31) holds. Therefore, all derivatives of the left hand side of (31) with respect to p have to equal zero at p^*.

The condition on the second derivative of the right hand side of (31) equal to zero reads

$$\int \frac{h(\mu)}{\lambda^2} e^{\frac{p^{-\theta}(p-\mu)}{\lambda}} p^{-2(1+\theta)} \Big(p^2(\theta-1)^2 - 2\mu p(\theta-1)\theta + \\ + \lambda p^{1+\theta}(\theta-1)\theta + \mu^2\theta^2 - \lambda\mu p^\theta \theta(1+\theta) \Big) d\mu = 0. \tag{32}$$

If the unit input cost is bounded by μ_{min} from below and $\mu_{min} + \Delta$ from above, then prices certainly lie in $\left[\frac{\theta}{\theta-1}\mu_{min}, \frac{\theta}{\theta-1}(\mu_{min} + \Delta)\right]$. Δ is the diameter of the prior's support. Therefore, all realizable μ and p satisfy

$$\mu = \frac{\theta-1}{\theta} p + \delta(\mu, p), \text{ where } |\delta(\mu, p)| \leq \Delta. \tag{33}$$

Plugging (33) into (32), we get

$$\int \frac{h(\mu)}{\lambda^2} e^{\frac{p^{-\theta}(p-\delta(\mu,p)\theta)}{\lambda\theta}} p^{-2(1+\theta)} \left[-\lambda p^\theta \left(p(\theta-1) + \delta(\mu,p)\theta(1+\theta) \right) + \delta(\mu,p)^2\theta^2 \right] d\mu = 0. \tag{34}$$

Let us focus on

$$\left[-\lambda p^\theta \left(p(\theta-1) + \delta(\mu,p)\theta(1+\theta) \right) + \delta^2(\mu,p)\theta^2 \right]. \tag{35}$$

If this is negative, then the whole integrand in (34) is negative too.

Notice that $p(\theta-1) \geq \theta\mu_{min}$. Since $|\delta(\mu,p)| \leq \Delta$, then $\Delta < \mu_{min}/(1+\theta)$ implies that $(p(\theta-1) + \delta(\mu,p)\theta(1+\theta)) > 0$. For $\Delta < \mu_{min}/(1+\theta)$, the first term of (35) is negative and proportional to λ.

$\delta(\mu, p)^2\theta^2$ is bounded by $\Delta^2\theta^2$. If $\Delta < \mu_{min}/(1+\theta)$, then there exists λ_0 such that (35) is negative for all $\lambda > \lambda_0$. For a specific solution to the seller's problem, the resulting Lagrange multiplier λ depends on the specific choice of the model's parameters. If we show that there exists information capacity κ generating a large enough multiplier, $\lambda > \lambda_0$, then we prove the proposition.

The uniform distribution, among all distributions over the same support, maximizes the entropy.[20] The variation of the distribution's entropy at the optimum is zero. When approaching the optimum, the shadow cost of information approaches infinity, $\lim_{\kappa \to 0+} \lambda(\kappa) = \infty$. Therefore, for all λ_0 there exists $\kappa_0 > 0$ such that for all $\kappa < \kappa_0$ the Lagrange multiplier $\lambda > \lambda_0$. In summary, for all $\Delta < \mu_{min}/(1+\theta)$ there exists $\kappa_0 > 0$ such that for all $\kappa < \kappa_0$ the left hand side of (34) is negative. The distribution of prices is discrete. $\square$

In fact, we showed that under the strong assumptions of Proposition 2, there are at most two price points. The assumptions imply that the second derivative of (31) is negative, the integral can therefore be equal to 1 at two points at most.

[20] See Cover and Thomas (2006).