

Rationally Inattentive Seller: Sales and Discrete Pricing

Filip Matějka*

CERGE-EI†

October 14, 2010

Abstract

This paper presents a model of a rationally inattentive seller. Price rigidities are driven by seller's inability to process all available information about shocks to input cost. The model generates a uniquely wide spectrum of observed price series properties. The single information constraint implies that prices move back and forth between a few rigid values, sales are short-lasting, or that responses to aggregate variables can be faster in volatile industries or in less stable aggregate conditions.

*I am especially grateful to Per Krusell and Chris Sims for invaluable insights and guidance. I would also like to thank to Nobuhiro Kiyotaki, Esteban Rossi-Hansberg, Ricardo Reis and Alisdair McKay for numerous helpful discussions.

†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, Politických veznu 7, Praha 1, Czech Republic, filip.matejka@cerge-ei.cz

The model is fully in the spirit of rational inattention with no simplifying assumptions on functional forms of processed signals.

Keywords: rational inattention, sticky prices, sales.

JEL: D83, D21, E31, E52

1 Introduction

This paper studies price movements. It proposes a mechanism leading to micro flexibility of prices together with their macro rigidity. Moreover, the flexibility in the form of frequent price changes is accompanied by rigidity of values that the prices take. The distinctive feature of the model, which is introduced in Section 2, is the limited information capacity of a seller, who sets prices. Small sellers or sellers with large portfolios of heterogeneous products are likely to devote only a very limited amount of attention to pricing of a single good. Large retailers often sell tens of thousands of products. Responding to new available information about every one of them would be too demanding. In fact, most agents do not pay attention to all information that is available.

In the model, the seller cannot observe input cost exactly and needs to process information about it, he does so by selecting pieces of information that are the most relevant to him. Once he acquires a signal about the input cost, he then decides what price to charge.

Results of the model agree with several findings in the data on both micro and macro levels. 1) Prices change frequently. 2) Responses of prices to aggregate variables are delayed. 3) Prices move back and forth between a few rigid values. 4) Sales are short-lasting. 5) Responses to aggregate shocks are faster in industries with more volatile idiosyncratic shocks. 6) Responses to aggregate shocks can be faster and losses higher when the aggregate variables are more volatile.

However, I do not introduce any heterogeneous cost structure of adjustments or a variety of other assumptions. All pricing behavior arises endogenously from preferences and one information constraint. The agent chooses what to pay attention to, how and when and thus responds differently to different shocks or under different circumstances. This is possibly due to the inner consistency and generality of rational inattention.

Rational inattention is a framework that I apply to model the processing of relevant information. It was introduced in (Christopher A. Sims 1998) and (Christopher A. Sims 2003). Rational inattention is based on a rigorous application of information theory. A few other authors, (Ricardo Reis 2006), (Bartosz Adam Mackowiak & Mirko Wiederholt 2007), and (Michael Woodford 2008), have recently studied pricing of inattentive agents. All three of these papers are very important contributions. However, as far as I am aware, this is the first model of pricing fully in the spirit of rational inattention, where agents are not ex ante constrained with a specific form of signals they can receive, i.e. what pieces of information they can pay attention to.

Agents in (Reis 2006) process information once in a while, but they receive perfect signals when they do process it. Agents in (Woodford 2008) on the other hand process information gradually, and in each period they make a decision whether to change prices or not. However, once they do choose to incur the cost of changing prices, they acquire perfect information and set prices optimally. The setup of (Mackowiak & Wiederholt 2007) is closer the true rational inattention formulated by Sims. They depart from the unrestricted choice of attention by assuming that agents can not process information about more variables jointly, but only about each one of them independently. The significant contribution is that they are able to introduce this form of rational inattention into a DSGE model.

The model presented in this paper does not make any simplifying assumptions about the information structure and is solved in its fully non-linear version, which has several implications. Perhaps the most striking is that to economize on information, the seller tends to choose signals that lead to a small number of distinct prices only. These few values of price are rigid, although the unit input cost varies in a continuous range. Whatever unit input cost is realized, the price takes one of the few rigid values. This issued is in a similar setup studied analytically in (Filip Matějka & Christopher A. Sims 2010).

Discreteness occurs, although no apparent preference towards it is present in the formulation of the model. It is optimal for the seller to pay attention to a source of information that provides a small number of different forms of signals only. For a decision maker in a large retail store, such a signal could be the first digit of cost of a leading product in each category, direction of cost movements, etc. Depending on the dynamics of input cost, prices can change frequently while attaining a few values only. The most frequently quoted price is often, but not always, the highest price. Furthermore, prices are likely to change when input cost changes. While markups are not always at their optimal levels, they are concentrated around them.

Assessing nominal rigidities is crucial for understanding the effects of monetary policy. Several studies found that aggregate price indices respond to monetary shocks with a delay of more than a year (see, e.g., (Lawrence J. Christiano, Martin Eichenbaum & Charles L. Evans 1999)). On the other hand, evidence on the micro level suggests that individual prices do not stay fixed for long periods of time, (Mark Bils & Peter J. Klenow 2002). Neither menu cost nor Calvo models, the most popular new-Keynesian models, can simultaneously generate the observed micro flexibility together with sufficient macro rigidity of prices.

(Martin Eichenbaum, Nir Jaimovich & Sergio Rebelo 2008) have recently carried out a detailed study of price as well as cost movements based on a database of another major US retailer, which sells close to 60 000 items in each of the stores.

A striking feature of most time series in these data sets is that prices change very frequently, while they often switch back and forth between a few different values. (Patrick J. Kehoe & Virgiliu Midrigan 2007) report that 33% of prices change every week.

Prices tend to spend a considerable amount of time at one level and most movements are reductions from this price, followed by a quick return to the original level. These movements are referred to as sales. Moreover, even the sales prices tend to be repeated, realizing the same sales prices over and over again. Rigidity of the

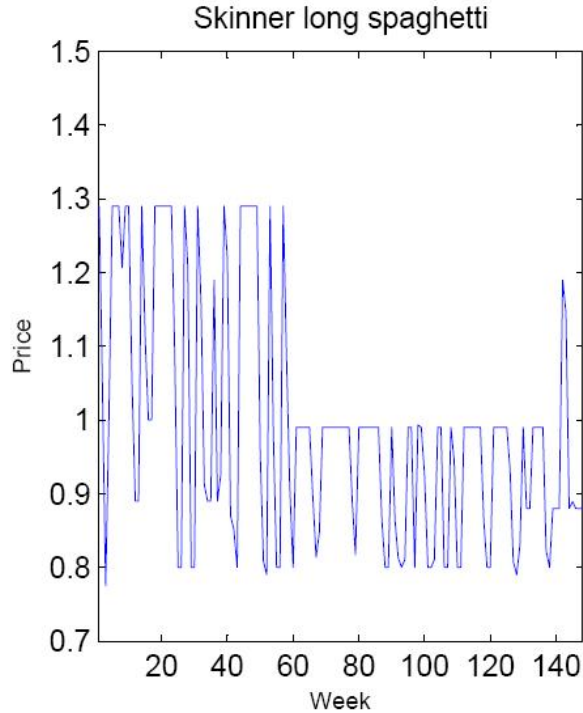


Figure 1: Time series of prices,(Eichenbaum, Jaimovich & Rebelo 2008)

values together with frequent switches between them were hard to reconcile with any of the existing pricing models.

(Kehoe & Midrigan 2007) propose a model leading to a rigid regular price and frequent price reductions. The model assumes that there is an adjustment cost of changing the regular price and a relatively smaller cost of renting a sales price for one period. While the results of the model agree well with several findings in the data, micro-foundations of the adjustment cost structure are not easy to motivate. Furthermore, this model also fails to generate any rigidity of price values below the regular price.

(Eichenbaum, Jaimovich & Rebelo 2008) emphasize a new form of nominal rigidity. They call the most quoted price during each quarter a reference price. While prices change very often, the reference price is quite rigid, with a duration of about 1 year. The price is at the reference price 60% of the time. The authors propose a hypothesis that duration of the reference price is set such that a variation in markups is kept

in a certain narrow range.

Results of the model presented in this paper correspond with the evidence quite well. However, the most quoted price does not have any special significance here, unlike the reference price in (Eichenbaum, Jaimovich & Rebelo 2008), or the regular price in (Kehoe & Midrigan 2007). It is merely a feature of the overall distribution of prices. When a distribution of input cost changes, then the distribution of prices changes too and its most quoted price with it.

This paper also studies implications of idiosyncratic and aggregate volatility levels for dynamics of price adjustments. Results, however, differ based on whether agents face a strict limit on their information capacity or, perhaps more naturally, they incur a cost from processing information and decide on the capacity. If they can choose how much information to process, then prices in more volatile industries respond to aggregate shocks faster. This result rests on the generality of rational inattention, which does not assume any specific form of received signals and simply allows agents to process those pieces of information they find the most useful.

Moreover, the model provides rationale for a steeper Philips curve and sub-optimal economic activity in less stable aggregate conditions. Overall, the presented model of a rationally inattentive seller can reconcile with several features of price dynamics at the micro as well as macro level.

2 The Model

A monopolistic seller processes information about a stochastic unit input cost μ and decides what price p to charge. A consumer can observe prices exactly.

I assume the unit input cost to be i.i.d., so the model can be solved on a period by period basis. In each period, a unit input cost is drawn from a fixed distribution $g(\mu)$. Due to limitations of the seller's abilities to process information, he can not

learn the input cost exactly. The seller selects what pieces of information about the input cost to process and chooses what price to charge.

A consumer is assumed to have unlimited information capacity and can thus observe the price set by the seller. Once the seller chooses a price, the consumer selects an amount to consume, which is given by the demand function $d(p)$. This demand function is known by the seller.

The seller maximizes expectation of profit Π .

$$\Pi(\mu, p) = d(p)(p - \mu). \quad (1)$$

Let $p_{opt}(\mu)$ denote the optimal price the seller would charge if he observed unit input cost μ perfectly.

In a framework of rational inattention, agents realize the limitations of their information processing capabilities. They are free to choose their information processing mechanisms as long as their information capacity is not exceeded. Agents in fact choose a form of noise in their observations, or pieces of information that are important to them. However, they are constrained with how much information they can process.

Knowledge about a random variable is expressed by its perceived probability distribution. The less dispersed the distribution is, the more precise the knowledge. In this model, the random variable of interest is the unit input cost μ . Prior knowledge about it is represented by a distribution $g(\mu)$. The seller is rational, therefore his prior knowledge coincides with the true distribution of costs. The degree of knowledge is then related to how concentrated the distribution is.

Posterior knowledge, represented by another distribution, is formed from the prior knowledge through an acquisition of signals. Posterior knowledge should be more informative and its entropy thus lower. Entropy is a measure of dispersion of distribution. A fundamental result in information theory states that if signals are acquired via a channel of a capacity κ , then the entropy of knowledge can not on

average decrease by more than κ . In other words, the seller can not learn too much about the input cost. He can however choose a form of noise in the signals, i.e. what parts of the knowledge distribution to narrow down.

The seller chooses what pieces of information about cost μ to process based on: i) what he knew in advance, $g(\mu)$, and ii) the relative importance of various pieces of information given by the shape of his profit function $\Pi(\mu, p)$.

Following a savings problem studied in (Christopher A. Sims 2006), the whole decision process can be modeled by an optimization problem (2)-(6). Its solution is a joint distribution $f(\mu, p)$ of cost and price, which is a compact way of representing the prior, the distribution of signals together with responses to those signals.

If the seller observed μ , he would simply choose an optimal price p . However, rationally inattentive sellers need to choose how to process information. The information processing mechanism is defined by what signals, s , can be received given a unit input cost, μ - this is specified by a conditional distribution, $f(s|\mu)$. A prior, $g(\mu)$, and the conditional distribution $f(s|\mu)$ define a joint distribution $f(\mu, s)$.

An acquired signal together with a prior distribution generate through Bayesian updating a unique posterior. Given a posterior about input cost, the seller chooses an optimal price, p . Since signals can be represented by prices they lead to¹, the process of information processing and responding to imperfect knowledge can be summarized by the joint distribution $f(\mu, p)$.

Conditional distribution $f(\mu|p)$ represents a posterior about cost leading to a selected price equal to p , while distribution $f(p|\mu)$ is a distribution of prices given the realized unit input cost is μ . If the seller had an unlimited information capacity, the conditional distribution $f(p|\mu)$ would be degenerate, it would be a delta function at $p = p_{opt}(\mu)$.

¹To economize on information capacity, it is never optimal to choose two signals leading to the same price

Moreover, the joint distribution allows for expressing the expected profit, which is to be maximized.

$$f(\mu, p) = \arg \max_{f'(\cdot, \cdot)} E[\Pi(\mu, p)] = \arg \max_{f'(\cdot, \cdot)} \int_{\mu} \int_p \Pi(\mu, p) f'(\mu, p) d\mu dp, \quad (2)$$

subject to

$$\int_p f'(\mu, p) dp = g(\mu) \quad \forall \mu \quad (3)$$

$$f'(\mu, p) \geq 0, \quad \forall \mu, p \quad (4)$$

$$I(\mu; p) \leq \kappa. \quad (5)$$

$I(\mu; p)$ is mutual information between μ and p , defined as

$$\begin{aligned} I(\mu; p) &= H(\mu) - H(\mu|p) = \\ &= \int_{\mu} \int_p f'(\mu, p) \log\left(\frac{f'(\mu, p)}{g(\mu)f'(p)}\right) d\mu dp. \end{aligned} \quad (6)$$

(3) requires consistency with prior knowledge and (4) states non-negativity of a probability distribution. (5) is the information constraint.

In case the seller could observe the cost μ , he would charge an optimal price $p = p_{opt}(\mu)$. However, if his information capacity does not allow him to learn the cost exactly², we should expect that pricing will not be optimal. For a given μ , the distribution of selected prices will be noisy and somehow dispersed around $p_{opt}(\mu)$. The higher the information capacity, the tighter this distribution should be. Pricing will also be more precise in regions where the seller's profit function is more concave. In such regions, processing information is relatively more valuable since the seller loses more from suboptimal pricing.

²An agent can learn the input cost exactly even with a limited information capacity, for example if the prior distribution consists of two points only. In such a case, the agent needs just 1 bit of information to observe the cost.

2.1 Solutions of the Model

The seller chooses an optimal strategy, which is a solution to (2)-(6). The demand function is assumed to be

$$d(p) = p^{-\theta}, \quad (7)$$

where θ is its price elasticity. Unit input cost will in all studied cases have a bounded support, it will typically be uniformly distributed. Unfortunately, I was not able to solve the model analytically. Thus far, no analytical solutions are available for most setups under rational inattention. However, most properties can still be discussed qualitatively using insights about the structure of the model.

To solve the problem numerically, I coupled an optimization language AMPL with a solver LOQO³. The solver is based on interior point methods, which are very efficient for large scale convex optimization problems, such as this one. The two dimensional domain of μ and p was discretized into 70×70 cells.

If the seller had unlimited information capacity, he would know the input cost exactly and would charge the optimal price

$$p_{opt}(\mu) = \frac{\theta}{\theta - 1} \mu. \quad (8)$$

The joint distribution of price and cost would thus be concentrated on the line given by (8). Since the seller can not learn the price exactly, his knowledge about the price will be noisy. Therefore, the joint distribution will be somewhat dispersed. However, the seller would still like to keep his pricing strategy as close to the optimal line as possible, realizing limits to his inattention.

Let us first inspect some basic properties of the seller's strategies under the constraints on his information capacity. Figure 2 shows the 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 $(0.8, 1.2)$. The dashed line represents the optimal pricing strategy arising under perfect information. $\theta = 3$ is a value used

³(Robert J. Vanderbei 1999)

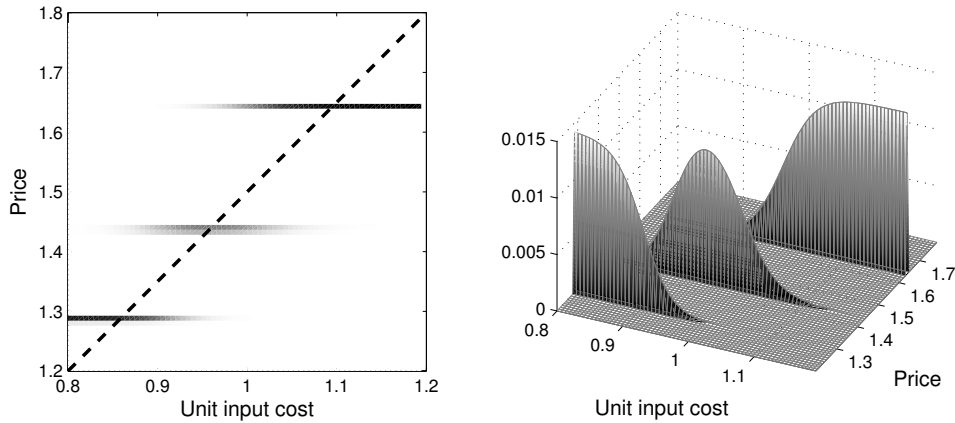


Figure 2: Joint distribution: $\theta = 3$, $\kappa = 1$ bit, uniform cost

in (Kehoe & Midrigan 2007). By the seller’s choice, there are only three different signals he acquires. Each of these signals then leads to a different choice of price. For low input costs, the seller is the most likely to realize the posterior knowledge on the left, that input cost is most probably somewhere between 0.8 and 0.95. This signal 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 1.64 .

In rational inattention, agents have complete freedom in acquiring all shapes of signals they like, as long as their information capacity is not exceeded. It assumes, that all signals are available. A signal on a unit input cost can for example be a noisy representation of the cost, an aggregate cost of a group of products or a first digit of a unit input cost of the best selling product in the same category. A first digit of another product’s input cost would provide signals similar to those in Figure 2. A digit takes a few different values only, so there would be a few different signals only, just like in Figure 2.

A simulated realization of a time series of prices corresponding to this solution is shown in Figure 3. The most apparent properties of solutions of the model are the following.

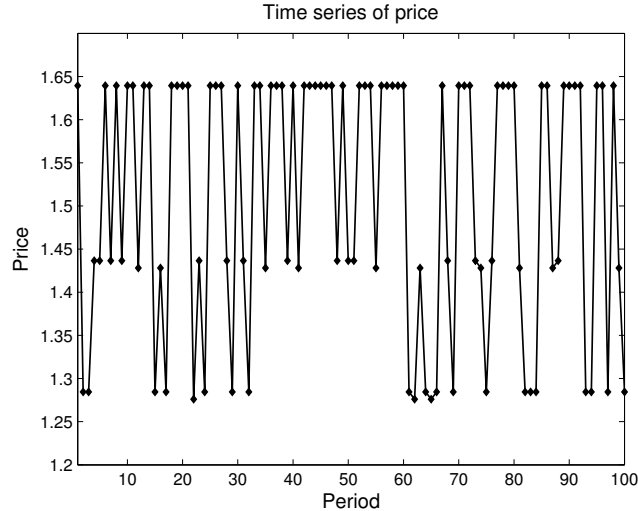


Figure 3: Simulated time series of price, $\theta = 3$, $\kappa = 1$ bit

1. Pricing strategies are sub-optimal. Figure 2 shows that the seller almost never charges the optimal price given by the dashed line. In case the seller processed more information, the joint distribution would be more closely concentrated around the dashed line of optimal pricing. Figure 4 presents a solution to a case with 2 bits of information capacity.

Markups deviate from the optimal markup $1/(\theta - 1)$. Standard deviations of log-markup are 0.16 for $\kappa = 1$ and 0.08 for $\kappa = 2$. The more information is processed, the less volatile markups are since prices are more tightly distributed around the optimal price. Due to the sub-optimality of pricing, the seller loses some fraction of expected profit. The higher the information capacity, the smaller the amount of the profit is lost. The marginal value of information is however decreasing. Already 2 bits of information are sufficient to recover 99.8% of expected profit under perfect information.

2. Posterior knowledge is imperfect. Due to limits to abilities of processing information, the seller cannot know the input cost exactly, and thus cannot respond optimally. According to the constraint (5), he can only acquire signals that imply

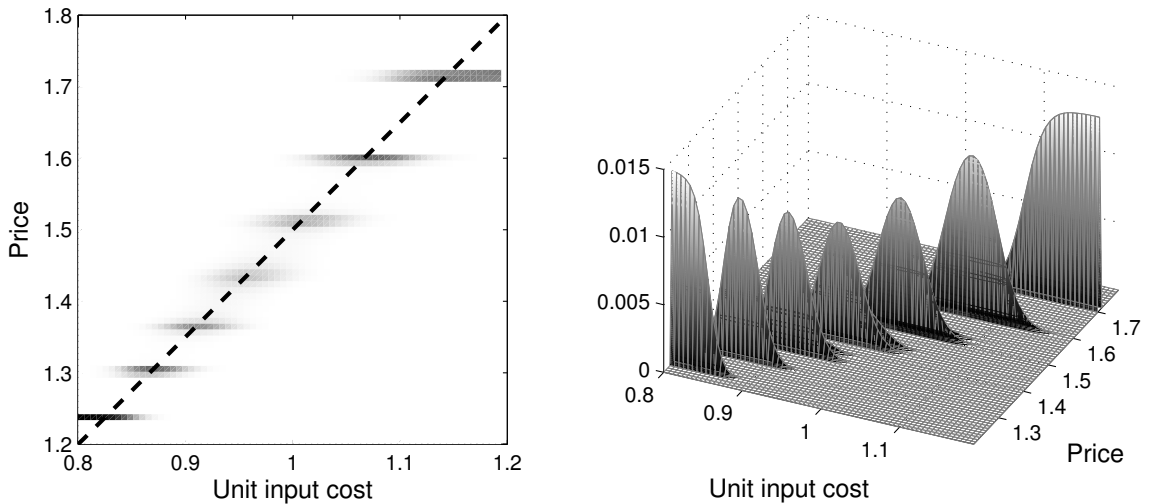


Figure 4: Joint distribution: $\theta = 3$, $\kappa = 2$ bits

posterior knowledge represented by a distribution that is not too concentrated.

Posterior knowledge coincides with the conditional probability distribution $f(\mu|p)$. In Figure 2, different forms of posterior knowledge are represented by the three clearly visible distributions along the lines of fixed prices. Figure 4 corresponds to a solution with higher information capacity and the posterior knowledge about input cost is therefore tighter, i.e. it is more precise.

If $f(p) > 0$, then the first order condition for the conditional distribution is:

$$f(\mu|p) = e^{\nu(\mu)} e^{\Pi(\mu,p)/\lambda}, \quad (9)$$

where $\nu(\mu) \in \mathbb{R}$ is a Lagrange multiplier on the constraint (3) and $\lambda > 0$ is a multiplier on the information constraint (5). Since $e^{\nu(\mu)} > 0$ for all μ such that $g(\mu) > 0$, posterior knowledge has the same support as the whole prior distribution of input cost. In other words, 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 signals overlap completely.

3. Signals and responses are noisy. Existence of noise in signals and responses is

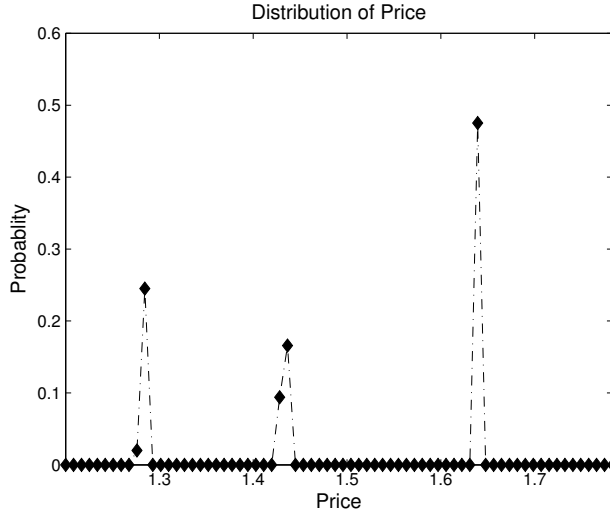


Figure 5: Overall distribution of price, $\theta = 9$, $\kappa = 1$ bit

related. Noise in signals means that given a unit input cost, the seller might acquire different signals and thus realize different posterior knowledge. Different forms of posterior knowledge lead to different responses in prices. Noise in responses is therefore generated by noise in signals. Since signals overlap, they can be realized for all different input cost.

4. Pricing is discrete. Inspecting the numerical solutions, it seems that the seller charges only a small number of different prices. The distribution $f(p)$ of prices is shown in Figure 5. It is a marginal distribution of $f(\mu, p)$, which is obtained after integrating the joint distribution over μ . The distribution of prices is supported by three narrow regions consisting of one or two cells only. However, the nature of numerical solutions does not allow us to infer with certainty whether the true analytical solution is completely discrete, or just highly concentrated.

However, complete discreteness tends to arise in various setups under rational inattention. This phenomenon is studied in (Matějka & Sims 2010). For example, we show that analytically that for quadratic utility functions, a prior distribution with a bounded support always leads to complete discreteness in responses.

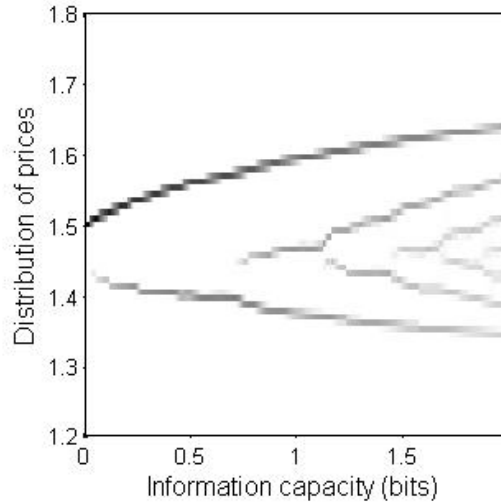


Figure 6: Bifurcation diagram of prices, $\theta = 3$

It is worth mentioning that the discreteness arises although all functional forms appearing in the model are completely continuous. Unfortunately, in the present settings, we have to rely on evidence from numerical solutions. By acquiring signals of discrete values, the seller can economize on information, which in turn allows him to respond more accurately. The more information the seller possesses the finer discretization occurs. The bifurcation diagram in Figure 6 shows price distributions as a function of information capacity. When the capacity increases, new price points emerge.

5. Pricing is asymmetric. Although the distribution of unit input costs is uniform, the resulting distribution of prices shown in Figure 5 is not symmetric. The highest realized price is chosen with higher probability than the lower prices.

The seller chooses to process the least amount of information in the region of high costs and thus charges the highest price very often to a wide range of different unit input costs. In this region, marginal utility from processing extra information is relatively lower.

Although I have not found an analytical solution to the model, some intuition for the

result of asymmetric pricing can be obtained if we abstract from the full generality of rational inattention. It is shown in Appendix A that the seller is likely to choose more precise signals in regions where a loss factor given by (22) is higher. The factor indicates magnitudes of losses due to imperfect knowledge. Change in profit due to misjudging unit input cost μ by a small amount ϵ is approximately equal to $-L(\mu)\epsilon^2/2$. For the profit function $p^{-\theta}(p - \mu)$,

$$L(\mu) = \theta \left(\frac{\theta}{\theta - 1} \right)^{-\theta} \mu^{-\theta-1}, \quad (10)$$

it is proportional to $\mu^{-\theta-1}$. For a fixed θ and the same amount of error ϵ , losses are larger for lower μ . The seller thus chooses to pay more attention to regions of low μ . The highest price has the highest probability of being realized - discretization is coarsest at high prices.

If the distribution of unit input cost is uniform, the most probable price is always the highest price. The distribution of prices approximates the distribution of optimal prices as responses to all costs. If the distribution of costs were not uniform, if it had a concentrated area with high density in the middle and tails with lower mass of probability, the resulting distribution of prices would also have less probable tails.

2.2 Choice of Information Amount

Thus far, we have addressed the question of attention distribution, while the total amount of information was kept fixed. However, we could assume that agents also choose how much information to process. Let RI_κ stand for the original model of a rationally inattentive seller with a fixed information and RI_λ denote a model with fixed unit cost of information. In RI_λ , sellers find the processing somewhat unpleasant and process more information only as long as its cost is lower than the marginal benefits from it. Instead of maximizing Π given by (1), they maximize expectation of

$$\tilde{\Pi}(\mu, p, \kappa) = \Pi(\mu, p) - \lambda\kappa, \quad (11)$$

where λ is the cost of processing 1 bit of information about μ , and κ is the amount of information the price-setter chooses to process. The RI_λ problem takes the same form as RI_κ , (2)-(6), except that (2) is replaced by

$$f(\mu, p) = \arg \max_{f'(\cdot, \cdot), \kappa} E[\Pi(\mu, p)] = \arg \max_{f'(\cdot, \cdot), \kappa} \int_{\mu} \int_p \tilde{\Pi}(\mu, p, \kappa) f'(\mu, p) d\mu dp. \quad (12)$$

For a given prior distribution, RI_κ and RI_λ are equivalent. Solutions can be parameterized by either of the two quantities, κ or λ . If κ is fixed, λ is a Lagrange multiplier on the information constraint (5). For instance, If $\theta = 3$ and input cost is uniformly distributed in $(0.8, 1.2)$, then $\kappa = 1$ corresponds to $\lambda = 3 \cdot 10^{-3}$, while $\lambda = 6 \cdot 10^{-4}$ when $\kappa = 2$. In these cases, expected profit is 0.15, marginal value of information is thus about 2% of the profit when $\kappa = 1$ and 0.4% for $\kappa = 2$.

However, given the same λ , different θ can lead to a different choice of the information amount. In fact, the higher elasticity of demand the more information the agent chooses to process. Therefore, the model predicts prices to be more flexible and conveying more information in industries with higher demand elasticity. In Hayek's words, competitive markets convey more information.

Appendix B discusses the relation between κ and λ in a linear-quadratic problem. In dynamic models, distribution of knowledge changes - a setup with a fixed λ can lead to a volatile κ since different knowledge results in different value of additional information.

3 Richer Dynamics

In this section, I will discuss an extension to the original model, (2)-(6), that will make it truly dynamic. Previously, the input cost was i.i.d. The price setter processed information about it and decided what price to charge. Since the input cost was not serially correlated, the processed information had no significance to agent's knowledge about next period's cost. Responses were imperfect, but they were not

delayed. In the following model, there are two stochastic variables to which the seller responds. Let one be an i.i.d. real unit input cost denoted by μ and the second one be a serially correlated nominal variable A . μ is supposed to be an idiosyncratic volatile part of the input cost specific to the seller, while A plays the role of a slowly moving aggregate variable, e.g. a price level. I will assume that the profit function has the following form.

$$\Pi(A, \mu, p) = p^{-\theta}(p - A\mu). \quad (13)$$

A is a price index shifting the distribution of the nominal input cost $A\mu$.

Solving fully dynamic problems under rational inattention is very demanding. A state variable is current knowledge, which is given by a distribution. For continuously distributed variables, the distribution is thus an infinitely dimensional object. For the sake of simplicity, I will assume that the aggregate variable takes two different values only and that it is Markov. Let the Markov process be symmetric with a probability of transition to the other state equal to t .

The seller maximizes discounted future profits. Since μ is i.i.d., the only state variable is the seller's knowledge about A . A is assumed to be binary, its distribution is determined by the probability of either one of the two states. Let the state variable be $x = \text{Prob}(A_L)$, where A_L stands for the lower value of A .

The recursive formulation of the seller's dynamic problem, a RI_κ version, is as follows.

$$V(x) = \max_f \int \left[\Pi(A, \mu, p) + \beta V(x') \right] f(A, \mu, p) dA d\mu dp, \quad (14)$$

subject to

$$x' = f(A_L|p)(1-t) + (1-f(A_L|p))t \quad (15)$$

$$\int f(A, \mu, p) dp = g(A, \mu) = g_1(A)g_2(\mu) \quad (16)$$

$$g_1(A_L) = x \quad (17)$$

$$I(A, \mu; P) \leq \kappa. \quad (18)$$

$f(A, \mu, p)$ is a joint distribution summarizing the seller's choice of signals and re-

sponses in the given period, (15) is law of motion for knowledge, it generates a prior on A in the following period from a posterior in the current period via a Markov process with the transition probability t . (16) is the constraint on a prior, μ and A are assumed to be independent. $g_2(\mu)$ is fixed, $g_1(A_L) = x$, (17), which implies $g_1(A_H) = 1 - x$. (18) is the traditional constraint on mutual information between the source variables μ and A , and a response variable p .

For computations, I used $\kappa = 1$, $\theta = 3$, μ uniformly distributed over $(0.8, 1.2)$, $A_L = 1$, $A_H = 1.1$, $t = 0.002$ and $\beta = 0.9992$. One period is supposed to be one week. $t = 0.002$ implies that the probability of changing a state (a 10% shock to the aggregate variable) at least once during a year is about 10%. The annual discount factor is 0.96.

Figure 7 shows the results of simulations over 120 periods. There is a shock to A in the period denoted as 1, when A switches from A_L to A_H . The top series in the figure presents one realization of a price series and the second one shows a time-series of knowledge about A in the same simulation.

The price setter processes information about $A\mu$ and responds to it, trying to target the optimal price $\theta/(\theta - 1)A\mu$. Although $A\mu$ is distributed over a continuous range in every period, prices again exhibit lots of rigidity of the values as well as in the i.i.d. case. Given prior knowledge about A , together with its true value, the distribution of prices as responses to realizations of μ is discrete. However, when knowledge about A changes, the distribution of prices changes too. For the used values of parameters (κ , t , A_H , etc.), knowledge adjustment is rather abrupt. The second series in Figure 7 shows a knowledge adjustment that is quite typical for all realizations of single simulations with these 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. The sudden change of knowledge is, however, not inherent to all solutions under rational inattention. The next subsection discusses this point in a little more detail.

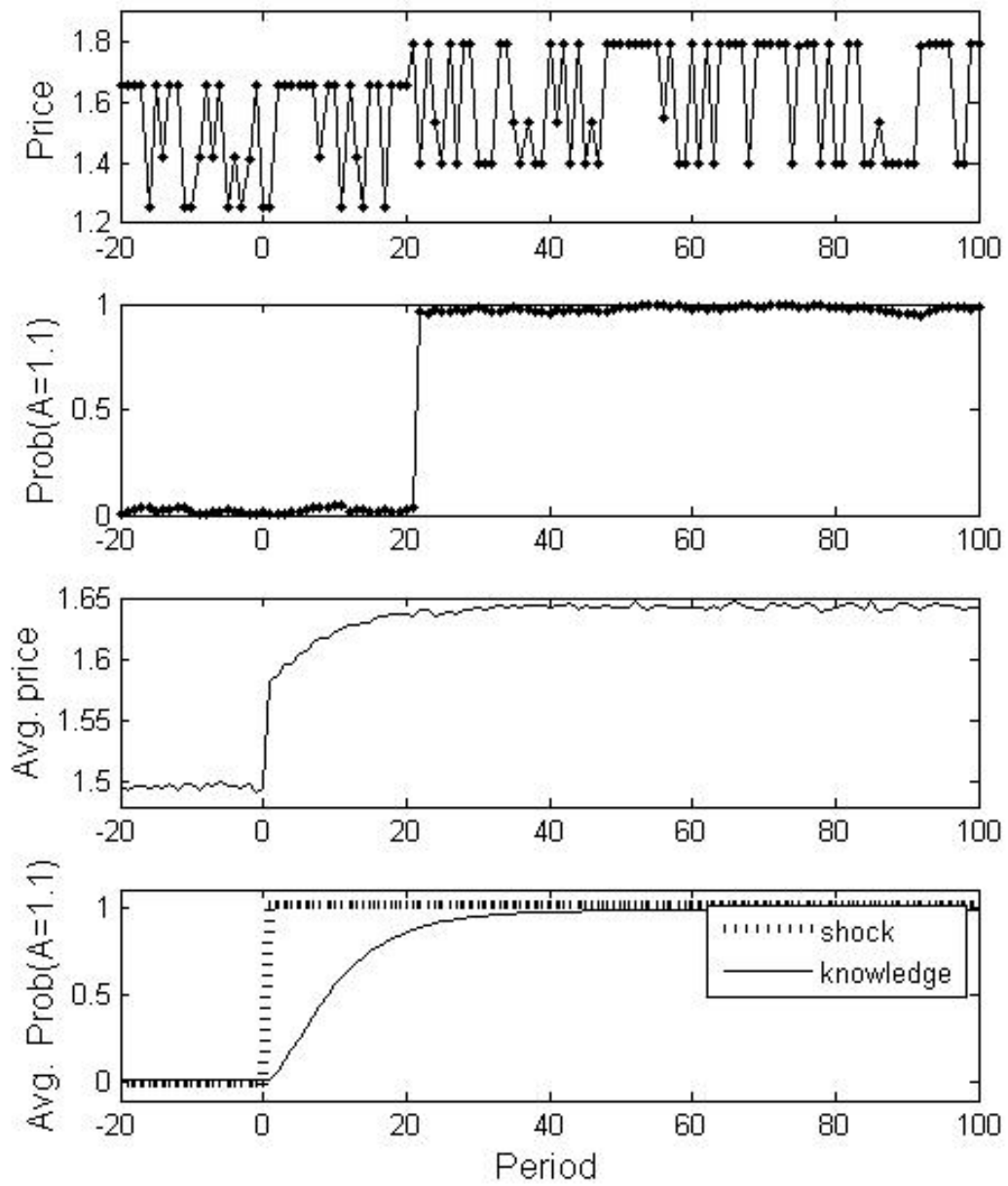


Figure 7: Two stochastic variables, sudden learning, $t = 0.002$, $\kappa = 1$

The bottom two series in Figure 7 are prices and knowledge averaged over 10 000 runs. The average knowledge about A shifts slowly, while the average price does actually change abruptly in period 1. The variable of interest to the seller is in fact $A\mu$, not values of A and μ separately. Due to different dynamical properties of A and μ , and a non-uniform prior on $A\mu$, the agent does not process information exactly about $A\mu$ only. Although, the seller does pay special attention to $A\mu$, he also refines knowledge about other regions in the whole $A \times \mu$ space. In period one, after a positive shock to A , the seller is likely to find out that the value of $A\mu$ is high and thus the probability that a distribution's top price is realized increases. Due to the prior knowledge that A is probably at the lower state, the agent underestimates the true value of $A\mu$. Expected price adjusts abruptly, but still less than optimally. Since A stays at the higher level, the agent obtains signals on a high $A\mu$ several periods in a row and slowly learns that it is not due to a streak of high μ , but rather due to a jump in A . The average price further increases towards the new optimal level. Prices change frequently, but responses to shocks to the aggregate variable are delayed.

3.1 Sudden or Gradual Learning and Responses of Averages

Figure 7 presents solutions for $t = 0.002$ and $\kappa = 1$. Since the probability of transition from A_L to A_H is very low, the seller pays attention mostly to the joint characteristics $A\mu$ only. When he processes a signal that the input cost is high, then it is usually attributed to a high μ . Once more such signals are processed in succession, the agent's prior changes a little bit. At that infrequent moment, the seller chooses to process more information specifically about A and very quickly finds out what its true value is, with little noise.

For different parameters, however, the knowledge dynamics can be different. If for instance the probability t were higher, then the transition would be smoother since the agent would be more willing to accept that A has changed.

To avoid confusion, let us emphasize that also the abrupt change of expected price in period 1 relies on the specific form of stochastic properties of μ and A . The seller pays attention to sudden changes of $A\mu$, since μ is very volatile. Such changes are thus 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 responses to the same shock depends on how unexpected is a resulting shock to the joint characteristics of interest. In such several-variable models, knowledge about joint characteristics that determines agent's decisions adjusts faster if the characteristics are more stable. On the other hand, knowledge about more volatile single variables is more up-to-date relatively to other variables. Decision responses and knowledge about the joint characteristics typically adjust faster than knowledge about single variables.

Figure 8 presents solutions to a dynamic problem with μ being fixed at 1, the only stochastic variable left is A . The seller's information capacity is $\kappa = 0.02$. In this case, responses to A and knowledge about A are perfectly synchronized. 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 is, however, not given deterministically.

Such a dynamics of knowledge resembles the one assumed in the sticky-information model, introduced in (N. Gregory Mankiw & Ricardo Reis 2002). They postulate that agents' information updating 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 8 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

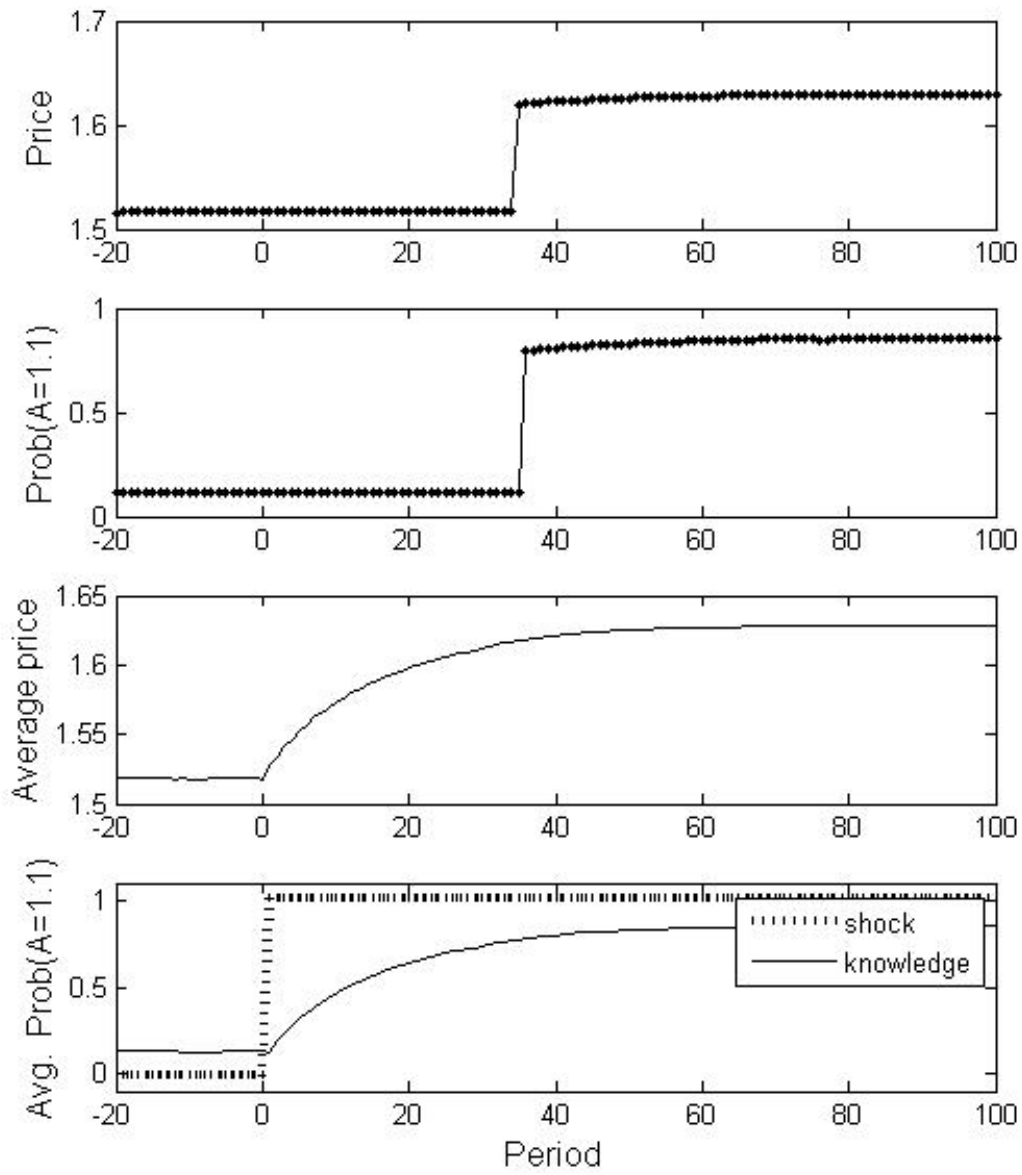


Figure 8: Serially correlated variable only

change of knowledge.

A model presented in (Woodford 2008) also generates similar time-series of knowledge and prices. The model assumes that agents process information to make a binary decision only: whether to update price or not. Prior knowledge is fixed, information processed in every period between the updates is immediately forgotten. The solutions in Figure 8 have a similar property to that of the rigid knowledge. The agent chooses to acquire one of two different signals only. If he keeps receiving a signal that price did not change, then his prior stays virtually unchanged. Knowledge stays pretty much constant for a while even after a shock occurs. Once the other signal is received, both knowledge and price jump discretely. Price then quickly stabilizes on a new level, which is again somewhat rigid. Unlike in the model of Woodford, this behavior is not assumed, but it can emerge endogenously.

3.2 Rational Inattention vs. Signal Extraction

In the presented model, the described dynamics of shocks, knowledge and responses rather resembles the results of Lucas's signal extraction model, (Robert Jr. Lucas 1972). In the spirit of signal extraction, an agent would gauge $A\mu$ through another variable. Let us say, he could observe $\xi \cdot A\mu$, where ξ is a random variable. In case $A\mu$ were not revealed after every period, impulse responses to a shock to A would be similar to those in Figure 7. At first, price would respond rapidly, but not completely since the agent would not know whether the shock came from ξ or $A\mu$. Since A is very persistent, the agent would perceive the shock coming from μ rather than from A - knowledge about A would adjust slowly, just like in Figure 7.

As noted in (Sims 2003) some implications of rational inattention and signal extraction go in the opposite directions. If $A\mu$ was very stable, most variations in $\xi \cdot A\mu$ would be attributed to a shock to ξ and responses would be very delayed. On the other hand, in rational inattention, stable $A\mu$ is easier to be tracked and responses to its shocks are thus swifter. In case a stochastic variable switches between two

states only, an agent needs 1 bit of information to track it perfectly and respond immediately. Moreover, if the variable is mostly at a state S_1 and once in a long while it switches to a state S_2 , the information necessary for perfect tracking is even lower, since the entropy of non-uniform binary distribution is lower than 1. Unlike signal extraction, rational inattention implies that responses are closer to the optimal ones at times when the underlining stochastic variables of interest are more stable.

In rational inattention, inference of μ from $A\mu$ has a flavor of Lucas' signal extraction, rather than inference of $A\mu$ itself. A rationally inattentive agent processes information about the optimal joint characteristics he chooses and then updates knowledge about the rest.

(Mackowiak & Wiederholt 2007) also studied pricing responses to two different stochastic variables. Their setup is partially in the sense of rational inattention. Unlike in rational inattention, an agent in their model has to process signals about the two variables separately, while the total amount of information capacity is fixed. Following their approach, the agent would desire to pay more attention to shocks to μ , and thus respond to innovations in A very slowly. Processing information purely about μ does not convey any information about A . In their setup, price responses to shocks to A would be as smoothed and as delayed as knowledge about A .

3.3 RI_κ and RI_λ : Implications for Aggregate Dynamics

A RI_λ version of the dynamic model takes the form of:

$$V(x) = \max_{f, \kappa} \int \left[\Pi(A, \mu, p) + \beta V(x') \right] f(A, \mu, p) dA d\mu dp - \lambda \kappa, \quad (19)$$

subject to (15)-(18).

Stochastic properties of the variables of interest (μ and A) influence the agent's choice of what pieces of information and potentially how much information to pro-

cess. Let us vary these properties and study their implications for responses to shocks to the aggregate variable A .

3.3.1 Idiosyncratic Volatility

Thus far, μ was uniformly distributed in $(0.8, 1.2)$. Table 1 summarizes numerical results for RI_κ with $\kappa = 1$, $\theta = 3$ and $t = 0.002$ for three different widths of the distribution of μ : for μ fixed at 1, uniformly distributed in $(0.9, 1.1)$ and in $(0.8, 1.2)$. Two characteristics of responses to a shock to A averaged over 10 000 runs are in columns 2 and 3. “1_{st} per. adj.” represents a portion of the average long-term adjustment that was realized during the first period, while “90% adjustment” denotes the number of periods it takes the average price until 90% of the full adjustment is realized.

μ	1 _{st} per. adj.	90% adjustment	profit loss
1	100%	1	0%
(0.9,1.1)	83%	2	0.28%
(0.8,1.2)	61%	11	1.09%

Table 1: Implications of idiosyncratic volatility for average responses, $\kappa = 1$

The more volatile is the seller’s idiosyncratic part of the input cost the slower he responds to aggregate shocks. When μ is fixed at 1, 1 bit of information is sufficient to track innovations of the binary variable A perfectly. The column “profit loss” presents seller’s losses in comparison with pricing under perfect information - this quantity was evaluated with both μ and A simulated according to their stochastic properties. As expected for RI_κ , losses are higher in more volatile environments.

Results of the similar experiments for RI_λ , $\lambda = 0.003$, are shown in Table 2 and in Figure 9. Unlike RI_κ , RI_λ generates faster responses to aggregate shocks when μ is more volatile. The table also presents average κ that was selected by the seller during the simulations. Agents in RI_λ choose to process more information when volatility

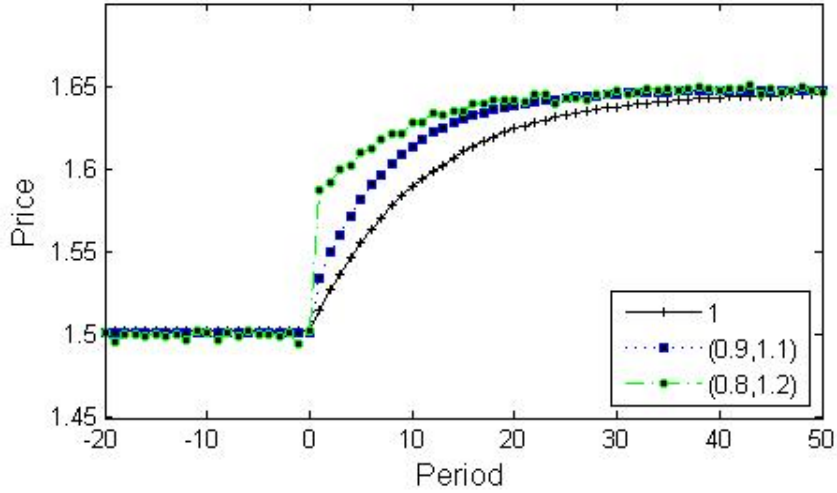


Figure 9: Average price response, 3 distributions of μ , $\lambda = 0.003$

μ	1 st per. adj.	90% adjustment	profit loss	mean κ
1	8%	25	0.08	0.0086
(0.9,1.1)	22%	17	1.34%	0.014
(0.8,1.2)	60%	12	1.34%	0.86

Table 2: Implications of idiosyncratic volatility for average responses, $\lambda = 0.003$

of input cost increases, which increases the marginal value of information. Since a rationally inattentive agent processes optimal joint signals about $A \times \mu$, then more total information also implies more information about A . However, faster average responses do not always mean more precise responses - profit loss is the same for $\mu \in (0.9, 1.1)$ and $\mu \in (0.8, 1.2)$. This is an analogy of the result for the RI_λ linear-quadratic problem in Appendix B: varied volatility of a prior influences the choice of κ , but not the resulting imperfections in tracking. However, the loss drops dramatically when μ is fixed at 1. Input cost becomes a binary variable, thus the intuition derived from the linear-quadratic approximation does not apply very well.

In RI_λ , the amount of processed information is not constant. It varies according to the expected value of information given a prior. Figure 10 presents average information amount as a function of time. Immediately after the shock occurs, a

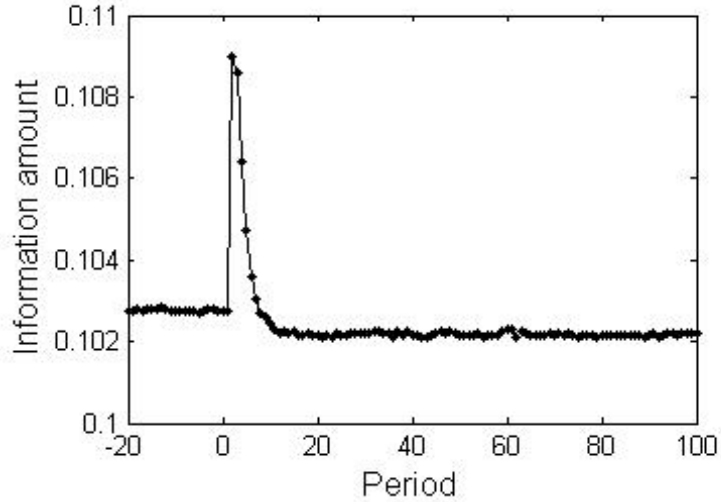


Figure 10: Average processed information, $\mu \in (0.9, 1.1)$, $\lambda = 0.003$, $t = 0.02$

large fraction of sellers realize there might have been a shock and they choose to process more. Later on, average selected capacity decreases. Once the transition period is over, the new equilibrium information capacity is actually below the initial level - the new average input cost is higher and the marginal value of information is thus lower.

The model of Mackowiak and Wiederholt would generate slower and less precise responses to A if volatility of μ were increased. Total κ is fixed, while marginal value of information devoted to μ increases and no joint signals are available. Less information capacity is therefore devoted to A .

3.3.2 Aggregate Volatility

Aggregate volatility can be adjusted by varying the Markov parameter t , the probability of transition between the two states. Tables 3 and 4, present characteristics of average responses for $\kappa = 1$ and $\lambda = 0.003$ for four different levels of t .

More volatile A generates faster responses to its innovations in both of the models, RI_κ and RI_λ . This is due to a higher marginal value of processing new information

t	1 _{st} per. adj.	90% adjustment	profit loss
0.001	56%	18	1.08%
0.002	61%	11	1.09%
0.006	73%	7	1.13%
0.02	80%	4	1.20%

Table 3: Implications of aggregate volatility for average responses, $\kappa = 1$

t	1 _{st} per. adj.	90% adjustment	profit loss	mean κ
0.001	55%	19	1.34%	0.86
0.002	60%	12	1.34%	0.86
0.006	73%	5	1.35%	0.88
0.02	83%	3	1.35%	0.92

Table 4: Implications of aggregate volatility for average responses, $\lambda = 0.003$

about A if A is more likely to vary and due to the signal extraction of A from $A\mu$. When volatility of A increases, shocks to $A\mu$ are more likely to be attributed to A . However, unlike in the Lucas' signal extraction of the whole $A\mu$, profit losses are higher when the aggregate environment is more volatile. Less stable $A\mu$ is more difficult to be tracked precisely.

The effect of accelerated average adjustment is slightly stronger in RI_λ , since higher volatility provides additional motive for processing more total information.

4 Confronting the Evidence

4.1 I.i.d. Input Cost

The model generates time series of prices that are appealing in various aspects. In agreement with the data, prices change frequently and jump between a few different values. The most quoted price is often the highest price. However, unlike in (Kehoe

& Midrigan 2007), the highest price and the most quoted price do not have any special significance and they do not always coincide.

Reference prices: the model generates asymmetry in pricing, with the highest price typically being the most quoted, which is in agreement with the data. (Eichenbaum, Jaimovich & Rebelo 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 $\kappa = 1$ and $\theta = 9$. If $\theta = 3$, then the fraction is about 50%.

Distribution of markups: Unlike flexible price models, the presented model does generate dispersion in markups. The standard deviations of markups however vary greatly. For $\theta = 3$, the standard deviation is between 0.04 and 0.16 if κ is between 0.5 and 4. Median standard deviation reported in (Eichenbaum, Jaimovich & Rebelo 2008) is 0.12. However, if the information capacity is fixed at $\kappa = 1$, the standard deviation of $\log(\text{markup})$ is as high as 0.7 once θ reaches 10.

State dependence: Solutions to the presented model do of course feature state dependence since price changes are driven by changes in cost. (Eichenbaum, Jaimovich & Rebelo 2008) measure the probability of changing a weekly price as a function of percentage deviation from a hypothetical markup. Given a realized input cost, a hypothetical markup is a markup that would be realized if price stayed constant. In the model, $\theta = 3$ and $\kappa = 2$ generate a state dependence that agrees with the evidence quite well.

Across industries: if information capacity can be chosen, RI_λ model, then sellers process more information if elasticity of demand is higher. Higher information capacity generates more flexible prices. It is well documented that prices are more

flexible in more competitive industries⁴.

4.2 Serially Correlated Input Cost

Price series on a micro level: (Eichenbaum, Jaimovich & Rebelo 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 the model does not have any special significance; it is just one of the characteristics of the whole distribution of prices.

At times when knowledge about the aggregate variable fluctuates only a little, then prices take values in a few very narrow regions only. Unlike in the purely i.i.d. case, the values need not be completely rigid. Figure 7 presents responses to an aggregate shock in an otherwise very stable aggregate environment. In this case, knowledge about the aggregate component of cost can adjust with a long delay, but once it does, it adjusts very swiftly. Such an environment generates rather rigid values of prices, similar to those in (Eichenbaum, Jaimovich & Rebelo 2008), where transitions between different distributions of prices are very rapid. See for example the price-series of Skinner long spaghetti between weeks 50 and 60 in Figure 1.

On the other hand, Figure ?? shows a resulting series when the aggregate variable is more volatile. After a 10% shock to the aggregate variable occurs, the formerly almost rigid values become much more flexible. In the data, prices do tend to be more flexible in periods of high inflation ⁵, but probably not as much as this model would predict. In the model, for instance, if the aggregate variable was deterministic with an upward trend, generated values of prices would be completely flexible.

However, it is difficult to conclude much since we do not know the true nature of shocks to retailers. It is possible that the abrupt transitions in the data correspond

⁴For instance, see (Luis J. Alvarez & Ignacio Hernando 2005).

⁵For example, (Luis J. Alvarez, Pablo Burriel & Ignacio Hernando 2010)

to weeks when contracts with suppliers are renegotiated. The sellers know when this is going to happen and pay extra attention to it.

This discrepancy could potentially be solved by the introduction of the information constraint on the consumer's side too. It is shown in (Filip Matějka 2010) that a seller considering consumer's limitations in processing information can under some conditions choose to keep prices rigid. He does so to equip the consumer with better knowledge about prices, from which the seller benefits by the consumer's higher consumption. Such an extension could imply that the seller would choose to modify the whole distribution of prices quite infrequently even if the nominal variable A was continuously increasing.

Sectoral volatility: Section 3.3.1 discusses the model's implications of volatility of μ on the dynamics of average prices. (?) found evidence that prices in sectors with more volatile idiosyncratic shocks react to aggregate shocks faster. This finding agrees with the results of the RI_λ model, Table 2, where agents can choose how much information to process. If volatility of the idiosyncratic variable increases, sellers decide to process more information. However, since they process joint signals about both variables, more information is processed about the aggregate variable too.

The RI_κ model, with a fixed information capacity, gives the opposite results, Table 1.

Aggregate volatility: Results presented in Tables 3 and 4 summarize the properties of dynamics of average prices with respect to volatility of the aggregate variable. Both models, RI_κ and RI_λ , generate faster responses in more volatile environments. (?) observed that Philips curve becomes steeper in more volatile aggregate environments. The models' results agree with this finding.

Moreover, both RI_κ and RI_λ generate higher losses if the aggregate variable is more volatile, when tracking it becomes a more demanding task. The effect is weaker

for RI_λ , losses from increased volatility are partially offset by more information processed by the seller.

4.3 Stochastic Demand

Although the model is formulated in terms of responses to the stochastic input cost, its results do not rest on this particular assumption. It is true that uncertainty for instance about demand can in some cases play a more significant role than about the input cost. However, all interesting properties of the solutions can occur even with different sources of shocks. The price setter processes information about a stochastic variable and targets some optimal price. This targeting is imperfect and delayed. Discreteness arises as an artefact of economizing on information, while the asymmetry of prices is driven by the shape of the demand curve. The profit function has higher curvature at lower prices, and thus losses from suboptimal pricing are potentially higher, etc.

5 Conclusion

This paper presents a model of a rationally inattentive seller responding to shocks to a unit input cost. The model generates price series simultaneously exhibiting all three of the following features that can be found in the data.

1. Prices change frequently.
2. Responses of prices to aggregate variables are delayed.
3. Prices move back and forth between a few rigid values.

Such behavior of prices is hard to reconcile with any of the existing models of pricing. Menu cost and Calvo models can not in general generate very frequent

price adjustments together with delayed responses. Moreover, no existing models fully explain the observed rigidity of price values.

In the presented model, discrete pricing arises even if the unit input cost varies in a continuous range. Whatever unit input cost is realized, the price takes one of only a few values. Results of the model also agree with the evidence that reductions in price, e.g. sales, are usually short-lasting and that the highest price in the sample tends to be the most quoted price. Discrete and asymmetric pricing is a seller's optimal response to his limited information capacity.

The proposed mechanism does not attribute any special significance to any value of price, whether it is the most quoted or the highest price. One price is merely a single feature of the overall distribution of prices. Asymmetry of pricing towards the highest price emerges endogenously. Once the distribution of unit input costs changes, the distribution of prices changes too, together with the most quoted price.

Two versions of the model were discussed. One with a fixed constraint on information capacity and the other with a fixed cost of processing a unit amount of information. The second version is perhaps more appealing - if agents can choose what information to process, why should not they be allowed to choose how much to process. In the best scenario, we would construct a model with inattentive agents processing information about all variables they need to track in their everyday lives. When considering a restricted model with a few economic variables only, a convex cost of information could be the most appropriate due to increasing marginal returns of information about variables not included in the restricted model.

The costly processing of information provides the rationale for:

4. More flexible prices in industries with a higher degree of competition.
5. Faster responses to aggregate shocks in industries with more volatile idiosyncratic shocks.

In such industries, sellers choose to process more information.

Similarly to signal extraction, rational inattention provides an explanation for the following.

6. Philips curve becomes steeper when aggregate conditions are more volatile.

However, unlike signal extraction, rational inattention can generate:

7. Higher losses from imperfect tracking when aggregate variables destabilize.

Rationally inattentive agents are not assumed to acquire signals of any specific form - they simply process the most useful pieces of information. This is a necessary ingredient for several implications of the model.

- Agents process different amounts of information about different levels of input cost. Implication: discrete and asymmetric pricing.
- Agents can process signals about a joint characteristic of interest⁶. Implication: faster responses to aggregate shocks in more volatile industries.
- Agents can process information about variables they find the most important. Implication: faster responses to aggregate shocks, when the aggregate environment is less stable.

Other models based on imperfect information processing generate solutions that typically lack some of these features. Most of the models can be thought of as proxies to the fully general setup of rational inattention.

Rational inattention seems to be a powerful framework to study patterns of pricing; it provides a consistent explanation for many different features of the time series of prices.

⁶Results are thus independent of a specific selection of variables in the model.

A Approximate Losses

This section provides an informal analytical justification for some properties of numerical solutions under rational inattention. Let Y be a stochastic variable, which an agent tracks by processing information about it, and let Z a variable of agent's responses to Z maximizing expectation of profit $\Pi(Y, Z)$. Solving a full rational inattention problem means that the agent selects a joint distribution $f(Y, Z)$ under additional constraints on conveyed information and prior knowledge. The agent can acquire signals from a fairly complicated collection. To gain further insight, I will for now depart from the full generality of rational inattention assuming something more specific about signals.

Let the true realized value of the source variable be y^* , the seller acquires a signal $y' = y^* + \epsilon$, where ϵ is a random error. We will concentrate on in what regions of y the seller is likely to acquire more information, and where he would thus prefer a smaller error in the signal. Let

$$z = Z_{opt}(y) \tag{20}$$

be the optimal response to y , if the agent had perfect information. Let us assume that once the agent observes the signal y' , his response z is $Z_{opt}(y')$. Whether this is actually optimal depends on a specific form of the noise. However, it is the optimal response for example if the agent's posterior knowledge is symmetric about y' and if profit is a quadratic function of z for a fixed y . Moreover, it is close to the optimal response if signals are tight and concavity of profit as a function of Z does not change very rapidly.

Let $\Pi_{y^*}(y') = \Pi(y^*, Z_{opt}(y'))$, its Taylor expansion about y^* is

$$\begin{aligned} \Pi_{y^*}(y') &= \Pi_{y^*}(y^*) + \frac{d\Pi_{y^*}(y')}{dy'}\epsilon + \frac{d^2\Pi_{y^*}(y')}{dy'^2}\epsilon^2/2 + \mathcal{O}(\epsilon^3) \\ &= \Pi(y^*, Z_{opt}(y^*)) + \left[\frac{d^2 Z_{opt}(y')}{dy'^2} \frac{d\Pi(y^*, z)}{dz} \right. \\ &\quad \left. + \left(\frac{dZ_{opt}(y')}{dy'} \right)^2 \frac{d^2\Pi(y^*, z)}{dz^2} \right] \epsilon^2/2 + \mathcal{O}(\epsilon^3) \Big|_{y'=y^*, z=Z_{opt}(y^*)}. \end{aligned}$$

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

$$\begin{aligned}\Delta\Pi &= \Pi_{y^*}(y') - \Pi_{y^*}(y^*) \\ &= \left(\frac{dZ_{opt}(y')}{dy'}\right)^2 \frac{d^2\Pi(y^*, z)}{dz^2} \epsilon^2/2 + \mathcal{O}(\epsilon^3) \Big|_{y'=y^*, z=Z_{opt}(y^*)}\end{aligned}\quad (21)$$

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

$$L(y^*) = -\left(\frac{dZ_{opt}(y^*)}{dy^*}\right)^2 \frac{d^2\Pi(y^*, z)}{dz^2}.\quad (22)$$

The formula recognizes that loss depends on a curvature of the profit function and also on how far away from the optimal response the realized response is. The more sensitively responses change with values of acquired signals, the further away can a realized z be from the optimal one.

If the seller could decide in what regions of Y to pay more attention and decrease the noise, he would do so for such y^* where the loss factor $L(y^*)$ is higher.

B Linear-quadratic Case: RI_κ vs. RI_λ

If a utility function is linear-quadratic and noise is Gaussian, then the rational inattention problem simplifies significantly⁷. Agents acquire normally distributed posterior knowledge with a constant variance. Let Y be drawn from $N(0, \sigma_Y^2)$, its entropy is

$$H(Y) = \frac{1}{2} \log(2\pi e \sigma_Y^2).\quad (23)$$

Information constrain $H(Y) - H(Y|Z) \leq \kappa$ takes the following form.

$$\frac{1}{2} \log\left(\frac{\sigma_Y^2}{\sigma_{Y|Z}^2}\right) \leq \kappa.\quad (24)$$

⁷See (Sims 2003) and (Mackowiak & Wiederholt 2007)

Variance of posterior:

$$\sigma_{Y|Z}^2 = \sigma_Y^2 2^{-2\kappa}, \quad (25)$$

the less an agent knows at the beginning the less he knows after he processes a given amount of information, κ . Let the agent minimize a quadratic loss, maximize the utility: $U = -a(Y - Z)^2$. Its expectations then is:

$$E[U] = -a\sigma_{Y|Z}^2. \quad (26)$$

If information capacity is fixed, RI_κ ,

$$E[U] = -a\sigma_Y^2 2^{-2\kappa}, \quad (27)$$

worse prior knowledge implies higher losses.

A Lagrange multiplier λ is a shadow cost of information,

$$\begin{aligned} \lambda &= \frac{\partial E[U]}{\partial \kappa} = 2 \ln(2) a \sigma_{Y|Z}^2 2^{-2\kappa} \\ &= 2 \ln(2) a \sigma_Y^2 2^{-4\kappa}. \end{aligned} \quad (28)$$

Higher information capacity and a more concentrated prior correspond to lower λ .

With a fixed cost of processing information, (RI_λ) ,

$$E[U] = -\frac{\lambda}{2 \ln(2)}, \quad (29)$$

losses and quality of posteriors are independent of initial knowledge, unlike in RI_κ .

References

- Alvarez, Luis J., and Ignacio Hernando.** 2005. "Price Setting Behaviour in Spain. Evidence from Survey Data." European Central Bank.
- Alvarez, Luis J., Pablo Burriel, and Ignacio Hernando.** 2010. "Price-setting behaviour in Spain: evidence from micro PPI data." *Managerial and Decision Economics*, 31(2-3): 105–121.

- Bils, Mark, and Peter J. Klenow.** 2002. “Some Evidence on the Importance of Sticky Prices.” National Bureau of Economic Research, Inc NBER Working Papers 9069.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans.** 1999. “Monetary policy shocks: What have we learned and to what end?” In *Handbook of Macroeconomics*. Vol. 1 of *Handbook of Macroeconomics*, , ed. J. B. Taylor and M. Woodford, Chapter 2, 65–148. Elsevier.
- Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo.** 2008. “Reference Prices and Nominal Rigidities.” National Bureau of Economic Research, Inc NBER Working Papers 13829.
- Kehoe, Patrick J., and Virgiliu Midrigan.** 2007. “Sales and the real effects of monetary policy.”
- Lucas, Robert Jr.** 1972. “Expectations and the neutrality of money.” *Journal of Economic Theory*, 4(2): 103–124.
- Mackowiak, Bartosz Adam, and Mirko Wiederholt.** 2007. “Optimal Sticky Prices under Rational Inattention.” C.E.P.R. Discussion Papers CEPR Discussion Papers 6243.
- Mankiw, N. Gregory, and Ricardo Reis.** 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, Filip.** 2010. “Rigid Pricing and Rationally Inattentive Consumer.” *CERGE-EI Working Papers Series*, 20(2): 1–40.
- Matějka, Filip, and Christopher A. Sims.** 2010. “Discrete Actions in Information-Constrained Tracking Problems.” Princeton University Technical report.
- Reis, Ricardo.** 2006. “Inattentive Producers.” *Review of Economic Studies*, 73(3): 793–821.

- Sims, Christopher A.** 1998. “Stickiness.” *Carnegie-Rochester Conference Series on Public Policy*, 49(1): 317–356.
- Sims, Christopher A.** 2003. “Implications of rational inattention.” *Journal of Monetary Economics*, 50(3): 665–690.
- Sims, Christopher A.** 2006. “Rational Inattention: Beyond the Linear-Quadratic Case.” *American Economic Review*, 96(2): 158–163.
- Vanderbei, Robert J.** 1999. “LOQO:an interior point code for quadratic programming.” *Optimization Methods and Software*, 11.
- Woodford, Michael.** 2008. “Information-Constrained State-Dependent Pricing.”