

# Rationally Inattentive Seller: Sales and Discrete Pricing\*

Filip Matějka

PACM, Princeton University and CERGE-EI<sup>†</sup>

WP 408

## Abstract

This paper presents a model of a rationally inattentive seller responding to shocks to unit input cost. The model generates price series imultaneously 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. Discrete pricing arises even if the unit input cost varies in a continuous range. 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 a sample tends to be the most quoted price. Discrete and asymmetric pricing is a seller's optimal response to his limited information capacity. Moreover, the model provides rationale for faster responses to aggregate shocks in industries with more volatile idiosyncratic shocks as well as for a steeper Phillips curve in less stable aggregate conditions.

## Abstrakt

Tento článek prezentuje model racionálně nepozorného prodejce, který reaguje na změny vstupních cen. Představený model generuje cenové řady mající všechny z následujících vlastností, jenž lze najít i v datech. 1) Ceny se mění často. 2) Reakce cen na změny makroekonomických veličin jsou zpožděné. 3) Ceny nabývají pouze několika různých hodnot, nemění se spojitě a také se vrací na stejné hodnoty. Nespojité, diskrétní, cenové strategie vznikají i v případě, kdy vstupní ceny nabývají všech hodnot ve spojitém intervalu. Výsledky modelu také souhlasí se skutečností, že slevy obvykle netrvají dlouho a nejčastější cena v možné všech realizovaných cen je často ta nejvyšší. Diskrétní a asymetrické ceny jsou prodejcovou optimální strategií, pokud má omezenou kapacitu zpracovávat informace. Model také poskytuje teoretickou podporu pro rychlejší reakce cen na makro-šoky prodejci v proměnlivějších odvětvích a také pro strmější Phillipsovu křivku v méně stabilních prostředích.

**Keywords:** rational inattention, nominal rigidity, sales.

**JEL classification:** D8, E3.

---

First draft: April 12, 2008, this version: January 20, 2010

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

<sup>†</sup> CERGE-EI is a joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of Academy of Sciences of the Czech Republic.  
Address: CERGE-EI, P.O. Box 882, Politických vězňů 7, Prague 1, 111 21, Czech Republic

# 1 Introduction

This paper studies price movements. It proposes a mechanism leading to frequent changes of price together with 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 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. 1) Prices change frequently. 2) Responses of prices to aggregate variables are delayed. 3) Prices move back and forth between a few rigid values. Moreover, the model generates short-lasting sales and provides rationale for faster responses to aggregate shocks in industries with more volatile idiosyncratic shocks as well as for a steeper Philips curve in less stable aggregate conditions. 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. This is possibly due to the generality of rational inattention.

Rational inattention is a framework I apply to model the processing of relevant information. It was introduced by Sims in [19] and [20]. Rational inattention is based on a rigorous application of information theory. A few other authors, Reis [18], Mackowiak and Wiederholt [11], and Woodford [22],

have also recently studied pricing of inattentive agents. However, as far as I am aware, this is the first model of pricing fully in the spirit of rational inattention.

It turns out, 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.

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., Christiano, Eichenbaum and Evans [5]). On the other hand, evidence on the micro level suggests that individual prices do not stay fixed for long periods of time. 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.

Bils and Klenow [3] found that individual prices change quite frequently,

about once every 4 months. Their data set included product categories covering approximately 70% of consumer spending. Several papers, such as Kehoe and Midrigan [8], used price series found in Dominick's Database<sup>3</sup>. This database consists of price movements for about 3 500 products in the 100-store retail chain in the Chicago area during 1989-1994. A thorough empirical study of prices is also provided by Nakamura and Steinsson [16] using CPI microdata.

Eichenbaum, Jaimovich and Rebelo [6] 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. Kehoe and Midrigan [8] report that 33% of prices change every week. A copy of six time series presented in Eichenbaum, Jaimovich and Rebelo [6] is shown in Figure 1.

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

Nakamura and Steinsson [15] provided an explanation for coexistence of a rigid regular price and frequent deviations from it. They study a model of customer markets. Consumers form habits in single goods and have imperfect information about firm's desired prices. The habit formation implies that firms face a time-inconsistency problem. Even at low prices, consumers might

---

<sup>3</sup><http://research.chicagogsb.edu/marketing/databases/dominicks/index.aspx>

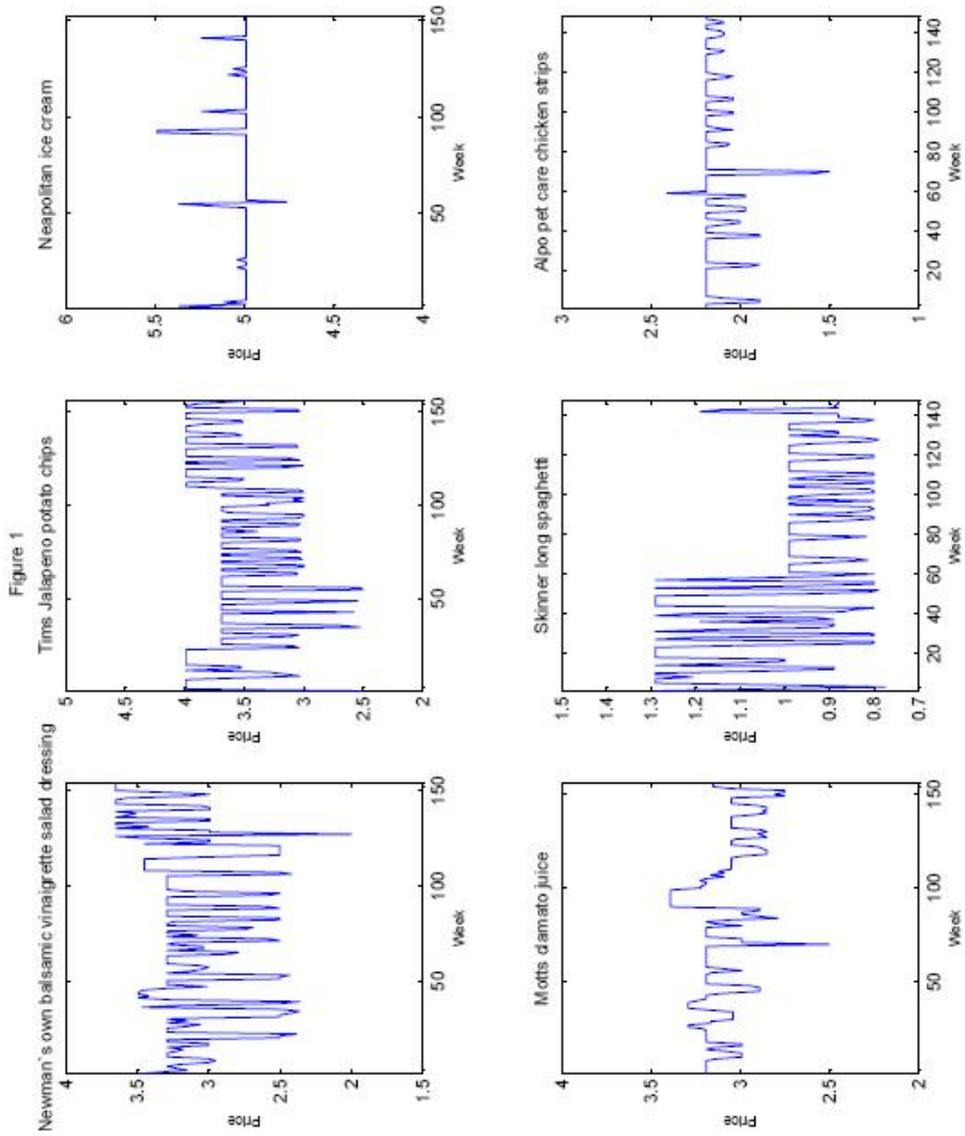


Figure 1: Time series of prices, copied from [6]

consume too little due to their fear of high prices in the future, unless the firm is able to make credible commitments to lasting low prices. It transpires that the most desirable sustainable path for the firm takes the form of a price cap above which it never sets its price. The firm's price can then spend a significant amount of time at the price cap and is fully flexible downward, when marginal cost or demand is relatively low.

The mechanism of implicit contracts is appealing. However, it does not explain why prices below the regular price seem to be rigid too and also why some deviations from the regular price are actually upwards.

Kehoe and Midrigan [8] also 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 and Rebelo [6] 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. Prices fluctuate around the reference price. A 1/3 of all price movements are movements to the reference price and 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 and

Rebelo [6], or the regular price in Nakamura and Steinsson [15] and Kehoe and Midrigan [8]. 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.

The rest of the paper is organized as follows. Section 2 describes the basic model and its solutions. An extended model is presented in Section 3. Section 4 confronts the results of the model with the evidence and Section 5 summarizes implications and findings of the paper.

## 2 The Model

A monopolistic seller processes information about a stochastic unit input cost  $\mu$  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  $\Pi$ .

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

Let  $p_{opt}(\mu)$  denote the optimal price the seller would charge if he observed unit input cost  $\mu$  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  $\mu$ . 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  $\kappa$ , then the entropy of knowledge can not on average decrease by more than  $\kappa$ . 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  $\mu$  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 by Sims in [21], 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  $\mu$ , 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,  $\mu$  - this is specified by a conditional distri-

bution,  $f(s|\mu)$ . 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 to<sup>4</sup>, 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  $\mu$ . If the seller had an unlimited information capacity, the conditional distribution  $f(p|\mu)$  would be degenerate, it would be a delta function at  $p = p_{opt}(\mu)$ .

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  $\mu$  and  $p$ , defined as

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

---

<sup>4</sup>To economize on information capacity, it is never optimal to choose two signals leading to the same price

(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  $\mu$ , he would charge an optimal price  $p = p_{opt}(\mu)$ . However, if his information capacity does not allow him to learn the cost exactly<sup>5</sup>, we should expect that pricing will not be optimal. For a given  $\mu$ , 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.

## 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  $\theta$  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 LOQO. The solver is based on interior point methods,

---

<sup>5</sup>An 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.

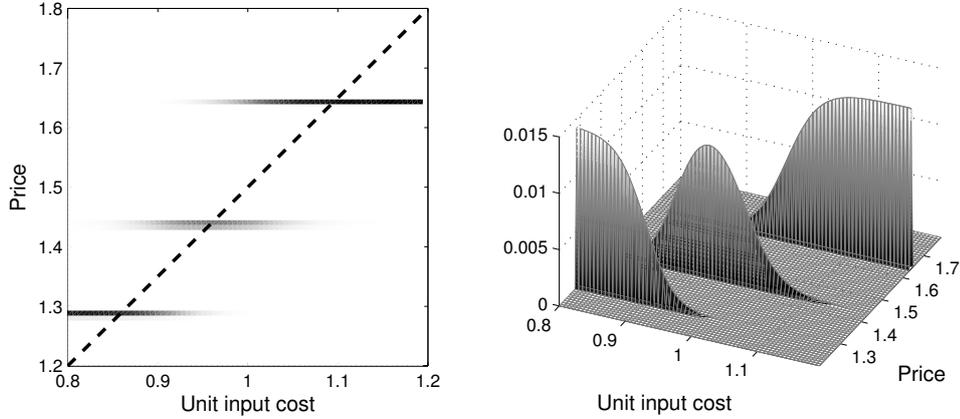


Figure 2: Joint distribution:  $\theta = 3$ ,  $\kappa = 1$  bit, uniform cost

which are very efficient for large scale convex optimization problems, such as this one. The two dimensional domain of  $\mu$  and  $p$  was discretized into  $70 \times 70$  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 in [8]. 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.

*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.

The distribution of percentage deviation of markups from the optimal markup  $1/(\theta - 1)$  is shown in Figure 5 for  $\kappa = 1$  and for  $\kappa = 2$ . Standard

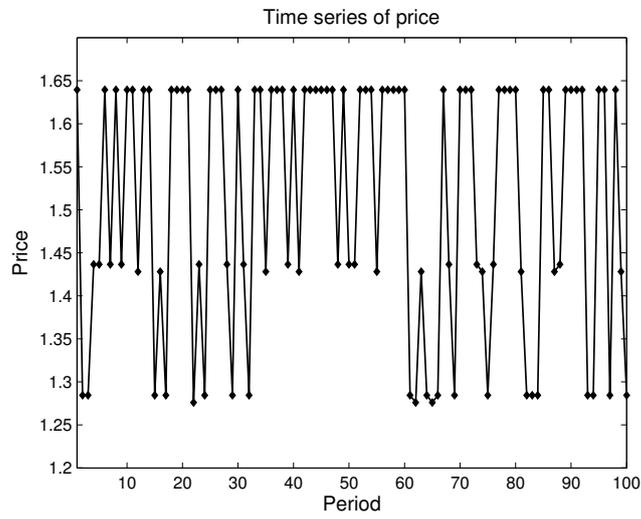


Figure 3: Simulated time series of price,  $\theta = 3$ ,  $\kappa = 1$  bit

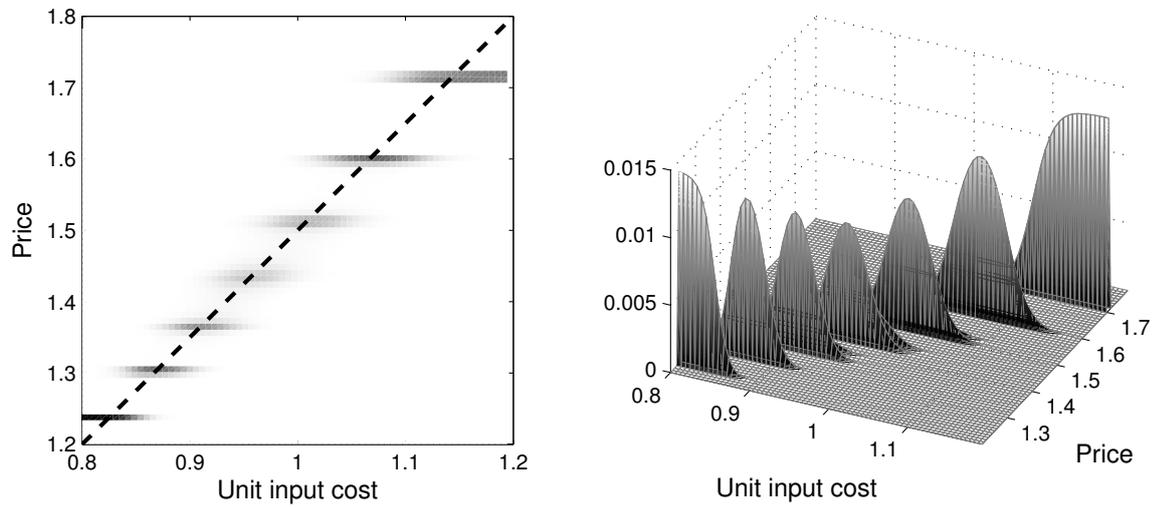


Figure 4: Joint distribution:  $\theta = 3$ ,  $\kappa = 2$  bits

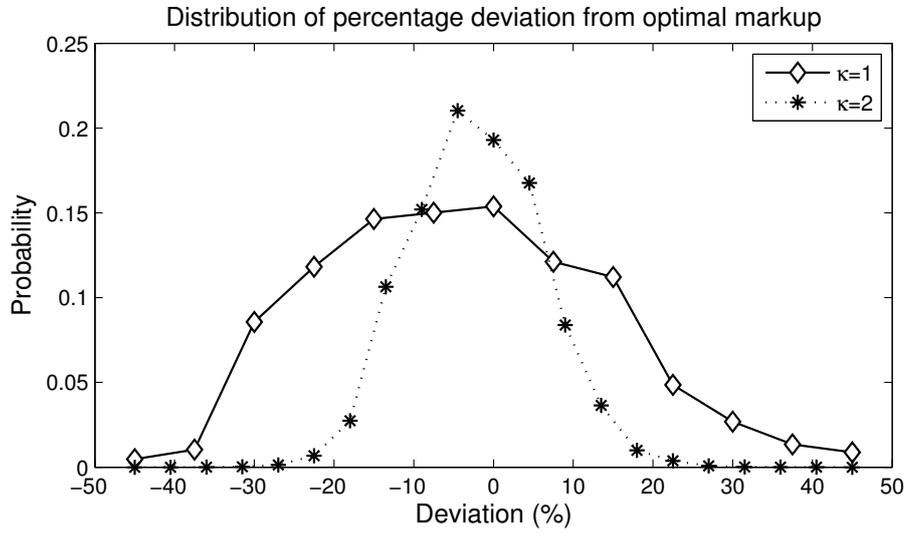


Figure 5:  $\theta = 3$

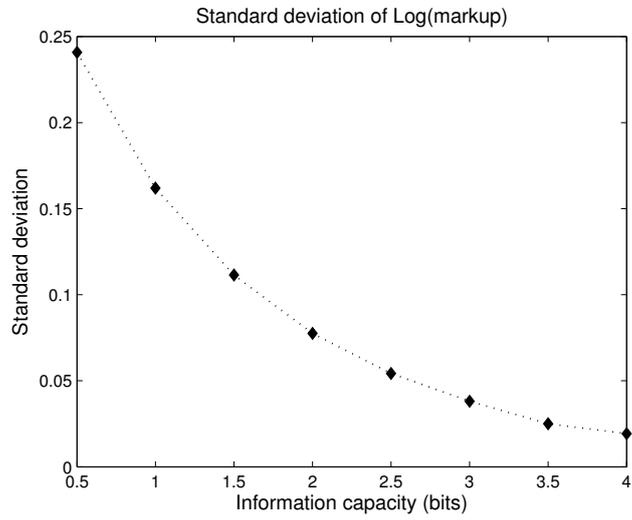


Figure 6:  $\theta = 3$

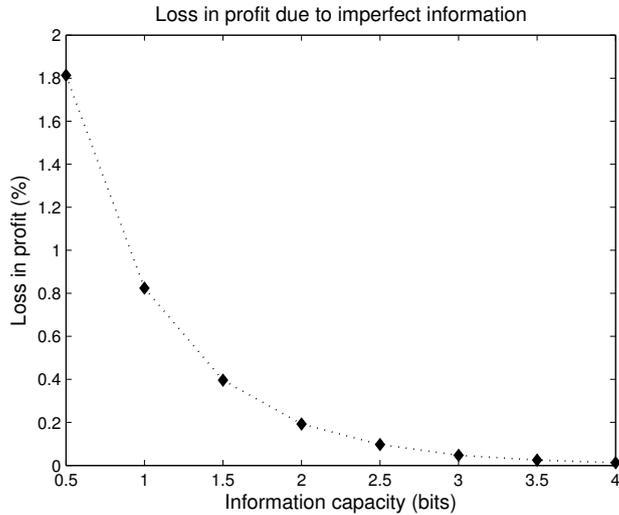


Figure 7: Decreasing marginal value of information,  $\theta = 3$

deviations of log-markup are 0.16 and 0.08 respectively. Figure 6 presents the dependance of the standard deviation of log-markup on the information capacity. 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.

Percentage loss in expected profit due to informational imperfections is shown in Figure 7. The loss is a difference between expected profit of the optimal pricing strategy (8) and expected profit of the informationally constrained strategy. The marginal value of information is 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 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.

Sims [21] expressed a first order condition for the conditional distribution. If  $f(p) > 0$ , then

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

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  $\mu$  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 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.

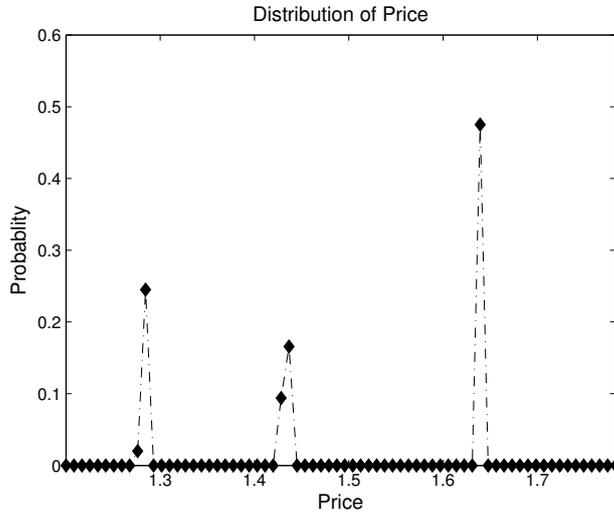


Figure 8: Overall distribution of price,  $\theta = 9$ ,  $\kappa = 1$  bit

*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 8. It is a marginal distribution of  $f(\mu, p)$ , which is obtained after integrating the joint distribution over  $\mu$ . 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 Sims and Matějka[14]. For example, we show analytically that for quadratic utility functions, a prior distribution with a bounded support always leads to complete discreteness in responses. The intuition is as follows. Discreteness occurs as an optimal response to a shape of prior knowledge. The agent would like to acquire

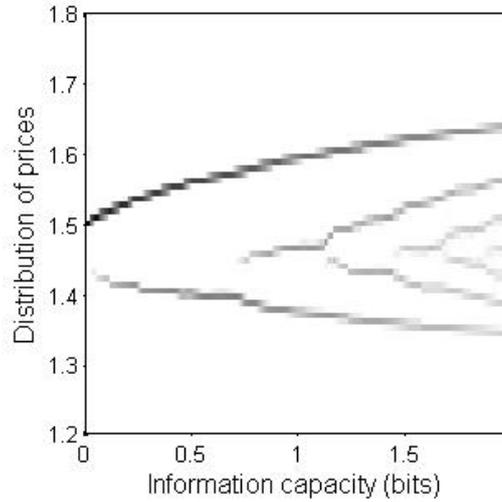


Figure 9: Bifurcation diagram of prices,  $\theta = 3$

signals with a specific form of noise, given by the first order condition <sup>6</sup>. 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. This effect then propagates towards the center of the support of prior knowledge and generates discreteness globally.

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

---

<sup>6</sup>The form of noise depends on the shape of utility function.

in Figure 9 shows price distributions as a function of information capacity. When the capacity increases, new price points emerge branching off the existing ones. There is 1 price only when  $\kappa = 0$ , the second one emerges immediately as  $\kappa > 0$ , the third at about  $\kappa = 0.7$ , etc. Low probability densities are not clearly visible, but the third point does in fact branch off the lower price of the first two.

5. *Pricing is asymmetric.* Although the distribution of unit input costs is uniform, the resulting distribution of prices shown in Figure 8 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 (28) is higher. The factor indicates magnitudes of losses due to imperfect knowledge. Change in profit due to misjudging unit input cost  $\mu$  by a small amount  $\epsilon$  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  $\theta$  and the same amount of error  $\epsilon$ , losses are larger for lower  $\mu$ . The seller thus chooses to pay more attention to regions of low  $\mu$ . Figures 10 and 11 show the probability of the highest

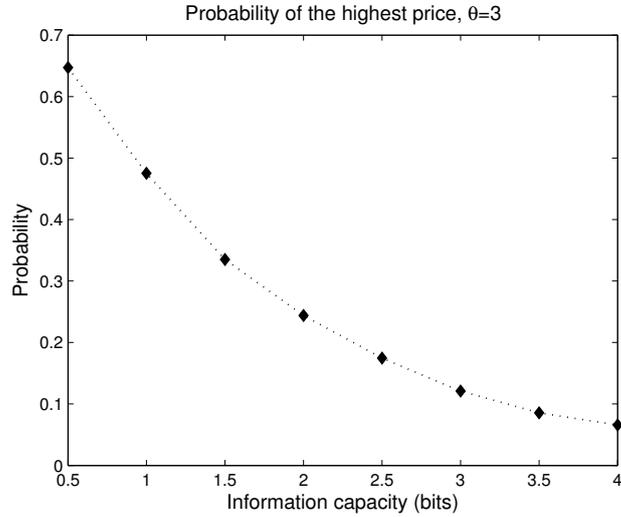


Figure 10: Asymmetry of pricing

price as a function of  $\kappa$  and  $\theta$ . 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. Figure 12 shows the distribution of prices arising as the solution to the problem with the symmetric triangular distribution of unit input costs. Although the highest price is not the most quoted in this example, the asymmetry towards higher prices is still apparent.

## 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

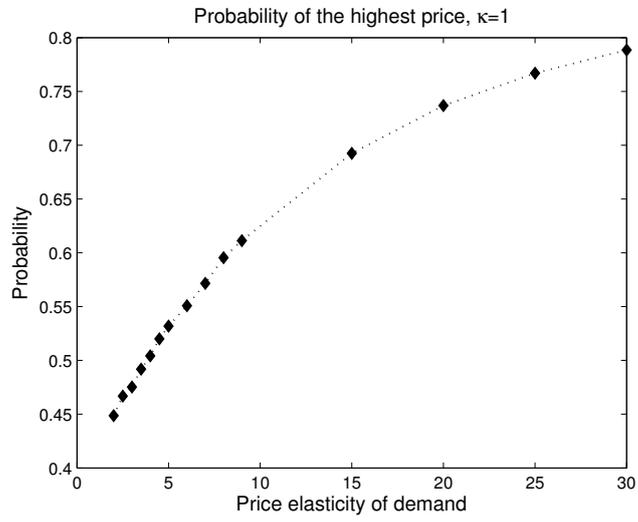


Figure 11: Asymmetry of pricing

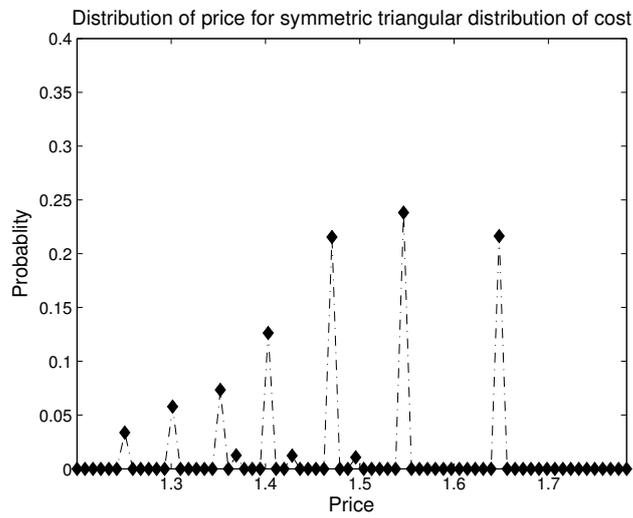


Figure 12: Triangular distribution of cost,  $\theta = 3$ ,  $\kappa = 1$  bit

that agents also choose how much information to process. Let  $RI_\kappa$  stand for the original model of a rationally inattentive seller with a fixed information and  $RI_\lambda$  denote a model with fixed unit cost of information. In  $RI_\lambda$ , 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  $\Pi$  given by (1), they maximize expectation of

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

where  $\lambda$  is the cost of processing 1 bit of information about  $\mu$ , and  $\kappa$  is the amount of information the price-setter chooses to process. The  $RI_\lambda$  problem takes the same form as  $RI_\kappa$ , (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,  $RI_\kappa$  and  $RI_\lambda$  are equivalent. Solutions can be parameterized by either of the two quantities,  $\kappa$  or  $\lambda$ . If  $\kappa$  is fixed,  $\lambda$  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  $\kappa = 2$ .

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

### 2.2.1 Implications Across Industries

Let us allow sellers to select  $\kappa$  - I shall call this model the  $RI_\lambda$  model. It turns out that sellers in highly competitive industries decide to process more information about input cost, because their profits are more sensitive to sub-optimal pricing. We will be comparing sellers of the same size that face demands with different elasticities. Elasticity of demand is a measure of the degree of competition in an industry. Magnitude of demand (size of the firm) determines shadow price of information and thus influences choices of how much information to process. Larger sellers decide to process more information. Normalizing the magnitude thus allows for unbiased comparisons of information choices across markets.

We could for instance normalize average demand or demand at one particular price. For the sake of simplicity, I fix demand equal to 1 when a seller charges optimal price at the mean input cost <sup>7</sup>.

Let the unit input cost  $\mu$  vary about its mean  $\mu_0$ , the normalized demand is:

$$d_1(p) = \left( \frac{p}{p_{opt}(\mu_0)} \right)^{-\theta}, \quad (13)$$

where  $p_{opt}(\mu_0)$  is the perfect information response to the average input cost,  $\mu_0$ . The seller maximizes

$$\tilde{\Pi}(\mu, p, \kappa) = \left( \frac{p}{p_{opt}(\mu_0)} \right)^{-\theta} (p - \mu) - \lambda\kappa. \quad (14)$$

Figure 13 shows computational results of the dependence of a selected  $\kappa$  on demand elasticity,  $\theta$ , for unit input cost uniformly distributed in (0.8, 1.2) and  $\lambda = 0.5 \cdot 10^{-3}$ . Optimal information capacity is an increasing function of  $\theta$ .

---

<sup>7</sup>Results of this section, however, do not rest on a particular choice of normalization.

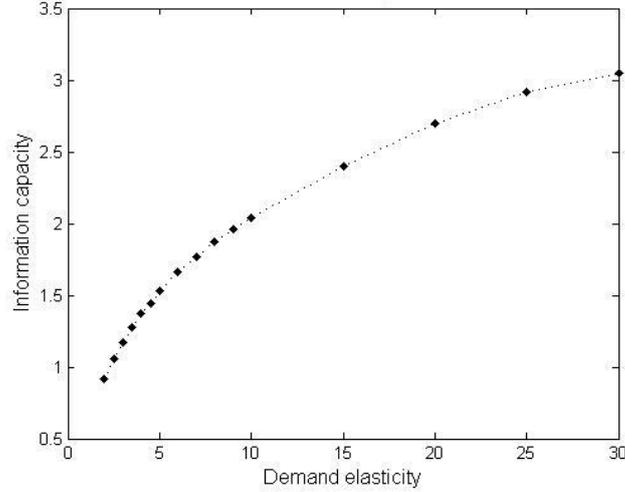


Figure 13: Comparisons across industries

To justify this result analytically, we can apply the approach developed in Appendix A. The corresponding approximative loss factor at  $\mu_0$  is

$$L(\mu_0, \theta) = \frac{\theta}{2\mu_0}. \quad (15)$$

It is a decreasing function of  $\theta$ . The higher elasticity of demand, e.g. degree of competition, the bigger the loss from imperfect information about the input cost. Given the same levels of input cost and the same firms' sizes, agents process more information in more competitive industries. Moreover, as it was discussed earlier, more information leads to more flexible prices.

Mutual information between two random variables is a measure of how much about one variable can be inferred from learning about the other. Even for an outside observer, the seller's higher information capacity implies that prices carry more information about the input cost.

This finding relates to Hayek's famous defense of free markets, [7], specifically on the grounds of markets' ability to convey information. Rational inattention implies that the more competitive a market is the more informa-

tion can be extracted by observing its prices.

## 2.3 Volatile Demand

This section presents another alternative model, a seller does not respond to shocks to unit input cost, but to volatile demand. It tries to address some concerns about the structure of the previous model (2)-(6). In that model, a rationally inattentive price setter responds every period to a realized unit input cost. However, it is not quite clear how realistic this setup is. The specific structure of contracts between retailers and their suppliers is not generally known. It is quite possible that retailers in fact know what input costs will be realized in every week in the next quarter, or maybe they know when they get a discount from their suppliers, but do not know the exact amount, etc. This issue is studied by Nakamura in [17]. On the other hand, it is known that retailers monitor prices of their competitors.

I will argue that most of the interesting properties of the studied solutions, such as discreteness and asymmetry of pricing, are inherent to the information constraint. Let us briefly visit another model. The cost schedule is fixed, while demand fluctuates. A specific realized level of unit input cost depends on what quantity is delivered. The quantity delivered is assumed to equal the quantity  $c$  that consumers decide to buy, stock of the product is automatically replenished. Let the total input cost be

$$TC(c) = \mu_0 c + \mu_1 c^2, \quad (16)$$

where  $\mu_0$  is the fixed component of unit input cost and  $\mu_1$  controls the increase of marginal input cost with quantity sold. I will assume that a demand factor  $D_0$  is i.i.d. and that demand  $d(D_0, p)$  is given by

$$d(D_0, p) = D_0 p^{-\theta}. \quad (17)$$

In every period, the price setter needs to process information about the level of demand  $D_0$ , then he decides what price to charge, while quantity sold and input cost are then automatically executed. This model takes a very similar form as (2)-(6), only  $D_0$  takes the place of an unknown stochastic variable and the profit function  $\Pi$  is

$$\Pi(D_0, p) = D_0 p^{-\theta+1} - \mu_0 p^{-\theta} - \mu_1 p^{-2\theta}. \quad (18)$$

I selected  $\mu_0 = 1$ ,  $\mu_1 = 0.4$  and the demand factor  $D_0$  to be uniformly distributed in (5, 20). The numerical solution for  $\theta = 3$  and  $\kappa = 1$  is shown in Figure 14. The dotted line represents optimal price for each level of the demand factor. The solution features the same properties as the solution of the previous model. Pricing is discrete and asymmetric. Only three price values are realized and the probability of the top price is 0.47. The distribution of markups over a marginal cost and the probability of changing price as a function of deviation from an average markup are shown in Figure 15. These results are also similar to the results of the former model in Figures 5 and 16.

The similarities are not a coincidence. The price setter processes information about a stochastic variable<sup>8</sup> and targets some optimal price. 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.

---

<sup>8</sup>This stochastic variable can be input cost, demand or even price of a competitor.

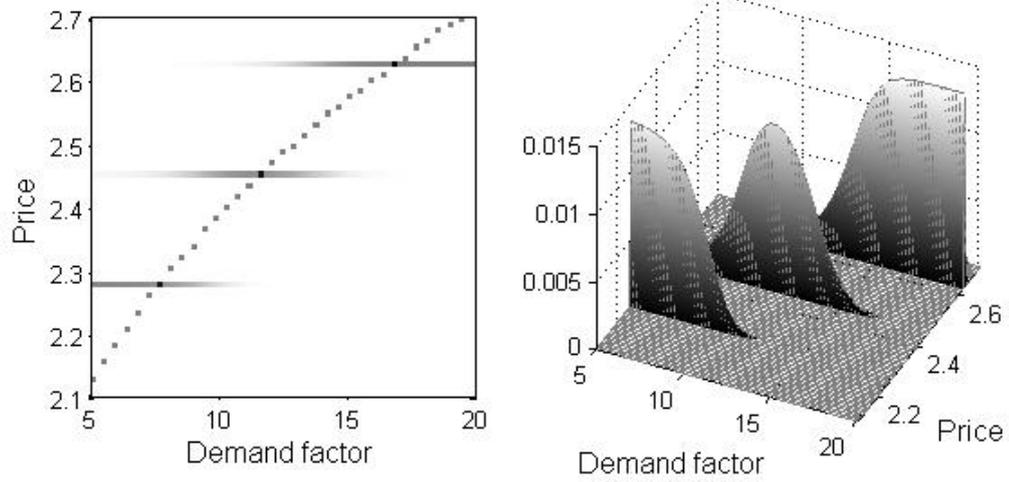


Figure 14: Volatile demand

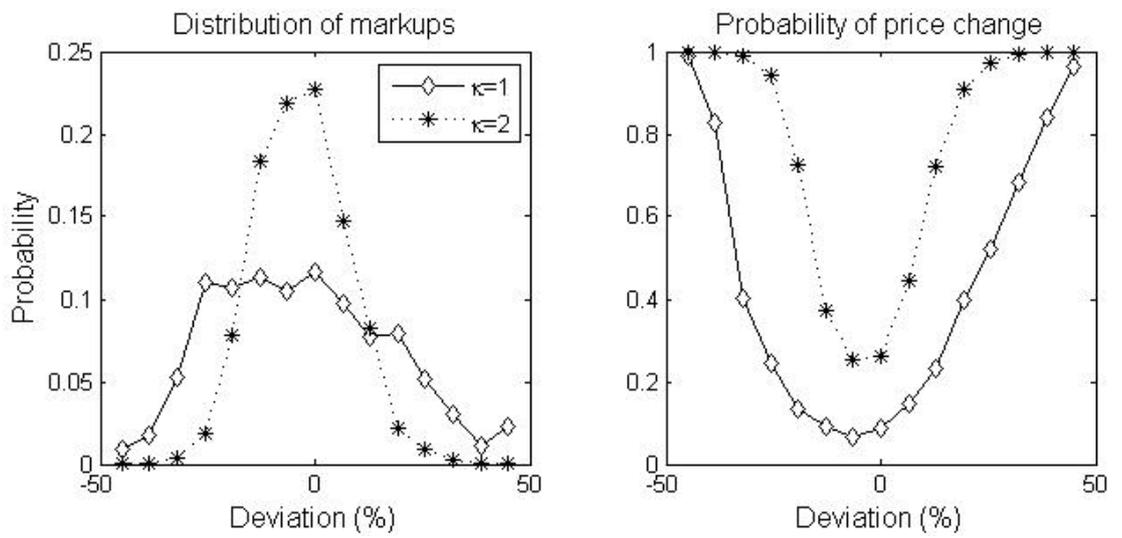


Figure 15: Volatile demand

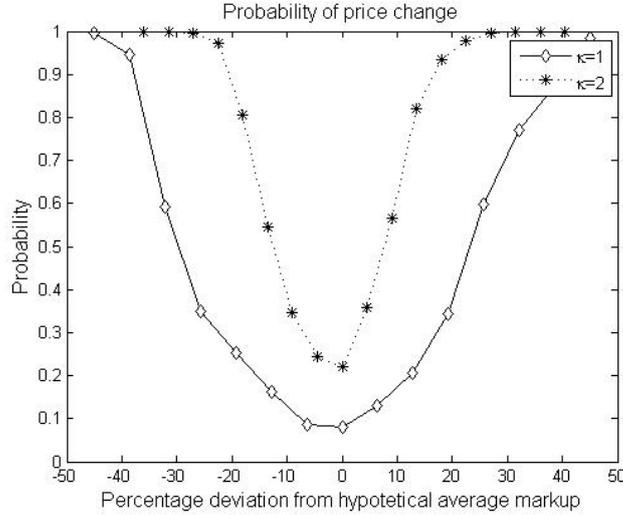


Figure 16:  $\theta = 3$

### 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  $\mu$  and the second one be a serially correlated nominal variable  $A$ .  $\mu$  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.

$$\Pi(A, \mu, p) = p^{-\theta}(p - A\mu). \quad (19)$$

$A$  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  $\mu$  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  $x = Prob(A_L)$ , where  $A_L$  stands for the lower value of  $A$ .

The recursive formulation of the seller's dynamic problem, a  $RI_\kappa$  version, is as follows.

$$V(x) = \max_f \int \left[ \Pi(A, \mu, p) + \beta V(x') \right] f(A, \mu, p) dA d\mu dp, \quad (20)$$

subject to

$$x' = f(A_L|p)(1-t) + (1-f(A_L|p))t \quad (21)$$

$$\int f(A, \mu, p) dp = g(A, \mu) = g_1(A)g_2(\mu) \quad (22)$$

$$g_1(A_L) = x \quad (23)$$

$$I(A, \mu; P) \leq \kappa. \quad (24)$$

$f(A, \mu, p)$  is a joint distribution summarizing the seller's choice of signals and responses in the given period, (21) 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  $t$ . (22) is the constraint on a prior,  $\mu$  and  $A$  are assumed to be independent.  $g_2(\mu)$  is

fixed,  $g_1(A_L) = x$ , (23), which implies  $g_1(A_H) = 1 - x$ . (24) is the traditional constraint on mutual information between the source variables  $\mu$  and  $A$ , and a response variable  $p$ .

For computations, I used  $\kappa = 1$ ,  $\theta = 3$ ,  $\mu$  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 17 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  $\mu$  is discrete. However, when knowledge about  $A$  changes, the distribution of prices changes too. For the used values of parameters ( $\kappa$ ,  $t$ ,  $A_H$ , etc.), knowledge adjustment is rather abrupt. The second series in Figure 17 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.

The bottom two series in Figure 17 are prices and knowledge averaged

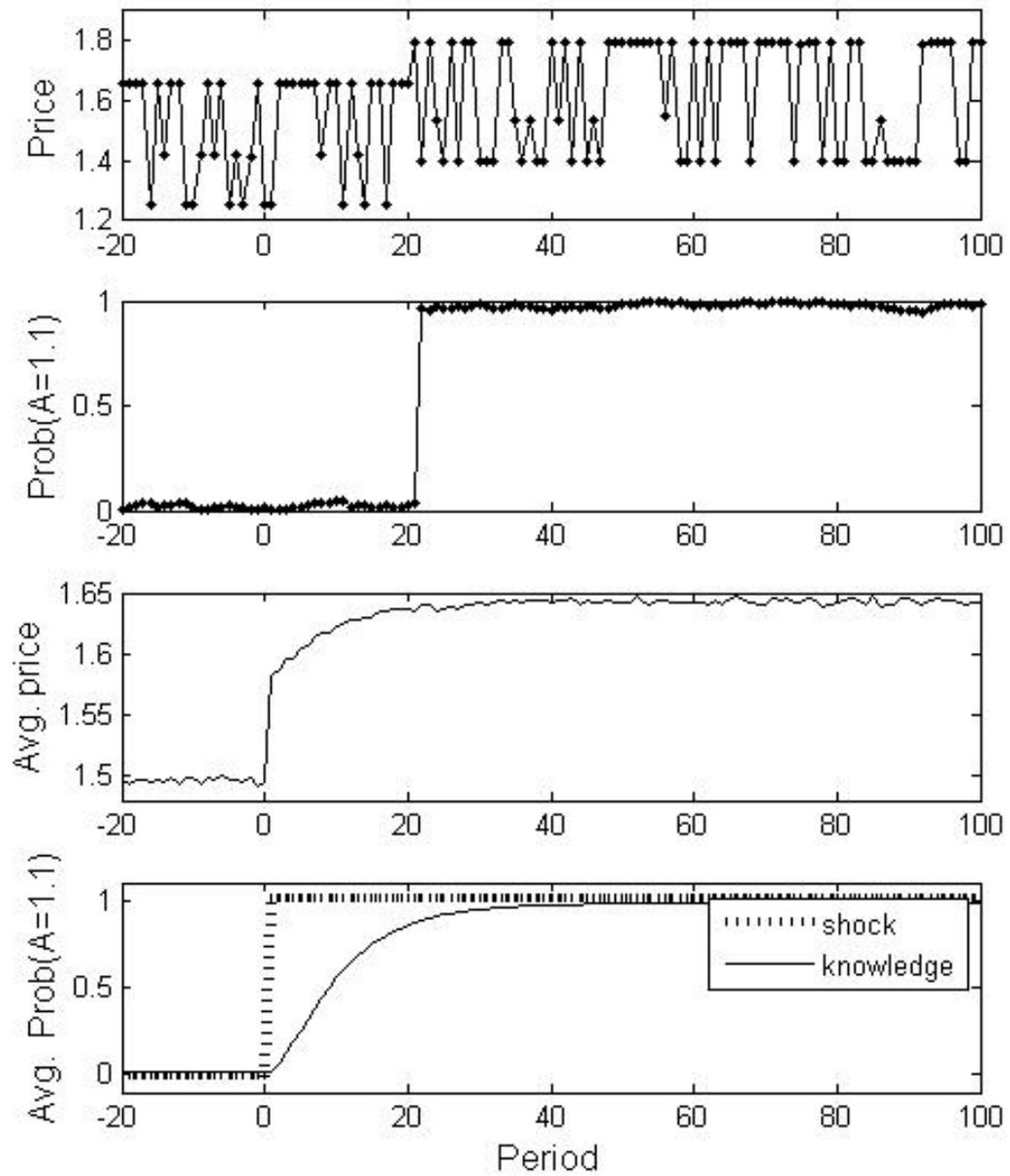


Figure 17: Two stochastic variables, sudden learning,  $t = 0.002$ ,  $\kappa = 1$

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  $\mu$  separately. Due to different dynamical properties of  $A$  and  $\mu$ , 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  $\mu$ , 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 17 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  $\mu$ . 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.

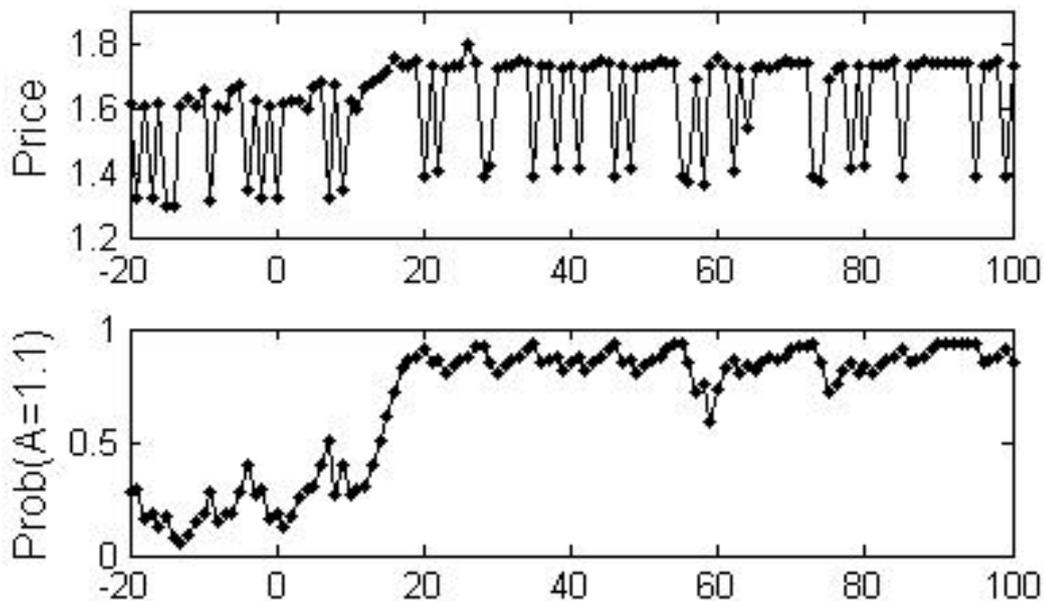


Figure 18: Two stochastic variables, gradual learning

For different parameters, however, the knowledge dynamics can be quite different. Figure 18 shows solutions for  $t = 0.02$  and  $\kappa = 0.5$ <sup>9</sup>. Knowledge starts adjusting almost immediately after the shock, but the rate of adjustment is much slower.

Some rigidity of price values is present again. For example, the top price seems to be fairly stable between period -20 and 0, and again after period 20. A shock to  $A$  occurs in period 1. In the following periods, the agent starts to realize that the value of  $A$  might have switched to the higher state and gradually learns about it. Since the prior knowledge changes more rapidly during these periods, even the price values are much more flexible.

When  $t = 0.02$ , the probability of at least one shock realized in a year increases to 65%. When the seller's knowledge starts shifting, he devotes less

---

<sup>9</sup> $\kappa = 0.5$  was chosen to keep the dynamics of average prices about the same as before.

of his information capacity specifically to  $A$  than if  $t = 0.002$  - with  $A$  more volatile, knowledge about it has relatively less value.

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  $\mu$  and  $A$ . The seller pays attention to sudden changes of  $A\mu$ , since  $\mu$  is very volatile. Such changes are thus at first attributed to shocks to  $\mu$  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 19 presents solutions to a dynamic problem with  $\mu$  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 of Mankiw and Reis, [12]. They postulate that agents' information updating is staggered; and when agents update they acquire per-

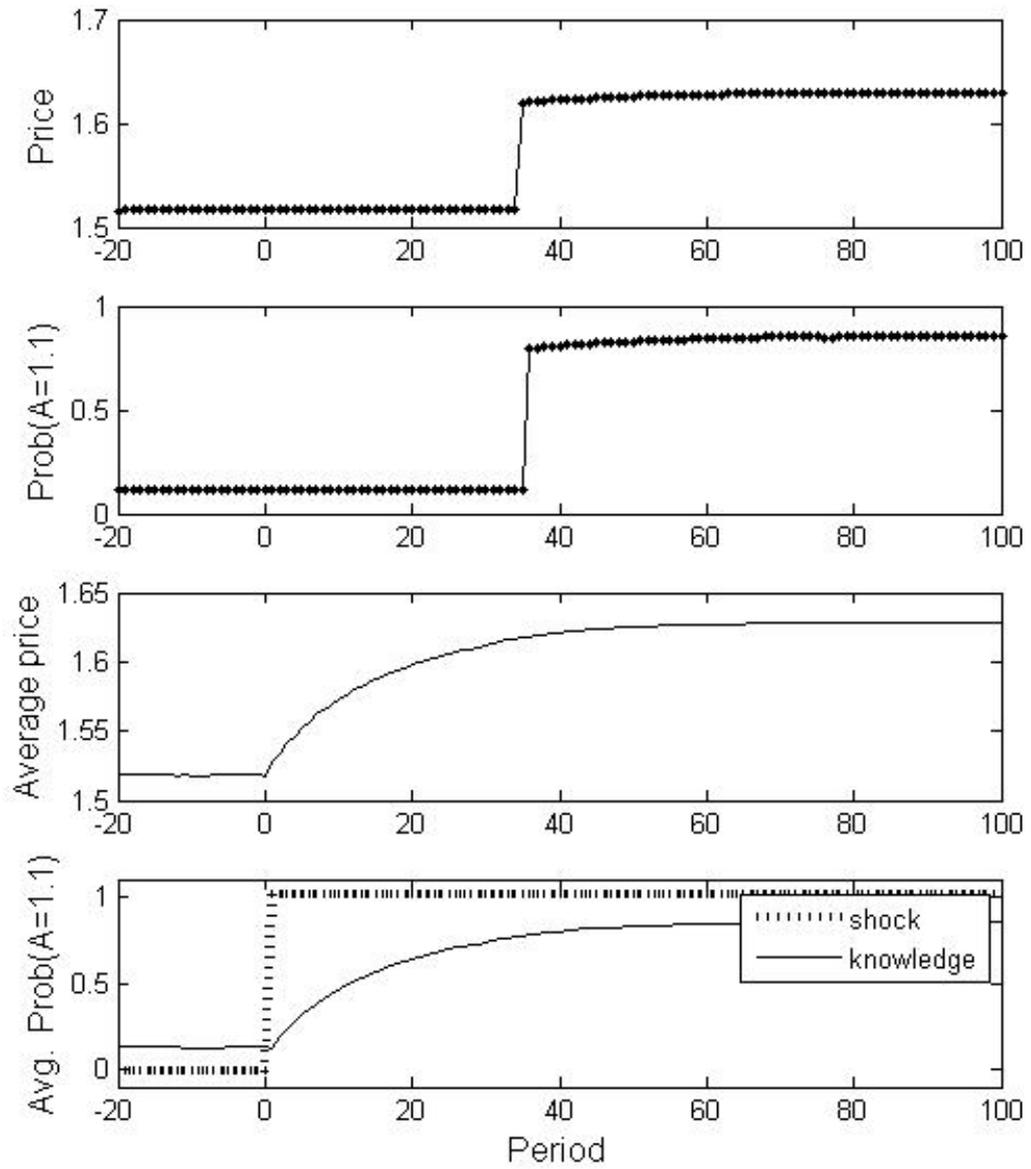


Figure 19: Serially correlated variable only

fect information. In each period, a firm updates with a certain probability  $\nu$ , i.e. only a fraction  $\nu$  of firms update information in that period. Figure 19 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.

Woodford's model, [22], 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 19 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 [9]. 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  $\xi$  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 17. At first, price would respond

rapidly, but not completely since the agent would not know whether the shock came from  $\xi$  or  $A\mu$ . Since  $A$  is very persistent, the agent would perceive the shock coming from  $\mu$  rather than from  $A$  - knowledge about  $A$  would adjust slowly, just like in Figure 17.

As Sims noted in [20], 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  $\xi$  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  $\mu$  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 and Wiederholt [11] 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  $\mu$ , and thus respond to innovations

in  $A$  very slowly. Processing information purely about  $\mu$  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 $RI_\kappa$ and $RI_\lambda$ : Implications for Aggregate Dynamics

A  $RI_\lambda$  version of the dynamic model takes the form of:

$$V(x) = \max_{f,\kappa} \int \left[ \Pi(A, \mu, p) + \beta V(x') \right] f(A, \mu, p) dA d\mu dp - \lambda \kappa, \quad (25)$$

subject to (21)-(24).

Stochastic properties of the variables of interest ( $\mu$  and  $A$ ) influence the agent's choice of what pieces of information and potentially how much information to process. 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,  $\mu$  was uniformly distributed in  $(0.8, 1.2)$ . Table 1 summarizes numerical results for  $RI_\kappa$  with  $\kappa = 1$ ,  $\theta = 3$  and  $t = 0.002$  for three different widths of the distribution of  $\mu$ : for  $\mu$  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<sub>st</sub> 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.

The more volatile is the seller's idiosyncratic part of the input cost the slower he responds to aggregate shocks. When  $\mu$  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

$\mu$	1 <sup>st</sup> 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,  $\kappa = 1$

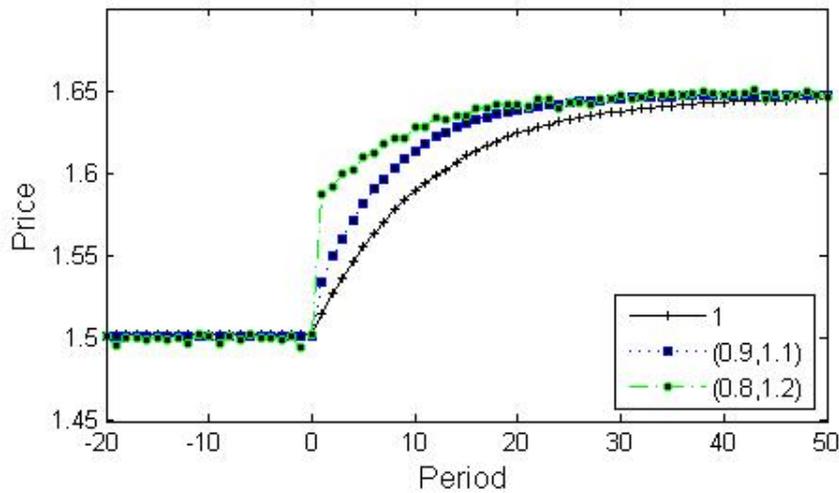


Figure 20: Average price response, 3 distributions of  $\mu$ ,  $\lambda = 0.003$

under perfect information - this quantity was evaluated with both  $\mu$  and  $A$  simulated according to their stochastic properties. As expected for  $RI_\kappa$ , losses are higher in more volatile environments.

Results of the similar experiments for  $RI_\lambda$ ,  $\lambda = 0.003$ , are shown in Table 2 and in Figure 20. Unlike  $RI_\kappa$ ,  $RI_\lambda$  generates faster responses to aggregate shocks when  $\mu$  is more volatile. The table also presents average  $\kappa$  that was selected by the seller during the simulations. Agents in  $RI_\lambda$  choose to process more information when volatility 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

$\mu$	1 <sub>st</sub> per. adj.	90% adjustment	profit loss	mean $\kappa$
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,  $\lambda = 0.003$

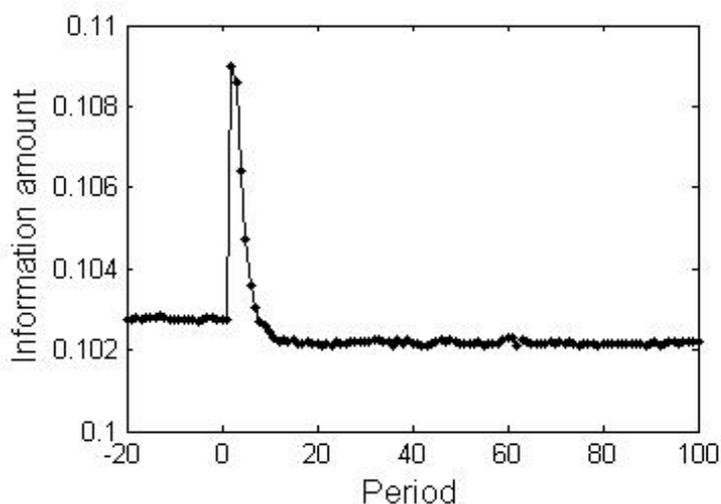


Figure 21: Average processed information,  $\mu \in (0.9, 1.1)$ ,  $\lambda = 0.003$ ,  $t = 0.02$

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  $RI_\lambda$  linear-quadratic problem in Appendix B: varied volatility of a prior influences the choice of  $\kappa$ , but not the resulting imperfections in tracking. However, the loss drops dramatically when  $\mu$  is fixed at 1. Input cost becomes a binary variable, therefore, the intuition derived from the linear-quadratic approximation does not apply very well.

In  $RI_\lambda$ , the amount of processed information is not constant. It varies

according to the expected value of information given a prior. Figure 21 presents average information amount as a function of time. Immediately after the shock occurs, a 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 [11] would generate slower and less precise responses to  $A$  if volatility of  $\mu$  were increased. Total  $\kappa$  is fixed, while marginal value of information devoted to  $\mu$  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  $t$ .

t	1 <sub>st</sub> 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,  $\kappa = 1$

More volatile  $A$  generates faster responses to its innovations in both of the models,  $RI_\kappa$  and  $RI_\lambda$ . This is due to a higher marginal value of processing new information about  $A$  if  $A$  is more likely to vary and due to the

t	1 <sub>st</sub> per. adj.	90% adjustment	profit loss	mean $\kappa$
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,  $\lambda = 0.003$

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  $RI_\lambda$ , 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 the models of Nakamura and Steinsson [15] and Kehoe and Midrigan [8], the highest price and the most quoted price do not have any special significance and they do not always coincide.

Results of the presented model rest on a selection of parameters. The parameters are: a distribution of unit input cost,  $g(\mu)$ , a price elasticity of demand,  $\theta$ , and an information capacity,  $\kappa$ . Unfortunately, the cost time series used in Eichenbaum, Jaimovich and Rebelo [6] is proprietary. For the sake of simplicity, I assume  $\mu$  to be i.i.d, uniformly distributed, each realization corresponding to a weekly cost. The width of uniform  $g(\mu)$  was set to match the standard deviation of costs reported in [6]. The authors found that the median standard deviation of log-unit input cost was 0.12.

Kehoe and Midrigan used  $\theta = 3$ . Estimates in the literature vary between 3 and 10. However, it is much trickier to pin the parameter  $\kappa$  down. We can imagine that especially in large retail stores with more than 10 000 of products, the price setters do not follow all the details in cost movements for all products, but what is the proper level of  $\kappa$  is however unclear. Time series of prices generated by the model appear similar to real time series for information capacity that is less than 3 bits.

*Reference prices:* the model generates asymmetry in pricing, with the highest price typically being the most quoted, which is in agreement with

the data. Figures 10 and 11 show the probability of the highest (and most quoted) price as a function of  $\kappa$  and  $\theta$ . Eichenbaum, Jaimovich and Rebelo [6] 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:* Eichenbaum, Jaimovich and Rebelo [6] study the distribution of markups. A copy of their graph for a median product is shown in Figure 22. The distribution has a similar shape as the distributions generated by the model, which are presented in Figure 5. The solution for  $\kappa = 2$  agrees with the evidence quite well<sup>10</sup>. Unlike flexible price models, the presented model generates 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  $\kappa$  is between 0.5 and 4. Median standard deviation reported in [6] 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  $\theta$  reaches 10.

*State dependance:* Eichenbaum, Jaimovich and Rebelo [6] 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. Their result is shown in Figure 23. The analogous dependance for the numerical solutions is presented in Figure 16. The results are qualitatively similar. Quantitative agreement, however, rests on the selected values of  $\theta$  and  $\kappa$ .

---

<sup>10</sup>Values on the y-axes in the two figures are quite different, because they denote probabilities, not probability densities. The values therefore depend on the number of bins on the x-axes.

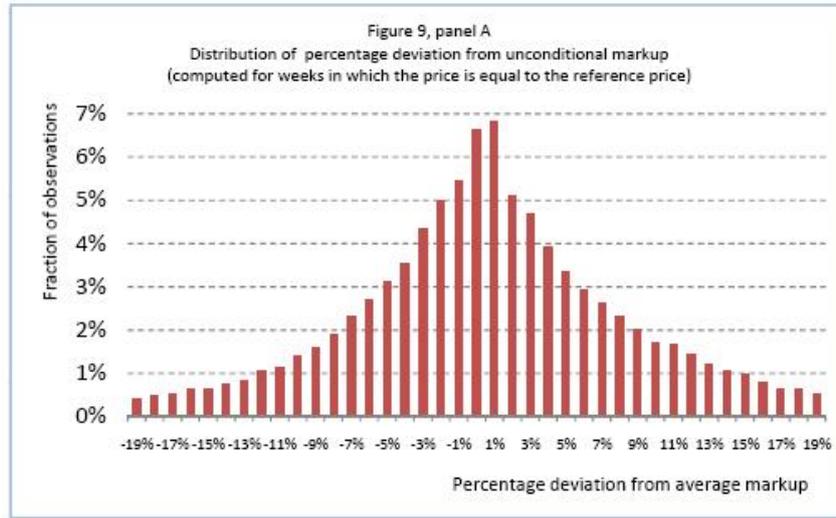


Figure 22: Copied from [6]

For  $\theta = 3$  and  $\kappa = 2$  generate a state dependence that more readily agrees with the evidence.

*Volatility of prices:* Eichenbaum, Jaimovich and Rebelo [6] find that standard deviation of logs of prices is slightly higher than deviation of logs of costs, 0.13 versus 0.12. In this model, it is the other way around, 0.10 versus 0.12. It is not surprising, since we studied effects of shocks to input costs only. If we considered other shocks too, price might have higher standard deviation than cost. Moreover, there are other drivers of volatility in prices, which are not modelled by the proposed mechanism. Such drivers are price discrimination, existence of loss leaders, etc.

*Across industries:* if information capacity can be chosen,  $RI_\lambda$  model, then sellers process more information if elasticity of demand is higher. Higher information capacity generates more flexible prices. It is well documented

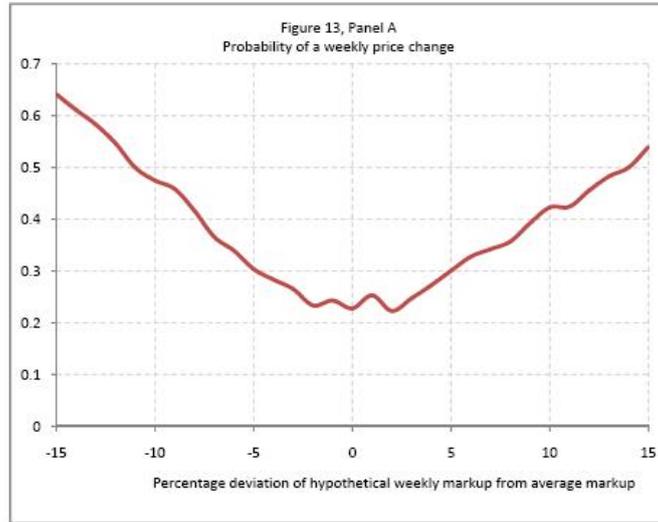


Figure 23: Copied from [6]

that prices are more flexible in more competitive industries<sup>11</sup>.

## 4.2 Serially Correlated Input Cost

*Price series on a micro level:* Eichenbaum, Jaimovich and Rebelo [6] 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 17 presents responses to an aggregate shock in an otherwise very stable aggregate environment. In this case, knowledge about the aggregate component

<sup>11</sup>For instance, see Alvarez and Hernando,[1].

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 and Rebelo [6], 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 18 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<sup>12</sup>, 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 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 Matějka [13] 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

---

<sup>12</sup>For example, Alvarez, Burriel and Hernando [2]

volatility of  $\mu$  on the dynamics of average prices. Boivin, Giannoni and Mihov [4] found evidence that prices in sectors with more volatile idiosyncratic shocks react to aggregate shocks faster. This finding agrees with the results of the  $RI_\lambda$  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  $RI_\kappa$  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,  $RI_\kappa$  and  $RI_\lambda$ , generate faster responses in more volatile environments. Lucas [10] observed that Philips curve becomes steeper in more volatile aggregate environments. The models' results agree with this finding.

Moreover, both  $RI_\kappa$  and  $RI_\lambda$  generate higher losses if the aggregate variable is more volatile, when tracking it becomes a more demanding task. The effect is weaker for  $RI_\lambda$ , losses from increased volatility are partially offset by more information processed by the seller.

## 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 more flexible prices in industries with a higher degree of competition and as well as for faster responses to aggregate shocks in industries with more volatile idiosyncratic shocks. In such industries, sellers choose to process more information.

Like signal extraction, rational inattention provides an explanation why the Philips curve becomes steeper when aggregate conditions are more volatile. However, unlike signal extraction, rational inattention generates 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 also a necessary ingredient for several implications of the model.

1. Agents process different amounts of information about different levels of input cost. Implication: discrete and asymmetric pricing.
2. Agents can process signals about a joint characteristic of interest<sup>13</sup>. Implication: faster responses to aggregate shocks in more volatile industries.
3. 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

---

<sup>13</sup>Results are thus independent of a specific selection of variables in the model.

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.

## 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  $\epsilon$  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

$$z = Z_{opt}(y) \tag{26}$$

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

$$\Pi_{y^*}(y') = \Pi_{y^*}(y^*) + \frac{d\Pi_{y^*}(y')}{dy'}\epsilon + \frac{d^2\Pi_{y^*}(y')}{dy'^2}\epsilon^2/2 + \mathcal{O}(\epsilon^3)$$

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

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:

$$\begin{aligned}
\Delta\Pi &= \Pi_{y^*}(y') - \Pi_{y^*}(y^*) \\
&= \left( \frac{dZ_{opt}(y')}{dy'} \right)^2 \frac{d^2 \Pi(y^*, z)}{dz^2} \epsilon^2 / 2 + \mathcal{O}(\epsilon^3) \Big|_{y'=y^*, z=Z_{opt}(y^*)} \quad (27)
\end{aligned}$$

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  $\epsilon$  is small, then the change in profit can be approximated by  $-L(y^*)\epsilon^2/2$ , where the approximative loss factor  $L(y^*)$  is

$$L(y^*) = - \left( \frac{dZ_{opt}(y^*)}{dy^*} \right)^2 \frac{d^2 \Pi(y^*, z)}{dz^2}. \quad (28)$$

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: $RI_\kappa$ vs. $RI_\lambda$

If a utility function is linear-quadratic and noise is Gaussian, then the rational inattention problem simplifies significantly<sup>14</sup>. Agents acquire normally

---

<sup>14</sup>See [20] and [11]

distributed posterior knowledge with a constant variance. Let  $Y$  be drawn from  $N(0, \sigma_Y^2)$ , its entropy is

$$H(Y) = \frac{1}{2} \log(2\pi e \sigma_Y^2). \quad (29)$$

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

$$\frac{1}{2} \log\left(\frac{\sigma_Y^2}{\sigma_{Y|Z}^2}\right) \leq \kappa. \quad (30)$$

Variance of posterior:

$$\sigma_{Y|Z}^2 = \sigma_Y^2 2^{-2\kappa}, \quad (31)$$

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

$$E[U] = -a\sigma_{Y|Z}^2. \quad (32)$$

If information capacity is fixed,  $RI_\kappa$ ,

$$E[U] = -a\sigma_Y^2 2^{-2\kappa}, \quad (33)$$

worse prior knowledge implies higher losses.

A Lagrange multiplier  $\lambda$  is a shadow cost of information,

$$\begin{aligned} \lambda &= \frac{\partial E[U]}{\partial \kappa} = 2 \ln(2) a \sigma_{Y|Z}^2 2^{-2\kappa} \\ &= 2 \ln(2) a \sigma_Y^2 2^{-4\kappa}. \end{aligned} \quad (34)$$

Higher information capacity and a more concentrated prior correspond to lower  $\lambda$ .

With a fixed cost of processing information,  $(RI_\lambda)$ ,

$$E[U] = -\frac{\lambda}{2 \ln(2)}, \quad (35)$$

losses and quality of posteriors are independent of initial knowledge, unlike in  $RI_\kappa$ .

## References

- [1] ALVAREZ, L., AND I. HERNANDO (2005). *Price Setting Behaviour in Spain. Evidence from Survey Data*, European Central Bank, working paper.
- [2] ALVAREZ, L., P. BURRIEL, AND I. HERNANDO (2005). *Price Setting Behaviour in Spain. Evidence from Micro PPI Data*, European Central Bank, working paper.
- [3] BILS, M., AND P.J. KLENOW (2004). *Some Evidence on the Importance of Sticky Prices*, Journal of Political Economy, 112.
- [4] BOIVIN J., M. GIANNONI, AND I. MIHOV(2007). *Sticky Prices and Monetary Policy: Evidence from Disaggregated U.S.Data*, working paper.
- [5] CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (1999). *Monetary policy shocks: what have we learned and to what end?* in Handbook of Macroeconomics, volume 1, ed, by J. B. Taylor and M. Woodford, Elsevier.
- [6] EICHENBAUM, M., N. JAIMOVICH, AND S. REBELO (2008). *Reference Prices and Nominal Rigidities*, Discussion paper, Northwestern University and Stanford University, NBER Working paper 13829.
- [7] HAYEK, F.A. (1945). *Use of Knowledge in Society*, The American Economic Review, 35(4).
- [8] KEHOE, P.J., AND V. MIDRIGAN (2007). *Sales and the Real Effects of Monetary Policy*, Working Paper 652,2007.
- [9] LUCAS, R.E. (1972).*Expectations and the Neutrality of Money*, Journal of Economic Theory.

- [10] LUCAS, R.E.(1973). *Some International Evidence on Output-Inflation trade-Offs*, American Economic Review.
- [11] MACKOWIAK, B., AND M. WIEDERHOLT (2004). *Optimal Sticky Prices under Rational Inattention*. working paper.
- [12] MANKIW N.G., AND R. REIS (2002). *Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve*, Quarterly Journal of Economics.
- [13] MATĚJKA, F.(2008). *Rigid Pricing and Rationally Inattentive Consumer*,working paper.
- [14] MATĚJKA, F., AND C.A. SIMS (2009) *Discrete Actions in Information-Constrained Tracking Problems*. working paper.
- [15] NAKAMURA, E., AND J. STEINSSON (2005). *Price Setting in Forward-Looking Customer Markets*, working paper.
- [16] NAKAMURA, E., AND J. STEINSSON (2006).*Five Facts About Prices: A Reevaluation of Menu Cost Models*, manuscript, Harvard University.
- [17] NAKAMURA, E. (2008). *Pass-Through in Retail and Wholesale*, working paper.
- [18] REIS, R. (2006) *Inattentive Producers*, Review of Economic Studies, 73.
- [19] SIMS, C.A. (1998). *Stickiness*, Carnegie-rochester Conference Series on Public Policy, 49(1).
- [20] SIMS, C.A. (2003). *Implications of Rational Inattention*, Journal of Monetary Economics, 50.

- [21] SIMS, C.A. (2006). *Rational Inattention: Beyond the Linear-Quadratic Case*, American Economic Review, 96(2), 158-163, Proceedings issue.
- [22] WOODFORD, M. (2009). *Information-Constrained State-Dependent Pricing*, working paper.