Rationally Inattentive Seller: Sales and Discrete Pricing

1

Filip Matějka[2]

CERGE-EI[3]

Abstract

Prices tend to remain constant for a period of time and then jump. In the literature, this "rigidity" is usually interpreted to reflect a cost of adjusting prices. This paper shows that price rigidity can alternatively reflect optimal price setting when there are no adjustment costs, namely, if the seller is rationally inattentive. The model generates non-trivial pricing patterns that are consistent with the data and that are hard to explain with the traditional adjustment-cost model. In particular, prices are adjusted frequently but move back and forth between a few given values, hazard functions are downward sloping, and responses to persistent shocks are sluggish. These results are obtained in a model that implements rational inattention without simplifying assumptions on the functional forms of the processed signals.

Keywords: rational inattention, nominal rigidity, sticky prices, sales.

JEL: D83, E31, E52

1 Introduction

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

An alternative line of the modeling of nominal rigidities is based on the assumption that agents cannot attend to all the available information about new shocks. This idea was proposed by Christopher Sims, formulated in a framework called "rational inattention"(Sims, 1998, 2003). The motivation for such assumptions can be found in cognitive psychology and managements science: an important problem that a decision maker in a firm faces is how to digest data and translate it into decisions (Simon, 1976, 1979; Mintzberg, 1973; O'Reilly, 1980). Decision makers simplify and synthesize the available information in such a way that its salient features come out and that the implied losses are minimized (Kiesler and Sproull, 1982; Levitt and March, 1988). I show in this paper that information frictions in the form of rational inattention can in fact generate nominal rigidities with several appealing properties for which other models cannot account. The most important results of this paper are about prices at the micro level.

I present a model where a rationally inattentive seller processes information about the volatile components of the environment, e.g., about competitors' prices, input costs, or elasticity of demand, and sets the price to maximize his profit. The question is then how prices respond to these shocks. The model generates the following results, which agree with the recent empirical literature: (A) Prices do not change all the time, but they can change frequently. (B) Prices tend to change back and forth between exactly the same values, thus appearing to follow a price plan based on a small number of values for the price. (C) Most price changes involve sales-like, short-term movements. Prices tend to spend a considerable amount of time at their modal values, and they are more likely to be below the mode than above it. (D) Hazard functions are downward sloping. (E) Price aggregates respond to aggregate shocks with a significant delay.

These results are driven by how the seller processes information about the shocks. The key

assumption is that rationally inattentive agents actively choose how to optimally allocate their limited attention. This choice then determines the nature of noise in observations and in beliefs, which affects the stochastic properties of prices.

Given that a seller cannot acquire perfect information, his pricing actions are imperfect and delayed. It turns out that, for instance, misjudging the elasticity of demand when it is high or input cost when it is low is more costly to the seller, so he pays more attention to shocks leading to low optimal prices, which then implies more flexible low prices and sales-like movements.

Perhaps most interestingly and surprisingly, the rationally inattentive seller chooses to price discretely, i.e., he sets up a price plan consisting of only a few prices that he chooses from, even when the shocks are continuously distributed. This implies that prices are likely to stay fixed when the shocks are small, although there are no explicit adjustment costs. The less information the seller is able to process, the lower the number of different prices he chooses to charge. This behavior looks as if the seller chooses to categorize data in some way, quite differently from how a seller would behave under the more standard assumption where the signal distribution is given exogenously as a fundamental plus additive noise, because such a setting generates continuous actions anytime the fundamentals change continuously. The discreteness result is derived from first principles here, based only on the assumption of rational inattention, and it is in line with cognitive psychology (Roach and Lloyd, 1978; Wilson et al., 1999) and some fields in economics, e.g., finance (Barberis and Shleifer, 2003), that acknowledge different forms of categorization as frequently observed strategies in decision making.

The reason for discreteness is cost-effectiveness: it allows the seller to pay attention to the "big picture", i.e., to isolate events that are quantitatively important for profits. The seller thus better distinguishes states that are far away from each other and avoids wasting information capacity on small movements. Discreteness in particular makes the likelihood of large errors small. The total amount of uncertainty in beliefs is given by the seller's level of attention, and a particular strategy only affects what the uncertainty is about, e.g., whether it is about nearby or more distant states. The seller can, for instance, choose to allocate his limited attention span to distinguishing between many similar situations in the market, leading to many slightly different prices (even a continuum); alternatively, the seller can devote all of the attention to discriminating between distant events and set only a very small set of prices. The lower the attention capacity, the smaller number of discrete

prices will turn out to be optimal. With very low attention capacity, the seller filters evidence to discriminate between distant events only.

The discreteness result derived here arises from forces that are quite different from the explicit adjustment frictions used in sticky-price models. This distinction might be particularly important when modeling rigidity under non-standard economic conditions. While, for instance, Calvo-style models fix the frequency of price adjustments and make prices fully flexible when they are adjusted, the form of nominal rigidity based on rational inattention emerges endogenously and is thus responsive to changes in economic conditions.

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

2 Related Literature

My paper is not the first model of pricing that uses rational inattention. Maćkowiak and Wiederholt (2009) were the first to do so. They show that under rational inattention, nominal aggregates respond sluggishly to money shocks. Their model is far more complex than the one presented in this paper.[4] However, its modeling of the allocation of attention is simpler. In particular, in their setup, the uncertainty is always Gaussian, and information about different shocks has to be processed separately.[5] While these can be realistic assumptions in some settings, the alternative model presented here allows for more flexibility of choices of attention. This model, for instance, allows the agents to pay more attention to lower input costs than to higher ones, and thus end up with somewhat asymmetric beliefs, or to processes information about some useful indices, such as sectoral aggregates, constructed from other variables of interest. Quite importantly, the model

---

[4]They even solve a DSGE model in Maćkowiak and Wiederholt (2015).

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

in this paper differs from Maćkowiak and Wiederholt (2009) by not assuming any specific form of posterior uncertainty.[6] In fact, this paper directly addresses Sims (2006), who claims that the nature, not just the level, of agents' uncertainty is subject to choice. The model of Maćkowiak and Wiederholt generates a very appealing result for price dynamics at the aggregate level. However, individual prices in their model change too often; in fact, they change all the time. The findings of my paper are thus important, because they show that when rational inattention is implemented in its general version, it can also account for evidence at the micro level, unlike sticky-price models. While the prices can change frequently, but not continuously, the aggregate response to shocks due to inattention is still delayed, just like in Maćkowiak and Wiederholt (2009).

Mankiw and Reis (2002) develop an alternative approach to the modeling of limited attention called "sticky information". There, agents are completely inattentive most of the time, and only once in a while they receive perfect information. While this approach is very tractable and can generate results that agree with aggregate data very well, individual prices under sticky information also change all the time—since there are predictable changes in aggregates relevant to the price setter that call for slow and predictable changes in his own price—and thus do not exhibit any rigidity on the micro level.

The discreteness result in this paper is very closely related to the findings in a technical study, Matějka and Sims (2010), where we explore the discreteness of actions under information constraints analytically in a class of tracking problems.[7] The present paper derives discreteness in more applied, price-setting contexts (which are not tracking models).

This study also relates closely to both the empirical and theoretical literatures, following the findings of Bils and Klenow (2004). Nakamura and Steinsson (2008) find that most price movements are price decreases, quickly followed by increases back to the original level, i.e., "sales".[8] They argue that these movements may be orthogonal to changes in macroeconomic conditions and thus may not have an effect on the flexibility of the price level. Eichenbaum et al. (2011) find that most prices, including sales, switch back and forth between exactly the same values.[9] They also infer that it is the rigidity of the "reference price", the quarter's most frequently quoted price, that is

---

[6]More accurately, the Gaussian posterior uncertainty in Maćkowiak and Wiederholt (2009) does result as an optimal choice if priors are Gaussian and the objective is quadratic.

[7]In tracking problems the objective is a function of a difference between state x and action y only.

[8]Nakamura and Steinsson use the CPI Research Database and also construct its analog for PPI.

[9]Eichenbaum et al. use the weekly scanner data of a major U.S. retailer.

the useful statistic for assessing aggregate nominal rigidity. The model presented here generates price series with these properties.

Kehoe and Midrigan (2014) build a model with heterogeneous menu cost where changing the regular price is more costly than renting a sale price for one period.[10] Eichenbaum et al. (2011) then propose a model with an assumed price plan, which is a finite set of prices between which adjustment is costless, while changing the plan is costly. Guimaraes and Sheedy (2011) were the first to embed a price-discrimination motive driving sales into a macroeconomic model. All these models have the feature that although prices do change frequently, the size of the real effects of money can be realistically large. The model with rational inattention presented here shares some implications with these models, and it endogenously generates some of their assumptions.

3   Model

3.1   Motivation and basic intuition

According to Simon (1976, 1979); Day and Nedungadi (1994), managers do not pay attention to all the details of their firms' market environment and instead consciously use selective attention and a simplified mental representation of the environment to be able to deal with all the available data on market conditions. They do this because they are cognitively constrained and not able to digest all the information that is presented to them. As documented in Kiesler and Sproull (1982); Levitt and March (1988), they moreover seem to digest those pieces of information that they find the most important.

As a starting example with which we can illustrate mechanisms, consider a large grocery store whose manager sets prices and makes many other decisions. The manager, however, operates under a constraint: he does not have enough time or mental capacity to digest all the details of the data relevant for setting the price of many products; the data includes anything relevant for judging the determination of supply (i.e., the cost of buying the product and the costs of making it reach the supermarket shelf) and of demand. Thus, the manager will have to make a decision under imperfect information. However, information is not an exogenous object and it is the manager's task to choose the relevant information to obtain and process. Here, the manager is subject to an

---

[10]This is the price inherited from the last period that can be charged without an additional cost.

"information processing capacity", a key element in rational inattention theory. This capacity is a constraint on how much the manager's prior information (about supply and demand) can feasibly be updated. Given this constraint, the manager chooses what to pay attention to. This choice can be thought of as what mental simplifications of the data will be used.

In the rational-inattention setting, the seller weighs the benefits and costs of obtaining various pieces of information. While the benefits from the information of a particular type are clear — they are given by the form of the profit function the seller faces—the costs of different ways of processing information are more difficult to assess. Rational inattention applies information theory to do so (Shannon, 1948; Cover and Thomas, 2006). There, entropy arises as a measure of achievable informativeness of information structures when the source information is provided in long sequences of its pieces, and the agent faces limitations on processing the individual pieces of information perfectly. We assume that the seller receives information in small pieces, perhaps from different sources, e.g., reads single digits or characters, or listens to pieces of speeches. We also assume that the information frictions come in the form of some elementary noise in the single pieces of information and that receiving each piece is costly. For instance, there is some probability that the seller misinterprets a given word that he hears or that he does not remember precisely a number presented to him. The exact form of the elementary pieces of information, or the exact form of the noise, is not important. Information theory applied to this setting implies that what information flow is achievable for the seller in this sequential problem can be described by a static consideration of entropy reduction in beliefs; see Appendix A.1 for more detail. The seller thus chooses the optimal information structure given a constraint based on a particular level of informativeness measured by entropy. The entropy constraint follows Sims (2003) and Appendix A.1 connects information theory with this constraint in more detail than has been done before.

The level of attention required to achieve a particular precision of belief is related to the number and precision of the elementary pieces of information that generate the total information. The number of such pieces is measured by the difference in uncertainty between the prior and posterior beliefs, and the measure of uncertainty is entropy. Naturally, more informative posterior beliefs are more difficult to achieve for any given prior—they require a higher attention capacity, i.e., a higher number of elementary pieces of information. In other words, with less attention, the difference between the posterior beliefs and the prior must, in an expectational sense, be smaller. This is the

key condition that restricts the attention and subject to which the rational decision maker must choose how to process the information: how many signals to obtain and their joint distribution with the state.

The main focus of this paper is on the coarseness of signals the seller chooses to process, i.e., on the number of predefined outcomes of information acquisition. It will turn out that the seller typically chooses a priori to distinguish between a finite number of different prices only, e.g., two prices, "high" and "low". While he is presented with many different pieces of information (on factors such as demand, costs, or competitors' prices), from every additional piece he only filters out the evidence discriminating between one of these two prices, and he does not keep track of any finer movements between them. The seller thus has reams of data at his disposal, but in the end he only has an impression of either "high" or "low".[11]

Discreteness allows the seller to pay attention to the big picture, i.e., to better distinguish states that are far away from each other, and not waste information capacity on small movements. Discreteness in particular lowers the likelihood of large errors and is therefore cost-effective. The higher the seller's attention is, the more useful it is to distinguish between more prices. Continuous prices typically emerge only when attention is unlimited. To understand these results intuitively, let us now briefly discuss two examples. We will then proceed to the formal analysis in Section 3.2.

---

[11]How is it possible that long sequences of elementary signals lead to a binary outcome? If there is a sequence of relevant and statistically independent pieces of information, then the number of possible outcomes of the whole sequence grows exponentially with the number of possible outcomes for any single piece of information. For instance, if the manager judged each of four different components of the product's total cost as either high or low, then there would be sixteen possible outcomes of this information acquisition process. But, using information theory, one can argue that it is suboptimal to process the information this way, i.e., when all pieces contribute jointly towards one cumulative stack of information (such as when estimating total cost). Instead, in general it is better to interpret each additional piece of information based on how it adds to the previously received information, such that the overall impression takes one of the predefined forms of the desired outcomes only. The impression from each component can be binary, but the threshold for the imperfect binary realization depends on the realization of the signal on the earlier components. For instance, if the impression from the first component is "high", then the cumulative impression after the second component is "low" only if the cost of the second component is more than 20% below average. And if the impression is "high" after the second component, then the cumulative impression after the third component switches to "low" only when the third component's cost is 40% below its annual average, etc. The choice of the meaning of signals is called in information theory "coding"; see Appendix A.1.

3.1.1 Example 1

We will first look at a very simple example that introduces the idea that finer signals can lead to larger errors if attention is limited. In this example, the set of possible types of attention is very limited, and the distinction between attention to small and large movements is obvious.

Consider a seller who processes information about the competitor's price, which can be anywhere between $10.00 and $29.90. The seller can spend all of his information capacity processing the first digit only, i.e., to gauge whether it is "1" or "2". Let's assume that the error rate in this assessment is 5%. Suppose, in contrast, that the seller instead chooses to distinguish between finer movements and processes the first two digits instead. Obviously, he can then devote less time and capacity to processing solely the first digit, and the probability of misjudging the first digit will be higher than 5%: this is the tradeoff he is facing. So, whereas processing the second digit may allow the seller to get a more precise idea about the optimal price, say, within $1, the likelihood that he misjudges it by a large amount ($10) also increases, due to a more likely misjudgment of the first digit. In the first case, the number of different signals is 2, while it is 20 in the second case. If the capacity is low, then by reading the second digit the error rate for the first digit increases significantly beyond 5%, and it is easy to show that it can be better to choose just 2 signals, i.e., focus on the first digit only.

The insights gained from this simple example extend to more general cases. Unlike in the example, this paper does not assume a particular set of available signals or a particular form of information acquisition. It instead allows for all forms of signals, and to assess how demanding it is to receive particular pieces of information it uses information theory; this is where entropy enters the picture.

3.1.2 Example 2

Let us now discuss the model with unconstrained forms of attention to illustrate the intuition behind why discreteness can emerge; the formal analysis appears later in the paper. The unconstrained form of attention is what the mechanisms based on long sequences of elementary pieces of information deliver, which is described in Appendix A.1.

Let p be the price set by the seller, let the state x represent the unknown profit-maximizing price (which the seller does not know when setting the price), and let $-(x - p)^2$ be the profit function.
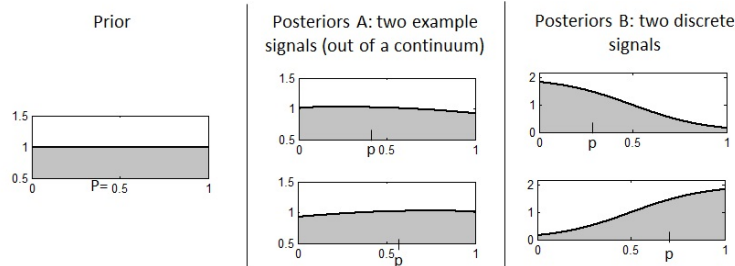
Figure 1: Prior and collections of posterior beliefs about state x for different strategies.

Given a posterior belief, the seller maximizes expected profits. This maximization straightforwardly leads to a maximizing price that is equal to the posterior expectation of x. The expected profit loss, moreover, then equals the posterior's variance.

Let us, for simplicity only, assume that the seller has a prior which is uniform on $(0, 1)$. This distribution thus is continuous, and one might expect that the signal that the seller will choose is continuous as well. That is, however, what we will argue is not optimal: the signal, and the posterior, will not in general be continuously distributed.

The first key element is to illustrate, in terms of the example, what the capacity constraint on attention means. This constraint says that the distance between the posterior and the prior—measured using entropy and itself a random variable—cannot be too large in expectation. We know that the prior is uniform in this case so we can illustrate how extracting a continuous signal would change the posterior under a tight capacity constraint. Figure 1 thus shows, in its left-side panel, the uniform prior and in its middle panel, how different outcomes for the continuous signal changes the posterior away from a uniform shape. Symbols "p" in the figure denote the locations of the expected state, and thus of optimal prices for each belief. If the seller did not process any information, he would set $p = 1/2$. In the middle panel, where posteriors differ little from the uniform prior, the prices are also close to $1/2$, they are about $1/2 \pm 0.1$.

The change is very slight because there are—given continuous support—many different signal outcomes. This means that the expected prior-to-posterior change, which is thus a probability-weighted sum over posteriors given all signal outcomes, cannot be too large.[12] Note now that entropy is a concave function of a distribution. This is a general property of measures of uncertainty:

---

[12]The posteriors are all non-negative-valued distributions and hence no "cancelation" can occur in the summation over signal values.
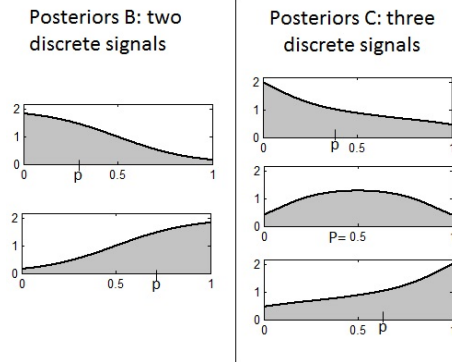
Figure 2: Collections of posterior beliefs about state x for 2-price and 3-price strategies.

mixing two beliefs increases uncertainty. If the expected entropy of posteriors is fixed by the information constraint, then the more posteriors there are, the more similar they must be not to exceed the capacity constraint. In other words, if there are many possible outcomes of information acquisition they must on average be very similar not to convey too much information.

The right-side panel shows, in contrast, the posteriors if a signal with two possible values is chosen subject to the same capacity constraint as for the middle panel. Thus, in this case the figure shows those two posteriors, which by definition, then, are the only ones that can occur. There could, of course, be many possible signals with only two outcome values, but the key here is to select one which focuses on extreme outcomes, as we shall argue in the next step. In this case, prices are also more extreme, about $1/2 \pm 0.2$. To conclude the first step of the argument, we now have two possible sets of posteriors, both satisfying the capacity constraint exactly, and the question is which of these sets of posteriors maximizes ex-ante expected profits.

The comparison between the continuous signal and the judiciously chosen two-value signal from an ex-ante perspective is straightforward, since profit losses are a convex function of the distance between x and p. Hence, the continuous signal will allow the seller to slightly improve on ex-ante profits as computed using the prior only—since the prior is very similar to each posterior. However, the discrete signal will increase profits significantly, since it focuses on eliminating large errors and hence large profit losses: differences between x and p are squared. The downside of the discrete signal is that it cannot improve matters by distinguishing between different, say, small values of x, which the continuous signals has some ability to do. But for ex-ante profits, this is not equally important.

11

Having argued that a discrete, two-valued signal can improve on a continuous signal under a tight capacity constraint, how can we continue the evaluation and comparison of different signals? For one, we can look at another discrete signal, but one with three outcome values. Can it improve matters? Figure 2 shows that this need not be the case. When the third signal is added, which resolves the middle range, then the initial two extreme signals must become less sharp. Note also that the corresponding prices are closer to the middle than in the two-price case. The average entropy of posteriors must be kept fixed, given by the information capacity, and so the posteriors must become more similar, again due to the concavity of entropy. The extreme states are thus resolved less well, which leads to a higher likelihood of large mistakes, and can lead to lower profit.

What if the capacity constraint is made looser? Infinite capacity, of course, would allow the manager to select a perfect signal, i.e., one which allows a perfect observation of x, and hence the ex-ante profits are the highest possible. So the focus of this paper, of course, is on limited capacity. Generally, higher but finite capacity will lead to more (and finite) signal values being feasible as well as desirable. Over some capacity ranges, however, it is possible that the number of signal values does not increase but rather improves in accuracy (based on observing "high", the probability of "low" falls). With higher capacity, the increase in the likelihood of large errors is relatively small, while a better resolution of the middle range increases profits.

In summary, a better resolution of middle states implies more frequent large mistakes in posteriors close to the prior's edge. Notice that the argument is based on the inherent asymmetry: some posteriors can lead to more extreme mistakes. In one case, however, this argument cannot be used, and that is when all the posteriors are of the same form. This can happen only when the prior is unbounded and different posteriors are translations of each other, and none of them is closer to extremal states than any other. We show in Section 4 that continuity with the quadratic profit function emerges if and only if the prior is unbounded and equal to a mixture of continuously distributed Gaussian distributions, e.g., is Gaussian.[13] This is exactly the special case studied in the earlier literature.

_____

[13]We discuss in Section 4 that in this case the optimal posterior is Gaussian because with a quadratic loss, the optimal noise balancing attention costs and profit losses is Gaussian in this model.

3.2 Formalization

A monopolistic seller sets a price p to maximize the expectation of his profit, Π.

$$\Pi = \Pi(\mathbf{x}, p), \tag{1}$$

where $\mathbf{x}$ is a vector of random variables that influence the objective. The stochastic properties of $\mathbf{x}$ are taken by the seller as given. In general, $\mathbf{x}$ can model any stochastic components of the environment, spanning micro variables such as all demand factors including competitors' prices or the input cost or stock levels, as well as macro variables such as the price level or interest rates.[14] For a fixed $\mathbf{x}$, $\Pi(\mathbf{x}, p)$ is a a single-peaked function in p.

The seller cannot observe the realized $\mathbf{x}$, but he first needs to process information about it. He makes two decisions. First, he chooses how to allocate his attention, which influences the nature of his posterior uncertainty about the variable. Second, conditional on receiving information and realizing some posterior with a pdf $k(\mathbf{x})$, the seller chooses the price maximizing the following expectation of profit.

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

Different strategies of attention allocation generate different collections of $k(\mathbf{x})$. In this model, we fix the amount of information the seller can process, a quantity κ called the information capacity, but we allow him to choose exactly how to process it, which is represented by choosing what "shapes" of posterior knowledge $k(\mathbf{x})$ can be realized. The fixed amount of processed information determines the achievable precision of knowledge in posteriors. For a given κ and thus a given level of posterior uncertainty, the seller chooses to resolve the uncertainty, i.e. selects the form of posteriors, in a way that is the most favorable with respect to his objective $\Pi(\mathbf{x}, p)$.

Time is discrete and the timing of events within each period is as follows. 1) Random vector $\mathbf{x}$ is drawn. 2) The seller processes information about the realized $\mathbf{x}$ and acquires some posterior

---

[14]What the true driving forces of price changes are, especially the sales-like movements on a short time scale, is still an open question. Eichenbaum et al. (2011) emphasize the important role of shocks to input costs and Kehoe and Midrigan (2014) build a model with technology shocks and money growth shocks. Nakamura (2008) finds evidence supporting the importance of dynamic pricing strategies similar to models with inventories (Lazear, 1986; Aguirregabiria, 1999) or price discrimination (Varian, 1980; Guimaraes and Sheedy, 2011). In this paper I study cases when $\mathbf{x}$ is strictly exogenous. Importantly, the model is not at odds with the other models of sales that do not rest on exogenous shocks. Rational inattention is a complement to the true drivers of profit-maximizing prices under perfect information.

knowledge $k(\mathbf{x})$. 3) The seller sets price p to maximize the expectation of the profit, given his knowledge about $\mathbf{x}$.

Both the choice of attention allocation and the pricing responses to posterior knowledge are described by a joint distribution $f(\mathbf{x}, p)$. Attention allocation is choosing what types of signals to receive, which is equivalent to deciding on the joint distribution of shocks and signals $f(\mathbf{x}, s)$. The distribution is subject to the information constraint as the signal cannot reflect $\mathbf{x}$ perfectly. Once a signal s is received and the posterior knowledge $f(\mathbf{x}|s)$ is realized, then the seller sets the price to $p[f(\mathbf{x}|s)]$ according to (2). The joint distribution of signals and $\mathbf{x}$ together with (2) determine the joint distribution of shocks and the selected prices $f(\mathbf{x}, p)$. The seller chooses how to process information, while considering what profit outcomes it leads to.

In this model, there is equivalence between prices and signals, which is a result. Appendix A and in more detail Matějka and McKay (2015) discuss a two-stage procedure where agents first choose a distribution of signals and then what price to set given that a particular signal was acquired. In the optimal strategy, there cannot be two signals leading to the same price, and thus the distribution of prices coincides with the distribution of signals.

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

$$f = \arg \max_{\hat{f}(\cdot, \cdot)} E[\Pi(\boldsymbol{x}, p)] = \arg \max_{\hat{f}(\cdot, \cdot)} \int_{\boldsymbol{x}} \int_{p} \Pi(\boldsymbol{x}, p) \hat{f}(\boldsymbol{x}, p) d\boldsymbol{x} dp, \tag{3}$$

subject to

$$\int_{p} \hat{f}(\boldsymbol{x}, p) dp = g(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \tag{4}$$

$$\hat{f}(\boldsymbol{x}, p) \geq 0, \qquad \forall \boldsymbol{x}, p \tag{5}$$

$$H[g(\boldsymbol{x})] - \mathbb{E}_p H\left[\hat{f}(\boldsymbol{x}|p)\right] \leq \kappa. \tag{6}$$

The seller maximizes the expected profit (3) by choosing the optimal $f(\mathbf{x}, p)$, which describes the distribution of prices charged for each $\mathbf{x}$, $\{f(p|\mathbf{x})\}$, but also the forms of posterior knowledge $\{f(\mathbf{x}|p)\}$. Constraint (4) requires consistency with prior knowledge. (5) states the non-negativity of a probability distribution.

(6) is the information constraint, see Appendix A for more detail. It requires that subjective uncertainty measured by entropy does not from the entropy of prior g on average decrease by more than a fixed quantity κ, the seller's information capacity. The entropy $H\left[g(\mathbf{x})\right]$ of a distribution with the pdf g is the following quantity:

$$H\left[g(\mathbf{x})\right] = -\int g(\mathbf{x})\log g(\mathbf{x})d\mathbf{x}. \tag{7}$$

For instance, for a normal distribution of variance σ^2, the entropy equals $\frac{1}{2}\log(2\pi e\sigma^2)$. The seller would like to observe $\mathbf{x}$ and concentrate all probability of $f(\mathbf{x},p)$ on the profit-maximizing price for each $\mathbf{x}$, but (6) states that the posterior distributions $\{f(\mathbf{x}|p)\}$ must be dispersed. The seller's problem is to allocate such entropy reductions to regions of $(\mathbf{x},p)$ where it has the highest benefit with respect to the objective $\Pi(\mathbf{x},p)$.

Definition 2. *Equilibrium.* *Let the random vector $\boldsymbol{x}$ be i.i.d. In each period t, $\boldsymbol{x}$ is drawn from a distribution whose pdf is $g(\boldsymbol{x})$. The equilibrium of the model is the joint distribution $f(\boldsymbol{x},p)$ and stochastic processes $\{\boldsymbol{x}_t\}$ and $\{p_t\}$, such that: (i) $f(\boldsymbol{x},p)$ is a solution to (3)-(6), (ii) $\{\boldsymbol{x}_t\}$ are drawn from $g(\boldsymbol{x})$, (iii) prices $\{p_t\}$ are drawn from $f(p|\boldsymbol{x}_t)$, and (iv) profits $\{\Pi(\boldsymbol{x},p)\}$ are realized.*

For the sake of simplicity, this part presents the version with i.i.d. random variables only. Section 5 then establishes that most properties of pricing often generalize to dynamic problems, too.

4   Solving the model

This section discusses the optimal strategies of the rationally inattentive seller. They generate random responses of prices to shocks in $\mathbf{x}$. I show, under some assumptions, that the support of the resulting distribution of prices consists of isolated points. This feature of the solution generates a rigidity of prices at the micro level. Then I show that the seller typically chooses to pay more attention to lower prices, which results in frequently-quoted high prices and flexible low prices. Then I show that the hazard rates of the prices are downward sloping. The final part of this section is devoted to a discussion of the empirical relevance of the findings.

When the information constraint (6) is not binding, then the seller observes $\mathbf{x}$, which is represented by posteriors with all probability mass concentrated at the true value of $\mathbf{x}$. The profit-

maximizing price

$$p_{opt}(\mathbf{x}) = \arg\max_p \Pi(\mathbf{x}, p) \tag{8}$$

is then set with certainty; the conditional $f(p|\mathbf{x})$ is degenerate at $p = p_{opt}(\mathbf{x})$.

The constraint (6) is binding and the seller acquires imperfect information any time the shocks $\mathbf{x}$ are continuously distributed and $\kappa < \infty$. The following proposition states how prices respond to changes in $\mathbf{x}$.

Proposition 1. (State dependence of prices) *If (6) is binding and if $g(\boldsymbol{x}) > 0$, then for a realized shock $\boldsymbol{x}$ the price is drawn from a conditional distribution with the following pdf:*

$$f(p|\boldsymbol{x}) = \frac{e^{\Pi(\boldsymbol{x}, p)/\lambda}}{\mathbb{E}_{\hat{p}}\left[e^{\Pi(\boldsymbol{x}, \hat{p})/\lambda}\right]} \frac{f(p)}{g(\boldsymbol{x})}, \tag{9}$$

where $\lambda \in \mathbb{R}$ is the Lagrange multiplier on the information constraint (6).

Proof: Appendix B.

Pricing is state dependent in a noisy fashion. For a realized shock $\mathbf{x}$, prices that generate higher profit tend to be realized with higher probability. This is given by the ratio of the term $\exp(\Pi/\lambda)$ for price p and the term's expectation across all prices. The seller's response function $f(p|\mathbf{x})$ is more closely concentrated about the profit-maximizing price where $\Pi(\mathbf{x}, p)$ is high, and more so if the information capacity is high, which makes λ low. Noise in signals that the seller receives generates a dispersion of the pricing response.

Notice that the probabilistic behavior of prices in (9) follows the logit model in Π/λ augmented for the unconditional $f(p)$, which is independent of the realized $\mathbf{x}$.[15] The effect of the unconditional distribution of prices $f(p)$ in (9) reflects prior beliefs and Bayesian updating; the more likely it is that some price is preferred by the seller a priori, the more likely it is for a specific shock, too. An analytical solution for $f(p)$ does not in general exist.

To allocate his information capacity efficiently, the seller never acquires signals that rule out some values of $\mathbf{x}$ with certainty, which is an immediate implication of Proposition 1. As long as the information constraint is binding, all posterior distributions overlap completely and all prices in the unconditional support can be realized for all possible $\mathbf{x}$.[16]

---

[15]Matějka and McKay (2015) study in detail how rational inattention leads the logit model.

[16]The variation of the entropy of the posterior $-\int f(\mathbf{x}|p) \log f(\mathbf{x}|p) d\mathbf{x}$ is infinite if $f(\mathbf{x}|p) = 0$ on a set of a positive measure, since the derivative of $d(-x \log x)/dx = \infty$ at $x = 0$. The marginal change of the entropy is infinite;

4.1 Price-plans

The price series that emerge have the properties of the price-plans observed in the data by Eichen-baum et al. (2011). Prices keep revisiting exactly the same values, while the set of these values is fixed and the top price tends to be visited more often than lower prices. There is no explicit cost of switching from one price value within the plan to another.

4.1.1 Discrete prices

The next proposition states that even when the profit-maximizing prices are continuously dis-tributed, the resulting distribution of prices can consist of isolated points only. While the propo-sition applies to profit functions that are quadratic in p only, numerical solutions suggest that the discreteness is very pervasive.

Proposition 2. *(Discreteness of prices) If $\Pi(\boldsymbol{x}, p)$ has a constant concavity in p,*

$$\frac{\partial^2 \Pi(\boldsymbol{x}, p)}{\partial p^2} = -a, \tag{10}$$

where $a > 0$ and $\kappa < \infty$, then the distribution of prices is discrete whenever there exists no pdf $f(p)$ such that

$$\hat{g}(x) = \int f(p) e^{-a(p-x)^2/\lambda} dp, \tag{11}$$

where $\hat{g}(x)$ is the distribution of profit-maximizing prices under perfect information, which is gen-erated from $g(\boldsymbol{x})$ and $x = p_{opt}(\boldsymbol{x})$, and $\lambda = 2\sigma^2/a$ such that σ^2 is the minimal variance of posteriors satisfying the information constraint.

Proof: Appendix C.

For all $f(p)$, the right-hand side of (11) is positive for all $x \in \mathbb{R}$ and is analytic. Therefore, if $g(x)$ is not positive on $\mathbb{R}$ or is not analytic, then (11) cannot hold, and the distribution of prices must be discrete. More generally, $g(x)$ with steep gradients and thin tails tend to generate discreteness.[17] Notice that if g is Gaussian, and thus analytic and positive on $\mathbb{R}$, then the solution can be continuous.

---

therefore the marginal value of the unit of such information completely ruling out some values of **x** is for a bounded profit function zero.

[17]See Widder (1951), who studies what functions can be expressed as the right-hand side of (11).

Corollary 1. *If $\hat{g}(x)$ has a bounded support or is not analytic, then the distribution of realized prices is always discrete.*

This result can be interpreted as if the seller chooses to use some form of categorization of data. Informationally constrained sellers prefer discretized summarizing signals, because they allow them to avoid large errors. The main mechanism is the interaction between additional signals and prior knowledge, which was discussed in Section 3.1.

Let us clarify why the waste from a small number of prices is dominated by the large-errors effect. The reason is that when the information constraint binds there is no direct waste from a low number of prices, e.g., two. It would seem that a higher number of prices always allows the seller to pick one closer to the optimum. But it is not the case in this model; the seller always sets the price at the optimum given his beliefs. More prices require signals with more realizations, but due to the information constraint the noise in beliefs and the errors in pricing do not vanish if more prices are added. Consider the following collections of uniform distributions representing three strategies of processing information, and let us abstract from the consistency with the prior.

$$S_1 = \left\{ U_{(0,1/2)}, U_{(1/2,1)} \right\}, S_2 = \left\{ U_{(0,1/2)}, U_{(1/4,3/4)}, U_{(1/2,1)} \right\}, S_3 = \left\{ U_{(d,1/2+d)} : d \in (0,1/2) \right\}.$$

With S_1, the signal takes two values and it leads to two posteriors and two prices, while for S_2 the signal takes an additional value in the middle and it leads to three prices, and with S_3 there is a continuous interval of prices. Each of the posteriors is a uniform distribution of the width $1/2$, and thus the expected posterior entropy as well as posterior variance, and thus also profit, are the same for the three strategies. The information constraint pins down the expected entropy of posteriors, which is then completely independent of the number of different posteriors and prices. Since profit equals the negative expected variance of posteriors, which is related to the entropy as both are measures of uncertainty, then from this basic perspective when not accounting for the prior the profit is independent of the number of prices.

In general, with bounded priors, posterior distributions with continuous prices are always inherently inhomogeneous in the sense that some posteriors must generate prices further from one of the edges of the prior than others, and thus imply a higher likelihood of large errors. Discreteness then helps to minimize this likelihood. This is not the case for unbounded priors, when all posteriors can in fact have exactly the same shape. Condition (11) for continuity is even more specific; it states

that continuity emerges only when the posteriors can all be Gaussian and of the same variance, which is for priors satisfying (11). Why Gaussian posteriors? For quadratic objective, Gaussian posteriors minimize variance given the level of entropy, i.e. they are optimal.[18] This is a special case when there is no discrepancy between the tails of different posteriors and information is utilized efficiently, and it is exactly what is assumed in the earlier literature on rational inattention (Sims, 2003; Maćkowiak and Wiederholt, 2009), where all the actions have continuous support. If the uncertainty were not a convolution of Gaussians, then the actions would be discrete. Or conversely, numerical solutions suggest that the discreteness arises for Gaussian uncertainty if the objective is not quadratic.

4.1.2 Asymmetric pricing and reference prices

The optimal pricing of a rationally inattentive seller can be asymmetric. The frequencies of the occurrence of different prices can be tilted towards high or low prices even if the distribution of profit-maximizing prices is uniform. In the data the asymmetry is such that the most frequently quoted prices tend to be high. The most-often-quoted price is what Eichenbaum et al. (2011) call the "reference price". In their sample, only 21% of prices away from the reference price are above it. Similarly, Kehoe and Midrigan (2014) find that prices are 2.5 times more likely to be below the annual mode than above it, and Nakamura and Steinsson (2008) found that most price movements are price decreases quickly followed by an increase to the original level.

The next proposition is concerned with the optimal allocation of attention across different levels of $\mathbf{x}$. The proposition is stated for low levels of noise, where it is a simple application of a Taylor expansion. The seller improves his knowledge the most in regions where the marginal profit from processing extra information is the highest and the seller's actions in these regions are therefore relatively more precise. First, the loss of profit due to a small departure ϵ from the profit-maximizing price is equal to $-1/2 \frac{d^2 \Pi(\mathbf{x},p)}{dp^2} \epsilon^2$. Second, misjudging a component $\mathbf{x}_i$ of the exogenous random vector $\mathbf{x}$ by a small η leads to mispricing by $\frac{\partial p_{opt}(\mathbf{x})}{\partial \mathbf{x}_i} \eta$.

Proposition 3. (Asymmetric prices) *When the information capacity is sufficiently close to*

---

[18] In general it is $\exp(\Pi(x,p)/\lambda)$, where λ is a constant. This finding can be derived in a similar way to equation (9) when one omits the constraint on the prior. λ is a Lagrange multiplier on the information constraint, which is given by g and κ.

$H(g)$, *then the expected entropy of posteriors is lower in regions of* $\boldsymbol{x}_i$ *where*

$$-\left(\frac{\partial p_{opt}(\boldsymbol{x})}{\partial \boldsymbol{x}_i}\right)^2 \frac{d^2\Pi(\boldsymbol{x}, p)}{dp^2}\Bigg|_{p=p_{opt}(\boldsymbol{x})} \tag{12}$$

is relatively high.

In other words, the seller pays closer attention to states with high (12). This implies that when the seller chooses to process less information in some regions of **x**, it forces him to discretize prices corresponding to such realizations of shocks more coarsely. The objective's asymmetry introduces a bias for prices being quoted more frequently when (12) is relatively low. For the specifications considered later, (12) is typically low at high profit-maximizing prices. If the distribution of profit-maximizing prices were not uniform, but for instance with a peak in the middle, then a mere discretization would generate the most frequently quoted middle price, and the asymmetric attention then shifts the weights towards higher prices.

For most of the standard profit functions that are discussed below, the seller chooses to pay more attention to **x** that lead to low prices. This implies that the top prices in the price plan are then the most rigid ones. When elasticity of demand is high, or input cost low, then the seller could lose relatively more of his potential profit by deviating from the profit-maximizing price.

4.1.3 Hazard rates

When prices take only a finite number of values, then there is a positive probability that the price does not change from one period to another.

Proposition 4. *(Decreasing hazard functions)*

Let the price take N different values only and let $\{f_i\}_{i=1}^N$ be their probabilities. Then the hazard function $H(T)$, the probability of a price being changed in period $T > 1$ conditional on its survival to period T, is:

$$H(T) = 1 - \frac{\sum_{i=1}^N f_i^T}{\sum_{i=1}^N f_i^{T-1}}. \tag{13}$$

$H(T)$ *is constant if and only if $\{f_i\}_{i=1}^N$ are uniform, otherwise it is decreasing.*

Proof: Appendix D.

Figure 3 presents the hazard function $H(T)$ for a two-point price-plan for different probabilities of the less frequent price f, while the reference price occurs with probability $(1 - f)$. The rates
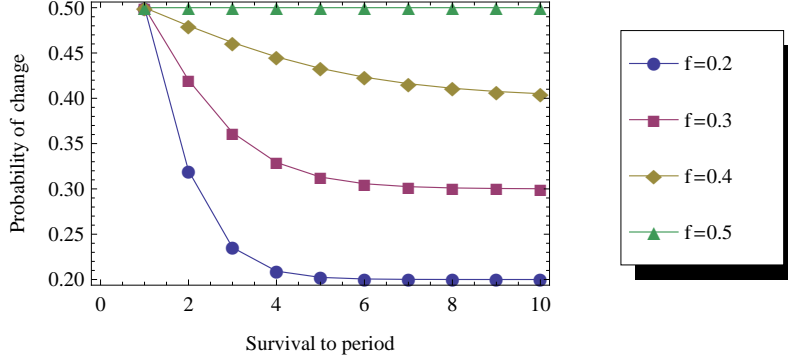
Figure 3: Hazard rates for a binary distribution of prices

are decreasing and converging to f. Conditional on survival to period T, the probability that the price is currently the frequently quoted price increases, since it is unlikely that a price associated with a low probability is drawn several periods in a row. Therefore, it is also more probable that a price surviving to high T is drawn in the next period. Rational inattention generates decreasing hazard functions naturally, while the quantitative agreement of the rates with data depends on the dynamics of the underlining shocks, too.

The plan's non-uniformity generates hazard rates very similar to those reported in Nakamura and Steinsson (2008); the hazard functions are first downward slopping and then flat.[19] This is an important finding, which is not easily generated by sticky-price models.[20]

4.2 Numerical solutions for different objectives and sources of shocks

Let us now inspect numerical solutions to (3)-(6) for the following profit function.[21] Its qualitative properties are shared by other specifications, too, which are shown later.

$$\Pi(d,p) = p^{-\frac{d+1}{d}}(p-1). \tag{14}$$

---

[19]See also Kehoe and Midrigan (2014) for a similar result.

[20]Alvarez et al. (2005) argue that decreasing hazard rates are found in the data due to aggregation.

[21]It is not difficult to solve the problem numerically, since it is a maximization of a linear objective on a convex set. I discretized the 2D domain of **x** and p by introducing 70×70 grid points, and used a solver called LOQO (Vanderbei, 1999) which is based on interior point methods.
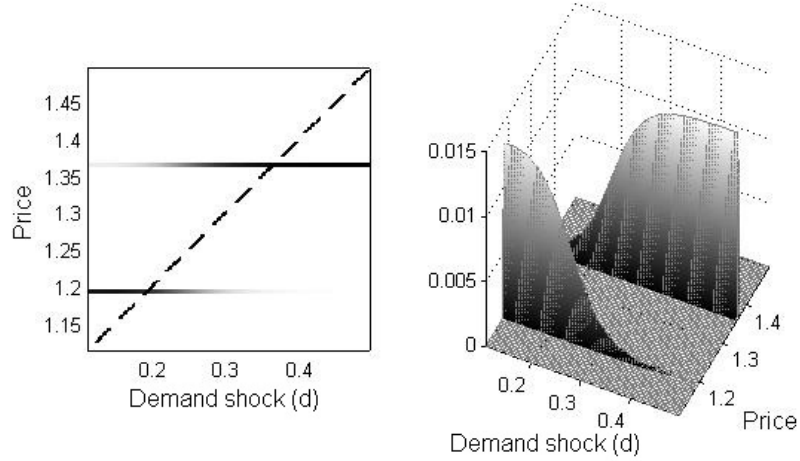
Figure 4: Joint distribution: $\kappa = 0.5$ bit, uniformly distributed optimal markup.

It is the elasticity of demand $\frac{d+1}{d}$ that is random, $\mathbf{x} \equiv d$, and the seller needs to process information about it. $p^{-\frac{d+1}{d}}$ is demand and $(p-1)$ is profit per unit sold; the unit input cost is 1.[22] d is uniformly distributed in $(\frac{1}{9}, \frac{1}{2})$. This is the prior, which makes the profit-maximizing prices uniformly distributed in $(\frac{10}{9}, \frac{3}{2})$ and the elasticity of demand distributed in $(3, 10)$.

Figure 4 shows the optimal joint distributions $f(d, p)$ for $\kappa = 0.5$ plotted in two different ways. The left exhibit depicts the support of $f(d, p)$ well, while the right one is more suitable for presenting the levels and shapes of the distribution. The dashed line in the graph on the left in Figure 4 is the profit-maximizing pricing strategy $p_{opt}(d) = 1 + d$ arising under perfect information, (8).

The optimal distribution of prices is discrete. Although the demand shocks as well as the profit-maximizing prices are continuously distributed, there are only two different prices being realized. For any d, the realized price is either 1.2 or 1.37. The unconditional probability of the higher price is 63%. Plugging in these values, $f(1.2) = 0.37$ and $f(1.37) = 0.63$, into (9) delivers the complete description of the responses of price to demand shocks.

$$f(1.2|d) = \frac{e^{\Pi(d,1.2)}}{0.37e^{\Pi(d,1.2)} + 0.63e^{\Pi(d,1.37)}} \frac{0.37}{g(d)}, \qquad f(1.37|d) = 1 - f(1.2|d), \qquad (15)$$

where $g(d) = 7/18$. In the right exhibit in Figure 4, it is apparent that the conditional distribution

---

[22]For illustrative purposes I specify the random component as d, and not directly as the elasticity $\frac{d+1}{d}$. The profit-maximizing price is linear in d, and thus prior distributions of d directly translate into distributions of profit-maximizing prices, e.g., uniform to uniform, and allow for a more straightforward understanding of the symmetry of solutions.
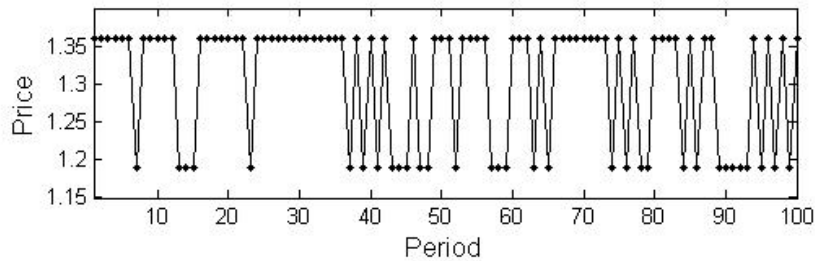
Figure 5: Simulated time series of prices.

$f(p|d)$ for d larger than about 0.3 places much more probability on the higher price, while with d close to 0 the lower price is more probable. This is what Proposition 1 states as at high d high prices are relatively more profitable. However, either price can be drawn for any realizations of the shock. Figure 5 presents the simulated time series of the resulting prices. The top price is realized more frequently, although the distribution of profit-maximizing prices is uniform, i.e. symmetric.

The rest of this section explores solutions to different specifications of objective Π and exogenous variable $\mathbf{x}$. It turns out that the solution's properties do not change much from one specification to another when the price changes are driven by exogenous shocks, whatever the shocks are. In all cases, the seller targets some profit-maximizing price, which is stochastic due to shocks to $\mathbf{x}$. What differs from case to case is exactly how losses from imperfect pricing depend on the magnitude and the type of mispricing. For smooth profit functions, the losses are always close to quadratic when mispricing is small, but can take other forms for larger deviations.

Let us study cases with exogenous shocks with a profit function in the form

$$\Pi(\mathbf{x}, p) = D(\mathbf{x}, p)\Big(p - AC(\mathbf{x}, p)\Big), \tag{16}$$

where $D(\mathbf{x}, p)$ is the demand for the seller's product and $(p - AC(x, p))$ is the markup with AC being the average input cost. The demand as a function of p can, for instance, change due to shocks to consumers' preferences or due to shifts competitor's prices. The average cost can be affected via shocks to unit input costs or through shifts in demand when the returns to scale are not constant.

I first discuss the properties of numerical solutions to separate setups with different sources of shocks and then also present the solution to the problem with several shocks being active. Table 1 presents the solutions to five representative specifications of $\Pi(\mathbf{x}, p)$, which span general small shocks, demand shocks, cost shocks, and shocks to competitors' prices. The profit-maximizing

prices under perfect information are uniformly and continuously distributed in all five cases.[23]

$$\Pi_1(x,p) = -(x-p)^2, \qquad \Pi_2(\theta,p) = p^{-\theta}(p-1), \qquad \Pi_3(\delta,p) = \delta p^{1-\theta} - \left(\delta p^{-\theta}\right)^{\frac{1}{a}},$$

$$\Pi_4(\mu,p) = p^{-\theta}(p-\mu), \qquad \Pi_5(p_2,p_1) = c_1(p_2,p_1)(p_1-1).$$

Π_1 represents an approximation to a general case with x summarizing all variables that affect the profit-maximizing price, while deviations of the selected price from the profit-maximizing price come at the quadratic cost. Such a specification naturally provides a good approximation for any smooth objective and small variability in profit-maximizing prices due to all kinds of shocks.

Π_2, which is equivalent to that in (14), and Π_3 model shocks to demand. In Π_2, it is the elasticity of demand and thus also the optimal markup that are stochastic with the unit input cost being constant and normalized. In Π_3 the source of shocks is the level of demand and the production technology has decreasing returns. The first term is revenue from selling $\delta p^{-\theta}$ units, while the second term is the total cost. The production function is L^a, where $a \in (0,1)$, $a = 2/3$, L is automatically adjusted labor, and the cost of labor is 1. When the demand increases, then more workers must be hired, but the marginal cost and the average cost per unit sold increase, thus the profit-maximizing price increases, too.

In Π_3, while the source of shocks is the demand level, the unit input cost is uncertain too, since it depends on the level of sales and thus in turn on the demand level. Π_4 models the case when it is the unit input cost that is random and drives shocks to the profit-maximizing price. As opposed to demand, the unit cost might seem likely to be known by the seller. However, this model does not assume that the information is not available within the price-setter's organization, it only assumes that it is costly for the decision maker to process the available information and thus reflect it in the resulting price. Cost shocks can also be driven by shocks to inventory levels.

The random variable entering Π_5 is the price of a competing product. $c_1(p_1,p_2)$ denotes the demand function for the seller's product; the seller's own price is p_1 and the competitor's price is p_2. I used a demand function derived from the consumer's problem with a utility function that is Dixit-Stiglitz in the consumption of the two products and linear in money spent, $U(c_1,c_2) =$

---

[23]In the first four cases, the profit-maximizing price can be expressed as proportional to some target, to markup $1/(\theta-1)$, $D^{2+\theta}$, and input cost μ, which were thus set to be uniformly distributed. The prior distribution of p_2 in Π_4 was more complicated, but it was adjusted such that profit-maximizing prices were uniform, too.

$\left(c_1^{1-1/\theta_1} + c_2^{1-1/\theta_2}\right)^{\frac{\theta_1(\theta_2-1)}{\theta_2(\theta_1-1)}} - c_1 p_1 - c_2 p_2$, where c_i is the consumption of product i and $\theta_{1,2}$ are the corresponding elasticities of substitution. The consumer's demand for the seller's product is:

$$c_1(p_1, p_2) = \left(\frac{\theta_2}{\theta_2 - 1} p_1 \left(1 + (p_1/p_2)^{\theta_1-1}\right)^{\frac{\theta_2-\theta_1}{\theta_2(\theta_1-1)}}\right)^{-\theta_2}. \tag{17}$$

Table 1 presents the properties of numerical solutions to $\Pi_1 - \Pi_5$; discussion is postponed to Sections 4.2.1 and 4.2.2. For all $\Pi_1 - \Pi_5$, the distribution of shocks were selected such that the distribution of profit-maximizing prices were uniform in $(1.1, 1.5)$. Π_2: $\theta \in (3, 10)$; Π_3: $\theta = 4$ and $D \in (0.43, 2.8)$; $\Pi_4 : \theta = 4$ and $\mu \in (0.85, 1.15)$; $\Pi_5 : \theta_1 = 10$, $\theta_2 = 3$, and $p_2 \in (0.9, 2.4)$.

	set of realized prices	% at top price	profit loss
Π_1	{1.22,1.39}	50%	N.A.
Π_2	{1.20,1.37}	63%	2.4%
Π_3	{1.22,1.39}	55%	2.3%
Π_4	{1.21,1.39}	63%	1.8%
Π_5	{1.21,1.37}	62%	2.9%

Table 1: Comparative statics for different specifications

Finally, let us include several shocks at once. I ran a computation for specifications Π_2, Π_3, and Π_4 combined.[24] This is a model with stochastic level and elasticity of demand, and also unit input cost conditional on a given production level stochastic.

$$\Pi_{2,3,4}(\theta, D, \mu, p) = Dp^{1-\theta} - \mu \left(Dp^{-\theta}\right)^{\frac{1}{a}}.$$

When the prior $g(\theta, D, \mu)$ is adjusted such that the profit-maximixing prices under perfect information are distributed uniformly, then for $\kappa = 0.5$ bit, the resulting distribution of prices were again discrete with values $\{1.10, 1.22, 1.38\}$ and the corresponding probabilities are $\{0.15, 0.23, 0.62\}$ with the top price being the most likely.

4.2.1 Rigid price values

Solutions to all the different specifications presented in Table 1 are very similar. Most importantly, all the distributions of prices are discrete, although Proposition 2 applies directly to Π_1 only. All

---

[24]Π_5 was not included simply because it differs more markedly from the other specifications, and the combination of all four specifications would thus be much less straightforward.
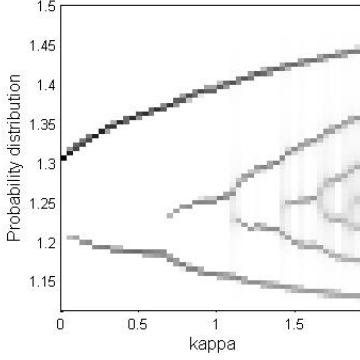
Figure 6: Bifurcation diagram of prices, Π_2

pricing strategies take the form of price-plans with a finite number of fixed price-points. Moreover, each solution to a problem with one shock consists of exactly two price-points. There are three prices when the shocks are combined, partly due to the fact that the range of profit-maximizing prices under perfect information is then larger.

The more information the seller processes, the finer the discretization. The bifurcation diagram in Figure 6 shows price distributions as a function of information capacity for specification Π_2. When the capacity increases, new price points emerge. There is only one price when $\kappa = 0$, the second price emerges immediately as $\kappa > 0$, the third price emerges at about $\kappa = 0.7$, etc. Very similar diagrams emerge also for the other profit functions. It is important to emphasize that while the resulting pricing strategies seem to be very crude approximations of the strategies under perfect information, the profit losses are fairly low. Already for the solutions presented in Table 1 to $\Pi_2 - \Pi_5$ with $\kappa = 0.5$, the loss in profit is lower than 2.5% and it further decreases for higher κ.

4.2.2   Reference prices

The solutions, however, differ in the probabilities of different prices, i.e. in the relative frequencies of different values of prices. While the profit-maximizing prices are uniform, the price-plans corresponding to Π_2, Π_3, and Π_4 with $\kappa = 0.5$ are highly asymmetric with the top price being more frequent. The solution to Π_3 is less asymmetric, and the solution to Π_1 is completely symmetric. This heterogeneity in price rigidity depends on the relative losses from imperfect pricing at different realizations of shocks; see Proposition 3.

For Π_1, pricing is symmetric since the objective is symmetric, i.e. its concavity is constant. On

26

the other hand, for $\Pi_2 - \Pi_4$ it is straightforward to prove that the seller pays more attention to shocks that generate low prices; the absolute value of the loss factor in (12) is decreasing in markup, demand level, and unit input cost. Top prices are more rigid, while sales are short-lasting. Finally, shocks to competitor's prices in Π_5 can generate asymmetry in both ways. Numerical solutions suggest that the top prices are rigid when competitor's prices are typically higher than the seller's and vice versa.

In this model, the top prices are more likely to be the rigid ones for most types of shocks when the demand's elasticity is not increasing too rapidly. For a demand with constant elasticity, $p^{-\theta}$, and a general markup $(p - AC(\mathbf{x}, p))$, the absolute value of the loss factor in (12), the concavity of profit at the profit-maximizing price, equals

$$p^{-\theta} \times \left| p^{-2}(\theta(1+\theta)AC(\mathbf{x},p)) + p^{-1}\theta((1+\theta) + 2AC_p'(\mathbf{x},p)) + AC_{pp}''(\mathbf{x},p) \right|.$$

$p^{-\theta}$ is decreasing in the price, while the second term can go either way depending on the specifics of $AC(\mathbf{x}, p)$; it is also decreasing for both Π_2 and Π_4. The decreasing convexity of the demand function is a force towards asymmetry with more rigid top prices.

In general, the asymmetry increases with the elasticity of demand θ, decreases with the information capacity κ, and increases with the variance of underlining shocks.

Eichenbaum et al. (2011) report that prices stay at their most frequent value 62% of the time, and that the weekly price changes with a probability of 43%. For $\Pi_2 - \Pi_4$ in Table 1 the probabilities of the top price are also about 60% and the implied probabilities of price changes are about 48%.

If the plan consists of two points only, then the probability $1 - f$ of the top price determines the probability of a price change in a given week, $2f(1 - f)$, and the implied expected price duration, $f(1 - f)\sum_{t=0}^{\infty} t(1 - f)^t + tf^t$.

Table 2 compares the implications of different f with data reported in Eichenbaum et al. (2011) and Kehoe and Midrigan (2014). When $1 - f$ equals 0.62, 0.77, or 0.58, which fits the data, then the implied probability of a price change is slightly above and the average price duration slightly below the one in the data. The fit is not bad, considering this is done for the simplest specification possible, i.e. for the i.i.d. model. The i.i.d. specification potentially biases the results in the observed direction, since the true driving processes are likely to be at least a little persistent. Calibrating the dynamic model in Section 5 with persistent A would bring the model closer to the data.

	1-f	Prob. of change	Avg. duration
Model	0.62	0.47	2.2
	0.77	0.35	3.6
	0.58	0.48	2.1
Eichenbaum et al. (2011), primary data-set	0.62	0.43	2.3
Eichenbaum et al. (2011), Dominicks data-set	0.77	0.24	4.1
Kehoe and Midrigan (2014)	0.58	0.33	3.0

Table 2: Dynamics of weekly prices

The model with i.i.d. shocks implies that the price decrease is followed by an increase with probability f, which is above 50%. Kehoe and Midrigan (2014) find this probability to be 46%, while it is 58% when the model fits their probability of the modal value. The model agrees with empirical findings that there is a high probability that the price increases after its decrease. The reason for the reversion is simply that higher prices are more probable.

5 Dynamic model

This section discusses the dynamic price-setting problem, which is an extension of the original model (3)-(6). Here, the exogenous variable $\mathbf{x}$ is serially correlated, which makes both the price-setting as well as processing of information dynamic.

The purpose of this section is threefold. First, it explains how such a model is formulated and solved. Second, it shows that the main insights from the static model, e.g., the discreteness of prices, are not driven by the i.i.d. structure of shocks. Third, it shows that while individual prices change frequently, the average, i.e. aggregate, prices respond to aggregate shocks on a much slower time scale.

The random vector $\mathbf{x}$ can have stochastic properties of any form. Similarly as in Sims (2003) or Maćkowiak and Wiederholt (2009), I study the responses to two distinct sources of shocks, which differ in their persistence. The other authors explore dynamic problems with quadratic objectives and autoregressive Gaussian processes for shocks, while I solve a case with a fully nonlinear objective and simpler stochastic properties of the shocks. The transient components of the shocks are

continuously distributed, and the persistent components are binary Markov.[25]

The binary structure of the persistent component suffices to highlight the model's implications. Here, the prices will again be discrete, while in models with Gaussian signals they would be continuous. There, continuity would emerge both due to the fact that the transient components of the shocks are continuously distributed, and thus the profit-maximizing prices are continuously distributed, too, as well as due to the continuously distributed noise in signals.

5.1 Formulation of the dynamic model

In each period t, the seller processes information about $\mathbf{x}$. The seller's decision strategy is again described by a joint pdf $f_t(\mathbf{x}, p)$. As in the i.i.d. case, the pdf needs to be consistent with the prior distribution g_t in that period, and the seller cannot in any period process more information than κ.

$$\int f_t(\mathbf{x}, p)dp = g_t(\mathbf{x}) \tag{18}$$

$$I\left[f_t(\mathbf{x}, p)\right] \leq \kappa. \tag{19}$$

If $\mathbf{x}$ is not i.i.d., then posterior knowledge $f_t(\mathbf{x}|p_t)$ translates into the next period's prior according to:

$$g_{t+1}(\mathbf{x}) = T\Big(f_t(\mathbf{x}|p_t)\Big), \tag{20}$$

where the law of motion $T(\cdot)$ is determined by the stochastic properties of $\mathbf{x}$.

Starting with a given prior g_0 at $t = 0$, the seller sequentially chooses strategies $\{f_t\}$ to maximize the discounted expected profits,

$$\sum_{t=0}^{\infty} \beta^t \int \Pi(\mathbf{x}, p) f_t(\mathbf{x}, p) d\mathbf{x} dp, \tag{21}$$

subject to (18)-(20).

This sequential problem can be formulated recursively. Let V be the maximum attainable value of the objective above. It is a function of the prior in period zero, $g = g_0$. The value function $V(g)$

---

[25]Solving fully-nonlinear dynamic problems under rational inattention is computationally demanding. The seller's state variable is his current knowledge of the cost's components, which is an infinitely dimensional object if the components are continuously distributed. For this reason, the persistent variable in the model I solve is binary Markov.

satisfies the Bellman equation

$$V(g_t) = \max_{f_t} \int \Big[\Pi(\mathbf{x}, p) + \beta V(g_{t+1})\Big] f_t(\mathbf{x}, p) d\mathbf{x} dp, \tag{22}$$

subject to (18)-(20).

Definition 3. *Equilibrium of the dynamic model.* *The equilibrium of the model is the value function $V(g)$ and stochastic processes $\{f_t(\boldsymbol{x}, p)\}$, $\{\boldsymbol{x}_t\}$, and $\{p_t\}$, such that: (i) $V(g)$ is a solution to (18)-(20) and (22), (ii) the initial prior $g = g_0$ and random vector $\boldsymbol{x}$ are given, (iii) a stochastic process $\{\boldsymbol{x}_t\}$ is generated, (iv) for a given g_t, $f_t(\boldsymbol{x}, p)$ is a solution to the right hand side of (22), (v) prices $\{p_t\}$ are drawn from $\{f_t(p|\boldsymbol{x}_t)\}$, and (vi) $\{g_t\}$ are generated by the low of motion (20).*

In each period, the seller maximizes the right-hand side of (22). Not only does the seller consider the current profit, but also the effects of the currently acquired knowledge on the future profits.

5.2 Numerical solutions

Let us solve the model for the specification Π_4 and $\mathbf{x} = (\mu, A)$. The unit input cost is then $A\mu$, a product of the two components of $\mathbf{x}$: a) an i.i.d. component μ and b) a Markov component A, which switches between two levels: A_L and A_H with a transition probability equal to τ. Similarly as in the static case, a particular choice of specification among the profit functions $\Pi_1 - \Pi_5$ does not change the results much.

Knowledge about a binary variable is described by the probability of each of the two states. It is therefore just a scalar, which is very convenient. Let the value function's state variable, which describes the prior, be the probability of $A = A_H$. μ is supposed to represent the volatile part of the input cost specific to the seller, while A plays the role of a stable aggregate variable, e.g., the price level. A is an index shifting the distribution of the nominal input cost $A\mu$.

The seller's prior takes the form

$$g_t(A, \mu) = g_{1,t}(A)g_2(\mu), \tag{23}$$

where g_2 is the fixed pdf of the i.i.d component, while the prior on A, $g_{1,t}$, can vary across periods. Shocks to the two components are independent of each other. The way the seller chooses to process information determines the collection of potential posteriors, and then priors in the following period.
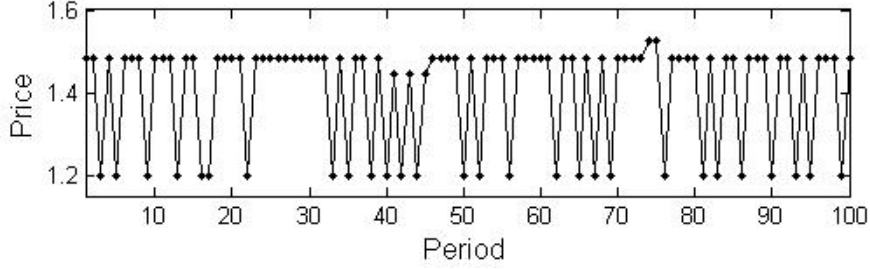
Figure 7: Simulated prices for unstable A, $\tau = 0.2$.

Since A is persistent, the posterior knowledge $f_t(A|p_t)$ translates into the next period's prior $g_{1,t+1}$. Considering that the probability of transition to the opposing state is τ, the law of motion for knowledge (20) takes the following form

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

The value function, the fixed point of (22), can be found by iterations, while the right hand side of (22) is solved using the same techniques as for the static model in Section 4.

Figures 7-9 present the realized price series for different levels of τ, the levels of instability of A. The figures show how the persistence of A affects responses to its shocks. The path of A is the same across all the simulations. A is at $A_L = 1.0$ for $t \in \{1..19\}$, then it switches to $A_H = 1.1$ at $t = 20$, and afterwards it stays there. The i.i.d. component, μ, is uniformly distributed over $(0.85, 1.15)$ and drawn randomly in each period, $\kappa = 0.5$, $\theta = 4$, and $\beta = 0.9992$. One period represents one week, which makes the annual discount factor equal to 0.96.

The price setter targets the profit-maximizing price $\frac{\theta}{\theta-1} A\mu$. He can choose to receive signals on $A\mu$ only, or discriminate among A and μ in some way. While for the profit in a given period, it is only the knowledge about the total marginal cost $A\mu$ that matters, it is knowledge about A that is useful in the following periods.

Prices again exhibit a rigidity of values. Given prior knowledge about A, the distribution of prices forming a current price plan is discrete. However, when knowledge about A changes, the distribution of prices determining the seller's price plan changes, too.

Unstable A, $\tau = 0.2$, Figure 7: there are two fixed levels of prices. The results can be interpreted as if the seller chose not to process any information about A separately from information about $A\mu$ and he acts as if the problem were a repeated static setup in the original model. Collecting
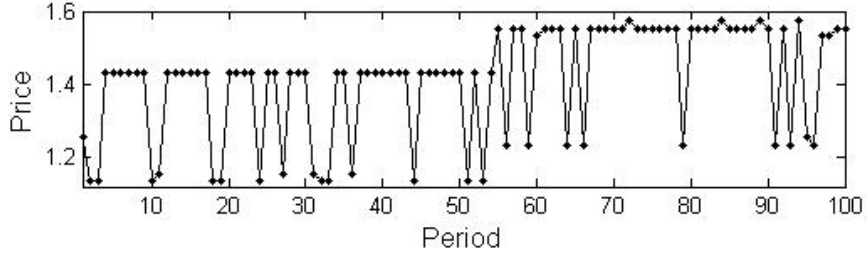
31

Figure 8: Simulated prices for very stable A, $\tau = 0.001$.

information about A would be inefficient since knowledge about A dissipates quickly, there is a 20% probability of a transition in A in each period.

In this model, some shocks can be mistakenly attributed to sources of more transient shocks, which has vast implications analogous to the signal extraction in Lucas (1973). The shock to A is after $t = 20$ attributed to μ, which is more volatile: the positions of the two price points are fixed, which reflects that knowledge about A is unchanged. This is one of the differences from the model in Maćkowiak and Wiederholt (2009), where the disjoint structure of signals implies that the price-setters never misinterpret the sources of the shocks, only their magnitude.

What, however, changes after the shocks to A is that the probability of the top price increases. In other words, the frequency of sales decreases. The seller receives imperfect signals on $A\mu$, which is more likely to be high after the shock. The seller does not adjust the whole price-plan as a response to shocks to A, but the average price does increase.

Very stable A, $\tau = 0.001$, Figure 8: in this simulation, there is a transition of the price plan after period 55. This means that the seller does collect some information specifically about A. The transition's timing varies from one simulation to another.

The transitions between price plans and thus forms of knowledge about A is immediate, which is typical any time τ is very low. There are two forces that affect attention allocation. First, the low probability of shocks to A make attention to A wasteful. Second, the persistence of the new state, i.e. the low probability of a follow-up shock, makes it beneficial to learn about A, since the knowledge is expected to be useful for several periods to come.

At first, the first force dominates, but once the seller starts realizing that A might have switched, this force weakens and he endogenously allocates more of his information capacity specifically to A, and his knowledge shifts quickly.
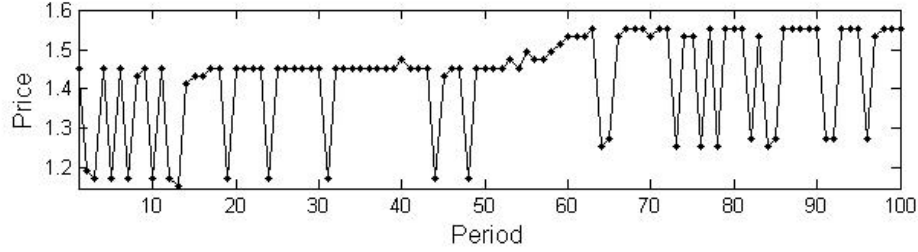
32

Figure 9: Simulated prices for semi-stable A, $\tau = 0.02$.

The jumpy adjustment of knowledge in period 55 in Figure 8 resembles the one often assumed in the literature on information frictions in the form of infrequent or costly reviews of economic conditions (Mankiw and Reis, 2002; Reis, 2006; Alvarez et al., 2011). While there the agents' information updating is perfect but infrequent, here the agents process information continually and imperfectly, only the realization of signals is discrete.

Nakamura and Steinsson (2008) argue that most high-frequency price movements may in fact be orthogonal to changes in the aggregate conditions. Similarly, Eichenbaum et al. (2011) find that while prices change frequently, the value of the reference price stays quite rigid. They claim that it is the rigidity of the reference price, not single prices, that is the useful statistics for macroeconomic analysis.

The presented model generates pricing patterns that in part explain such findings. The dynamics of the reference in the model does indeed mimic the seller's current knowledge about the state of the slowly moving index A. The current knowledge of the slowly moving variables determines the selection of a price plan together with its reference price. While the high-frequency price movements are predominantly driven by the highly volatile components of $\mathbf{x}$ only, their frequency is affected by the persistent components, too. Aggregate shocks thus affect the frequency of sales even when the reference price is unchanged.

Semi-stable A, $\tau = 0.02$, Figure 9: the price plan's adjustment is gradual, which implies continuously changing prices. I find this to happen in some medium range of τ. Here, the seller devotes less of his information capacity specifically to A even after his knowledge about A becomes less sharp. The flexibility is only in shifting the plan by an amount caused by the persistent aggregate shock, while no new price points arise within the plan. When the persistent shock is small in magnitude, then the range of flexibility is small, too. Below, in Section 5.3, I show that
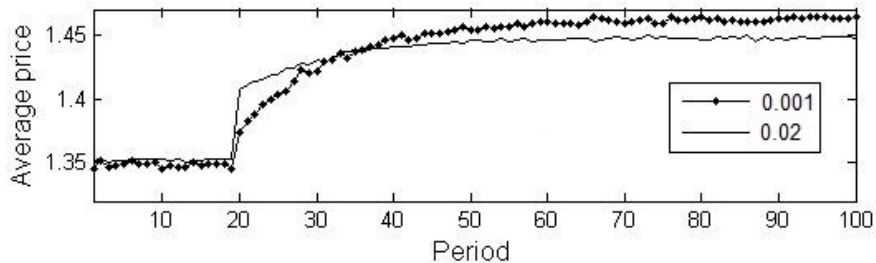
33

Figure 10: Average prices for stable and semi-stable A, $\tau = 0.001$ and $\tau = 0.02$.

an introduction of a tiny adjustment cost makes the transition discrete, as for stable A in Figure 7.

Average prices: Figure 10 presents the responses of prices averaged over 10,000 runs for $\tau = 0.001$ and $\tau = 0.02$. This exhibit is supposed to model the responses of the aggregate price level to an aggregate shock to the stable variable A. While the individual prices change frequently, the price aggregate responds very slowly to the shock in the persistent variable. It takes about one year for the response of average prices to achieve two-thirds of the full response.

The response is swifter when A is less stable. This is an effect that is in a slightly different form present in Maćkowiak and Wiederholt (2009), too. While in Maćkowiak and Wiederholt (2009) the price-setter pays more attention to the less stable A, here he updates his knowledge about A even if he pays attention directly to the product $A\mu$. If A is less stable, then more shocks are attributed to A.[26] The range of adjustment is, on the other hand, lower for less stable A. This is since the benefit of finding out about A precisely is lower if A changes often.

5.3 Epsilon menu cost

The previous sections showed that the adjustment of price plans is often discrete, but not always. In this section I briefly discuss that even in the remaining cases, such as in Figure 9, the gradual adjustment of price plans is likely to disappear completely once a very small explicit adjustment cost is introduced into the model.

Figure 11 presents a simulated price series for semi-stable A, $\tau = 0.02$, with a seemingly

---

[26]The described interdependence of shocks, knowledge, and price responses rather resembles the implications of the signal extraction model in Lucas (1972), but some implications of rational inattention and signal extraction go in the opposite direction; see Sims (2003). In signal extraction, highly volatile variables are tracked with the least amount of error, while in rational inattention it is the stable variables that are easy to follow.
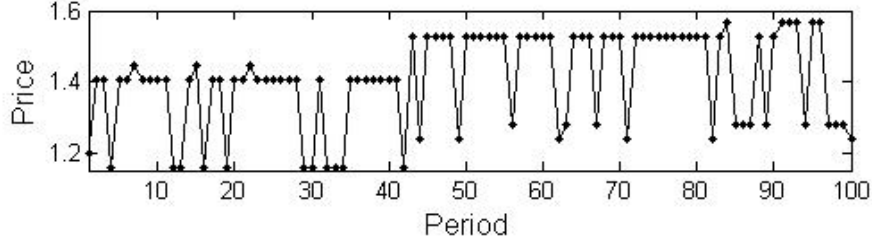
Figure 11: Simulated prices for semi-stable A, $\tau = 0.02$ and menu cost of 0.05% of revenue.

negligible menu cost for changing the charged price. I used the menu cost of 0.05% of average weekly revenue, which is almost two orders of magnitude lower than what is used in standard menu cost models. The menu cost is so low that absent the information friction the price would in each period adjust with a probability of almost 1. The adjustment of a price plan, and thus also of knowledge about A, is discrete.

The reason that such a tiny adjustment cost enforces discrete adjustment is because of the multidimensional flexibility in choosing the price plan, and because of the endogeneity of information processing.

Rational inattention implies that in each period prices are drawn from a finite set of values. Each price plan is defined by locations of a finite number of prices as well as by their frequencies. The error caused by fixing one price, which has a second order effect on profit, is partially mitigated by the possible adjustment of frequencies within the plan.

Let the last period's price be p_{t-1}. If the knowledge about A changes a little, then this period's optimal price plan may include the value $(p_{t-1} + \epsilon)$ instead. When $|\epsilon|$ is very small, as it would be when periods are short and the adjustment of knowledge is gradual, as in Figure 9, then the menu cost may suffice to select the current price plan such that p_{t-1} is included in the price plan instead of $(p_{t-1} + \epsilon)$. The small menu cost may thus perhaps stabilize one of the plan's values, when the knowledge changes a little. But how can the cost stabilize the value once A shifts more, and how does it stabilize the whole price plan?

Once the seller knows he would not respond to small changes in the profit-maximizing prices, then he selects his strategy of processing information such that he does not collect signals that would lead to small changes of knowledge about A. This implies that a small menu cost fixing one price stabilizes the price values of the whole price plan, only affecting the prices' frequencies.

35

In particular, the channel of signal extraction from processing information about $A\mu$ is weakened significantly. Furthermore, imperfect endogenous information motivates the seller to collect signals that would lead to larger changes only, as information on small changes can be populated by noise in information processing.

6   Conclusion

This paper shows that rational inattention can account for several empirical findings about nominal rigidities. Most importantly, I find that some stylized facts at the micro level that constitute puzzles for sticky-price models can be accounted for by information frictions in the form of rational inattention. Rational inattention can drive sluggish responses to new shocks. Moreover, the rationally inattentive seller chooses to follow a price plan with a finite number of distinct prices.

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

References

Aguirregabiria, V. (1999). The dynamics of markups and inventories in retail firms. *Review of Economic Studies*, 66:275–308.

Alvarez, F. E., Lippi, F., and Paciello, L. (2011). Optimal price setting with observation and menu costs. *Quarterly Journal of Economics*, 126(4).

Alvarez, L. J., Burriel, P., and Hernando, I. (2005). Do decreasing hazard functions for price changes make any sense? Working Paper Series 461, European Central Bank.

Barberis, N. and Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, 68(2):161–199.

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

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

Day, G. S. and Nedungadi, P. (1994). Managerial representations of competitive advantage. *The Journal of Marketing*, pages 31–44.

Eichenbaum, M., Jaimovich, N., and Rebelo, S. (2011). Reference prices, costs, and nominal rigidities. *American Economic Review*, 101(1):234–62.

Guimaraes, B. and Sheedy, K. D. (2011). Sales and monetary policy. *American Economic Review*, 101(2):844–76.

Kehoe, P. and Midrigan, V. (2014). Prices are sticky after all. *Journal of Monetary Economics*.

Kiesler, S. and Sproull, L. (1982). Managerial response to changing environments: Perspectives on problem sensing from social cognition. *Administrative Science Quarterly*, pages 548–570.

Lazear, E. P. (1986). Retail pricing and clearance sales. *American Economic Review*, 76:14–32.

Levitt, B. and March, J. G. (1988). Organizational learning. *Annual review of sociology*, pages 319–340.

Lucas, Robert E, J. (1973). Some international evidence on output-inflation tradeoffs. *American Economic Review*, 63(3):326–34.

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

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

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

Maćkowiak, B. A. and Wiederholt, M. (2015). Business cycle dynamics under rational inattention. *Review of Economic Studies*.

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

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.

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

Mintzberg, H. (1973). The nature of managerial work. *New York.*

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

Nakamura, E. (2008). Pass-through in retail and wholesale. *American Economic Review*, 98(2):430–37.

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

O'Reilly, C. A. (1980). Individuals and information overload in organizations: is more necessarily better? *Academy of management journal*, 23(4):684–696.

Reis, R. (2006). Inattentive producers. *The Review of Economic Studies*, 73(3):793–821.

Roach, E. and Lloyd, B. B. (1978). *Cognition and categorization.* Hillsdale, New Jersey.

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

Simon, H. A. (1976). *Administrative behavior*, volume 3. Cambridge Univ Press.

Simon, H. A. (1979). Information processing models of cognition. *Annual review of psychology*, 30(1):363–396.

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

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

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

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

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

Varian, H. R. (1980). A model of sales. *The American Economic Review*, 70(4):pp. 651–659.

Widder, D. (1951). Necessary and sufficient conditions for the representation of a function by a weierstrass transform. *Transactions of the American Mathematical Society*, 71(3).

Wilson, R. A., Keil, F. C., et al. (1999). *The MIT encyclopedia of the cognitive sciences*, volume 134. MIT Press.

A   Formulation of the model

A.1   Information processing

This part establishes what forms of knowledge are achievable by a seller who processes a given amount of information. Let us recall that the seller collects information in sequences of small pieces, perhaps from different sources, and about many things, among which are those that are important for the pricing decision of the one product. A "given amount of information" means a given number of elementary pieces of information.

The rational inattention framework applies the results of information theory, which was introduced in Shannon (1948). Information theory provides an understanding of what information can be passed through so-called noisy channels that transmit blocks of symbols at a limited rate. In our setting, we do not apply these results to communication, but reinterpret them and study achievable information processing by a seller who has limited attention.

Before the seller starts processing information about the realized random variable $\mathbf{x}$, he possesses some prior knowledge about it given by a pdf $g(\mathbf{x})$. The aim is to express a constraint on the forms of achievable posterior knowledge, i.e., express what the seller can learn. Any information processing mechanism can be formally defined by the resulting $f(s|\mathbf{x})$, a distribution of "signals" s received conditional on a realized state $\mathbf{x}$. The signal s is a cumulative contribution of the many pieces of

information the seller collects. $f(s|\mathbf{x})$ together with the prior $g(\mathbf{x})$ form the posterior knowledge $f(\mathbf{x}|s)$ through Bayes law:

$$f(\mathbf{x}|s) = \frac{f(s|\mathbf{x})g(\mathbf{x})}{\int f(s|\mathbf{x})g(\mathbf{x})d\mathbf{x}}. \tag{25}$$

Entropy and mutual information One possible measure of information content carried by a signal s is the expected reduction of entropy about $\mathbf{x}$ due to an observation of a realization of s. Entropy is a measure of uncertainty, and its expected reduction is called the "mutual information" I between the random variable $\mathbf{x}$ and the signal s. In particular, we define it by

$$
\begin{aligned}
I[f(\mathbf{x}, s)] &= H\left[g(\mathbf{x})\right] - \mathbb{E}_s H\left[f(\mathbf{x}|s)\right] \\
&= \int f(\mathbf{x}, s) \log\left(\frac{f(\mathbf{x}, s)}{g(\mathbf{x})f(s)}\right) d\mathbf{x}ds,
\end{aligned}
\tag{26}
$$

where $H\left[f(\mathbf{x}|s)\right]$ is the entropy of the posterior distribution conditional on the observation of s, and the joint pdf $f(\mathbf{x}, s) = f(s|\mathbf{x})g(\mathbf{x})$.[27]

In information theory, mutual information emerges as the fundamental measure of information flow. Below, we discuss why this is the case, and conclude that it provides a useful constraint on the information flow in our setting.

Noisy information channels Information theory asks what information can be transmitted over a noisy channel. A noisy channel is a device that transfers an input random variable $z^I \in Z^I$ into an output random variable $z^O \in Z^O$ with a given probability transition matrix $p(z^O|z^I)$. For instance, the channel can be a wire transmitting 0's and 1's with some noise, or a device transmitting Gaussian signals. The question now is: when can such a given channel transmit a random state $\mathbf{x} \in X$ (encoded in a sequence of z^I) such that a signal $s \in S$ (reconstructed from a sequence of z^O) is received with a desired probability $f(s|\mathbf{x})$? In information theory, the agent's information processing mechanism is determined by the channel and the coding he uses. While the channel is defined by $p(z^O|z^I)$, coding provides meaning to the input z^I, i.e., it is a mapping $X \to \{Z^I\}_i$ describing how the state is encoded into the input alphabet. In our setting, the channel will describe the seller's ability to comprehending signals, i.e., how many pieces of data he can comprehend in a unit of time, and coding describes what the seller pays attention to, i.e., what the elementary

---

[27]To be mathematically precise, one would use probability measures instead of pdfs; pdfs are merely chosen to facilitate the exposition.

signals are about. For instance, it is the meaning of a particular digit the seller looks at, what variable it belongs to, or what questions he asks.

The coding theorem   The coding theorem (see Shannon (1948) and Cover and Thomas (2006)), which is one of the cornerstones of information theory, states the sufficient and necessary conditions for what information can be passed through an information channel. The theorem says that any information channel can transmit any message with an information content less than the channel's capacity. The agent thus only needs to choose a code $X \to \{Z^I\}_i$ optimally, i.e., the seller needs to choose the proper sequences of pieces of data to look at; this connects well with our particular application. If the channel's information capacity per period is κ, then $f(s|\mathbf{x})$ is achievable with arbitrary precision if and only if

$$I[f(\mathbf{x}, s)] \leq \kappa. \tag{27}$$

The coding theorem holds when the agent obtains information about many inputs and can encode it into a large number of imprecise pieces of information.[28] The channel's information capacity κ equals the mutual information of $p(z^I, z^O) \equiv p(z^I)p(z^O|p^I)$ maximized over $p(z^I)$. It is a fixed quantity given by the channel's properties and is independent of what message is to be transmitted.

One extremely useful implication of the coding theorem is the fact that we do not need to specify the elementary information constraints the seller faces, i.e., what $p(z^o|z^i)$ is. The coding theorem thus allows us to characterize the information processing constraints abstractly by a single quantity, the information capacity κ. Therefore, we can model the sequential problem of information acquisition by sidestepping inputs and outputs (z^I, z^O) completely; in the case of the seller, this means that we do not need to specify these objects, i.e., exactly how the seller collects and codes signals. We can instead thus work with $f(\mathbf{x}, s)$ constrained only by a given level of mutual information. What emerges is the mutual information constraint that is independent of the form of elementary pieces of information that the seller collects as long as those pieces are small relative to all information that the seller collects (including the information for other decisions the seller makes).

The seller in our model thus does exactly what the coding theorem describes. In our setting,

---

[28]In other cases, when the number is small, then while the agent can also choose very coarse summarizing signals, the entropy constraint holds approximately.

the seller's choice coding is performed by the choice of what pieces of information to look at and what mental representations to make of them.

Summary:

Choosing how to process information is equivalent to selecting the conditional distribution $f(s|\mathbf{x})$. $f(s|\mathbf{x})$ together with the prior knowledge about the random components $g(\mathbf{x})$ form $f(\mathbf{x}, s) = f(s|\mathbf{x})g(\mathbf{x})$, which in turn determines the posterior knowledge $f(\mathbf{x}|s)$ conditional on any received signal s. The seller observes a signal s, which is a cumulative contribution of many elementary pieces of information, and then infers $\mathbf{x}$. Typically, he would like to receive signals generating posteriors $f(\mathbf{x}|s)$ that are Dirac delta functions: that would represent perfect knowledge about $\mathbf{x}$. However, the channel's information capacity limits how tight the signals can be on average. Using the coding theorem, we know that any $f(s|\mathbf{x})$ is achievable if and only if the average reduction of the entropy of the seller's knowledge in one period does not exceed the seller's information capacity κ.

A.2 Putting together the choices of how to process information and what price to set

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

1. There exists a sequence of elementary pieces of information for any $f(s|\mathbf{x})$ that the seller chooses to achieve, as long as it satisfies (27).

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

$$\tilde{P}(s) = \arg\max_p \int \Pi(\mathbf{x}, p) f(\mathbf{x}|s) d\mathbf{x}, \tag{28}$$

   where the posterior knowledge $f(\mathbf{x}|s)$ is given by Bayes law (25).

The two decisions, 1. and 2., are not independent. While deciding the optimal mechanism of processing information, $f(s|\mathbf{x})$, the agent is aware of his policy function, $\tilde{P}(s)$. Choosing how to

process information thus takes the following form:

$$f(s|\mathbf{x}) = \arg\max_{\hat{f}(S|\mathbf{x})} E[\Pi] = \arg\max_{\hat{f}(s|\mathbf{x})} \int \Pi\Big(\mathbf{x}, \tilde{P}(s)\Big) \hat{f}(s|\mathbf{x}) g(\mathbf{x}) d\mathbf{x} ds, \qquad (29)$$

subject to (27) and (28). Notice that $f(s|\mathbf{x})g(\mathbf{x})$ is the perceived probability density of a realization of $\Pi\Big(\mathbf{x}, \tilde{P}(s)\Big)$.

The profit function is a single-peaked function of p; there is only one profit-maximizing price. Moreover, since entropy is a concave function of the distribution, it is never optimal to choose a collection with two posteriors of different forms resulting in the same profit-maximizing price. An averaged distribution of the two posteriors would necessarily lead to the same expected profit but with a lower utilization of the information capacity. We can therefore equivalently describe the information processing mechanism in terms of the resulting distribution of prices that the received signals lead to, instead of in terms of the distribution of signals. We thus substitute $f(s|\mathbf{x})$ by $f(p|\mathbf{x})$, where $p = \tilde{P}(s)$. The whole optimization problem can be formulated in terms of a joint distribution of $\mathbf{x}$ and p; see Definition 1.

B First order condition

The first order condition, equation (9), is derived here. The mutual information can be written as

$$I(\mathbf{x}; p) = \int f(\mathbf{x}, p) \log \frac{f(\mathbf{x}|p)}{\int f(p_1) f(\mathbf{x}|p_1) dp_1} d\mathbf{x} dp. \qquad (30)$$

The Lagrangian of (3)-(6) is:

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

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

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

If $f(p) > 0$ and $\lambda > 0$ we obtain the first order condition (9):

$$f(\mathbf{x}|p) = h(\mathbf{x}) e^{\Pi(\mathbf{x}, p)/\lambda}, \qquad (31)$$

43

where $h(\mathbf{x}) = e^{-\nu(\mathbf{x})/\lambda}g(\mathbf{x})$. Now, since $f(\mathbf{x}, p) = f(\mathbf{x}|p)f(p)dp = g$, from (4) we get

$$h(\mathbf{x}) = g(\mathbf{x})/\int f(\hat{p})e^{\Pi(\mathbf{x},\hat{p})/\lambda}d\hat{p},$$

which together with (31) implies (9). □.

C Discreteness

Proof of Proposition 2:

Let us first transform the setup into a quadratic tracking problem that has the same distribution of prices.

$$\Pi'(\mathbf{x}, p) = \Pi'\Big(\mathbf{x}, p_{opt}(\mathbf{x})\Big) + \Big(\Pi'(\mathbf{x}, p) - \Pi'\Big(\mathbf{x}, p_{opt}(\mathbf{x})\Big)\Big). \tag{32}$$

The first term on the right-hand side is independent of the charged price p; it does not enter the decision problem. The second term is a contribution from a suboptimal selection of p:

$$\Pi''(\mathbf{x}, p) = \Pi'(\mathbf{x}, p) - \Pi'\Big(\mathbf{x}, p_{opt}(\mathbf{x})\Big) = -a(p_{opt}(\mathbf{x}) - p)^2. \tag{33}$$

Let us perform a substitution of variables: $\mathbf{x} \to x$, s.t. $x = p_{opt}(\mathbf{x})$, therefore $g(\mathbf{x}) \to \hat{g}(x)$. The original problem is transformed into a problem with the objective $U(x, p) = -a(x - p)^2$ and the distribution of profit-maximizing prices $g(x)$, while the distribution of p is the same as in the original problem.

Let us recall that Hermite polynomials can be defined in the following way:

$$H_n(x) = (-1)^n e^{x^2}\frac{d^n e^{-x^2}}{dx^n} \qquad \forall n \geq 0, \tag{34}$$

while Hermite functions are:

$$\Psi_n(x) = \frac{1}{\sqrt{n!2^n\sqrt{\pi}}}H_n(x)e^{-x^2/2}, \qquad \forall n \geq 0. \tag{35}$$

We will use the fact that Hermite functions $\{\Psi_n(x)\}_{n=0}^{\infty}$ form an orthonormal basis of $L^2(\mathbb{R})$.

Now, we show by contradiction that the set of price points p such that $f(p) > 0$ must be discrete. If $f(p) > 0$, then the first order condition (9) holds. Since $f(x|p)$ is a pdf, it integrates to 1,

$$\int h(x)\exp\Big(-\frac{a(p-x)^2}{\lambda}\Big)dx = 1. \tag{36}$$

44

If the distribution of prices is not discrete, then there exists a limit point p^* of points, where (36) holds. Therefore, all derivatives of the left-hand side of (36) with respect to p have to equal zero at p^*.

$$\frac{d^n}{dp^n} \int h(x) \exp\left(-\frac{a(p-x)^2}{\lambda}\right) dx \mid_{p=p^*} = 0 \qquad \forall n > 0. \tag{37}$$

Under the following substitution of variables: $x' = \sqrt{a/\lambda}x$, $p' = \sqrt{a/\lambda}p$ and $\xi(x') = h(\sqrt{\lambda/a}x') \exp(-(p^{*'} - x')^2)$, equations (36) and (37) take the following forms:

$$\int \xi(x')\Psi_0(x' - p^{*'})dx' = 1, \tag{38}$$

$$\int \xi(x')\Psi_n(x' - p^{*'})dx' = 0, \qquad \forall n > 0. \tag{39}$$

Since $h(x) \in L^\infty(\mathbb{R})$, then $\xi(x') \in L^2(\mathbb{R})$, therefore, we can expand $\xi(x')$ in the series of Hermite functions centered at $p^{*'} = \sqrt{a/\lambda}p^*$.

$$\xi(x') = \sum_i \xi_i \Psi_i(x' - p^{*'}), \tag{40}$$

where ξ_i are the series' coefficients. Therefore, (39) implies

$$\int \sum_i \xi_i \Psi_i(x' - p^{*'})\Psi_n(x' - p^{*'})dx' = \xi_n = 0 \qquad \forall n > 0,$$

while $\xi_0 = 1$ from (38). $h(x)$ equals a positive constant c on $\mathbb{R}$ anytime there exists such a limit point of prices p^*. Therefore, the following has to hold.

$$\hat{g}(x) = c \int f(p) \exp\left(-\frac{a(p-x)^2}{\lambda}\right) dp, \tag{41}$$

which cannot hold for instance when $g(x)$ has bounded support. $\qquad \qquad \square$

D   Hazard Function

Proof of Proposition 4: The probability of a price change in period $T > 1$, $H(T)$, equals $1 - \sum_i f_i \mathbf{P}(i|T-1)$, where $\mathbf{P}(i|T-1)$ is the probability that the price is in state i conditional on the fact it survived $T-1$ periods, which can be found by a simple application of Bayes law:

$$\mathbf{P}(i|T-1) = \frac{f_i^{T-1}}{\sum_{j=1}^N f_j^{T-1}}. \tag{42}$$

Therefore,

$$H(T) = 1 - \frac{\sum_{i=1}^{N} f_i^T}{\sum_{i=1}^{N} f_i^{T-1}}. \tag{43}$$

Now, show that $H(T)$ is non-increasing. Let us start with the following inequality,

$$2f_i f_j \leq f_i^2 + f_j^2, \qquad \forall i \neq j, \tag{44}$$

where the strict inequality holds any time $f_i \neq f_j$. Multiplying (44) by $f_i^{T-1} f_j^{T-1}$ yields

$$2f_i^T f_j^T \leq f_i^{T+1} f_j^{T-1} + f_i^{T-1} f_j^{T+1}, \qquad \forall i \neq j. \tag{45}$$

Summing up (45) over all $i \neq j$ and adding $\sum_i f_i^{2T}$ to both sides leads to:

$$\left(\sum f_i^T \right)^2 \leq \left(\sum f_i^{T+1} \right)^2 \left(\sum f_i^{T-1} \right)^2, \tag{46}$$

which implies $H(T) \geq H(T+1)$, and the equality holds only for uniform $\{f_i\}_{i=1}^N$. $\square$