Paralyzed by Fear: Rigid and Discrete Pricing under Demand Uncertainty *

Cosmin Ilut — Duke University & NBER

Rosen Valchev — Boston College

Nicolas Vincent — HEC Montreal

March 2016

Abstract

Price rigidity is central to many predictions of modern macroeconomic models, yet, standard models are at odds with certain robust empirical facts from micro price datasets. We propose a new, parsimonious theory of price rigidity, built around the idea of demand uncertainty, that is consistent with a number of salient micro facts. In the model, the monopolistic firm faces Knightian uncertainty about its competitive environment, which has two key implications. First, the firm is uncertain about the shape of its demand function, and learns about it from past observations of quantities sold. This leads to kinks in the expected profit function at previously observed prices, which act as endogenous costs of changing prices and generate price stickiness and a discrete price distribution. Second, the firm is uncertain about how aggregate prices relate to the prices of its direct competitors, and the resulting robust pricing decision makes our rigidity nominal in nature.

*Email addresses: Ilut cosmin.ilut@duke.edu, Valchev: valchev@bc.edu, Vincent: nicolas.vincent@hec.ca.

We would like to thank Etienne Gagnon, Han Hong, Nir Jaimovich, John Leahy, Virgiliu Midrigan, Martin Schneider and Luminita Stevens as well as seminar and conference participants at the Workshop on "Ambiguity and Robustness in Macroeconomics and Finance", on "Ambiguity and its Implications in Finance and Macroeconomics", "Behavioral Aspects of Macroeconomics and Finance" Conference, Boston Macro Juniors Workshop, Duke, ESWC, NBER Summer Institute, NBER EFG meeting, Northwestern, SED, Stanford, UC San Diego, U Penn for helpful discussions and comments.

1 Introduction

Macroeconomists have long recognized the crucial role played by the speed of adjustment of prices in the amplification and propagation of macroeconomic shocks. In particular, there is ample evidence that inflation responds only slowly to monetary shocks (e.g. Christiano et al. (2005)). In an attempt to better understand the price adjustment frictions underpinning these aggregate findings, numerous studies have turned their attention to micro-level price datasets and have extracted a variety of additional salient pricing facts. In this paper, we propose a parsimonious new theory of price rigidity that revolves around a simple reality faced by firms: the demand for their product is uncertain and potentially complex. Coupled with ambiguity aversion, this single mechanism endogenously creates a cost of moving to a new price. Not only does the model naturally generate sticky prices, but its parsimony also yields a number of overidentifying restrictions that are consistent with pricing facts from micro data.

One of the earliest documented empirical findings in the micro price literature is that prices at the product level tend to be sticky, that is do not change for long periods of time (Bils and Klenow (2004)). Yet, if one plausibly believes that firms are regularly hit by demand and cost shocks, in turn altering the profit-maximizing price, then firms would be expected to update posted prices more often.[1] This robust stylized fact led to the widespread use of both time-dependent (e.g. Calvo (1983), Taylor (1980)) and state-dependent (e.g. menu cost) price rigidity mechanisms. However other facts, such as the surprising coarseness and stickiness of the set of prices chosen by firms over time (Eichenbaum et al. (2011)), are more difficult to generate without expanding the standard models.

In our framework, the economy is composed of a continuum of industries, each populated with monopolistic firms that face Knightian uncertainty about their competitive environment. In particular, an intermediate good firm does not know the production function of the final good of its respective industry, which leads to two important implications. First, there is uncertainty about the shape of the demand function the firm faces, and second, there is uncertainty about the relevant relative price, and how it relates to the aggregate price index.

Firms understand that the quantity sold is the sum of an unknown, time-invariant component, and a temporary demand shock. They use their observations of past prices and quantities to learn about the time-invariant component, but cannot observe the two components separately, only the total quantity sold, and thus face a signal extraction problem. Furthermore, firms are not confident that demand belongs to a single parametric

[1] Eichenbaum et al. (2011), for example, argue that the large fluctuations in quantities sold in weekly grocery store data in the absence of any price change are indicative of sizable demand shocks.

family, but rather entertain potentially complex demand shapes. Firms use their noisy signals to reduce uncertainty and build estimates of their demand curves. We thus put the economic agent on the same footing as an econometrician outside the model that attempts to estimate demand in a complex environment.[2]

We assume that the firm has enough prior knowledge to put some loose prior bounds on the possible demand schedules, but not enough to impose functional form restrictions or to assign a single probability measure to the space of admissible demand functions. Thus, the firm faces Knightian uncertainty about the shape of its demand function. The agent owning the firm is ambiguity averse in the sense that it acts *as if* the true distribution of the demand at a given price yields the lowest possible total quantity sold. Ambiguity aversion is described by recursive multiple priors preferences, axiomatized in Epstein and Schneider (2003), that capture the agents' lack of confidence in probability assessments.

Since demand is not restricted to a particular parametric family, uncertainty reduction is local, not global. Unlike updating beliefs about the parameters of a given function, by observing a noisy demand signal at a given posted price, the firm primarily reduces demand uncertainty at that price, but remains uncertain about the quantity it could sell at other prices. This generates kinks in demand uncertainty at previously observed prices, and an uncertainty averse price-setter is reluctant to move to a new price since it would lead to a sharp rise in uncertainty.

For our ambiguity-averse firms, the kinks show up in expected demand. A firm that entertains switching to a higher price is worried that demand becomes more elastic in the region above its current information set, maybe because a price increase could trigger an exodus towards competing products. At the same time, the higher uncertainty at lower prices generates the opposite fear that demand is in fact more inelastic in that region, and a price cut might undermine profit margins without increasing sales much. This endogenous switch in the worst-case scenario about the demand schedule, depending on whether the firm is considering a price increase or decrease, leads to kinks in expected demand, which in turn generate price stickiness. The kinks create a cost, in terms of expected profits, associated with changing the price, which in turn compels the firm to abstain from changing its price, unless it faces a sufficiently large shock. The higher is the uncertainty in the unexplored regions of the price space, relative to the uncertainty at previously observed prices, the steeper are the kinks in expected demand and the stronger is the stickiness.

A corollary implication is that the firm is not only reluctant to change its current price,

[2]The equal footing between the uncertainty faced by agents inside the model and econometricians outside the model addresses a desideratum proposed in Hansen (2007) for time-series models and more generally in Hansen (2014).

but is in general inclined to repeat a price it has already posted in the recent past. These previously observed prices become ‘reference’ points at which there are kinks in the profit function. The pricing policy function then includes step-like regions of flatness around the reference prices. When a shock moves the optimal price within such a flat area, the posted price will be exactly equal to one of these reference prices. The steps in the policy function also imply that each of those reference prices is associated with a positive measure of shocks that map to it. Thus, the model is consistent with the optimal policy having ‘price memory’, characterized by discrete price changes between a set of previously posted prices.

Moreover, since signals are noisy, the uncertainty across the previously posted prices is not equal. Prices that have been observed more frequently have accumulated more signals and thus greater uncertainty reduction. Hence, optimal prices would not necessarily bounce randomly around the set of ‘reference prices’, but will exhibit a greater propensity to stay put and return to prices that have been observed more often. Among other things, this has the implication of endogenously generating a decreasing hazard of price change. Lastly, since not all kinks are necessarily deep, the policy function is not exclusively a step-function, but has regions in which the optimal price adjusts flexibly. Thus, the price series of this model can look both flexible and sticky at the same time, and the unconditional distribution of price changes features non-trivial density around zero.

Our mechanism has two key modeling ingredients. The first is the uncertainty about the demand *shape*, which makes uncertainty reduction local, and the second is some form of uncertainty aversion – i.e. uncertainty should ultimately matter. We have implemented these ingredients in a model of learning under ambiguity, but qualitatively similar results can be obtained in a model where uncertainty is only in the form of risk. As long as the prior over the admissible demand functions does not rule out non-differentiable functions, observing noisy signals would generate kinks in the posterior variance of demand, which would have a similar effect on pricing decisions under risk-aversion. Intuitively, since risk aversion is a smooth operator, there can be no kinks in the certainty equivalent if the prior rules out non-differentiability. In contrast, with ambiguity we do not need to allow for non-differentiability in the set of admissible functions. Instead, the kink in expected demand arises endogenously, from the switch in the worst-case beliefs.

Fundamentally, this demand uncertainty represents a real rigidity: it does not, in itself, generate money non-neutrality. Nominal rigidity is the result of the interaction of demand uncertainty with the uncertainty about the relevant relative price. The firm does not know the final good technology of its industry, hence it does not know the appropriate industry price level, nor how it relates to the aggregate price and sees that relationship as ambiguous. It conducts periodic marketing reviews that reveal the industry price, but in between reviews the firm updates beliefs using the ambiguous relationship with the observed aggregate prices. Thus, the firm's beliefs about the industry prices are anchored by the value of the last review, and evolve in an ambiguous way with the observed aggregate inflation.

In this context, the firm understands that its demand is uncertain in two dimensions – both the demand function and its argument, the relative price, are ambiguous. The firm chooses an action robust to this two-dimensional uncertainty, and acts *as if* nature draws the true Data Generating Process (DGP) to be the relationship between aggregate prices and industry prices that implies the lowest possible demand, given the non-ambiguous choice of the firm – own nominal price versus the last observed industry price level. The resulting worst-case relationship is that aggregate prices are not informative about industry prices, and this defines a worst-case demand schedule as a function of own nominal price relative the last observed industry price, that the firm can then estimate via the process described above. Since the review signals arrive periodically, the real rigidity created by the perceived kinks in demand becomes a nominal one, as in order to keep the relevant relative price constant, the firm needs to keep nominal prices constant. This results in nominal price paths that are sticky, and also resemble infrequently updated "price plans".

Our setup has stark implications about price-setting behavior. The model's key outcome is that it endogenously produces a cost of adjusting prices in the form of a higher perceived uncertainty away from previously posted prices. This is different from standard models where there is an assumed, exogenous fixed cost of adjustment. Moreover, the single, uncertainty-based mechanism behind this endogenous cost generates many additional features observed in micro price data that have proven challenging, if not impossible, for standard price-setting models to replicate. On one hand, our mechanism is also compatible with the evidence that firms appear to select from a small set of unique prices, and tend to revisit past price levels. On another, because the cost of moving away from a price is negatively related to how much information was gleaned from posting it in the past, it is by nature inherently history and state dependent. As a result, our mechanism not only predicts a decreasing hazard function of price changes (i.e. the probability of observing a price change is decreasing in the time since the last price movement), but it can also rationalize the coexistence of small and large price changes in the data.

The paper is organized as follows. In Section 2, we discuss its relation to the relevant literature. In Section 3 we present motivating empirical evidence. Sections 4 describes a simplified model that studies learning under demand uncertainty, and explains the real rigidity mechanism. Section 5 derives analytical results. Section 6 introduces the full model, and the interaction that generates nominal rigidity. Section 7 presents a quantitative version.

2 Relation to literature

By connecting learning under ambiguity to the problem of a firm setting prices, this paper relates to multiple literature strands. The question of price rigidity has generated a very large empirical and theoretical literature. On the empirical side, the recent analysis on micro-datasets, such as Bils and Klenow (2004), Klenow and Kryvtsov (2008), Nakamura and Steinsson (2008), Klenow and Malin (2010) or Vavra (2014), attempts to uncover stylized pricing facts whose aim is to act as overidentifying restrictions on theoretical models of price rigidity. Of particular motivating interest for us are the empirical findings in Eichenbaum et al. (2011), Kehoe and Midrigan (2014) and Stevens (2014), who find evidence of 'reference prices', i.e. the set of prices chosen by the firm is surprisingly sticky over time.

Our mechanism produces kinks in expected demand and as such is related to theoretical work on real price rigidity based on kinked demand, such as Stigler (1947), Stiglitz (1979), Ball and Romer (1990) and Kimball (1995). While in these models the kinks are a feature of the true demand curve, in our setup they arise only in the beliefs of the firm, as a result of the uncertainty about demand, and an econometrician would not need to find evidence of actual kinks in demand. Moreover, in our model the size and the location of the kinks are endogenous, and are a function of the information accumulated at observed prices.

In terms of theories of nominal stickiness, our mechanism does not rely on any actual impediment to adjusting prices. This distinguishes our contribution from a large literature specifying either a fixed length of a price contract (Taylor (1980)), an exogenous chance of resetting the optimal price (Calvo (1983)), a physical cost of price adjustment (Barro (1972), Rotemberg (1982))[3], or a cost of information acquisition present in more recent models of rational inattention (Woodford (2009)).[4] Instead, our model is based on the firm's uncertainty about demand as a source of what looks like an endogenous cost of changing prices. Moreover, the emerging cost is also time-varying, with properties that are state and history-dependent. It is this dependence that allows our single, parsimonious mechanism to rationalize a set of otherwise puzzling pricing facts, such as price discreteness, memory,

[3]The large "menu cost" literature that followed includes recent contributions such as Golosov and Lucas (2007), Gertler and Leahy (2008), Nakamura and Steinsson (2008, 2010), Alvarez et al. (2011), Midrigan (2011), and Vavra (2014).

[4]Imperfect information models, such as Mankiw and Reis (2002), Sims (2003), Woodford (2003), Reis (2006), Lorenzoni (2009) and Mackowiak and Wiederholt (2009), predict sluggish adjustment to shocks. However, in order to generate nominal prices that are constant for some periods, as we see in the data, they typically require additional nominal rigidities. Bonomo and Carvalho (2004), Nimark (2008) and Knotek and Edward (2010) are early examples of merging information frictions with a physical cost or an exogenous probability of price adjustment. Our model instead not only generates a partial response of a firm's price to a monetary policy shock, but also actual nominal stickiness.

small and large price changes and a decreasing hazard function.[5]

We also relate to theoretical work on firm pricing under demand uncertainty. The standard approach has been to study this uncertainty in the context of an expected utility model and analyze learning about a parametric demand curve. An early contribution is Rothschild (1974), who frames the learning process as a two-arm bandit problem,[6] while more recent work includes Balvers and Cosimano (1990), Bachmann and Moscarini (2011) and Willems (2011). Different from our environment, learning about parametric functions, such as linear demand curves, does not produce kinks from uncertainty reduction since the latter reflects the estimation risk of the whole function.

Lastly, we connect to the literature on ambiguity aversion. We use the multiple priors preferences to capture the notion that the firm is not confident in the probability assessments over various demand curves, and as such we build on previous contributions that include Gilboa and Schmeidler (1989), Dow and Werlang (1992), Pires (2002) and Epstein and Schneider (2003). Some recent work analyzes a firm pricing problem under a related ambiguity-aversion preference, namely maxmin regret (Handel et al. (2013) and Bergemann and Schlag (2011)), but does not analyze learning about the distributions.

3 Empirical motivation

In response to the marked interest of modelers in identifying the most appropriate way to model nominal rigidities, a large empirical literature developed around micro level price datasets. While case studies such as Carlton (1986) and Cecchetti (1986) had given researchers some insights into the extent of price rigidity, their scope was limited, generally focusing on very specific products or markets. In their seminal work, Bils and Klenow (2004) leveraged the broad coverage of the U.S. Bureau of Labor Statistics' consumer price index (CPI) dataset to gain general insights into the dynamics of prices at the micro level. Numerous other studies have followed, producing results from CPI (Nakamura and Steinsson (2008), Klenow and Kryvtsov (2008)) or scanner datasets (Eichenbaum et al. (2011)).

Macroeconomic modelers have made extensive use of the findings from these studies to calibrate or estimate their models. To do so, they have generally relied on a subset of moments, most frequently the frequency and average size of price increases and decreases.

[5]Recent modeling advances address the challenge of obtaining a discrete distribution of prices out of continuous shocks using a combination of physical adjustment costs to regular and sales price (Kehoe and Midrigan (2014)) or information costs (Matějka (2010) and Stevens (2014)). In the latter case, given some restrictions on the curvature in the objective function and the prior uncertainty, the firm chooses a discrete price distribution to economize on the costs of acquiring information about the unobserved states.

[6]See Bergemann and Valimaki (2008) for a survey of related applications of bandit problems studied under expected utility.

One issue from relying on a small number of moments is that researchers have had a very difficult time discriminating between the various price-setting mechanisms that have been put forward in the literature. Yet, there exist a number of robust findings that have received much less attention and remain a challenge for standard price-setting models. In this section, we describe some of them using the IRI Marketing Dataset. It consists of scanner data for the 2001 to 2011 period collected from over 2,000 grocery stores and drugstores in 50 U.S. markets. The products cover a range of almost thirty categories, mainly food and personal care products. A more complete description of the dataset is available in Bronnenberg et al. (2008). For our purposes, we focus on nine markets and six product categories.[7]

We start by highlighting a finding ubiquitous across price datasets: firms appear to favor choosing from a sticky, discrete set of prices even when given a chance to pick a brand new price. For example, the median number of unique prices in a window of 26 weeks (half a year) is only 3. Another way to describe this empirical property is to look at the degree of price memory. To do so, we compute the probability that when a firm resets the price of its product, the new price is one that was visited within the last six months. This statistic is equal to 62% when we consider all price changes. Arguably such a high degree of memory may be due to the tendency of retailers to post similar-sized discounts on a frequent basis. Yet, even when we filter out temporary sales, memory probabilities still range between 31% and 64% across market/category combinations, with a weighted average of 48%.

Another feature is the declining hazard function found in many micro price datasets: the probability of a price change decreases with the time since the last price reset. As highlighted for example by Nakamura and Steinsson (2008) and Campbell and Eden (2014), this characteristic represents a challenge to many popular price-setting mechanisms. Despite the fact that declining hazards can be found across numerous datasets, some have argued that the finding could be a by-product of not taking proper care of heterogeneity: as noted by Klenow and Kryvtsov (2008), "[t]he declining pooled hazards could simply reflect a mix of heterogeneous flat hazards, that is, survivor bias." We find, however, that the declining hazard remains a robust finding in our dataset, even once we aggressively control for heterogeneity. We start by computing the hazard function for each single product in our sample, pooling across retailers within a specific market. Then, we took the median probability of a price change across all products for each duration. We find that the resulting function is clearly downward sloping, as we show in more detail when we compare the data with a quantitative model in Section 7. This downward slope is not only an artifact of

[7] The markets are Atlanta, Boston, Chicago, Dallas, Houston, Los Angeles, New York City, Philadelphia and San Francisco, while the categories are beer, cold cereal, frozen dinner entrees, frozen pizza, salted snacks and yogurt.

temporary discounts: the hazard declines beyond the first few weeks, and the overall slope remains negative even if we focus on regular prices.

Standard state-dependent pricing models tend to predict that firms only reprice when the optimal price change is sufficiently large. Yet, while it is true that the typical price change tends to be large in absolute value, this statistic masks the pervasive coexistence of small and large prices in the data, as documented for example by Klenow and Kryvtsov (2008).[8] We document the same phenomenon in our dataset by computing the fraction of price changes less than 5% and greater than 15% in absolute value, across all products and markets. We find that price changes smaller than 5% in absolute value account for 14% of all price changes, and prices changes larger than 15% for 56% of all price changes. Hence, both small and large price changes are pervasive in the data. Next, we turn to a model whose predictions are consistent with the empirical regularities described above.

4 Analytical Model

In this section we lay out and analyze the key mechanism in a smaller, analytically tractable real model. We present the full, explicitly nominal model in Section 6.

We study the problem of a monopolistic firm that each period sells a single good at price P_t expressed in real terms. Denoting logs by lower-case, the firm's demand is given by:

$$q(p_t) = x(p_t) + z_t, \tag{1}$$

It consists of two components, the price sensitive part $x(p_t)$, and the price-insensitive z_t. Having posted the price P_t, the firm's time t realized profit is:

$$\upsilon_t = (P_t - e^{c_t})\, e^{q(p_t)} \tag{2}$$

where we have assumed a linear cost function, with c_t denoting the time t log marginal cost. Crucially, the firm does not know the functional form of $x(.)$, and has to learn about it from past observations of quantity sold.

The decomposition of demand in (1) serves two purposes. First, it generates a motive for signal extraction. In this respect we assume that the firm only observes total quantity sold, $q(p_t)$, but not the underlying $x(p_t)$ and z_t. Furthermore, we model z_t as iid, and thus past demand realizations $q(p_t)$ are noisy signals about the unknown function $x(p)$.

The second purpose is to differentiate between risk and ambiguity. We model z_t as purely

[8]See also Midrigan (2011) and Campbell and Eden (2014).

risky, and give the firm full confidence it is iid and drawn from a known Gaussian distribution:

$$z_t \sim N(0, \sigma_z^2)$$

On the other hand, the $x(p_t)$ component is ambiguous, meaning that the firm is not fully confident in the distribution from which it has been drawn, and does not have a unique prior over it. Instead, the firm entertains a whole *set* of possible priors, Υ_0, which is defined on the general space of measurable functions and is not restricted to a given parametric family.

Each individual prior in the set Υ_0 is a Gaussian Process distribution, $GP(m(p), K(p, p'))$, with mean function $m(p)$ and covariance function $K(p, p')$. A Gaussian Process distribution is the generalization of the Gaussian distribution to infinite-sized collections of real-valued random variables, and is a convenient choice of a prior for doing Bayesian inference on function spaces. It has the defining feature that any finite subcollection of random variables has a multivariate Gaussian distribution.[9] Thus, for any finite vector of prices $\mathbf{p} = [p_1, ..., p_N]'$, the corresponding vector of demands $x(\mathbf{p})$ is distributed as

$$x(\mathbf{p}) \sim N\left(\begin{bmatrix} m(p_1) \\ \vdots \\ m(p_N) \end{bmatrix}, \begin{bmatrix} K(p_1, p_1) & \dots & K(p_1, p_N) \\ \vdots & \ddots & \vdots \\ K(p_N, p_1) & \dots & K(p_N, p_N) \end{bmatrix}\right)$$

where the mean function $m(\cdot)$ controls the average slope of the underlying functions $x(p)$, and the covariance function $K(\cdot, \cdot)$ controls their smoothness. In other words, this distribution is a cloud of functions dispersed around $m(p)$, according to the covariance function $K(\cdot, \cdot)$.

We model ambiguity by assuming that that all priors have the same covariance function, but possibly different mean functions. In particular, the set Υ_0 is the collection of all Gaussian Process with a fixed covariance function $K(\cdot, \cdot)$, and a continuous mean function that is weakly downward sloping, i.e. $m(p_1) \leq m(p_0)$ for any $p_1 > p_0$, and satisfies

$$m(p) \in [\gamma_l - bp, \gamma_h - bp]. \tag{3}$$

Figure 1 provides an illustration of the set of admissible $m(p)$. The overall interpretation is that the firm has some a-priori information on the true demand, but is not confident in a single probabilistic weighting of the potential demand schedules (i.e. a single prior), nor is it able to restrict attention to a particular parametric family of demand functions.

[9]Intuitively, we can think of a function as an infinite collection of variables, and the GP distribution defines a measure over such infinite length random vectors by defining the distribution of any finite sub-collection.

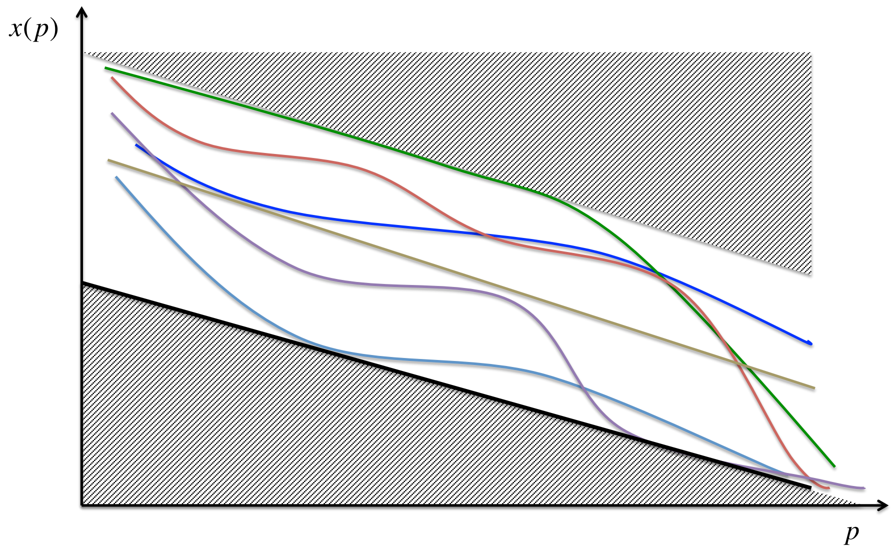


Figure 1: Set of potential $m(p)$ defining the initial set of priors

For the covariance function we specify a simple constant function

$$K(p, p') = \sigma_x^2.$$

The parameter σ_x^2 controls the variance of the GP prior at any given price and thus σ_x^2/σ_z^2 is the signal-to-noise ratio for the demand signals the firm observes. A constant covariance function means that an observation at some particular price p, $q(p)$, is equally informative about the demand function at that p or at some other different price p'. We focus on this case because of its analytical tractability, and because it showcases the minimal complexity of the learning environment that is needed for our main point. The *as if* kinked behavior that will emerge from our analysis does not require kinks in the covariance function or unequal degrees of informational content of the signals about different points on the demand schedule.[10]

Finally, we assume that the true DGP is a standard log-linear demand with no kinks that lies in the middle of the interval for prior mean functions $m(p)$, defined in (3):

$$x^{DGP}(p) = \overline{\gamma} - bp \tag{4}$$

[10]The assumption of constant K implies that there is no probabilistic uncertainty about the shape of $x(p)$, so that signals are equally informative about demand at all prices, and hence probabilistic uncertainty (i.e. the posterior variance) shrinks globally. We shut it down because it is not needed here – the Knightian uncertainty about the shape of $m(p)$ is sufficient. However, our analysis can be extended to more general covariance functions where $K(p, p) \neq K(p, p')$, which would turn on the probabilistic uncertainty about the shape of demand. Lastly, note that in that case we could obtain our main results through risk-aversion alone and without ambiguity, but the mechanism would operate through kinks the posterior variance instead.

with $\overline{\gamma} = \frac{\gamma_l + \gamma_h}{2}$. We also find it useful for analytical and parsimony reasons to parametrize the lower and upper bound of the prior set relative to the true DGP in (4), as

$$\gamma_l = \overline{\gamma} - \nu\sigma_z; \;\; \gamma_h = \overline{\gamma} + \nu\sigma_z \tag{5}$$

4.1 Information and Preferences

The timing of choices and revelation of information is the following. We assume that c_t is known at the end of period $t-1$ and that it follows a Markov process with a conditional distribution $g^c(c_t|c_{t-1})$. The firm enters the beginning of period t with information on the history of all previously sold quantities $q^{t-1} = [q(p_1), ...q(p_{t-1})]'$ and the corresponding prices at which those were observed $p^{t-1} = [p_1, ...p_{t-1}]'$, where a superscript denotes history up to that time. It updates its beliefs about demand conditional on $\varepsilon^{t-1} = \{q^{t-1}, p^{t-1}\}$, observes c_t and posts a price p_t that maximizes its objective, which we further specify below. At the end of period t the idiosyncratic demand shock z_t is realized, and the firm updates its information set with the observed realized quantity sold $q(p_t)$ and marginal cost c_{t+1}.

4.1.1 Learning: prior-by-prior Bayesian updating

The firm uses the available data ε^{t-1} to update the set of initial priors Υ_0. Learning occurs through standard Bayesian updating, but measure-by-measure to account for the initial ambiguity.[11] Thus, for each prior in the inital set Υ_0 the firm uses the new information and Bayes' Rule to obtain a posterior distribution. Given that there is a set of priors, the Bayesian update results in a set of posteriors. As new data is observed, Bayesian updating means that the role of each prior decreases in forming the corresponding posterior.

We denote by $x_{t-1}(p_t)$ the posterior distribution of $x(p_t)$ conditional on end of period $t-1$ information. We denote the conditional mean and variances as:

$$\widehat{x}_{t-1}(p_t; m(p)) := E\left[x(p_t)|\varepsilon^{t-1}; m(p)\right] \tag{6}$$

$$\widehat{\sigma}^2_{t-1}(p_t) := Var\left[x(p_t)|\varepsilon^{t-1}\right] \tag{7}$$

where $m(p)$ is one particular prior on $x(p)$, from the set of priors Υ_0. Thus, conditional on each prior there is a corresponding time t posterior belief about average demand given by

$$x_{t-1}(p_t) \sim N(\widehat{x}_{t-1}(p_t; m(p)), \widehat{\sigma}^2_{t-1}(p_t)) \tag{8}$$

The evolution of beliefs about average demand, $\widehat{x}_{t-1}(p_t, m(p))$, follow the standard

[11] See Jaffray (1994) and Pires (2002) for early axiomatizations of Bayesian updating for multiple priors.

Bayesian updating formulas, as detailed in the Online Appendix A. The analytical derivation is standard and is facilitated by the assumption of Gaussian shocks and the linear state space.

4.1.2 Preferences: recursive multiple priors

The monopolist firm is owned by an agent that is ambiguity-averse and has recursive multiple priors utility[12], so that the value of the firm's profits is defined by the recursion:

$$V\left(\varepsilon^{t-1}, c_t\right) = \max_{p_t} \min_{m(p)} E^{\widehat{x}_{t-1}(p_t;m(p))} \left[\upsilon(\varepsilon_t, c_t) + \beta V\left(\varepsilon^{t-1}, \varepsilon_t, c_{t+1}\right)\right], \tag{9}$$

where $\upsilon(\varepsilon_t, c_t)$ is the per-period profit defined in (2), a function of the beginning-of-period t posted price and end-of-period realized demand $q(p_t)$. The firm forms its conditional expectations and evaluates expected profits and continuation utility using the worst-case conditional expected demand $\widehat{x}_{t-1}(p_t; m^*(p))$, given the available information ε^{t-1} and the prior $m^*(p)$ that achieves that worst-case belief. The maximization step is over the action of what price p_t to post, which affects demand and profit today, but also affects the information set in the future, and hence enters as a state variable for next period's value function.

There are two aspects worth emphasizing about the *min* operator in (9). First, the assumed aversion to ambiguity amounts to minimization over the set of conditional distributions for $x_{t-1}(p_t)$. As detailed by equation (8) the set is formed by updating the set of initial priors Υ_0, measure-by-measure, with the available data ε^{t-1} via Bayes' rule. Because the set of posteriors is indexed by the choice of the initial prior $m(p)$, and in turn this only affects the conditional mean $\widehat{x}_{t-1}(p_t; m(p))$, the minimization problem over the set of posterior distributions becomes equivalent to selecting the worst-case prior. As such, we have stated the preference in (9) as directly minimizing over the initial set of priors.

Second, the minimization is conditional on an entertained choice of p_t. We conjecture that the minimizing belief $m^*(p)$ is such that, for a given price p_t and history ε^{t-1}, it implies the lowest possible expected demand $\widehat{x}_{t-1}(p_t; m^*(p))$ at that price p_t.[13] Thus, for any price p_t, the firm worries that the underlying demand is low, given the data it has seen. The outcome is that the firm maximizes over p_t under the worst-case belief $\widehat{x}_{t-1}(p_t; m^*(p))$.

After solving for the optimal policy rule, including the value function, we can verify the conjecture on $m^*(p)$. In this case, it is sufficient to establish that the profit function $\upsilon(\varepsilon_t, c_t)$ and the continuation utility are both increasing in $x(p_t)$. The former is straightforward by (2). The latter needs to be verified, but it is also intuitive: a higher persistent component of demand increases not only current profits but also future expected profits.

[12] Epstein and Schneider (2003) develop axiomatic foundations for the recursive multiple priors utility.

[13] The worst-case $m^*(p; p_t)$ is conditional on p_t, however, for notational simplicity we simply use $m^*(p)$.

4.2 Kinks from learning

To build intuition for the updating formulas, suppose that the demand history only contains observations of demand at a single price p_0, that has been seen N_0 times. The firm uses the average signal $y_0 = x(p_0) + \frac{1}{N_0}\sum_{i=1}^{N_0} z_i$ to update beliefs about the unknown demand function $x(.)$. For a given prior $m(p)$, the joint distribution of the signal and $x(.)$ at any price p is:

$$\begin{bmatrix} x(p) \\ y_0 \end{bmatrix} \sim N\left(\begin{bmatrix} m(p) \\ m(p_0) \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 + \sigma_z^2/N_0 \end{bmatrix} \right)$$

The conditional distribution $x(p)|y_0$ is also Normal, and its expectation and variance are given by the familiar formulas:

$$E(x(p)|y_0; m(p)) = m(p) + \alpha\left[y_0 - m(p_0)\right] \tag{10}$$

$$Var(x(p)|y_0) = \frac{\sigma_x^2\sigma_z^2/N_0}{\sigma_x^2 + \sigma_z^2/N_0}, \tag{11}$$

where $\alpha = \frac{\sigma_x^2}{\sigma_x^2+\sigma_z^2/N_0}$. Thus, the Bayesian update of the conditional expectation combines the prior for demand at that price, $m(p)$, with the information revealed by the difference between the observed signal y_0, and the prior expected demand at that price, $m(p_0)$.

4.2.1 Worst-case prior

The firm minimizes the conditional expectation of demand over the priors $m(p) \in \Upsilon_0$. Using equation (10), and since $\alpha \in (0, 1)$, it follows that when updating demand at $p = p_0$ the worst case prior is simply the lowest possible $m(p_0)$ – i.e. the lower bound of the set Υ_0:

$$m^*(p_0) = \gamma_l - bp_0$$

When updating demand at a price $p' \neq p_0$, the firm minimizes over both $m(p')$ and $m(p_0)$. The problem can be represented more intuitively as minimizing over the level of demand at p', and the likely change in demand between p' and the observed p_0. We can re-write (10) as:

$$E(x(p')|y_0; m(p)) = \underbrace{(1-\alpha)m(p')}_{\text{Prior demand at } p'} + \underbrace{\alpha(y_0 + m(p') - m(p_0))}_{\text{Signal at } p_0 \,+\, \Delta \text{ in Demand between } p' \text{ and } p_0}$$

The firm's uncertainty about the shape of the demand function implies a lack of confidence in how the information about the level of demand at p_0 translates into information about the level of demand at p'. Clearly, the worst-case prior is that $m(p') = \gamma_l - bp'$, i.e. demand at the considered price p' is low overall. However, the worst-case demand shape depends on whether the firm considers a price or a price decrease.

For a price, $p' > p_0$, $m(p') - m(p_0) \leq 0$ and hence the worst-case is that demand falls a lot between p_0 and p'. The largest possible change in demand is restricted by the initial set of priors Υ_0, and given that $m(p')$ is at the lower bound, the solution is to pick $m(p_0)$ at the upper bound and hence:[14]

$$m^*(p') = \gamma_l - bp'; \; m^*(p_0) = \gamma_h - bp_0 \tag{12}$$

Intuitively, the firm is worried that increasing the price from p_0 to p' would lead it into a particularly elastic part of the demand curve, so that the price increase results in a significant fall in average quantity demanded.

For a price $p' < p_0$, on the other hand, $m(p') - m(p_0) \geq 0$. The firm understands that demand is weakly downward sloping, and hence given a price decrease the worst-case prior is that demand does not change – i.e. the demand curve is inelastic to the left of p_0 and the price cut does not generate an increase in demand. Given the downward sloping restriction on $m(p)$ and the prior set Υ_0, the resulting worst-case prior for $p' < p_0$ is:[15]

$$m^*(p') = \gamma_l - bp'; \; m^*(p_0) = \min(\gamma_l - bp', \gamma_h - bp_0) \tag{13}$$

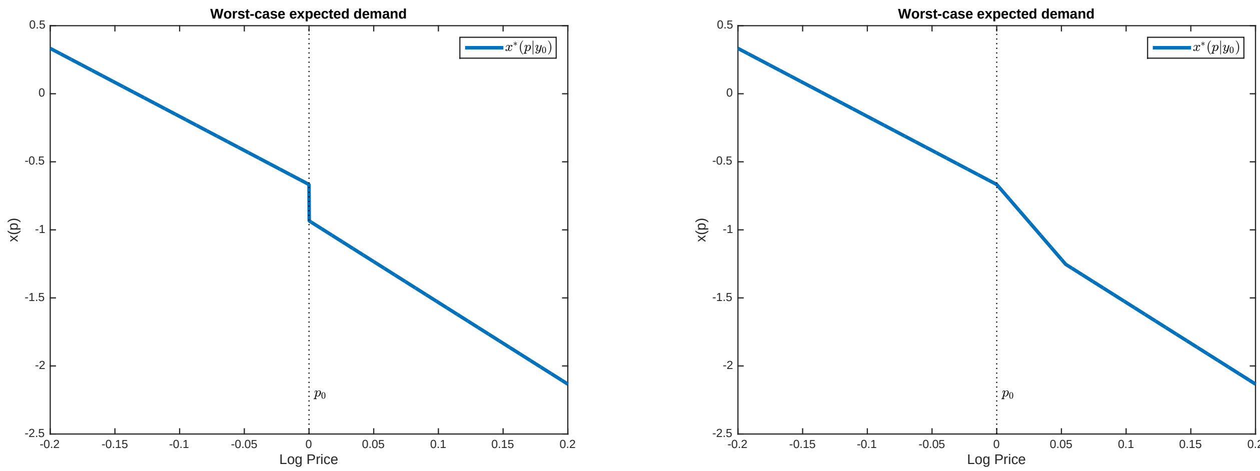
Thus, the worst-case prior when considering a switch to p' is characterized by two features. The firm is concerned that demand at p' is low in general and that it has only changed for the worst from its previously observed price p_0. This leads to an endogenous switch in the worst-case, where the firm worries that demand is relatively elastic when considering a price increase, but worries about the opposite, an inelastic demand, when considering a price decrease. As a result, the firm acts *as if* the prior is locally flat for downward price movements, and *as if* the prior is steep for price increases, generating a kink in its beliefs.

4.2.2 *As if* kinked expected demand

Having characterized the worst-case prior, we can now plug it in equation (10) to obtain the worst-case conditional expectation at any price p'. Since the worst-case prior changes depending on whether p' is above or below p_0, as per equations (12) and (13), the conditional

[14] Because there are no signals observed at other prices, the rest of the prior demand $m^*(p)$ does not enter the conditional mean at p' and as such is not uniquely determined out of the prior set Υ_0.

[15] Since Υ_0 is downward sloping, it could be the case that $\gamma_l - bp'$ is bigger than the upper bound of Υ_0 at p_0, hence the worst-case prior is defined as the minimum of the two.



(a) No restrictions on admissible $m(p)$ (b) Admissible $m(p)$ must satisfy $m'(p) \leq b_{\max}$

Figure 2: Worst-case Expected Demand

expectation becomes the following piecewise function:

$$E(x(p')|y_0; m^*(p)) = \begin{cases} \gamma_l - bp' + \alpha\left[y_0 - (\gamma_h - bp_0)\right] & \text{for } p' < \underline{p} \\ \gamma_l - bp' + \alpha\left[y_0 - (\gamma_l - bp')\right] & \text{for } p' \in [\underline{p}, p_0] \\ \gamma_l - bp' + \alpha\left[y_0 - (\gamma_h - bp_0)\right] & \text{for } p' > p_0 \end{cases} \tag{14}$$

where the first line shows the case when p' is sufficiently lower than p_0, so that $\gamma_l - bp' > \gamma_h - bp_0$. The critical value at which this inequality flips is $\underline{p}$, defined as $\gamma_l - b\underline{p} = \gamma_h - bp_0$.

Thus, the multiple priors endogenously generate a kink in expected demand at the price p_0, even though there is no kink in the DGP. Intuitively, this happens because the worst-case depends on the considered price p', and it switches around the observed price p_0. In the case of a price increase, the firm worries that demand is elastic, but in case of a price decrease it worries of that demand is inelastic. In essence, the overall worst-case is the result of splicing two different priors together – an elastic one to the right of p_0, and an inelastic one to the left. Panel (a) in Figure 2 illustrates the resulting, kinked worst-case expected demand, conditional on seeing a signal equal to the true DGP: $y_0 = \bar{\gamma} - bp_0$.

Continuous expected demands

The worst-case expected demand in (14) is not only kinked, but also discontinuous. However the jump is not an integral part of the mechanism, the firm need not consider the possibility of a drastic change in demand. We have found it most straightforward to impose only limited restrictions on the set of admissible priors, but we could easily impose further restrictions on the smoothness of the possible priors that would ensure expected demand is continuous, and still obtain the main results.

In particular, we can require that admissible priors $m(p)$ must have a derivative no bigger than some $b_{\max}$. Without this restriction, the worst-case prior is discontinuous to the right of p_0 due to equation (12). If we impose it, however, the worst-case prior becomes:

$$m^*(p_0) = \min\left[\gamma_h - bp_0, \gamma_l - bp + b_{max}(p - p_0)\right]$$

As before, the worst-case prior picks $m(p')$ equal to the lower bound and seeks the maximal fall in demand between p' and p_0. But now there is a restriction on how high $m(p_0)$ can be, given the value for $m(p)$ and the understanding that $m(p)$ cannot have an infinite derivative. As illustrated in panel (b) of Figure 2, this constraint rules out jumps in $m(p)$, which makes the worst-case demand continuous, but it still has a kink at p_0.

The kink is the fundamental feature of the mechanism: it is generated by the endogenous switch in the worst-case elasticity to the left and to the right of p_0 and does not depend on discontinuities in the admissible priors. Indeed, the derivative of the worst-case expected demand is equal to $(\alpha - 1)b$ to the left of p_0 and equal to $(\alpha - 1)b - \alpha b_{\max}$ to its right. When faced with ambiguity about the shape of the underlying demand, the ambiguity averse firm acts as if demand is relatively more elastic for price increases than for price decreases.

Updating with more observed prices

The Bayesian update given a vector of signals is standard, and leverages our Gaussian framework. Online Appendix A describes the general formulas and an analytical approach to finding the worst-case prior. This involves sorting the observed prices from smallest to largest and fully characterizing the worst-case prior for all prices recursively. For any entertained price p' the worst-case prior is obtained through three steps: (i) the priors on demand signals at observed prices to the left of that p' are at the upper bound of the prior set; (ii) the prior at p' is the lowest bound, and (iii) the priors on demand signals at observed prices to the right of p' to be as large as possible, while still respecting a downward sloping $m(p)$. The intuition is similar as before: the firm worries that demand at price p' is low, while the observed signals can be attributed to high prior demand at those other prices.

The main observation is that the switch in the worst-case priors now applies more generally at all observed prices. For example, Figure 3 shows the worst-case expectation when the firm has observed demand signals at two distinct prices, both equal to the corresponding true DGP value. As we show in the next section, the emergence of the two concave kinks at the previously observed prices in this as if expected demand leads not only to stickiness in the pricing actions, but also to discreteness and memory of the optimal price.

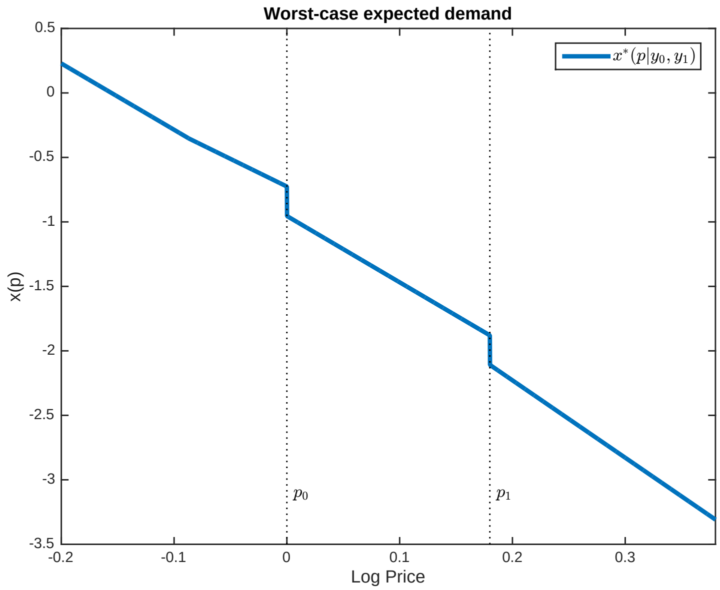


Figure 3: Worst-case Expected Demand, 2 previously observed prices

5 Optimal pricing

The firm's problem is to choose the optimal price that maximizes expected utility as if the worst-case probability distribution is the true data generating process. The problem is specified in equation (9). In the previous section we have analyzed how the worst-case prior endogenously changes, depending on the entertained pricing action.

The pricing problem of the ambiguity averse firm is dynamic. Posting a price today does not only affect the current profit, but also affects next period's information set. Solving fully optimal learning problems while allowing for experimentation is a difficult numerical task. The main computational burden here is that the state space explodes as the number of posted prices increases with time. For this reason we focus on studying a two-period model, where in the second period there are only static profits to be gained and no continuation utility. We believe that while simple, this two-period model transparently captures the most important effects of the infinite horizon version of the model.

In the second period, the firm observes the cost shock, c_2, and the price-quantity history, ε^1, which includes the first period's realized quantity sold, $q(p_1)$, and some initial information inherited from period 0. We start by analyzing the static maximization problem in this last period, and provide analytical results. We then proceed backwards and study numerically the dynamic problem of the first period, where the firm takes into account the effect of its optimal price on the information set in the last period.

5.1 Second period: a static optimization problem

In the second period, the firm chooses a price p_2 to maximize the end-of-period profits under the worst-case expected demand, conditional on the observed ε^1:

$$v(\varepsilon^1, c_2) = \max_{p_2} \min_{m(p)} E^{\widehat{x}_1(p_2;m(p))} \left(e^{p_2} - e^{c_2}\right) e^{x(p_2)+z_2} \tag{15}$$

where the posterior distribution of demand at some price p_2 is a Normal distribution, as shown in equation (8).

5.1.1 Price rigidity

Stickiness with one previously observed price

To highlight the analytical mechanics of the model, we start with the case where the firm has only observed a single price p_0 in the past, for N_0 times and with an average realized quantity sold y_0. Evaluating the worst-case expectation, the static problem becomes:

$$v(\varepsilon^1, c_2) = \max_{p_2} \left(e^{p_2} - e^{c_2}\right) e^{0.5(\sigma_z^2 + \widehat{\sigma}_1^2)} e^{\widehat{x}_1(p_2;m^*(p))}$$

where the posterior variance evaluates to $\widehat{\sigma}_1^2 = \frac{\sigma_x^2 \sigma_z^2/N_0}{\sigma_x^2 + \sigma_z^2/N_0}$ and applying equation (14), the worst-case expectation is given by the piece-wise function

$$\widehat{x}_1(p_2; m^*(p)) = \begin{cases} \gamma - bp_2 - \nu\sigma_z + \alpha\left[y_0 - (\gamma - bp_0 + \nu\sigma_z)\right] & \text{for } p < \underline{\text{p}} \text{ and } p > p_0 \\ (1-\alpha)(\gamma - bp_2 - \nu\sigma_z) + \alpha y_0 & \text{for } p \in [\underline{\text{p}}, p_0] \end{cases}$$

where $\underline{p} = p_0 - \frac{2}{b}\nu\sigma_z$.

Thus, for higher and significantly lower prices than p_0, the firm acts as if it perceives a demand curve with a slope b (same as the DGP) that has been shifted from the actual DGP curve $\gamma - bp_2$ by two components. The first, $-\nu\sigma_z$, is a shift down resulting from the lower-bound on the set of priors. The second component, $\alpha\left[y_0 - (\gamma - bp_0 + \nu\sigma_z)\right]$, is the result of the informative signal y_0. On the other hand, for prices $p \in [\underline{p}, p_0]$ the firm perceives a flatter demand curve with a slope $-b(1-\alpha)$.

There are three potential local maxima that need to be checked: (i) $p_2 = p_0$ since that is a kink point; (ii) the optimal price for a demand curve with slope $-b$, given by the standard expected utility choice $p_2^{RE,b} = \ln\left(\frac{b}{b-1}\right) + c_2$, and (iii) the optimal price for a demand curve with slope $-b(1-\alpha)$ or $p_2^{RE,b(1-\alpha)} = \ln\left(\frac{b(1-\alpha)}{b(1-\alpha)-1}\right) + c_2$.

Solving this problem, we can show that there is a positive interval of cost shock realizations for which it is optimal to stick with the previously posted price p_0, making that price sticky. We formally establish and characterize the stickiness in Proposition 1.

Proposition 1. *If the firm has posted a single price p_0 in the past then,*

(i) the price p_0 is sticky. There are values $\underline{c}_0 < \overline{c}_0$ such that p_0 is the optimal price for all cost realizations $c_2 \in [\underline{c}_0, \overline{c}_0]$

(ii) the inaction region around p_0 (i.e. stickiness) increases with α (more precise signal) and $\nu\sigma_z$ (more ambiguity).

Proof. Follows from the kink in $\widehat{x}_1(p_2; m^*(p))$ at p_0. For details, see Online Appendix C. □

The proposition showcases several important features of the mechanism. First, this is a mechanism of rigidity – there is a positive probability that the firm does not change its price, even if costs change. This is in contrast with the rational expectations firm, which adjusts the price one-to-one with cost movements. Second, the perceived cost of changing the price is endogenous and varies with the amount of information the firm has about demand at the price p_0 – the more signals the firm has seen, the more confident it is in demand at p_0, and the more apprehensive about leaving that price. Third, more initial ambiguity makes the kink more prominent and thus the perceived cost of moving larger.

Stickiness for two previously observed price

The previous analysis can be extended to the case of many observed prices. In our two-period model we focus on the situation where the firm has seen two distinct prices in the past, arising potentially from different observations at time 0 and time 1. Similarly to the case of one observed price, the emergence of kinks in the as if expected demand naturally lead to inaction around both previously observed prices. As a counterpart to Proposition 1, we establish the following:

Proposition 2. *If the firm has previously posted two distinct prices $p_1 \neq p_0$, then*

1. *there is a kink in the* as if *expected demand at each p_i and each has an associated inaction region, such that p_i is the optimal price for all cost realizations $c_2 \in [\underline{c}_i, \overline{c}_i]$*
2. *the inaction region around each p_i (i.e. stickiness) increases with α_i (the precision of the signal at price p_i)*

Proof. See Online Appendix C. □

5.1.2 An endogenous, time-varying cost of price changes from learning

Our theory predicts an endogenous time-varying cost of price changes. New kinks are formed at newly observed prices, and old kinks change their importance as the firm obtains repeated observations of certain prices. In this section we go beyond stickiness, and characterize other important features of the optimal price series. The results are formalized in Proposition 3.

Proposition 3. *Optimal prices have the following characteristics:*

(i) ***Discreteness and Memory.*** *If the two previously observed prices are distinct $p_1 \neq p_0$, then there is a positive probability that a price change results in a discrete move within the set of observed prices, exhibiting both discreteness and memory.*

(ii) ***Declining Hazard.*** *Increasing the number of times the firm has observed the price p_1 increases its region of inaction and hence the probability that the firm remains at p_1.*

(iii) ***Large and Small Changes.*** *Optimal price adjustment is characterized by both discrete jumps and arbitrarily small price movements.*

Proof. (i) and (ii) follow from Proposition 2, (iii) obtains because the worst-case expected demand is continuous to the left at kinks. For details, see Online Appendix C. □

The proposition establishes several key results. The firm is not only reluctant to change its current price, but is in general inclined to repeat a price it has already seen in the recent past. These previously observed past prices become 'reference' prices at which there are kinks in the profit function. The existence of kinks at these prices means that both are associated with a positive measure of shocks that map to it. Intuitively, the perceived cost of switching between the two of them is lower than the cost of changing to a wholly new price, thus the model is consistent with the optimal policy having 'price memory', characterized by discrete price changes between a set of previously posted prices. Moreover, the perceived cost of changing the price varies with the amount of information about demand at that price – the more signals the firm has seen about p_0, the more confident it is in demand at p_0, and the more apprehensive about leaving that price. The endogenous cost of price changes is also central in generating a price distribution that features both small and large price changes – this is a model in which prices can simultaneously look both sticky and flexible.

5.2 A dynamic problem

Having solved the last period problem, we now analyze the first period. Here, the firm observes its marginal cost c_1 and the initial price-quantity history, ε^0, and chooses p_1 to maximize the worst-case expectation of the discounted sum of this and next period's profits:

$$\max_{p_1} \min_{m(p)} E^{\widehat{x}_0(p_1;m(p))} \left[\left(e^{p_1} - e^{c_1}\right) e^{x(p_1)+z_1} + \beta \upsilon(\varepsilon^1, c_2) \right]$$

where $\upsilon(\varepsilon^1, c_2)$ is the period two profit given by equation (15).

This is a dynamic problem because the next period's state variable, the price-quantity history ε^1, includes the quantity sold at the price chosen this period, $q(p_1)$. That observation is a noisy signal on demand that the firm would use next period to further update its beliefs. As a consequence, when the firm chooses its price today it is not only maximizing over this period's profit, but also taking into account the effect on the next period's information set.

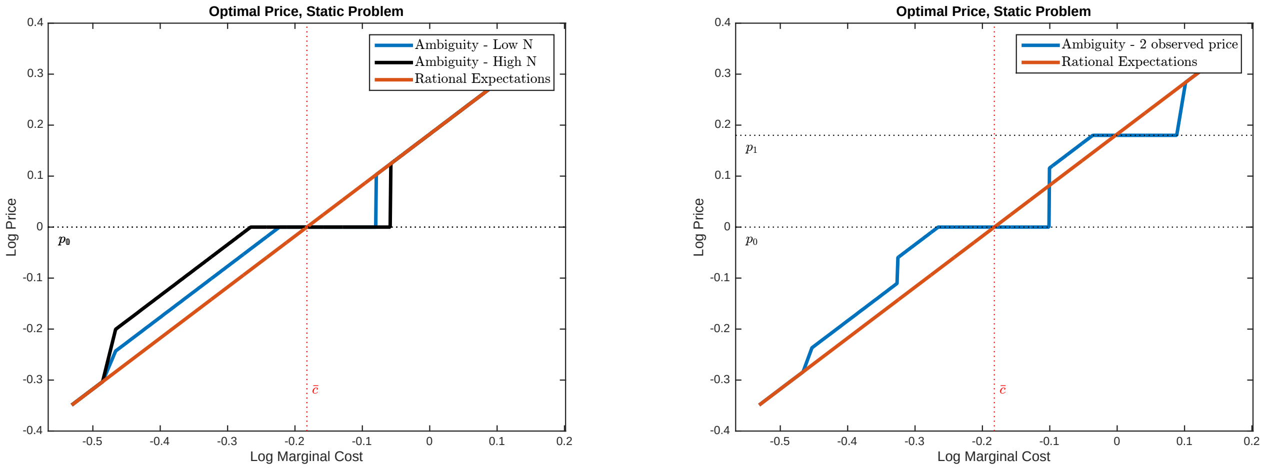
We now investigate the optimal policy functions in the context of an illustrative parametrization. We first note that we are interested in a continuous distribution for the cost shocks as otherwise that may mechanically generate discreteness in prices even in a standard model. The Markov process $g^c(c_t|c_{t-1})$ for the cost shock is

$$c_t - \overline{c} = \rho_c \left(c_{t-1} - \overline{c}\right) + \sigma_c \eta_t^c$$

where η_t^c is white noise. We set $b = 6$, the constant $\overline{\gamma} = 0$ and the discount factor $\beta = 0.97^{(1/52)}$. We normalize $\overline{c} = (b-1)/b$ so that $P^{RE} = 1$. We set the cost shock parameters ρ_c and σ_c to values calculated by Eichenbaum et al. (2011), at 0.14 and 0.11, respectively. We set $\nu = 2$, argued in Ilut and Schneider (2014) as a reasonable upper bound on ambiguity, and illustrate the mechanisms by setting $\sigma_z = 0.4$ and a signal to noise ratio $\sigma_x^2/\sigma_z^2 = 0.2$.

5.2.1 Static policy functions

For comparison purposes, we begin by illustrating the static problem's optimal price policy that we characterized analytically in section 5.1. Using the parametrization above, the left panel in Figure 4 plots the static problem's policy under RE in red, and in blue the case of ambiguity for one previously observed price p_0. For the latter there is a clear area of inaction at p_0, for which the firm finds it optimal not to change its price. Outside that area the optimal price is: (i) for $p < p_0$ equal to $p^{RE,b(1-\alpha)}$, the RE optimal price when demand elasticity is equal to $-b(1-\alpha)$; (ii) for $p < \underline{p}$ or $p > p_0$, it is equal to $p^{RE,b}$, the RE optimal price under the true elasticity of $-b$. The black line shows the case where the price p_0 has been observed more often. Importantly, the higher confidence accumulated at this price leads to a larger inaction area, and it is now the optimal price for a larger mass of cost shock realizations, i.e. the price is stickier. This panel illustrates the stickiness result of Proposition 1 and the declining hazard property of Proposition 3.



(a) One previously observed price

(b) Two previously observed prices

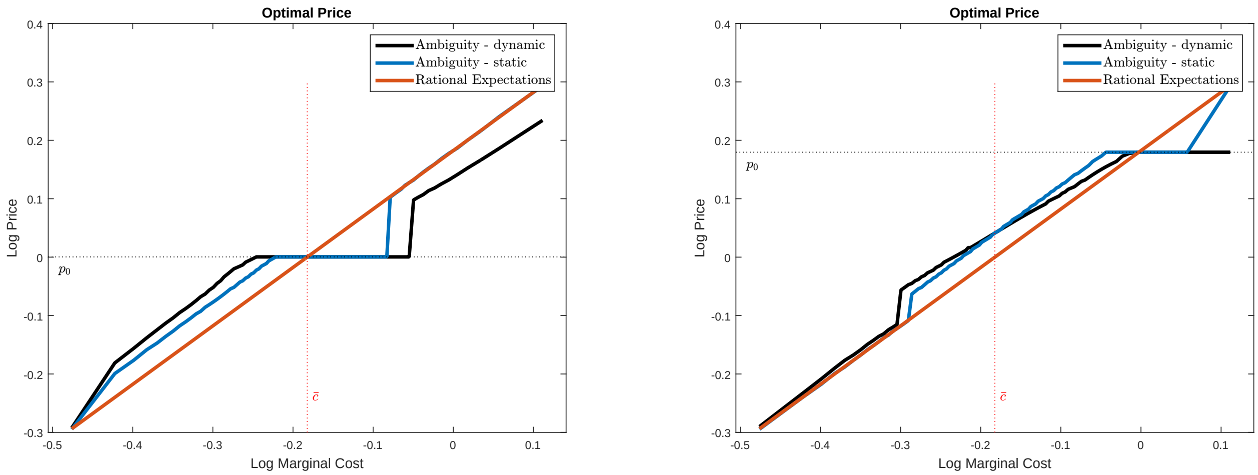
Figure 4: Policy Function, Static problem

The right panel of the figure plots the optimal price for the case where the firm has also seen a second price $p_1 > p_0$. The two kinks in expected demand manifest themselves as areas of inaction around these two previously observed prices. This captures the discreteness of the policy function: previously observed prices become 'focal points'. Notice that there is a whole range of cost shocks, that would have previously resulted in setting p^{RE}, but now lead to setting p_1. There is now a high probability that conditional on a price change the price adjusts discretely and not proportionally with the cost. This panel illustrates the inaction result of Proposition 2 and the additional properties analyzed in Proposition 3, namely discreteness and memory as well as price changes being potentially small or large.

5.2.2 Dynamic policy functions

The dark solid line in the top left panel of Figure 5 plots the period one pricing policy of the two-period model, where the firm has an initial signal at price p_0 and takes into account the effect of its current price choice on the future. In comparison to a static optimization, the dynamic one features even more stickiness, especially for higher cost shocks.

Accounting for active learning has two competing effects. On the one hand, by sticking to the same price, the firm gets to learn more about it. On the other, by moving to another price it can expect to learn something new and potentially valuable. Which force dominates is state-dependent. The left panel is an example of the former effect being stronger, which leads to more stickiness than the static policy function. This is because the observed price p_0 is the optimal price for the mean cost shock $\bar{c}$. The firm expects future cost shocks to be



(a) One previously observed price

(b) One previously observed price, high p_0

Figure 5: Policy Function, Dynamic problem

close to it, and hence realizes that it is likely to post the price p_0 in the future with a high probability. Hence, learning more about this part of the demand curve is particularly useful. In the right panel, we plot the different case where the observed p_0 is significantly higher and would generally be optimal only for high cost values. In this case, the experimentation motive dominates, as it is not very useful to learn about this relatively unusual price p_0. The firm is not very likely to revisit such a high price again, and thus finds it optimal to move earlier away from it and in particular explore prices closer to the more likely region. This leads to the optimal price featuring less stickiness than the static solution,.

Our results suggest that there is an inherent tension between the incentive to experiment and that of acquiring further information at a previously observed action. In general, we find that dynamic learning does not negate the price stickiness results from the static model, and that it typically further amplifies inaction. The local nature of learning is key for the result that experimentation may lead to additional stickiness of actions.[16]

6 Nominal Rigidity

The model presented so far was one of real rigidity, in which p is interpreted as a real price, and nothing prevented nominal adjustments. For example, if the firm knew that the

[16] Consistent with the behavior that our model predicts, Anderson (2012) documents that in laboratory experiments subjects undervalue information from experimentation but are willing to pay more than the ambiguity neutral agents to learn the true mean of the payoff distribution.

aggregate price level had shifted, it could similarly change its nominal price to achieve the same "safe" real price. In this section we enrich the model so as to make a distinction between real and nominal prices and show how nominal rigidity arises as a result of the interaction of demand uncertainty with the uncertainty about the relevant relative price.

The model consists of a continuum of industries populated by monopolistically competitive firms. The firm's demand is thus a function of the aggregative technology of its industry and of the relevant relative price, equal to the ratio of its nominal price against the industry price index. We assume that the monopolistically competitive firm faces ambiguity about the technology of its industry. This results in the firm not knowing both its demand function as well as the appropriate relative price argument of this demand function. The ambiguity averse firm sets an optimal nominal pricing action that is robust to both sources of ambiguity, and this turns the real rigidity generated in the previous section into nominal rigidity.

6.1 Economic Framework

There is a continuum of industries indexed by j and a representative household that consumes a CES basket of the goods produced by the different industries. The final good basket and the associated aggregate price index are:

$$C_t = \left(\int C_{jt}^{\frac{b-1}{b}} dj \right)^{\frac{b}{b-1}} \quad , \quad P_t = \left(\int P_{jt}^{1-b} dj \right)^{\frac{1}{1-b}} \tag{16}$$

where P_{jt} are the price indices of the separate industries.[17]

Each industry j has a representative final goods firm that produces by aggregating over intermediate goods i with the technology

$$C_{jt} = f_j^{-1} \left(\int f_j(C_{ijt}) v_j(z_{it}) di \right) \tag{17}$$

where z_{it} is an idiosyncratic demand shock for the good i, distributed as $N(0, \sigma_z^2)$. Each industry j has potentially different functions f_j and v_j, and a price index P_{jt} such that

$$P_{jt} C_{jt} = \int P_{it} C_{ijt} di$$

where C_{ijt} is the amount purchased of good variety i by industry j. Solving the cost

[17] An equivalent alternative interpretation of our setup is that the economy is composed by a continuum of households j with different preferences, which share risk and aggregate according to the basket C_t.

minimization problem of the representative firm in industry j yields

$$C_{ijt} = f_j'^{-1}\left(\frac{P_{it}}{P_{jt}}\frac{f'(C_{jt})}{v(z_{it})}\right) \equiv H_j\left(\frac{P_{it}}{P_{jt}}, C_{jt}, z_{it}\right) \tag{18}$$

The demand of industry j for a given intermediate good i is a function of the relevant relative price, $\frac{P_{it}}{P_{jt}}$, overall industry output C_{jt}, and demand shocks z_{it}. We denote this function by H_j and note that it is a transformation of the functions f_j and v_j. The intermediate goods consumed by an industry j are produced by a continuum of monopolistic firms i. Each firm i sells to only one industry j, hence $Y_{it} = C_{ijt}$.[18]

6.2 Information structure and learning

The information of the intermediate good firms is imperfect in two ways. First, they do not know the functional forms of the industry-level production technologies f_j and v_j, and in fact the uncertainty over the production functions cannot be described by a single probability measure – firms face Knightian uncertainty (or ambiguity) about their industry structure. Second, they do not observe all variables every period. They see their own prices and quantities, P_{it} and Y_{it}, and the aggregate output and price level, C_t and P_t, every period. However, they observe industry level prices and quantities, C_{jt} and P_{jt}, infrequently, only every T periods. Lastly, the firms never see the demand shock z_{it}.

6.2.1 Demand uncertainty

A firm does not know the specific functional form of its demand, but rather needs to estimate it using its observables. For tractability, we assume the firm understands that the aggregate industry demand C_{jt} and the demand shocks z_{it} enter multiplicatively so that[19]

$$C_{ijt} = H_j\left(\frac{P_{it}}{P_{jt}}\right) C_{jt}\exp(z_{it})$$

The firm can then use the known structure of aggregate demand

$$C_{jt} = \left(\frac{P_{jt}}{P_t}\right)^{-b} C_t \tag{19}$$

[18]As a result, firms are indexed by both i and j, however, we for ease of notation we drop the j subscript with the understanding that each firm i is unique to a given industry.

[19]Our learning framework extends to the case of learning about demand as a function fo multiple variables without conceptual differences. We make this assumption to transparently focus on the main mechanism.

to write its demand schedule as

$$C_{ijt} = H_j\left(\frac{P_{it}}{P_{jt}}\right)\left(\frac{P_{jt}}{P_t}\right)^{-b} C_t \exp(z_{it}) \tag{20}$$

Thus, the firm understands how the aggregates affect its individual demand through their effect on average industry demand C_{jt}. However it does not have complete information on the specific competitive environment it faces, and hence does not know the function $H_j(.)$. Taking logs and denoting logged variables as lower-case letters, we obtain a linear expression in an unknown function, h_j, an unknown variable, p_{jt}, known effects, c_t and bp_t, and an unobserved shock z_{it}:

$$y_{it} = h_j(p_{it} - p_{jt}) + c_t - b(p_{jt} - p_t) + z_{it}. \tag{21}$$

The uncertainty about the unknown function h_j is modeled as before - there is a set of multiple priors Υ_0, where each prior is a GP distribution with a weakly decreasing mean function $m(r)$ such that

$$m(r) \in [-\gamma - br, \gamma - br],$$

Learning about this unknown function proceeds as before, and next we turn our attention to the uncertainty about p_{jt}.

6.3 Uncertainty about the relationship with aggregate prices

The firm has two sources of information on p_{jt}. First, every T periods, it conducts marketing reviews that reveal the current industry price. The idea is that reviews are costly and time consuming, but since they are useful, they are done on a regular basis.[20] Second, in between reviews, the firm attempts to filter p_{jt} out of the aggregate information it observes. Since the firm's direct competitors form only a small portion of the overall economy, the firm knows that $p_{jt} \neq p_t$, where p_t is the aggregate, fully-observable price level.

Even though the industry price p_{jt} is not equal to the observed aggregate price, the firm can use the latter to extract information about p_{jt}. Indeed, the firm understands that prices are cointegrated and that there is a link between industry prices and aggregate prices. However, since the firm does not know the exact structure of industry demand (i.e. the

[20] As long as reviews do not happen every period, introducing state-dependent reviews would not significantly change our analysis. For simplicity we are implicitly assuming that the firm either does not want to perform reviews more frequently, or there are some technological constraints on the ability to perform frequent reviews (e.g. the necessary data is not observed every period).

production functions f_j), it does not know the exact functional form of that relationship.[21] In fact, the ambiguity about the industry's production structure transfers to this issue as well – different industry production functions imply different structural relationships between aggregate and industry level prices. Due to this ambiguity, the firm is not confident in any single relationship, and entertains a whole set of potential relationships such that

$$p_{jt} = p_{js} + \phi(p_t - p_{js}) + \nu_{jt}, \tag{22}$$

where p_{js} is the last perfectly revealing signal the firm has seen. Thus, in between reviews the firm is trying to forecast the industry prices p_{jt} with the aggregate price p_t, but is not certain what is the correct structure of that signal.

Ambiguity is modeled through multiple priors on the co-integrating relationship $\phi(.)$ and the transitory term ν_{jt}. The priors on ν_{jt} are Gaussian white noise, but with different, possibly time-varying variances. The uncertainty about the cointegrating function is modeled in a similar fashion to the uncertainty about the demand function $h(.)$. As such, we assume that the priors on $\phi(.)$ are Gaussian Process distributions that put non-zero probability on all functions that lay in a set Ω_ϕ around the true DGP $\phi(p_t - p_{js}) = p_t - p_{js}$. Lastly, for tractability, we focus on the limiting case where the variance function of the Gaussian Processes distributions for the functions $\phi(.)$ goes to zero, so conditional on a prior, one function $\phi(.)$ has probability 1 and all others probability zero.

The set of potential cointegrating functions allows for a weak relationship between industry and aggregate inflation in the short-run. We model this by specifying that for small $|p_t - p_{jt}|$, i.e. small inflationary pressure, the function $\phi(.)$ lies in an interval around 0

$$\phi(p_t - p_{js}) \in [-\gamma_p, \gamma_p], \text{ for } |p_t - p_{js}| \leq \Gamma. \tag{23}$$

This allows for functions that imply weak short-run relationship between aggregate and industry inflation. The firm realizes, however, that the two are cointegrated in the long-run, and for that reason, away from zero, the set of potential $\phi(.)$ grows linearly with $p_t - p_{jt}$

$$\phi(p_t - p_{js}) \in [p_t - p_{js} - \gamma_p + \Gamma, p_t - p_{js} + \gamma_p + \Gamma], \text{ for } |p_t - p_{js}| \geq \Gamma.$$

The particular boundaries of Ω_ϕ are chosen to define an analytically tractable set of priors, but this is done solely for convenience, and has no bearing on the rest of the argument. The magnitude of Γ is chosen to be high enough so that in between reviews the function

[21] In essence, the firm does not know the functional form of the relevant industry price index, and how it relates to the aggregate price index.

$\phi(.)$ belongs to the set described by (23). Our empirical evidence discussed in subsection 6.4 supports the notion that it is reasonable for the firm to consider a lack of precise relationship between aggregate and industry prices for horizons of up to several years.

Note that all admissible priors imply that the price ratio $p_{jt} - p_t$ is stationary with probability 1, but allow for potentially complex, non-linear relationships locally. Intuitively, the firm understands price levels are co-integrated in the long-run, however, it is not confident in extrapolating this long-run relation to short-run fluctuations, and entertains functions $\phi(.)$ which allow for a variety of local, possibly time-varying relationships. This is meant to capture the empirical regularity that estimates of the short-run relationship between disaggregated inflation indices and overall inflation are imprecise and appear to be time-varying, but estimates on long-run inflation series confidently point towards cointegration. The firm has no advantage over real-world econometricians and cannot eliminate the uncertainty in the short-run inflation relationship by postulating a single, linear cointegrating relationship with full certainty. Thus, the set of priors explicitly allows for the possibility that the short-run relationship is weak, even though in the long-run the firm expects prices to rise in lock-step.

6.3.1 Worst-case beliefs

The unknown portion of the firm's demand can be written as

$$h(\hat{r}_{it} - \phi(p_t - p_{js}) - \nu_{jt}) - b(\phi(p_t - p_{js}) + \nu_{jt}),$$

where $\hat{r}_{it} = p_{it} - p_{js}$, and it includes two unknown functions: $h(.)$ and $\phi(.)$. The firm understands that its demand is ambiguous in two dimensions. First, the functional form of demand, $h(.)$, is ambiguous, and second the argument of that function itself is also ambiguous, due to the uncertainty about $\phi(.)$. The firm chooses an optimal pricing action, $\hat{r}_{it}$, that is robust to both sources of ambiguity. This amounts to choosing a profit maximizing price, under the worst-case demand schedule, where worst-case demand is determined price-by-price, i.e. conditional on any given pricing action $\hat{r}_{it}$.

For each admissible demand shape $h(.)$ and pricing action $\hat{r}_{it}$, we can find a worst-case cointegrating relationship $\phi(.)$ that yields the worst demand:[22]

$$h^*(\hat{r}_{it}, \nu_{jt}) = \min_{\phi} h(\hat{r}_{it} - \phi(p_t - p_{js}) - \nu_{jt}) - b(\phi(p_t - p_{js}) + \nu_{jt}) \tag{24}$$

This is the demand level that would prevail if nature draws the worst possible $\phi(.)$, conditional on a particular $h(.)$ and price $\hat{r}_{it}$. Since in the short run $\phi(p_t - p_{js}) \in [-\gamma_p, \gamma_p]$,

[22]Here we are able to minimize over ϕ directly due the assumption of Delta priors.

variation in p_t does not change the set of possible numerical values that could be realized through $\phi(p_t - p_{js})$. Hence we can recast the optimization in terms of minimizing over a parameter, $\bar{\phi} \in [-\gamma_t, \gamma_p]$, which represents the short-run conditional expectation of p_{jt}. The solution to the minimization can then be written as $\phi^*(p_t - p_{js}) = \bar{\phi}^*$. Intuitively, the worst-case cointegrating relationship implies that movements in the aggregate price are not informative about industry prices in the short-run. This is because when there is no such informative relationship, nature has the greatest flexibility in choosing the worst-case expectation of p_{jt}, given a demand function $h(.)$ and a price choice $\hat{r}_{it}$.

Since the transitory shocks ν_{jt} are not observed, we can take an expectation over them and define the expected demand under the worst-case cointegrating relationship:

$$x(\hat{r}_{it}) = E_t(h^*(\hat{r}_{it}, \nu_{jt}))$$

This is the object that the firm can learn about through its past prices and quantities because, according to the optimal behavior under ambiguity, it believes that nature has minimized demand in this same fashion at any point in time. For tractability, we assume that the implied expectational errors follow a normal distribution,

$$h^*(\hat{r}_{it}, \nu_{jt}) = x(\hat{r}_{it}) + \varepsilon_{it};\ \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2). \tag{25}$$

6.3.2 Nominal rigidity from real rigidity

The firm uses past signals to learn about the worst-case demand. Putting together (21) and (25), the demand facing the firm is

$$y_{it} = x(\hat{r}_{it}) + c_t + b(p_t - p_{js}) + \varepsilon_{it} + z_{it} \tag{26}$$

which is a known function of the observed aggregates, namely price p_t and quantity c_t, an unknown function $x(.)$ of its perceived relevant relative price and Gaussian noise. This forms a well-defined learning problem that the firm approaches in the way described in Section 4.

The kinks are formed in the space of relative prices $\hat{r}_{it}$. However the base of this relative price, i.e. the last review signal p_{js}, does not change every period. To keep this relative price constant then, and thus take advantage of the kinks, the firm needs to keep its nominal price constant. Hence, the model generates both nominal stickiness and memory in nominal prices. In essence, all results from the analytic section go through, but their effects are now primarily on nominal prices. In addition, since the firm does update its beliefs about p_{jt} regularly, the stickiness in nominal prices appears as stickiness in "price plans". The price series tends to bounce around a few common prices that look like a "price plan", and then when new review signals arrive the firm shifts that price plan accordingly.

6.4 Empirical link between aggregate and industry prices

Here we use US CPI data to show that the relationship between aggregate and industry prices is time-varying and unstable over short-horizons. In particular, an econometrician would generally have very little confidence that short-run aggregate inflation is related to industry-level inflation, even though he can be confident that the two are cointegrated in the long-run. Thus, our assumption on the uncertainty over $\phi(.)$ above again puts the firm on an equal footing with an econometrician outside of the model.

Our analysis uses the Bureau of Labor Statistics' most disaggregated 130 CPI indices as well as aggregate CPI inflation. The empirical exercise consists of the following regression method. For a specific industry j, we define its inflation rate between $t-k$ and t as $\pi_{j,t,k}$ and similarly $\pi^a_{t,k}$ for aggregate CPI inflation. For each industry j, we run the rolling regressions:

$$\pi_{j,t,k} = \beta_{j,k,t}\pi^a_{t,k} + u_t$$

over three-year windows starting in 1995 and ending in 2010.[23] We repeat this exercise for k equal to 1, 3, 6, 12 and 24 months. Finally, for each of these horizons we compute the fraction of regression coefficients $\beta_{j,k,t}$ (across industries and 3-year regression windows) that are statistically different from zero at the 95% level.

We find that for 1-month inflation rates, only 11.4% of the relationships between sectoral and aggregate inflation are statistically significant. For longer horizons k, these fractions generally remain weak but do rise over time: 26.4%, 40.6%, 58.5% and 69.1% for the 3-, 6-, 12- and 24-month horizons respectively. This supports our assumption that while disaggregate and aggregate price indices might be cointegrated in the long run, their short-run relationship is weak.

In fact, not only is the relationship statistically weak in general, but it is highly unstable. This can be seen in Figure 6 that shows the evolution of the coefficient $\beta_{j,k,t}$ for $k = 3$ for 3-year-window regressions starting in each month between 1995 and 2010, for four industries. Not only are there large fluctuations in the value of this coefficient over our sample, but sign reversals are common. In general, at any given date, there is little confidence that the near-future short-horizon industry-level inflation would be highly correlated with aggregate inflation, even though the data is quite clear that the two are tightly linked over the long-run.

[23] Results are very similar if we use windows of 2 or 5 years instead.

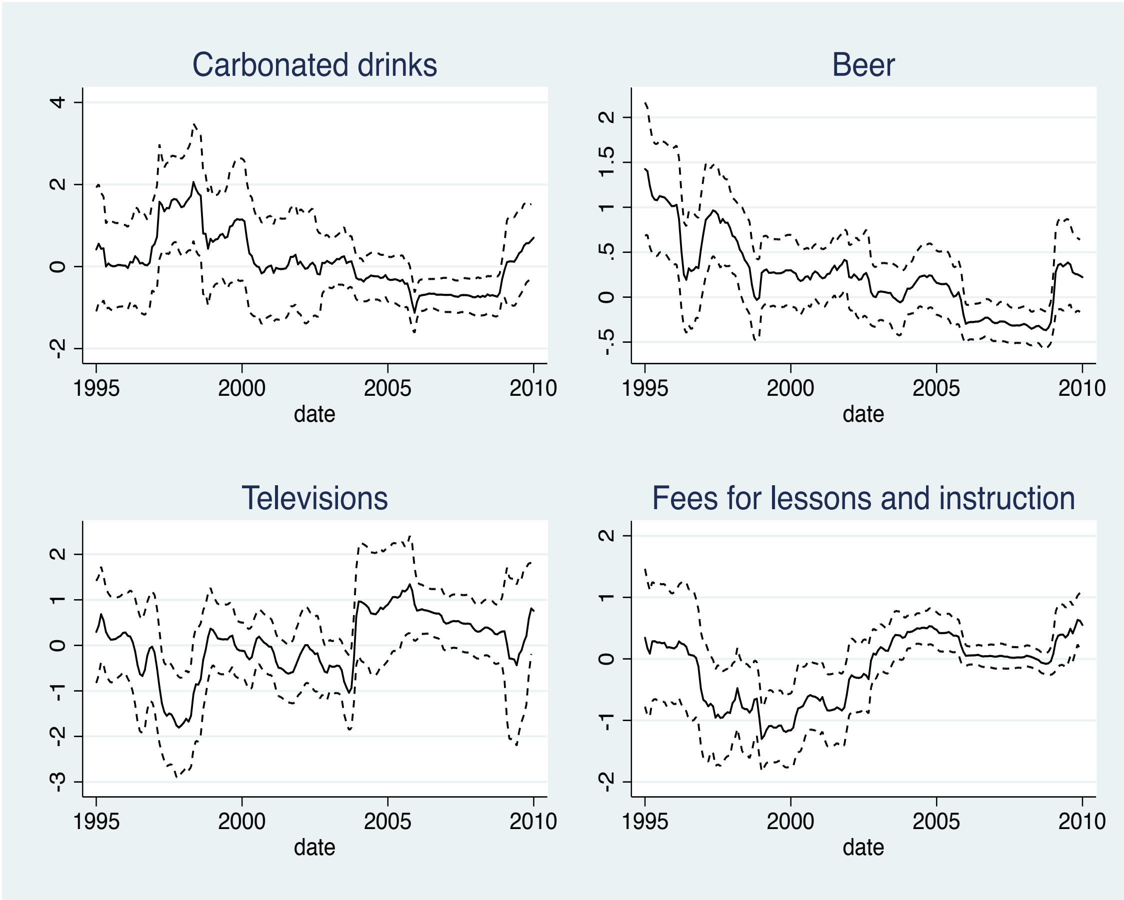


Figure 6: 3-year rolling regressions of 3-month industry inflation on 3-month aggregate inflation for four categories. The solid line plots the point estimate of regression coefficient on aggregate inflation. The dotted lines plot the 95% confidence intervals.

7 Quantitative model

We build a quantitative version of the model in the previous section, that endogenizes marginal cost and introduces a law of motion for the aggregate price level. The objective is to quantitatively study the individual decision problem of a firm that faces demand uncertainty. A precise way to view the setup proposed here is to consider it as general equilibrium model with a measure zero of myopic, ambiguity averse firms. This means that the aggregate variables follow their flexible, rational expectations law of motion.

7.1 Model setup

As described in section 6, there are three layers of production. A representative household purchases consumption from a competitive final good producer, who buys from a continuum of industries indexed by j. Each industry itself is composed of a competitive final good producer, that aggregates over a continuum of intermediate monopolistic firms index by i.

The representative household consumes and works according to

$$\sum_{k=0}^{\infty} E_t \left(\beta^{t+k} \left[\log C_{t+k} - \chi \int L_{i,t+k} di \right] \right)$$

subject to the budget constraint

$$\int P_{j,t} C_{jt} dj + E_t q_{t+1} b_{t+1} = b_t + W_t \int L_{i,t} di + \int \upsilon_{i,t} di$$

where q_{t+1} is the stochastic discount factor, b_{t+1} is state contingent claims on aggregate shock, $\upsilon_{i,t}$ is the profit from the monopolistic intermediaries and consumption integrates over the varieties produced by competitive industries j with a CES aggregator with elasticity b as shown in (16). The solution to the cost minimization problem of the representative agent is to demand from each industry the amount given by (19). The technology and resulting cost minimization solution of the j-th industry are described by equations (17) and (18).

The demand for the monopolistic firm i comes from the industry j in the form of (18) which we have further restricted to be described in (20). The firm produces variety i using the production function:

$$Y_{i,t} = \omega_{it} A_t L_{it}$$

where ω_{it} and A_t are an idiosyncratic and aggregate productivity shock, respectively, and L_{it} is hours hired by firm i at wage W_t. The processes for these shocks are:

$$\log \omega_{it} = \rho_\omega \log \omega_{it-1} + \varepsilon^{\omega}_{i,t};\ \ \log A_t = \rho_a \log A_{t-1} + \varepsilon^a_t$$

where $\varepsilon^{\omega}_{i,t}$ is iid $N(0, \sigma^2_{\xi})$ and ε^a_t is iid $N(0, \sigma^2_a)$.

Monopolistic firms are owned by the representative agent, and thus they discount profits using the agent's stochastic discount factor. The economy-wide price index and aggregate output are defined as

$$P_t = \int_0^1 P_{j,t} \frac{Y_{j,t}}{Y_t} di;\ \ Y_t = \int_0^1 \left(Y_{j,t}^{\frac{b-1}{b}} \right)^{\frac{b}{b-1}} dj$$

Finally, nominal aggregate spending $S_t = P_t C_t$ follows a random walk with drift

$$\log S_t = \mu + \log S_{t-1} + \epsilon^s_t$$

where ϵ^s_t is iid $N(0, \sigma^2_s)$. Using the household's hours decision $W_t/P_t = \chi C_t$ to substitute out for W_t, the real flow profits can be written as

$$\upsilon_{i,t} = \left(\frac{P_{it}}{P_t} - \frac{\chi S_t}{\omega_{it} A_t P_t} \right) Y_{i,t} \tag{27}$$

7.2 Demand uncertainty

As in section 6, we assume that the firm observes the aggregate P_t and C_t, but not its demand function. The learning process is the same as described in section 6, where equation (26) gives the demand to be estimated as

$$y_{it} = x(\hat{r}_{it}) + c_t + b(p_t - p_{js}) + z_{it} + \varepsilon_{it} \tag{28}$$

and $\hat{r}_{it} = p_{it} - p_{js}$ is the price relative to the last observed p_{jt} and the set of priors consists of Gaussian Processes with a weakly decreasing mean function

$$m(\hat{r}_{it}) \in [\gamma_l - b\hat{r}_{it}, \gamma_h - b\hat{r}_{it}],$$

and a covariance function $K(\hat{r}, \hat{r}') = \sigma_x^2$.

The firm enters period t with knowledge of the history of previous realized demand and corresponding prices, denoted by ϵ^{t-1}; the current productivity ω_{it} and the aggregate state variables: current productivity A_t, nominal spending S_t and aggregate price P_t; and an incomplete history of past $P_{j,t}$, where it has observed the industry price level only once every T periods. Based on the state variables, the firm chooses its price. Demand shocks are realized at the end-of-period and the firm fulfills demand at that price. The firm then updates its information set.

The firm does not observe the distribution of idiosyncratic states, but needs to conjecture how the aggregate price is formed. Here we use the assumption that there is a measure zero of ambiguity averse firms while the rest of the economy is populated by flexible price firms that have full confidence in their knowledge. This is the flexible price, rational expectations (RE) general equilibrium version of our economy.[24]

To characterize the RE version, we assume a simple true DGP: each industry j has the same CES functions f_j and v_j in (17): $f_j(C_{ijt}) = C_{ijt}^{\frac{b-1}{b}}$; $v_j(z_{it}) = z_{it}^{1/b}$. These aggregators imply the standard demand $C_{j,i,t} = C_{jt}\varepsilon_{it} \left(P_{i,t}/P_{j,t}\right)^{-b}$. Thus, under the true DGP, the demand function is simply $y_{it} = -bp_{i,t} + c_t + bp_t + z_{it}$ and the RE firms know that the

[24] A similar approach of a flexible aggregate price level is taken for example by Stevens (2014) in the context of a rational inattention model. This benchmark provides an upper bound for the degree of price neutrality compared to the case of a measure one of ambiguity averse firms.

underlying demand is $x(\hat{r}_{it}) = -b\hat{r}_{it}$.[25] The aggregate price solution of this economy is, up to a log-linear term:

$$p_t^{flex} = \log \frac{b\chi}{b-1} + \log S_t - \log A_t \tag{29}$$

and the optimal price for the RE firms' price is to subtract $\log \omega_{it}$ from p_t^{flex}.

The ambiguity averse firm has all the knowledge about aggregate equilibrium relationships of a RE economy, except knowing its demand function. For the quantitative model of this section we solve for decision rules of the firm by assuming that the firm is myopic, so that it solves a static optimization of end-of-period profit $\upsilon_{i,t}$:[26]

$$\max_{\hat{r}_{it}} \min_{\widehat{x}(\hat{r}_{it}|\varepsilon^{t-1})} E^{\widehat{x}(\hat{r}_{it}|\varepsilon^{t-1})} \upsilon_{i,t} \tag{30}$$

The optimal decision rule is characterized as follows: the ambiguity-averse firm takes as given the aggregates, and maximizes the objective given by (30), where profits are defined in (27), subject to the demand uncertainty in (28) and the assumed information structure.

7.3 Results

7.3.1 Calibration

The model period is a week. We calibrate $\beta = 0.97^{(1/52)}$ to match an annual interest rate of 3%. The mean growth rate of nominal spending $\mu = 0.00046$ is set to match an annual inflation of 2.4% and we set the standard deviation $\sigma_s = 0.0015$ to generate an annual standard deviation of nominal GDP growth of 1.1%. Following the calibration in Vavra (2014) we set the persistence and standard deviation of aggregate productivity $\rho_a = 0.9785$ and $\sigma_a = 0.003$ to match the quarterly persistence and standard deviation of average labor productivity, as measured by non-farm business output per hour.

We are left with seven parameters that refer to the firm's problem. We choose an elasticity of substitution of $b = 6$, implying a markup of 20%. We set the interval between reviews, given by the parameter T, to be equal to 31 weeks, which is the average duration of a pricing regime documented by Stevens (2014). For the other parameters we use pricing and quantity moments based on the IRI Marketing Dataset, as described in section 3.

First, we calibrate the standard deviation of demand shocks σ_z by using empirical evidence on the accuracy of predicting one-period-ahead quantity. In particular, using our

[25] Notice that the whole layer of industry demand has dissapeared in this case. This was done on purpose for the simplicity of the model. However, the monopolistic firm retains all the uncertainty about the direct competitors, reflected in the unknown, relevant price $p_{j,t}$.

[26] This simplifying assumption allows us to compute easier a larger model such as this. We have investigated more forward-looking problems in the exogenous cost section 4, which produce an incentive to experiment.

dataset we run linear regressions of $\log(Q)$ on a vector of controls X, that include: 2 lags of $\log(Q)$, $\log(P)$ plus its own 2 lags, the weighted average of weekly prices in that category and its 2 lags as well as item and store dummies. We compute the absolute in-sample prediction error $(Q - X\widehat{\beta})/\overline{Q}$, where $\widehat{\beta}$ are the regression coefficients and $\overline{Q}$ is the mean quantity.[27] We calibrate the size of noise shocks to $\sigma_z = 0.5$, corresponding to a median forecast error of $0.50 * 0.675 = 0.3375$, matching our sample average.[28]

We set symmetric bounds on the prior set Υ_0, such that $-\gamma_l = \gamma_h = \nu\sigma_z$, normalized by a parameter ν, which we set equal to 2 following Ilut and Schneider (2014). We calibrate the remaining three parameters, the persistence and volatility of idiosyncratic productivity (ρ_w and σ_ω), and the signal-to-noise ratio in demand signals ($\frac{\sigma_x^2}{\sigma_z^2}$), by targeting three salient pricing moments: the frequency of price changes, the frequency of 'reference price' changes, and the fraction of price increases.[29] Table 1 presents the whole set of parameters.[30]

Table 1: Parameters

β	μ	σ_s	ρ_a	σ_a	T	σ_z	ρ_ω	σ_ω	ν	σ_x^2/σ_z^2
$0.97^{(1/52)}$	0.00046	0.0015	0.9785	0.003	31	0.5	0.9	0.0975	2	0.2

7.3.2 Pricing behavior

Pricing moments

Table 2 presents pricing moments generated by the model against their empirical counterparts. Only the first three moments are targeted by the calibration. As in the data, the model produces posted prices that look as if they change frequently but at the same time reference prices that are relatively sticky.

In section 5 we analyzed the potential of our proposed mechanism to be consistent with a range of stylized pricing facts. We follow the description of price characteristics in Proposition 3 and report in Table 2 the model implied moments along four dimensions. The emerging message is that the mechanism operates as if firms face an endogenous, state-dependent cost of price changes, that is not only consistent with observed stickiness of posted and reference prices, but also with additional empirical overidentifying restrictions.

[27] We do this across all items within a category/market and also for the item with most sales in its category. Table B.1 in the Online Appendix reports prediction errors for these various regressions.

[28] Here we used that $\Phi(-0.6745) = 0.25$, with $\Phi(.)$ denoting the standard normal cdf. In addition, we note that, with a slight abuse of notation, we use σ_z to denote the standard deviation of the sum $z_{it} + \varepsilon_{it}$. From the firm's perspective either source of disturbance amounts as noise in the demand equation (28).

[29] As in Gagnon et al. (2012), a 'reference price' is the modal price within a rolling window of 13-weeks.

[30] Moments are based on a simulation of 1000 firms for 5000 periods. For computational purposes, we need to limit the proliferation of useful past information the firm carries over, which we do by imposing that only information in the last 200 periods is used in the Bayesian update.

Table 2: Pricing moments

Moment	Data	Model
(1) Fraction of price increases	51%	52.6%
(2) Frequency of posted price changes	22.85%	22.93%
(3) Frequency of 'reference price' changes	5.96%	5.58%
(4) Probability of revisiting a price	62.1%	50.9%
(5) Average number of unique prices (13 weeks)	2.62	2.77
(6) Fraction of price changes $\leq 5\%$	13.9%	17.3%
(7) Fraction of price changes $\geq 15\%$	56.3%	55.1%

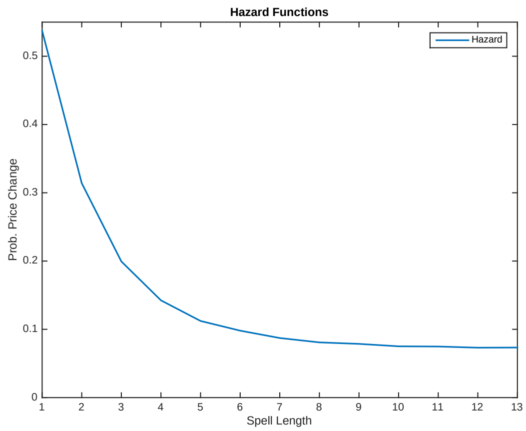
First, there is strong memory in prices: conditional on a price change, the probability of selecting the same price in the last 26 weeks is about 51%.[31] This in particular is a challenging moment to match for a benchmark menu cost model, as it typically features no incentives for firms to revisit prices, conditional on changing. Second, there is discreteness in prices: a window of 13 weeks experiences a relatively small number of unique prices. Memory and discreteness arises from the multiple kinks in the as if expected demand, produced by the lower perceived cost in terms of uncertainty of moving back to previously observed prices.

Third, as in the data, there are both small and large price changes: the model implies that 17.3% of all price changes are less than 5% in absolute terms, while 55.1% of all changes are greater than 15% in absolute terms. The existence of kinks in the policy function result in the potential for frequent, large price changes as the firm switches between the prices at those kinks. Small price changes arise because the policy function also has parts where the firm adjusts flexibly, as discussed in the analytical model. On the one hand, this can happen when the history of shocks is such that kinks in the policy function are small, for instance because of little accumulated previous information at some prices. On the other hand, the ambiguity price policy also has regions of flexibility outside the kinks. Thus, because of the endogeneity of what appears as a cost of changing a price in the ambiguity model, large and small price changes co-exist. This endogeneity makes the model behave differently than a model with a fixed cost of a price change which would typically not feature small price changes, as they are not worth paying that fixed cost.[32]

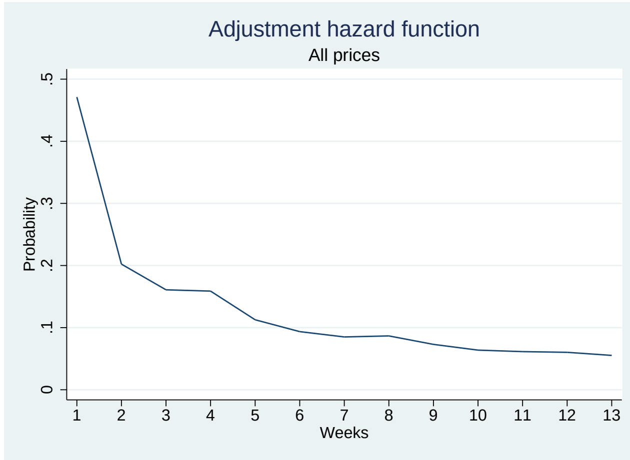
Fourth, the model produces a declining hazard function. As the firm accumulates information at some price, the kink in the expected demand deepens and the cost of changing that price increases. Figure 7 plots the probability of a price change, given that the price

[31] Note that if we filter sales out of the data, the probability of revisiting a price seen in the last 26 weeks is still 48%, which is even closer to the model.

[32] Midrigan (2011) uses a multiproduct firm and assumes economies of scope in price adjustment to generate small price changes in a menu cost model. A reduced form is to assume the random possibility of a much smaller menu cost, as used for example in Vavra (2014).



(a) Model



(b) Data

Figure 7: Price Change Hazard

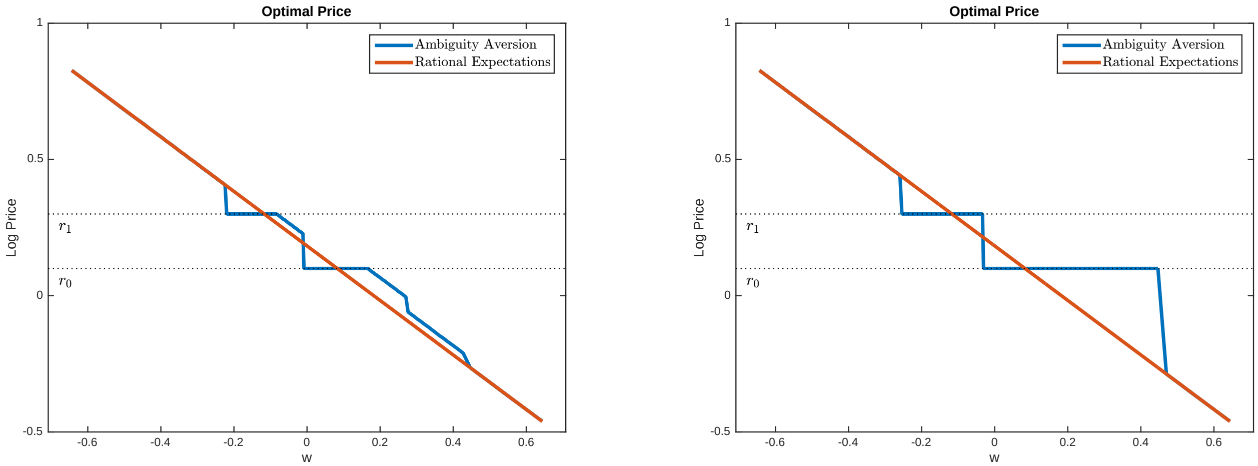
has stayed fixed for n periods, with n on the X-axis. We find that the model (left panel) matches the data very well (right panel).[33] The model implies a probability of a price change, given that the current price has been posted for just one week so far, of about 50%, and the probability steadily declines to about 7% for prices that have stayed constant for 13 weeks.

Policy functions

Next, we examine the underlying optimal price policy functions. Figure 8 plots the price policy as a function of idiosyncratic productivity. The left panel shows the case of two previous prices observed only once each. The resulting kinks are relatively small and the policy function resembles the flexible price one – it is characterized by large flexibility and likely small price changes. However, the right panel shows that as the number of observations at those same prices increases (to five in this case), the kinks become deeper. In this situation we will mostly observe few and large (discrete) price changes, as the firm switches between the two kinks. Moreover, even in this situation, the firm may choose small price changes in the areas further away from the kinks.

Of particular interest is the optimal pricing behavior as a function of monetary policy shocks. We are specifically interested in the implied degree of monetary non-neutrality, defined as the effect of the monetary policy shock on the quantity sold, which can be read off from the deviation of the optimal price from its flexible version. The left panel of Figure 9 plots the price policy when the firm has observed a single signal in the past. The resulting kink and inaction region are small, and monetary non-neutrality is relatively weak. In the

[33]The empirical hazard function is computed product by product, pooling over retailers within a single market, and then we report the median probability across products.



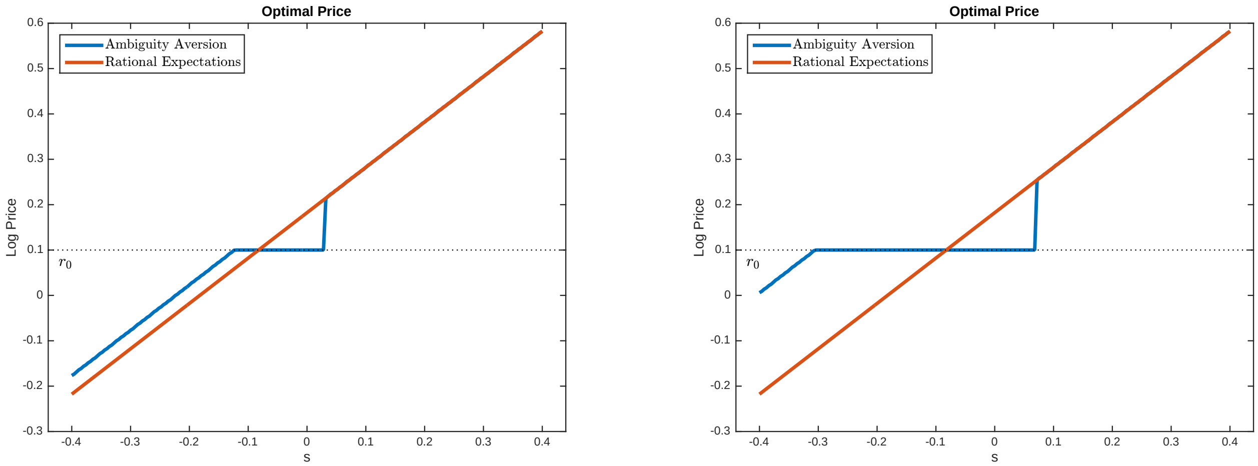
(a) Two past prices, low N (b) Two past prices, high N

Figure 8: Optimal Price Policy, idiosyncratic shocks

right panel, we plot the price policy for a firm that has seen the single reference price five times, and we see a much larger region of inaction and stronger monetary non-neutrality. Moreover, both policy functions show that even conditional on a price change, the ambiguity averse firm is likely to deviate from the flexible price, and thus preserve non-neutrality.

Having multiple observed prices leads to different, potentially non-linear effects of monetary shocks, as we illustrate in Figure 10. The left panel plots the case in which two previous prices have been observed once each. We see that there are two flat areas in the policy function, corresponding to the two past prices, and consequently there is more total inaction compared to the case of single past price. Furthermore, we see that in general there is now a higher probability that the price deviates from the flexible benchmark.

In particular, a monetary policy shocks can have small, large and even negative effects. Consider for example a contractionary monetary shock, starting at $s = -0.1$. As we move to the left, there is strong monetary non-neutrality as prices at first do not change, and then remain above the flexible price level, even conditional on changing. On the other hand, an expansionary monetary shock behaves differently. Initially (i.e. for smaller shocks) there is a significant amount of inaction and thus monetary non-neutrality, but then the optimal price jumps up to a level *above* the flexible price. Thus, instead of under-reacting and generating a positive quantity effect, the price would over-react, and in fact have a negative effect on the average quantity. For even larger shocks, the price would eventually settle at the other kink, and the positive quantity effects can be restored, at least for a while. This is an example of the possibly non-linear monetary shock effects, where small and large shocks can have the



(a) One past price, low N

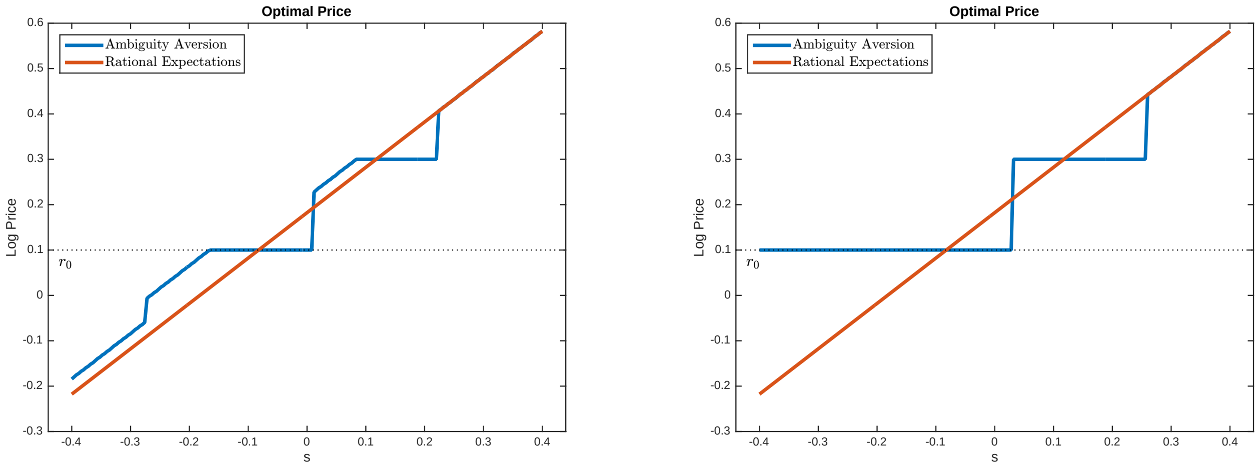
(b) One past price, high N

Figure 9: Optimal Price Policy, monetary shocks

expected positive effects, but moderate shocks might in fact have negative effects.

The right panel plots the policy function in the case where the firm has observed each of the two past prices five times each. We can again see that this results in deeper kinks and larger regions of inaction. Moreover, this panel also illustrates that monetary policy shocks can have asymmetric effects. In general, contractionary shocks might have quite strong monetary non-neutrality, because of the deepness of the lower kink. However, positive shocks (starting from $s = -0.1$ again) would relatively quickly incentivize the firm to change price to its other kink. This generally bring the price close to the flexible price, and thus there appears to be less monetary non-neutrality to the right than to the left.

To summarize, monetary policy shocks have effects that are history and size dependent. History matters because it affects where in the state space the kinks are formed and how large they are. For example, there may be a history of shocks, either idiosyncratic or aggregate, that has generated larger kinks, and in that case the firm will behave *as if* there are significant costs of changing its nominal price, together with potentially strong memory in its price. Alternatively, the firm may find itself in a situation where these kinks are much smaller, and as such monetary non-neutrality is likely to be small. At the same time, for a given history, the current size of the shock matters through the standard effect of pulling the optimal price out of an inaction area. However, when there are multiple kinks, the qualitative and quantitative effect on the sign on the average quantity sold depends on the interaction between the size of the shock and the history-dependent kink formation.



(a) Two past prices, low N (b) Two past prices, high N

Figure 10: Optimal Price Policy, monetary shocks

8 Conclusion

Despite its central role in modern macroeconomic models, a price-setting mechanism that happens to be both plausible and in line with the numerous pricing facts that have been documented in the literature remains elusive. In this paper, we model an uncertainty-averse firm that learns about the demand it faces by observing noisy signals at posted price. The limited knowledge allows the firm to only characterize likely bounds on the possible demand schedules. Since the firm is ambiguity-averse, it acts *as if* the true demand is the one that yields the lowest possible total quantity sold at a given price. In other words, for a price decrease the firm is worried that there will be very little expansion in demand; while it fears a drop in quantity sold if it were to raise its price. This endogenous switch in the worst-case scenario leads to kinks in the expected profit function. This is akin to acting *as if* there is a cost, in terms of expected profits, associated with moving to a new price.

A corollary implication is that because signals are noisy, repeated observations are useful to learn about demand at a specific price. The firm thus finds it beneficial to stick with prices that it has less uncertainty about by having repeatedly posted them in the past. This discrete set of previously observed past prices become 'reference prices' at which there are kinks in the profit function. In addition, we show that if publicly available indicators such as aggregate inflation are ambiguous signals of the price aggregate most relevant for the firm, then our real rigidity becomes nominal in nature and money shocks can have real effects.

Our model naturally predicts that prices should be sticky, unless shocks are sufficiently large. In addition, the proposed mechanism is parsimonious in the sense that it produces a set of overidentifying restrictions that are consistent with stylized facts from micro data: prices exhibit 'memory' as firms find it optimal to stick to a discrete distribution of prices; the probability of observing a price change is decreasing in the time since the last price movement; and small and large price changes coexist in the data.

References

Alvarez, F. E., F. Lippi, and L. Paciello (2011): "Optimal Price Setting With Observation and Menu Costs," *The Quarterly Journal of Economics*, 126, 1909–1960.

Anderson, C. M. (2012): "Ambiguity aversion in multi-armed bandit problems," *Theory and decision*, 72, 15–33.

Bachmann, R. and G. Moscarini (2011): "Business cycles and endogenous uncertainty," Manuscript, Yale University.

Ball, L. and D. Romer (1990): "Real rigidities and the non-neutrality of money," *The Review of Economic Studies*, 57, 183–203.

Balvers, R. J. and T. F. Cosimano (1990): "Actively learning about demand and the dynamics of price adjustment," *The Economic Journal*, 882–898.

Barro, R. J. (1972): "A theory of monopolistic price adjustment," *The Review of Economic Studies*, 39, 17–26.

Bergemann, D. and K. Schlag (2011): "Robust monopoly pricing," *Journal of Economic Theory*, 146, 2527–2543.

Bergemann, D. and J. Valimaki (2008): "Bandit problems," The New Palgrave Dictionary of Economics, 2nd ed. Macmillan Press.

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

Bonomo, M. and C. Carvalho (2004): "Endogenous time-dependent rules and inflation inertia," *Journal of Money, Credit and Banking*, 1015–1041.

Bronnenberg, B., M. Kruger, and C. Mela (2008): "Database paper: The IRI Marketing Data Set," *Marketing Science*, 27, 745–748.

Calvo, G. (1983): "Staggered prices in a utility-maximizing framework," *Journal of Monetary Economics*, 12, 383–398.

Campbell, J. R. and B. Eden (2014): "Rigid prices: Evidence from US scanner data," *International Economic Review*, 55, 423–442.

Carlton, D. W. (1986): "The Rigidity of Prices," *American Economic Review*, 76, 637–658.

Cecchetti, S. G. (1986): "The Frequency of Price Adjustment: A Study of the Newsstand Prices of Magazines," *Journal of Econometrics*, 31, 255–274.

Christiano, L., M. Eichenbaum, and C. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113.

Dow, J. and S. Werlang (1992): "Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio," *Econometrica*, 60, 197–204.

Eichenbaum, M., N. Jaimovich, and S. Rebelo (2011): "Reference Prices, Costs, and Nominal Rigidities," *American Economic Review*, 101, 234–62.

Epstein, L. G. and M. Schneider (2003): "Recursive Multiple-Priors," *Journal of Economic Theory*, 113, 1–31.

Gagnon, E., D. López-Salido, and N. Vincent (2012): "Individual Price Adjustment along the Extensive Margin," *NBER Macroeconomics Annual*, 27, 235–281.

Gertler, M. and J. Leahy (2008): "A Phillips Curve with an Ss Foundation," *Journal of Political Economy*, 116, 533–572.

Gilboa, I. and D. Schmeidler (1989): "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

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

Handel, B. R., K. Misra, and J. W. Roberts (2013): "Robust firm pricing with panel data," *Journal of Econometrics*, 174, 165–185.

Hansen, L. P. (2007): "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97, 1–30.

——— (2014): "Nobel lecture: uncertainty outside and inside economic models," *Journal of Political Economy*, 122, 945–967.

Ilut, C. and M. Schneider (2014): "Ambiguous Business Cycles," *American Economic Review*, 104, 2368–2399.

Jaffray, J.-Y. (1994): "Dynamic Decision Making with Belief Functions," in *Advances in the Dempster-Shafer Theory of Evidence*, ed. by Y. E. Al., Wiley.

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

Kimball, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit, and Banking*, 27.

Klenow, P. J. and O. Kryvtsov (2008): "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent US Inflation?" *The Quarterly Journal of Economics*, 863–904.

Klenow, P. J. and B. A. Malin (2010): "Microeconomic Evidence on Price-Setting," *Handbook of Monetary Economics*, 3, 231–284.

Knotek, I. and S. Edward (2010): "A Tale of Two Rigidities: Sticky Prices in a Sticky-Information Environment," *Journal of Money, Credit and Banking*, 42, 1543–1564.

Lorenzoni, G. (2009): "A Theory of Demand Shocks," *American Economic Review*, 99, 2050–84.

Mackowiak, B. and M. Wiederholt (2009): "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99, 769–803.

Mankiw, N. G. and R. Reis (2002): "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *The Quarterly Journal of Economics*, 117, 1295–1328.

Matějka, F. (2010): "Rationally inattentive seller: Sales and discrete pricing," Cerge-EI Working Paper Series, number 408.

Midrigan, V. (2011): "Menu costs, multiproduct firms, and aggregate fluctuations," *Econometrica*, 79, 1139–1180.

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

——— (2010): "Monetary Non-Neutrality in a Multi-Sector Menu Cost Model," *Quarterly Journal of Economics*, 125, 961–1013.

Nimark, K. (2008): "Dynamic pricing and imperfect common knowledge," *Journal of Monetary Economics*, 55, 365–382.

Pires, C. P. (2002): "A rule for updating ambiguous beliefs," *Theory and Decision*, 53, 137–152.

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

Rotemberg, J. J. (1982): "Monopolistic price adjustment and aggregate output," *The Review of Economic Studies*, 49, 517–531.

Rothschild, M. (1974): "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 9, 185–202.

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

Stevens, L. (2014): "Coarse Pricing Policies," Manuscript, Univ. of Maryland.

Stigler, G. J. (1947): "The kinky oligopoly demand curve and rigid prices," *The Journal of Political Economy*, 432–449.

Stiglitz, J. E. (1979): "Equilibrium in product markets with imperfect information," *The American Economic Review*, 339–345.

Taylor, J. B. (1980): "Aggregate dynamics and staggered contracts," *The Journal of Political Economy*, 88, 1.

Vavra, J. (2014): "Inflation Dynamics and Time-Varying Volatility: New Evidence and an Ss Interpretation," *The Quarterly Journal of Economics*, 129, 215–258.

Willems, T. (2011): "Actively Learning by Pricing: A Model of an Experimenting Seller," Manuscript.

Woodford, M. (2003): "Imperfect Common Knowledge and the Effects of Monetary Policy," *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 25.

——— (2009): "Information-constrained state-dependent pricing," *Journal of Monetary Economics*, 56, S100–S124.

Online Appendix

A Updating with more observed prices

We can readily expand the updating formulas that we have developed in Section 4.2 for one observed price. Assume that firm has seen a whole vector of T previous signals, $\mathbf{y_0}$, with the corresponding vectors of prices $\mathbf{p_0}$ and number of times $\mathbf{N_0}$. The joint distribution with demand at any price p is again jointly Normal

$$\begin{bmatrix} x(p) \\ \mathbf{y_0} \end{bmatrix} \sim N\left(\begin{bmatrix} m(p) \\ m(\mathbf{p_0}) \end{bmatrix}, \Sigma(p, \mathbf{p_0})\right)$$

with

$$\Sigma(p, \mathbf{p_0}) = \begin{bmatrix} \sigma_x^2 & (\sigma_x^2, \ldots, \sigma_x^2) \\ (\sigma_x^2, \ldots, \sigma_x^2)' & \Sigma_x + diag(\frac{\sigma_z^2}{\mathbf{N_0}}) \end{bmatrix}$$

where $(\sigma_x^2, \ldots, \sigma_x^2)$ is a $1xT$ vector, and Σ_x is a TxT matrix with all entries equal to σ_x^2.

The resulting conditional expectation follows from applying the standard formula for conditional Normal expectations:

$$E(x(p)|\mathbf{y_0}) = m(p) + [\sigma_x^2, \ldots, \sigma_x^2](\Sigma_x + diag(\frac{\sigma_z^2}{\mathbf{N_0}}))^{-1}(\mathbf{y_0} - m(\mathbf{p0}))$$

The conditional expectation is again linear in the prior and a weighted sum of the demeaned signals. Expanding the above formula, we obtain

$$E(x(p)|\mathbf{y_0}) = m(p) + \alpha_0(y_{0,1} - m(p_{0,1})) + \cdots + \alpha_T(y_{0,T} - m(p_{0,T}))$$

where $y_{0,i}$ is the i-th element of the vector $\mathbf{y_0}$, and $\alpha_i \in (0, 1)$ is the i-th element of the $1xT$ vector $[\sigma_x^2, \ldots, \sigma_x^2](\Sigma_x + diag(\frac{\sigma_z^2}{\mathbf{N_0}}))^{-1}$.

Without loss of generality, assume the prices in $\mathbf{p_0}$ are sorted and that the last element is the largest price. In building the worst case expectation, one can work from right to left and start with $p' > p_{0,T}$. This is the easiest case, since the firm wants $m^*(p')$ to be the lowest possible so it sets it equal to the lower bound of the prior set, but sets the priors on all observed signals to the upper bound of the prior set

$$m(p) = \begin{cases} \gamma_h - bp & \text{for } p \leq p_{0,T} \\ \gamma_l - bp & \text{for } p > p_{0,T} \end{cases}$$

Next consider, $p' \in (p_{0,T-1}, p_{0,T}]$. As for the case of one observed price, the worst case is when

$m^*(p')$ is low, but the priors on the observed prices are high. So we set $m^*(p_{0,t})$ equal to the upper bound for all $t \leq T-1$, and set $m^*(p_{0,T})$ to the highest admissible value that satisfies the downward sloping restriction, so again $m^*(p_{0,T}; p') = \min(\gamma_l - bp', \gamma_h - bp_{0,T})$. As a result

$$m(p; p') = \begin{cases} \gamma_h - bp & \text{for } p < p' \\ \min(\gamma_l - bp', \gamma_h - bp) & \text{for } p \in (p', p_{0,T}] \\ \gamma_l - bp & \text{for } p > p_{0,T} \end{cases}$$

We can work recursively to the left (i.e. $p' \in (p_{0,T-2}, p_{0,T_1}]$ and so on) and fully characterize the worst-case prior for all possible price choices p'. The general rule is that for any p', the worst-case $m(p)$ for all signals to the left of p' to be at the top of the tunnel, the prior at p' to be at the bottom of the tunnel, and priors on signals to the right of p' to be the highest admissible value that respects a downward sloping $m(p)$.

B Additional Table: Predicting Demand

Table B.1: Predicting demand

		Median	p10	p25	p75	p90
		(1) Across all items				
Spaghetti sauce	Detroit	0.26	0.05	0.12	0.5	0.95
Beer	Boston	0.3	0.05	0.14	0.5	0.87
Frozen pizza	Dallas	0.46	0.07	0.2	0.91	1.63
Peanut butter	Seattle	0.45	0.08	0.2	0.83	1.36
		(2) Item with most sales in category/market				
Salted snacks	Seattle	0.3	0.04	0.11	0.65	1.16
Beer	NYC	0.46	0.17	0.3	0.71	1.23
Frozen dinner	LA	0.48	0.09	0.23	0.84	1.35
Spaghetti sauce	Dallas	0.28	0.05	0.13	0.53	0.9

The dependent variable is $\log(Q)$. Independent variables are: 2 lags of $\log(Q)$, $\log(P)$ + 2 lags; $\log(P)^2$; $\overline{\log(P)}$ + 2 lags; $\overline{\log(P)}^2$; item/store and week dummies, where $\overline{\log(P)}$: weighted average of weekly prices in category/market. The Table reports the moments on the absolute in-sample prediction error: $(Q - X\widehat{\beta})/\overline{Q}$.

C Proofs

Proposition 1. *If the firm has posted a single price p_0 in the past then,*

(i) the price p_0 is sticky. There are values $\underline{c}_0 < \overline{c}_0$ such that p_0 is the optimal price for all cost realizations $c_2 \in [\underline{c}_0, \overline{c}_0]$

(ii) the inaction region around p_0 (i.e. stickiness) increases with α (more precise signal) and $\nu\sigma_z$ (more ambiguity).

Proof. **(i)** Given the piece-wise form of the worst-case expectation in (14), there are three potential local maxima that we need to check: p_0, $p_t^{RE,b}$, and $p_t^{RE,b(1-\alpha)}$. We start by comparing p_0 and $p_t^{RE,b}$. It is useful to define

$$\theta = \ln(\frac{b}{b-1})$$

and also the cost value c_0 such that p_0 would be the optimal price for a RE firm facing a demand curve with slope $-b$, i.e. $p_0 = p_t^{RE,b}(c_0)$:

$$p_0 = \theta + c_0$$

We also define the difference between the current cost and c_0 as

$$\hat{c}_t = c_t - c_0$$

Now we can write the expected profit at p_0 as

$$\begin{aligned}
E^*(\pi(p_0)) &= (\exp(p_0) - \exp(c_t))\exp(\frac{1}{2}\sigma_z^2 + \hat{x}^*(p_0)) \\
&= (\exp(\theta + c_0 - c_t) - 1)\exp(c_t + \frac{1}{2}\sigma_z^2 + (1-\alpha)(\gamma - b(\theta + c_0) - \nu\sigma_z) + \alpha y_0) \\
&= (\exp(\theta - \hat{c}_t) - 1)\exp(c_t - b(1-\alpha)c_0 + \frac{1}{2}\sigma_z^2 + (1-\alpha)(\gamma - b\theta - \nu\sigma_z) + \alpha y_0)
\end{aligned}$$

and write the expected profit at $p_t^{RE,b}$ as:

$$\begin{aligned}
E^*(\pi(p_t^{RE,b})) &= (\exp(p_t) - \exp(c_t))\exp(\frac{1}{2}\sigma_z^2 + \hat{x}^*(p_t^{RE,b})) \\
&= (\exp(\theta) - 1)\exp(c_t + \frac{1}{2}\sigma_z^2 + \gamma - b(\theta + c_t) - \nu\sigma_z - \alpha(\gamma - b(\theta + c_0) + \nu\sigma_z) + \alpha y_0) \\
&= (\exp(\theta) - 1)\exp((1-b)c_t + b\alpha c_0 + \frac{1}{2}\sigma_z^2 + (1-\alpha)(\gamma - b\theta) - (1+\alpha)\nu\sigma_z + \alpha y_0)
\end{aligned}$$

Dividing the two

$$\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b}))} = \frac{(\exp(\theta-\hat{c}_t)-1)\exp(c_t - b(1-\alpha)c_0 + \frac{1}{2}\sigma_z^2 + (1-\alpha)(\gamma - b\theta - \nu\sigma_z) + \alpha y_0)}{(\exp(\theta)-1)\exp((1-b)c_t + b\alpha c_0 + \frac{1}{2}\sigma_z^2 + (1-\alpha)(\gamma - b\theta) - (1+\alpha)\nu\sigma_z + \alpha y_0)}$$
$$= \frac{(\exp(\theta-\hat{c}_t)-1)}{(\exp(\theta)-1)}\exp(b\hat{c}_t + 2\alpha\nu\sigma_z)$$

Notice that this is a continuous function of $\hat{c}_t$ and that at $\hat{c}_t = 0$

$$\left.\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b}))}\right|_{\hat{c}_t=0} = \exp(2\alpha\nu\sigma_z) > 0$$

Moreover, as $\hat{c}_t \to -\infty$ the ratio grows without bound, and as $\hat{c}_t \to \infty$ the ratio turns negative. Since this is a continuous function of $\hat{c}_t$, there exist $\underline{c}_b < \overline{c}$ such that

$$E^*(\pi(p_0)) \geq E^*(\pi(p_t^{RE,b}))$$

for all $c_t \in [\underline{c}_b, \overline{c}]$.

Next we compare the profit at p_0 to the profit at $p_t^{RE,b(1-\alpha)}$. This is straightforward since $p_t^{RE,b(1-\alpha)}$ could be an optimal price only if $p_t^{RE,b(1-\alpha)} \in [p_0 - \frac{2}{b}\nu\sigma_z, p_0]$ – the region in which the demand curve has a slope of $b(1-\alpha)$. Since this section of the demand curve includes p_0 itself, it follows that

$$E^*(\pi(p_t^{RE,b(1-\alpha)})) > E^*(\pi(p_0))$$

$$\Longleftrightarrow$$

$$p_t^{RE,b(1-\alpha)} \in [p_0 - \frac{2}{b}\nu\sigma_z, p_0)$$

There are two possible cases for $p_t^{RE,b(1-\alpha)}$. First, if α is so large that

$$b(1-\alpha) < 1$$

then the slope is less than 1, and hence the optimal price is the maximum admissible price and hence

$$p_t^{RE,b(1-\alpha)} = p_0.$$

In that case, we only need to compare profits at p_0 and $p_t^{RE,b}(c)$ which we had already done before, and can conclude that p_0 is the optimal price for all cost shocks

$$c_t \in [\underline{c}, \overline{c}]$$

where $\underline{c} = \underline{c}_b$.

On the other hand, if $b(1-\alpha) > 1$, let

$$\tilde{\theta} = \ln(\frac{b(1-\alpha)}{b(1-\alpha)-1}$$

and then $p_t^{RE,b(1-\alpha)} = \tilde{\theta} + c_t$. Then the condition $p_t^{RE,b(1-\alpha)} \in [p_0 - \frac{2}{b}\nu\sigma_z, p_0)$ is satisfied if and only if

$$c_t \in [c_0 - (\tilde{\theta} - \theta) - \frac{2}{b}\nu\sigma_z, c_0 - (\tilde{\theta} - \theta))$$

Notice that since $(1-\alpha) < 1$,

$$\tilde{\theta} > \theta$$

and hence

$$c_0 - (\tilde{\theta} - \theta)) < c_0$$

Putting this together with the above result comparing $E^*(\pi(p_0))$ and $E^*(\pi(p_t^{RE,b}))$, we have that p_0 is the optimal price for all cost shocks

$$c_t \in [\underline{c}, \overline{c}]$$

where $\underline{c} = \max(\underline{c}_b, c_0 - (\tilde{\theta} - \theta))$, and we have that $\underline{c} < c_0 < \overline{c}$.

To sum things up, p_0 is the optimal price for all $c_t \in [\underline{c}, \overline{c}]$ where $\overline{c}$ is defined above as the cutoff point at which the price $p_t^{RE,b}$ starts yielding higher profits than p_0 for $c_t > c_0$ and,

$$\underline{c} = \begin{cases} \underline{c}_b & \text{for } b(1-\alpha) < 1 \\ \max(\underline{c}_1, c_0 - (\tilde{\theta} - \theta)) & \text{for } b(1-\alpha) > 1 \end{cases}$$

(ii) From the proof above we know that there exist $\underline{c} < c_0$ and $\overline{c} > c_0$ such that $\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b}))} = 1$. Let $\hat{c} = \overline{c} - c_0$ and take logs of the expected profits ratio when cost equals $\overline{c}$ so that:[34]

$$\ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b}))}) = \ln(\frac{(\exp(\theta - \hat{c}) - 1)}{(\exp(\theta) - 1)}) + b\hat{c} + 2\alpha\nu\sigma_z = 0$$

Applying the implicit function theorem we get

$$\frac{\partial \hat{c}}{\partial \alpha} = -\frac{2\nu\sigma_z}{-\frac{\exp(\theta-\hat{c})}{exp(\theta-\hat{c})-1} + b}$$

[34]Implicitly we are assuming $\hat{c} < \theta$ which is true since at $\hat{c} = \theta$ we have $E^*(\pi(p_0)) = 0$ and thus is clearly below the profit at the price $p_t^{RE,b}$. Hence $\hat{c} < \theta$.

Since 1) $-\frac{\exp(\theta)}{exp(\theta)-1} = -b$ and 2) $\frac{\partial \frac{\exp(\theta-\hat{c})}{\exp(\theta-\hat{c})-1}}{\partial \hat{c}} < 0$, it follows that the denominator is negative when $\hat{c} > 0$, and positive otherwise:

$$-\frac{\exp(\theta-\hat{c})}{\exp(\theta-\hat{c})-1} + b < 0 \iff \hat{c} > 0$$

And since $2\nu\sigma_z > 0$ it follows that

$$\frac{\partial \hat{c}}{\partial \alpha} > 0 \iff \hat{c} > 0$$

Hence increasing α increases $\bar{c}$ and decreases $\underline{c}_b$, which increases the region over which p_0 dominates $p_t^{RE,b}$. If $b(1-\alpha) < 1$ then we are done, since then $\underline{c} = \underline{c}_b$. If $b(1-\alpha) > 1$, notice that

$$\begin{aligned}\frac{\partial(c_0 - (\tilde{\theta}-\theta))}{\partial \alpha} &= -\frac{\partial(\tilde{\theta}-\theta)}{\partial \alpha} \\ &= -\frac{\partial(\ln(\frac{b(1-\alpha)}{b(1-\alpha)-1}))}{\partial \alpha} \\ &= -\frac{1}{(1-\alpha)(b(1-\alpha)-1)} < 0\end{aligned}$$

and hence $\underline{c} = \max(\underline{c}_b, c_0 - (\tilde{\theta}-\theta))$ unambiguously decreases with α as well. Thus, we have shown that the inaction region increases in α.

To prove that the inaction region increases with $\nu\sigma_z$, notice that we can apply the implicit function theorem in a similar way to get

$$\frac{\partial \hat{c}}{\partial(\nu\sigma_z)} = -\frac{2\nu\sigma_z}{-\frac{\exp(\theta-\hat{c})}{exp(\theta-\hat{c})-1} + b} > 0 \iff \hat{c} > 0$$

and we can complete the proof following the same steps as above. □

Proposition 2. *If the firm has previously posted two distinct prices $p_1 > p_0$, then*

1. *there is a kink in the* as if *expected demand at each p_i and each has an associated inaction region, such that p_i is the optimal price for all cost realizations $c_2 \in [\underline{c}_i, \bar{c}_i]$*
2. *the inaction region around each p_i (i.e. stickiness) increases with α_i (the precision of the signal at price p_i)*

Proof. Without loss of generality, we assume $p_1 > p_0$.

(i) First, we characterize the worst-case expectation. Let $\underline{p}_0 = p_0 - \frac{2}{b}\nu\sigma_z$ and $\underline{p}_1 = p_1 - \frac{2}{b}\nu\sigma_z$ be the prices such that the lower bound on the prior of $\underline{p}_i$ equals the upper-bound of the prior tunnel at p_i:

$$\gamma - b\underline{p}_i - \nu\sigma_z = \gamma - bp_i + \nu\sigma_z$$

Then, if $\underline{p}_1 > p_0$ the worst-case expectation is

$$\hat{x}^*(p_t) = \begin{cases} \gamma - bp_t - \nu\sigma_z + \alpha_1(y_1 - (\gamma - bp_1 + \nu\sigma_z)) + \alpha_0(y_0 - (\gamma - bp_0 + \nu\sigma_z)) & \text{for } p < \underline{p}_0,\ p \in (p_0, \underline{p}_1), \text{ or } p > p_1 \\ (1-\alpha_0)(\gamma - bp_t - \nu\sigma_z) + \alpha_0 y_0 + \alpha_1(y_1 - (\gamma - bp_1 + \nu\sigma_z)) & \text{for } p_t \in (\underline{p}_0, p_0] \\ (1-\alpha_1)(\gamma - bp_t - \nu\sigma_z) + \alpha_1 y_1 + \alpha_0(y_0 - (\gamma - bp_0 + \nu\sigma_z)) & \text{for } p_t \in (\underline{p}_1, p_1] \end{cases}$$

and if $\underline{p}_1 < p_0$ we have

$$\hat{x}^*(p_t) = \begin{cases} \gamma - bp_t - \nu\sigma_z + \alpha_1(y_1 - (\gamma - bp_1 + \nu\sigma_z)) + \alpha_0(y_0 - (\gamma - bp_0 + \nu\sigma_z)) & \text{for } p < \underline{p}_0,\ p \in (p_0, \underline{p}_1), \text{ or } p > p_1 \\ (1-\alpha_1)(\gamma - bp_t - \nu\sigma_z) + \alpha_1 y_1 + \alpha_0(y_0 - (\gamma - bp_0 + \nu\sigma_z)) & \text{for } p_t \in (p_0, p_1] \\ (1-\alpha_1-\alpha_0)(\gamma - bp_t - \nu\sigma_z) + \alpha_1 y_1 + \alpha_0 y_0 & \text{for } p_t \in (\underline{p}_1, p_0] \\ (1-\alpha_0)(\gamma - bp_t - \nu\sigma_z) + \alpha_0 y_0 + \alpha_1(y_1 - (\gamma - bp_1 + \nu\sigma_z)) & \text{for } p_t \in (\underline{p}_0, \underline{p}_1] \end{cases}$$

In both cases there is a jump and a kink at both p_1 and p_0. As we show below this leads to regions of inaction around both of those prices. The proof is constructed in a way similar to Proposition 1.

Case 1: $\underline{p}_1 > p_0$. We have 5 candidate optima: $p_0, p_1, p_t^{RE,b}, p_t^{RE,b(1-\alpha_0)}, p_t^{RE,b(1-\alpha_1)}$.

It is again helpful to define the cost values c_0 and c_1 such that

$$p_0 = \theta + c_0$$

$$p_1 = \theta + c_1$$

and

$$\hat{c}_{it} = c_t - c_i$$

Start by comparing the expected profits at $p_t^{RE,b} = \theta + c_t$ and p_0 and p_1 respectively:

$$\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b}))} = \frac{\exp(\theta - \hat{c}_{0t}) - 1}{\exp(\theta) - 1} \exp(b\hat{c}_{0t} + 2\alpha_0\nu\sigma_z) \tag{31}$$

$$\frac{E^*(\pi(p_1))}{E^*(\pi(p_t^{RE,b}))} = \frac{\exp(\theta - \hat{c}_{1t}) - 1}{\exp(\theta) - 1} \exp(b\hat{c}_{1t} + 2\alpha_1\nu\sigma_z) \tag{32}$$

The same analysis as in the proof of Proposition 1 tells us that there exist $\underline{c}_{bi} < \bar{c}_{bi}$ such that $E^*(\pi(p_0)) \geq E^*(\pi(p_t^{RE,b}))$ for all $c_t \in [\underline{c}_{b0}, \bar{c}_{b0}]$, and $E^*(\pi(p_1)) \geq E^*(\pi(p_t^{RE,b}))$ for all $c_t \in [\underline{c}_{b1}, \bar{c}_{b1}]$.

Similarly, if $b(1-\alpha_i) < 1$, then $p_t^{RE,b(1-\alpha_i)} = p_i$, and otherwise $E^*(\pi(p_i)) \geq E^*(\pi(p_t^{RE,b(1-\alpha_i)}))$ if and only if $c_t \geq c_i + (\tilde{\theta}_i - \theta)$, where we define

$$\tilde{\theta}_i = \ln(\frac{b(1-\alpha_i)}{b(1-\alpha_i)-1}$$

Next, we need to compare the profits at p_0 and p_1. That ratio of expected profits is:

$$\begin{aligned}
\frac{E^*(\pi(p_0))}{E^*(\pi(p_1))} &= \frac{\exp(\theta - \hat{c}_{0t}) - 1}{\exp(\theta - \hat{c}_{1t}) - 1} \exp(b(p_1 - p_0) + 2(\alpha_0 - \alpha_1)\nu\sigma_z) \\
&= \frac{\exp(\theta - \hat{c}_{0t}) - 1}{\exp(\theta - \hat{c}_{1t}) - 1} \exp(b(c_1 - c_0) + 2(\alpha_0 - \alpha_1)\nu\sigma_z) \\
&= \frac{\exp(\theta - \hat{c}_{0t}) - 1}{\exp(\theta - \hat{c}_{0t} + (c_1 - c_0)) - 1} \exp(b(c_1 - c_0) + 2(\alpha_0 - \alpha_1)\nu\sigma_z)
\end{aligned}$$

Notice that this is a continuous function of $\hat{c}_{0t}$, that equals 0 for $\hat{c}_{0t} = \theta$, which tells us there are sufficiently high values of c_t such that $E^*(\pi(p_0)) < E^*(\pi(p_1)))$. Next, take logs and derivate in respect to $\hat{c}_{0t}$:

$$\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_1))})}{\partial \hat{c}_{0t}} = -\frac{\exp(\theta - \hat{c}_{0t})}{\exp(\theta - \hat{c}_{0t}) - 1} + \frac{\exp(\theta - \hat{c}_{0t} + (c_1 - c_0))}{\exp(\theta - \hat{c}_{0t} + (c_1 - c_0)) - 1} < 0 \tag{33}$$

where the inequality follows from $c_1 > c_0$ and the fact that $\frac{\exp(\theta+x)}{\exp(\theta+x)-1}$ is a decreasing function of x. Hence the ratio of expected profits is a monotonically decreasing function of $\hat{c}_{0t}$ and thus there can be at most only one crossing point $\tilde{c}$ such that the profits at p_0 exceed the profits at p_1. The limit as $\hat{c}_{0t} \to -\infty$ is

$$\lim_{\hat{c}_{0t} \to -\infty} \left(\frac{\exp(\theta - \hat{c}_{0t}) - 1}{\exp(\theta - \hat{c}_{1t}) - 1} \exp(b(p_1 - p_0) + 2(\alpha_0 - \alpha_1)\nu\sigma_z) \right) = \exp((b-1)(c_1 - c_0) + 2(\alpha_0 - \alpha_1)\nu\sigma_z)$$

Which is greater than 1 if and only if

$$(b-1)(c_1 - c_0) + 2(\alpha_0 - \alpha_1)\nu\sigma_z > 0$$

If this condition holds, then there indeed exists a $\tilde{c}$ such that

$$E^*(\pi(p_0)) > E^*(\pi(p_1)) \iff c_t < \tilde{c}$$

otherwise, if $(b-1)(c_1 - c_0) + 2(\alpha_0 - \alpha_1)\nu\sigma_z \leq 0$ then

$$E^*(\pi(p_0)) \leq E^*(\pi(p_1)) \text{ for all } c_t$$

It is possible to be in a situation where p_0 is never an optimal price, and p_1 always dominates it.

Next, we compare p_0 with $p_t^{RE,b(1-\alpha_1)}$ and p_1 with $p_t^{RE,b(1-\alpha_0)}$. Starting with the first pair, recall that $p_t^{RE,b(1-\alpha_1)}$ is a potential optimum only for $c_t \in [c_1 - (\tilde{\theta}_1 - \theta) - \frac{2}{b}\nu\sigma_z, c_1 - (\tilde{\theta}_1 - \theta)]$ and thus if $\tilde{c} < c_1 - (\tilde{\theta}_1 - \theta) - \frac{2}{b}\nu\sigma_z$, then

$$E^*(\pi(p_0)) < E^*(\pi(p_1)) \leq E^*(\pi(p_t^{RE,b(1-\alpha_1)}))$$

for all c_t for which $p_t^{RE,b(1-\alpha_1)}$ is a potential optimum. Next, consider $\tilde{c} \geq c_1 - (\tilde{\theta}_1 - \theta) - \frac{2}{b}\nu\sigma_z$. The ratio between the expected profits is

$$\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_1)}))} = \frac{\exp(\theta - \hat{c}_{0t}) - 1}{\exp(\tilde{\theta}_1) - 1} \exp((1-\alpha_1)b(\tilde{\theta}_1 - \theta) + b(1-\alpha_1)\hat{c}_{0t} + \alpha_1 b(c_1 - c_0) + 2\nu\sigma_z(\alpha_0 - \alpha_1))$$

The first and second derivatives are:

$$\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_1)}))})}{\partial \hat{c}_{0t}} = -\frac{\exp(\theta - \hat{c}_{0t})}{\exp(\theta - \hat{c}_{0t}) - 1} + b(1 - \alpha_1)$$

$$\frac{\partial^2 \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_1)}))})}{(\partial \hat{c}_{0t})^2} = -\frac{\exp(\theta - \hat{c}_{0t})}{(\exp(\theta - \hat{c}_{0t}) - 1)^2} < 0$$

Since this is a concave function, we can find it's maximum by setting the first derivative equal to zero. This achieved at

$$\hat{c}_{0t} = \ln(\frac{b(1 - \alpha_1) - 1}{(b - 1)(1 - \alpha_1)}) < 0$$

Evaluating the profits ratio at that cost value we get

$$\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_1)}))} = \exp(\alpha_1(b(c_1 - c_0) - 2\nu\sigma_z) + 2\nu\sigma_z\alpha_0) > 1$$

Where the inequality follows from that fact that $\underline{p}_1 > p_0$, which implies that

$$0 < \underline{p}_1 - p_0 < c_1 - c_0 - \frac{2}{b}\nu\sigma_z$$

Since this is a convex function with a maximum above zero, it crosses zero at two distinct points, $\underline{c}_{(1-\alpha_1)b}$ and $\overline{c}_{(1-\alpha_1)b}$. And because the maximum is obtained at a cost value below c_0, we know that $\underline{c}_{(1-\alpha_1)b} < c_0$, but since $p_t^{RE,b(1-\alpha_1)}$ is a relevant maximum only for $c_t \geq c_1 - (\tilde{\theta}_1 - \theta) - \frac{2}{b}\nu\sigma_z > c_0$, then $E^*(\pi_0) < E^*(p_t^{RE,(1-\alpha_1)b})$ if and only if $c_t > \overline{c}_{b(1-\alpha_1)}$ and $\overline{c}_{b(1-\alpha_1)} \in [c_1 - (\tilde{\theta}_1 - \theta) - \frac{2}{b}\nu\sigma_z, c_1 - (\tilde{\theta}_1 - \theta)]$. To ease notation, we adopt the convention that $\overline{c}_{b(1-\alpha_1)} = \infty$ if $\overline{c}_{b(1-\alpha_1)} \notin [c_1 - (\tilde{\theta}_1 - \theta) - \frac{2}{b}\nu\sigma_z, c_1 - (\tilde{\theta}_1 - \theta)]$.

Similarly, when comparing p_1 and $p_t^{RE,b(1-\alpha_0)}$, we note that $p_t^{RE,b(1-\alpha_0)}$ can only be the optimal price for $c_t \in [c_0 - (\tilde{\theta}_0 - \theta) - \frac{2}{b}\nu\sigma_z, c_0 - (\tilde{\theta}_0 - \theta)]$. If $\tilde{c} > c_0 + (\tilde{\theta}_0 - \theta)$ then $E^*(\pi(p_1)) < E^*(\pi(p_t^{RE,b(1-\alpha_0)}))$ for all $c_t \in [c_0 - (\tilde{\theta}_0 - \theta) - \frac{2}{b}\nu\sigma_z, c_0 - (\tilde{\theta}_0 - \theta)]$. Otherwise, the the ratio of the expected profits is:

$$\frac{E^*(\pi(p_1))}{E^*(\pi(p_t^{RE,b(1-\alpha_0)}))} = \frac{\exp(\theta - \hat{c}_{1t}) - 1}{\exp(\tilde{\theta}_0) - 1} \exp((1-\alpha_0)b(\tilde{\theta}_0 - \theta) + b(1-\alpha_0)\hat{c}_{1t} - \alpha_1 b(c_1 - c_0) - 2\nu\sigma_z(\alpha_0 - \alpha_1))$$

which we can again show that is strictly concave in $\hat{c}_{1t}$ (and as a result in c_t) and thus there exist at most two points, $\underline{c}_{b(1-\alpha_0)}$ and $\overline{c}_{b(1-\alpha_0)}$, where the ratio crosses 1. And adopting a similar convention as above, that we set $\underline{c}_{b(1-\alpha_0)} = -\infty$ if $\underline{c}_{b(1-\alpha_0)} \notin [c_0 - (\tilde{\theta}_0 - \theta) - \frac{2}{b}\nu\sigma_z, c_0 - (\tilde{\theta}_0 - \theta)]$, and hence

$$E^*(\pi_1) \geq E^*(p_t^{RE,(1-\alpha_0)b}) \iff c_t \geq \underline{c}_{(1-\alpha_0)b}$$

Putting everything together, we conclude that p_0 is optimal for all

$$c_t \in [\underline{c}_0, \overline{c}_0]$$

where $\underline{c}_0 = \max(\underline{c}_{b0}, c_0 - (\tilde{\theta}_0 - \theta))$ and $\overline{c}_0 = \min(\overline{c}_{b0}, \tilde{c}, \overline{c}_{b(1-\alpha_1)})$. And p_1 is optimal for all

$$c_t \in [\underline{c}_1, \overline{c}_1]$$

where $\underline{c}_1 = \max(\underline{c}_{b1}, c_1 - (\tilde{\theta}_1 - \theta), \tilde{c}, \underline{c}_{b(1-\alpha_0)})$ and $\overline{c}_1 = \overline{c}_{b1}$.

Case 2: $\underline{p}_1 \leq p_0$. Everything is the same except for the fact that we need to re-work the comparison between p_0 and p_1 and between p_0 and $p_t^{RE,b(1-\alpha_0)}$, and also need to compare p_0 and p_1 with $p_t^{RE,(1-\alpha_0-\alpha_1)}$.

We can show that the ratio of expected profits at p_0 and p_1 yields:

$$\frac{E^*(\pi(p_0))}{E^*(\pi(p_1))} = \frac{\exp(\theta - \hat{c}_{0t}) - 1}{\exp(\theta - \hat{c}_{0t} + (c_1 - c_0)) - 1} \exp(b(1-\alpha_1)(c_1 - c_0) + 2\alpha_0\nu\sigma_z)$$

which is again a decreasing function of $\hat{c}_{0t}$ and by similar analysis as above, we can conclude that if $(b(1-\alpha_1) - 1)(c_1 - c_0) + 2\alpha_0\nu\sigma_z < 0$ then $E^*(\pi(p_0)) < E^*(\pi(p_1))$ for all c_t, and otherwise there exists a $\tilde{c}$ such that

$$E^*(\pi(p_0)) < E^*(\pi(p_1)) \iff c_t > \tilde{c}$$

.

In comparing p_0 to $p_t^{RE,b(1-\alpha_0-\alpha_1)}$, first notice that $p_t^{RE,b(1-\alpha_0-\alpha_1)}$ is a relevant potential maximum only for $c_t \in [c_0 - (\tilde{\theta}_{01} - \theta) - \frac{2}{b}\nu\sigma_z, c_0 - (\tilde{\theta}_{01} - \theta)]$, where we define

$$\tilde{\theta}_{01} = \ln(\frac{b(1-\alpha_0-\alpha_1)}{b(1-\alpha_0-\alpha_1)-1})$$

and we again have that $p_0 = p_t^{RE,b(1-\alpha_0-\alpha_1)}$ for $c_t = c_0 - (\tilde{\theta}_{01} - \theta)$. Then, by a similar analysis to the above we conclude that

$$E^*(\pi(p_0)) < E^*(\pi(p_t^{RE,b(1-\alpha_0-\alpha_1)})) \iff c_t \in [c_0 - (\tilde{\theta}_{01} - \theta) - \frac{2}{b}\nu\sigma_z, c_0 - (\tilde{\theta}_{01} - \theta))$$

To compare p_1 and $p_t^{RE,(b(1-\alpha_0-\alpha_1)}$, we look at the ratio of their expected profit:

$$\frac{E^*(\pi(p_1))}{E^*(\pi(p_t^{RE,b(1-\alpha_0-\alpha_1)}))} = \frac{\exp(\theta - \hat{c}_{1t}) - 1}{\exp(\tilde{\theta}_{01}) - 1} \exp((1-\alpha_0-\alpha_1)b(\tilde{\theta}_{01} - \theta) + b(1-\alpha_0-\alpha_1)\hat{c}_{1t} - \alpha_0 b(c_1 - c_0) - 2\nu\sigma_z\alpha_0)$$

Which is again a concave function of $\hat{c}_{1t}$, and by similar analysis to the above we can show that at the maximum $\hat{c}_{1t}$ the profits ratio is equal to

$$\left.\frac{E^*(\pi(p_1))}{E^*(\pi(p_t^{RE,b(1-\alpha_0-\alpha_1)}))}\right|_{\hat{c}_{1t}=\hat{c}_{1t}^*} = \exp(-\alpha_0 b(c_1 - c_0) - 2\nu\sigma_z\alpha_0) < 1$$

So p_1 is always dominated by $p_t^{RE,b(1-\alpha_0-\alpha_1)}$, but that is a relevant comparison only for $c_t \in [c_0 - (\tilde{\theta}_{01} - \theta) - \frac{2}{b}\nu\sigma_z, c_0 - (\tilde{\theta}_{01} - \theta)]$. For values of c_t greater than this interval, the relevant comparison is between p_0 and p_1 which we have already addressed above.

Lastly, we turn our attention to comparing p_0 and $p_t^{RE,b(1-\alpha_0)}$. We can express the ratio of those profits as

$$\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_0)}))} = \frac{\exp(\theta - \hat{c}_{0t}) - 1}{\exp(\tilde{\theta}_0) - 1} \exp((1-\alpha_0)b(\tilde{\theta}_0 - \theta) + b(1-\alpha_0)\hat{c}_{0t} - \alpha_1 b(c_1 - c_0) + 2\nu\sigma_z\alpha_1)$$

which is strictly concave in $\hat{c}_{0t}$ and at the maximum:

$$\left.\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_0)}))}\right|_{\hat{c}_{0t}=\hat{c}_{0t}^*} = \exp(\alpha_1(2\nu\sigma_z - b(c_1 - c_0)) > 1$$

Then, by the strict concavity there can exist at most two values $\underline{c}_{b(1-\alpha_0),0} < \overline{c}_{b(1-\alpha_0),0}$ such that the ratio crosses 1, and similarly to above, only the lower value $\underline{c}_{b(1-\alpha_0),0}$ is relevant in comparing profits, since $p_t^{RE,b(1-\alpha_0)}$ is a potential maximum only for $c_t \in [c_0 - (\tilde{\theta}_0 - \theta) - \frac{2}{b}\nu\sigma_z, c_0 - (\tilde{\theta}_0 - \theta)]$.

Thus,

$$E^*(\pi(p_0)) > E^*(\pi(p_t^{RE,b(1-\alpha_0)})) \iff c_t \geq \underline{c}_{b(1-\alpha_0),0}$$

Putting everything together, we conclude that p_0 is optimal for all

$$c_t \in [\underline{c}_0, \overline{c}_0]$$

where $\underline{c}_0 = \max(\underline{c}_{b0}, c_0 - (\tilde{\theta}_{01} - \theta), \underline{c}_{b(1-\alpha_0),0})$ and $\overline{c}_0 = \min(\overline{c}_{b0}, \tilde{c}, \overline{c}_{b(1-\alpha_1)})$. And p_1 is optimal for all

$$c_t \in [\underline{c}_1, \overline{c}_1]$$

where $\underline{c}_1 = \max(\underline{c}_{b1}, c_1 - (\tilde{\theta}_1 - \theta), \tilde{c}, \underline{c}_{b(1-\alpha_0),1}, c_0 - (\tilde{\theta}_{01} - \theta))$ and $\overline{c}_1 = \overline{c}_{b1}$.

Part (ii) **Case 1:** $\underline{p}_1 > p_0$. The steps are similar to part (ii) of the proof of Proposition 1. Using the ratio of profits at p_0 and $p_t^{RE,b}$ (equation (31)) and the Implicit Function Theorem, it follows that:

$$\frac{\partial(\overline{c}_{b_0} - c_0)}{\partial \alpha_0} = -\frac{2\nu\sigma_z}{-\frac{\exp(\theta - (\overline{c}_{b_0} - c_0)}{\exp(\theta - (\overline{c}_{b_0} - c_0) - 1} + b} > 0$$

$$\frac{\partial(\underline{c}_{b_0} - c_0)}{\partial \alpha_0} = -\frac{2\nu\sigma_z}{-\frac{\exp(\theta - (\underline{c}_{b_0} - c_0)}{\exp(\theta - (\underline{c}_{b_0} - c_0) - 1} + b} < 0$$

Thus the inaction in comparison with $p_t^{RE,b}$ increases in α_0. Moreover,

$$\frac{\partial(c_0 + (\tilde{\theta}_0 - \theta))}{\partial \alpha_0} = -\frac{1}{(1-\alpha_0)(b(1-\alpha_0) - 1)} < 0$$

and hence inaction in respect to $p_t^{RE,b(1-\alpha_0)}$ is increasing in α_0.

Next, turning to the comparison between p_0 and p_1, notice that

$$\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_1))})}{\partial \alpha_0} = 2\nu\sigma_z > 0$$

and by equation (33) $\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_1))})}{\partial \hat{c}} < 0$ and hence, by the implicit function theorem

$$\frac{\tilde{c}}{\alpha_0} = -\frac{\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_1))})}{\partial \alpha_0}}{\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_1))})}{\partial \hat{c}}} > 0$$

we see that the tipping point at which $E^*(p_1) = E^*(p_0)$ increases with α_0. Lastly, we turn to comparing p_0 and $p_t^{RE,b(1-\alpha_1)}$. The derivative of the log of the ratio of expected profits is

$$\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_1)}))})}{\partial \hat{c}_{0t}}\Bigg|_{\hat{c}_{0t}=\bar{c}_{b(1-\alpha_1)}-c_0} = -\frac{\exp(\theta-\hat{c}_{0t})}{\exp(\theta-\hat{c}_{0t})-1}\Bigg|_{\hat{c}_{0t}=\bar{c}_{b(1-\alpha_1)}-c_0} + b(1-\alpha_1) < 0$$

since $\bar{c}_{b(1-\alpha_1)} - c_0 > 0$. At the same time,

$$\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_1)}))})}{\partial \alpha_0} = 2\nu\sigma_z > 0$$

so, by applying the Implicit Function Theorem again, we see that inaction is increasing in α_0. Putting all of this together, we see that

$$\frac{\partial \underline{c}_0}{\partial \alpha_0} < 0 \; ; \; \frac{\partial \bar{c}_0}{\partial \alpha_0} > 0$$

and hence the inaction region around p_0 is increasing in α_0. We can show the symmetric result for p_1 following the same steps as above.

<u>Case 2:</u> $\underline{p}_1 \le p_0$. Only a few things change. First, the ratio of profits at p_0 and p_1 is slightly different,

$$\frac{E^*(\pi(p_0))}{E^*(\pi(p_1))} = \frac{\exp(\theta-\hat{c}_{0t})-1}{\exp(\theta-\hat{c}_{0t}+(c_1-c_0))-1}\exp(b(1-\alpha_1)(c_1-c_0)+2\alpha_0\nu\sigma_z)$$

but the derivatives in respect to $\hat{c}$ and α_0 remain the same, so the above analysis again implies that the tipping point is increasing in α_0. Next, recall that $p_t^{RE,b(1-\alpha_0-\alpha_1)}$ dominates p_0 for all $c_t \le c_0 - (\tilde{\theta}_{01} - \theta)$, and notice that

$$\frac{\partial \tilde{\theta}_{01}}{\alpha_0} = \frac{1}{1-\alpha_0-\alpha_1} > 0$$

and hence inaction increases with α_0. Lastly, we need to compare p_0 and $p_t^{RE,b(1-\alpha_0)}$. Recall that the log of the ratio of their expected profits is:

$$\ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_0)}))}) = \ln(\frac{\exp(\theta-\hat{c}_{0t})-1}{\exp(\tilde{\theta}_0)-1}) + (1-\alpha_0)b(\tilde{\theta}_0-\theta) + b(1-\alpha_0)\hat{c}_{0t} - \alpha_1 b(c_1-c_0) + 2\nu\sigma_z\alpha_1$$

The derivative in respect to $\hat{c}$ is

$$\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_0)}))})}{\partial \hat{c}} = -\frac{\exp(\theta-\hat{c})}{\exp(\theta-\hat{c})-1} + b(1-\alpha_0)$$

But since $p_t^{RE,b(1-\alpha_0)}$ is a potential maximum only for $c_t \leq c_0 - (\tilde{\theta}_0 - \theta)$, it follows that

$$\hat{c} = c_t - c_0 \leq -(\tilde{\theta}_0 - \theta) < 0$$

and since $\frac{\exp(\theta-\hat{c})}{\exp(\theta-\hat{c})-1}$ is a decreasing function of $\hat{c}$, it follows that

$$\frac{\exp(\theta - \hat{c})}{\exp(\theta - \hat{c}) - 1} \leq \frac{\exp(\tilde{\theta}_0)}{\exp(\tilde{\theta}_0) - 1} = b(1 - \alpha_0)$$

and hence for the relevant cost values:

$$\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_0)}))})}{\partial \hat{c}} \geq 0$$

On the other hand, the derivative in respect to α_0 is

$$\begin{aligned}
\frac{\partial \ln(\frac{E^*(\pi(p_0))}{E^*(\pi(p_t^{RE,b(1-\alpha_0)}))})}{\partial \alpha_0} &= -\frac{\exp(\tilde{\theta}_0)\frac{1}{1-\alpha_0}}{\exp(\tilde{\theta}_0) - 1} - b(\tilde{\theta}_0 - \theta) + b - b\hat{c} \\
&> -b(\tilde{\theta}_0 - \theta) + b(\tilde{\theta}_0 - \theta) \\
&> 0
\end{aligned}$$

where the first equality follows from the fact that $\hat{c} \leq -(\tilde{\theta}_0 - \theta)$. Applying the inverse function theorem again, we have that

$$\frac{\partial \underline{c}_{b(1-\alpha_0)}}{\alpha_0} < 0$$

and hence the inaction region is increasing in α_0. Putting it all together,

$$\frac{\partial \underline{c}_0}{\partial \alpha_0} < 0 \ ; \ \frac{\partial \overline{c}_0}{\partial \alpha_0} > 0$$

and hence the inaction region around p_0 is increasing in α_0. We can again apply symmetric arguments to obtain the corresponding result that the inaction around p_1 is increasing in α_1. The only difference is in the comparison between p_1 and $p_t^{RE,b(1-\alpha_1-\alpha_0)}$. Recall that p_1 is always dominated by $p_t^{RE,b(1-\alpha_0-\alpha_1)}$, which is a relevant comparison only for $c_t \leq c_0 - (\tilde{\theta}_{01} - \theta)$, and since $\tilde{\theta}_{01}$ is increasing in α_1, it follows that the range of cost shocks for which p_1 might be an optimal price increases. □

Proposition 3. *Optimal prices have the following characteristics:*

(i) ***Discreteness and Memory.*** *If the two previously observed prices are distinct $p_1 \neq p_0$, then there is a positive probability that a price change results in a discrete move within the set of observed prices, exhibiting both discreteness and memory.*

(ii) ***Declining Hazard.*** *Increasing the number of times the firm has observed the price* p_1 *increases its region of inaction and hence the probability that the firm remains at* p_1.

(iii) ***Large and Small Changes.*** *Optimal price adjustment is characterized by both discrete jumps and arbitrarily small price movements.*

Proof. **Part (i)** By Proposition 2, part (i), both past prices have associated intervals of cost shocks, $[\underline{c}_i, \overline{c}_i]$ for $i \in \{0, 1\}$, such that p_i is the optimal choice for all cost shocks $c_t \in [\underline{c}_i, \overline{c}_i]$. Let c_1 be the particular marginal cost the firm faced at time 1 and $g(c_2|c_1)$ be the conditional pdf of the marginal cost at time 2. Thus, the probability that at time 2 the firm finds it optimal to switch form p_1 back to p_0 is simply

$$Prob(p_2^* = p_0|c_1) = \int_{\underline{c_0}(c_1)}^{\overline{c}_0(c_1)} g(c_2|c_1)dc_2 > 0$$

where $[\underline{c_0}(c_1), \overline{c}_0(c_1)]$ is the particular region of inaction associated with p_0, given that the firm has faced a cost shock c_1 at time 1. Hence, conditional on a cost value c_1, there is a positive probability that the optimal price at time 2, p_2^*, switches back to p_0. In other words, the distribution of price changes at time 2 features a mass point at $p_0 - p_1$, and price changes display discrete changes. Moreover, there is memory, since the discrete change reverts back to a price posted in the past.

This analysis was conditional on a particular cost value c_1, but it is straightforward to extend it by integrating over the possible values of c_1:

$$Prob(p_2^* = p_0) = \int_{(-\infty, \underline{c}_0) \cup (\overline{c}_0, \infty)} \left(\int_{\underline{c_0}(c_1)}^{\overline{c}_0(c_1)} g(c_2|c_1)dc_2 \right) g(c_1)dc_1 > 0$$

where we integrate only over cost values c_1 that would result in $p_1 \neq p_0$. But the basic result is the same – the optimal price at time 2 has a positive probability of reverting back to p_0, implying now that the unconditional distribution of prices at time 2 is discrete and displays memory.

Part (ii) By Proposition 2, part(ii), the regions of inaction associated with the observed prices p_i is increasing in α_i. Notice that if N_1 is the number of times the firm has observed signals at the price p_1 in the past, and N_0 is the number of times the firm has observed the price p_0, then the resulting signal to noise ratio of the average signal at p_1 is:

$$\alpha_1 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2/N_1 + \frac{N_0}{N_1}\sigma_x^2}$$

Clearly, increasing N_1 increases α_1 as it decreases the variance of the error in the average signal. As a result, by the results in Proposition 2 part (ii), increasing N_1 increases the inaction region $[\underline{c}_1, \overline{c}_1]$, and hence the probability that the firm remains at p_1.

Part (iii) The fact the price distribution features discrete jumps follows from **(i)**. To complete the proof, we'll show that for any $\varepsilon > 0$, there are situations in which the firm finds it optimal to change its price by less than ε. Let $p_1 < p_0$, and be far enough apart so that

$$\frac{E^*(\pi(p_1))}{E^*(\pi(p_0))} > 1$$

for all $c_t < \tilde{c}$ where $\tilde{c} > c_0 - (\tilde{\theta}_0 - \theta) - \frac{2}{b}\nu\sigma_z$. In that case, we know that there is a cost value $\underline{c}_{b(1-\alpha_0)}$ such that

$$\frac{E^*(\pi(p_1))}{E^*(\pi(p_t^{RE,b(1-\alpha_0)})} > 1$$

for all $c_t < \underline{c}_{b(1-\alpha_0)}$. Lastly, assuming that $\underline{c}_{b1} < c_1 - (\tilde{\theta}_1 - \theta)$ it follows that there exists a $\underline{c} < c_1 - (\tilde{\theta}_1 - \theta)$ such that $p_t^{RE,b(1-\alpha_1)}$ is the optimal price for all $c_t \in [\underline{c}, c_1 - (\tilde{\theta}_1 - \theta)]$. As a result, for any $\varepsilon > 0$, we can find a $c_t > \underline{c}$ such that

$$c_1 - (\tilde{\theta}_1 - \theta) - c_t < \varepsilon$$

and thus the optimal price switches from p_1 to

$$p_t^{RE,b(1-\alpha_1)} = \tilde{\theta}_1 + c_t$$

However, notice that

$$p_1 - p_t^{RE,b(1-\alpha_1)} = \theta + c_1 - (\tilde{\theta}_1 + c_t) < \varepsilon$$

Thus, there are situations in which the optimal price changes by less than an arbitrary ε, and hence the price distribution features arbitrarily small price changes.

$\square$