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

Cosmin Ilut
Duke University & NBER

Rosen Valchev
Boston College

Nicolas Vincent
HEC Montreal

October 2015

Abstract

The degree of rigidity of nominal variables is central to many predictions of modern macroeconomic models. Yet, standard models of price stickiness 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. It learns non-parametrically about the underlying, uncertain demand and makes robust pricing decisions. The non-parametric learning leads to kinks in the expected profit function at previously observed prices, which generate price stickiness and a discrete price distribution. In addition, we show that when the ambiguity-averse firm worries that aggregate inflation is an ambiguous signal of the prices of its direct competitors in the short run, our rigidity becomes nominal in nature.

*Email addresses: Ilut cosmin.ilut@duke.edu, Valchev: valchev@bc.edu, Vincent: nicolas.vincent@hec.ca.
We would like to thank 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", "Behavioral Aspects of Macroeconomics and Finance" Conference, Boston Macro Juniors Workshop, ESWC, NBER Summer Institute, SED, Stanford, UC San Diego, 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 stylized facts that can help us build more realistic and robust macroeconomic models. 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. Coupled with ambiguity aversion, this single mechanism does not only endogenously generate price stickiness, but can also rationalize a number of other salient pricing facts.

One of the earliest documented empirical facts 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)). 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. Yet other facts, such as the surprising 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 this paper, we propose a single, parsimonious mechanism, built around the idea of demand uncertainty, that can rationalize these robust empirical facts. In our model, the economy is composed of a continuum of industries, each populated with monopolistic firms that face Knightian uncertainty about their competitive environment. In particular, a firm does not know the production function that produces the final product of its industry, which leads to two important implications. First, there is uncertainty about the shape of the firm's demand function, 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 a temporary, price-insensitive demand shock and an underlying, time invariant price-sensitive component. 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

[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.

a signal extraction problem. Furthermore, firms are not confident that demand belongs to a single parametric family, but rather use their noisy signals to build non-parameteric estimates of their demand curves. We thus put the decision maker in charge of the pricing action on the same footing as an econometrician outside the firm that attempts to estimate non-parametrically the demand curve.[2]

We assume that the firm has enough prior knowledge to put some loose bounds on the possible demand schedules, but not enough to impose functional form restrictions or to put a single probability measure on the admissible demand functions. Thus, the firm faces Knightian uncertainty about the probability distribution of demands inside these bounds and is averse to this ambiguity. The firm (or 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 posted 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.

The non-parametric nature of the learning implies that uncertainty reduction is local, not global, and this generates kinks in the subjective beliefs about demand. Unlike the case of updating beliefs about the parameters of a given function, for example, the arrival of new information does not change the perception of demand at all prices. By observing noisy signals about the underlying demand at a given posted price, the firm primarily reduces the demand uncertainty associated with the price-sensitive component at that price, while remaining uncertain about the quantity it could sell at other prices. In essence, that signal is given increasingly smaller weight in updating beliefs about demand at prices further away from that posted price. This generates kinks in the subjective beliefs about demand 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

[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).

generates 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.[3]

A corollary implication is that 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 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 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 tracks the optimal flexible price. 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.

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 on worst-case basis, 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

[3]We emphasize that these effects are present for any uncertainty-averse agent, and not only in the case of ambiguity aversion. The kinks in uncertainty are a result of the non-parametric learning, and not due to the preferences of the agent. Under risk-aversion, for example, the kinks are in the posterior variance, but still have the same effect of introducing regions of inaction.

inflation. Hence, in addition to not knowing the demand function, the firm is also uncertain about its appropriate argument.

In this context, the firm understands that its demand is ambiguous in two dimensions. First, the demand function itself is ambiguous, and second, the relative price argument of the function is now also ambiguous. The firm sets an optimal nominal pricing action that is robust to this two-dimensional uncertainty. The firm thus acts *as if* nature draws the true DGP to be the relationship between aggregate prices and industry prices that implies the lowest possible demand for any given combination of the non-ambiguous choice of the firm – own nominal price versus the last observed industry price level. The nature reaction defines a worst-case demand schedule that is a function of this non-ambiguous relative price. The firm uses its past sales and unambiguously defined price relative to the latest signal to estimate this worst-case demand. Intuitively, the ambiguity about the industry prices makes one of the arguments of demand ambiguous, and the robust action is to consider the worst case conditional on the non-ambiguous arguments of demand. Since the review signals arrive periodically, and when they are unchanged 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. First, prices in the model can be rigid in the face of shocks despite the absence of ad hoc costs to changing prices. Second, the firm finds it optimal to stick to a discrete distribution of prices. This implies that unlike standard models, our mechanism is also compatible with the pricing strategy of many retail firms to alternate between a regular and a sale price. Finally, 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 relation to literature. In Section 3 we present some motivating evidence. Sections 4 presents 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. In Section 7 we present a general equilibrium version.

2 Relation to literature

By connecting learning under ambiguity to the problem of a firm setting prices, this paper is related to a range of literature strands. The economic question that we address in this paper, of price rigidity, has generated a huge 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 produce kinks in the expected demand and as such is related to theoretical work on real price rigidities 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 from the uncertainty about demand. Thus, in our model, while an econometrician would not need to find evidence of actual kinks in the demand curve, the firm behaves as if they are. A further important economic distinction is that in our model the size and the location of the kinks are a function of information accumulated at observed prices.

Some papers generate sticky prices as a result of consumers' reaction to price changes, such as those based on kinked demand, on consumer-anger (Rotemberg (2005), Rotemberg (2011)), or on consumers' imperfect information (L′Huillier (2011), Nakamura and Steinsson (2011) and Matějka (2015)). In these models, by the assumption of expected utility, firms have full confidence in their view on the consumers' reaction functions. The different environments proposed in these models can be seen as examples of possible consumer demand functions that are taken as unknown and potentially complex by the firm in our model.

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),

Sheshinski and Weiss (1977), Rotemberg (1982))[4], or a cost of information acquisition present in more recent models of rational inattention (Woodford (2009)).[5] 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 gives rise to the economic mechanisms that can help rationalize, within a parsimonious explanation, a set of otherwise puzzling pricing facts, such as price discreteness, memory, small and large price changes and a decreasing hazard function.[6]

The other major strand of literature that we relate to is the 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 within the Bayesian updating framework. An early contribution is that of Rothschild (1974), who frames the learning process as a two-arm bandit problem,[7] while more recent work includes Balvers and Cosimano (1990), Bachmann and Moscarini (2011) and Willems (2011). 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. Different from this approach, in our model the firm learns about non-parametric functions and establishes likely bounds by reducing uncertainty mostly locally, around the observed prices. Coupled with aversion to this uncertainty, our learning mechanism generates kinks in the uncertainty-adjusted expected profits.

Lastly, we connect to the literature on ambiguity aversion. We use the recursive multiple priors preferences to capture the notion that the firm is not confident in the probability assessments over various demand curves. Some recent work analyzes a firm pricing problem under a related ambiguity-aversion preference, namely maxmin regret (Handel et al. (2013)

[4]The large "menu cost" literature that followed includes recent contributions such as Burstein (2006), Golosov and Lucas (2007), Gertler and Leahy (2008), Nakamura and Steinsson (2008), Nakamura and Steinsson (2010), Alvarez et al. (2011), Midrigan (2011), and Vavra (2014).

[5]The 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.

[6]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.

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

and Bergemann and Schlag (2011)), but does not analyze learning about the distributions. In our focus on learning under ambiguity, we also extend the decision-theoretical framework of Epstein and Schneider (2007) to learning about functions rather than single parameters.

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. 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 and Mela (2008). For our purposes, we focus on nine markets and six product categories.[8]

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

[8]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.

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 by Nakamura and Steinsson (2008) and others, 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. To construct Figure 1, we computed 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. The resulting hazard function is clearly downward sloping. This is not only an artifact of temporary discounts: the hazard continues to decline 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. This is an empirical fact that has been documented in multiple studies. Klenow and Kryvtsov (2008), for example, find that 40% and 11.3% of all price changes in the CPI dataset are less than 5% and 1% in absolute value, respectively. We performed a related exercise in our dataset: For each market/category, we computed the number of price changes that are between 0% and 5% in absolute value, and did the same for the 10%-to-15% interval. We then computed the ratio of these two numbers. The weighted average across all market/category pairs of this ratio is 1.15, implying that small price changes (0% to 5%) are 15% more numerous than larger ones. Next, we turn to a model whose predictions are consistent with the empirical regularities described above.

4 Model

In this section we focus on describing the non-parametric learning framework, under demand uncertainty. To focus on this, we present a simplified model that does not specify whether prices are nominal or real. In Section 6 we present the full model that differentiates between

nominal and real prices.

There is a single monopolist firm that each period sells a single good at price P_t. To focus squarely on the main mechanism, here P_t is expressed in real terms. Later we will extend the model to account for nominal prices. Denoting by lower-case logs, the firm's demand is determined as

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

where we detail below the distributional assumptions on the two components. 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 log time t marginal cost.

The decomposition of demand in (1) serves two modeling purposes. One is to generate a motive for signal extraction. In this respect, we assume that the firm only observes total demand $q(p_t)$., that z_t is iid and that $x(p_t)$ is constant through time. Thus the role of the former component is to act as noisy demand realizations that will require the firm to use the history of demand realizations $q(p_t)$ to learn about $x(p_t)$.

The second differentiating property is that we assume that the firm views z_t as risky so that the firm is fully confident that this component is drawn from a unique distribution. In particular we assume that the firm knows the true distribution of z_t

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

and that z_t are independent across time.

On the other hand, the $x(p_t)$ component is ambiguous. This means that the firm is not fully confident in the distribution from which this demand has been drawn. The firm starts at time 0 by considering a set of such possible distributions. This set of priors is defined on the space of measurable functions and is not restricted to any given parametric family of demand functions. Instead, the firm uses the noisy realizations of demand it observes to learn non-parametrically about the underlying demand structure.

The firm entertains an initial *set* of priors, Υ_0, over the space of non-increasing functions, i.e. $p \geq p' \Rightarrow x(p) \geq x(p')$. The individual priors are Gaussian Processes, $GP(m(p), K(p, p'))$, with mean function $m(p)$ and covariance function $K(p, p')$.[9] We assume that all priors have the same covariance function, but possibly different mean functions. In

[9]A Gaussian process distribution is the generalization of the Normal distribution to functional spaces. It is a distribution over functions, rather than over numbers or vectors.

particular, we consider all Gaussian Processes with a mean function that satisfies

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

This set is motivated as a limit on the ambiguity the firm faces, and its size will be calibrated based on what the firm could reject at standard 5% levels with a small sample of observations. Intuitively, the interpretation is that while the firm's marketing department provides it with some possible DGPs, it is not confident enough to restrict itself to probabilistically weighing different demand schedules. Moreover, it has no information on the particular functional form of the possible demand functions, but rather needs to learn about them by combining a prior from the set Υ_0 with observed signals.

We further specify the set of Υ_0 by studying the limiting case when the covariance function K goes to zero almost surely. In that case, Υ_0 consists entirely of Dirac measures, on the space of measurable, downward sloping functions. So that for any given prior $\phi_0 \in \Upsilon_0$, there is a unique function $x(p)$ which has probability one, and all other downward sloping functions have probability zero. In addition, all possible functions fall within the tunnel (3). Focusing on the limiting case of Dirac measures allows us to achieve a great deal of analytical tractability and transparency.[10]

4.1 Uncertainty

The timing of choices and revelation of information is the following. First, the firm enters the beginning of period t with information on the history of all previously sold quantities. We also assume that c_t is known at the end of period $t-1$ and that it follows a Markov process with a conditional distribution $g^c(c_t|c_{t-1})$. Thus the relevant past information is the vector of quantities $q^{t-1} = \{(q(p_1), ...q(p_{t-1}))\}$ and the vector of corresponding prices at which those were observed $p^{t-1} = \{(p_1, ...p_{t-1})\}$, where a superscript denotes history up to that time. Let ε^{t-1} denote the quantity-price vector $\{q^{t-1}, p^{t-1}\}$. The firm does not observe separately the history of realizations $(z_1, ...z_{t-1})$ and the underlying conditional means $(x(p_1), ...x(p_{t-1}))$ that have generated the history of sold quantities.

At the beginning of period t the firm posts a price that maximizes its objective, which we will further specify below, conditional on q^{t-1} and on the observed c_t. At the end of period t the idiosyncratic demand shock z_t is realized but is not observed by the firm. The firm

[10] At the same time, modeling Υ_0 as a set of more general Gaussian Process distributions yields qualitatively similar results and we are interested in studying it further in future research.

observes the total demand at its posted price $q(p_t)$. In addition, the firm also observes c_{t+1} and the information set ε^t is updated with the realization $q(p_t)$.

The history of quantities sold acts as noisy signals about the underlying conditional mean demand $x(p)$. The key distinguishing feature of our filtering problem is that we allow for the uncertainty faced by the firm to be both in the form of risk, i.e. the agent fully trusts probability distributions, and ambiguity, or Knightian uncertainty, in which the agent does not have full confidence in her probability assessments.

4.2 Preferences

The monopolist firm is owned by an agent that is ambiguity-averse and has recursive multiple priors utility. The agent values the profits produced by the firm such that conditional valuation is defined by the recursion

$$V\left(\varepsilon^{t-1}, c_t\right) = \max_{p_t} \min_{\pi \in \mathcal{P}_{t-1}(\varepsilon^{t-1})} E^{\pi}\left[\upsilon(\varepsilon_t, c_t) + \beta V\left(\varepsilon^{t-1}, \varepsilon_t, c_{t+1}\right)\right], \tag{4}$$

where $\upsilon(\varepsilon_t, c_t)$ is the per-period profit defined in (2), being a function of the beginning-of-period t posted price and end-of-period realized demand $q(p_t)$. The firm builds its conditional expectations and evaluates expected profits and continuation utility using the worst possible prior, ϕ_0, from the set of admissible priors Υ_0. However, the firm knows the true transition process for cost shocks $g^c(c_{t+1}|c_t)$. The recursive formulation ensures that preferences are dynamically consistent. Axiomatic foundations are in Epstein and Schneider (2003).

The maximization step is over the action of what price P_t to post. The firm cares about profit which is a function of demand. The firm also takes into account that the price posted today reveals information about demand, information that enters as a state variable for next period's value function.

The minimization is over the admissible priors, ϕ_0, of the demand function $x(p_t)$, and hence over the conditional expectation of demand, further denoted by $\widehat{x}(p_t|\varepsilon^{t-1};\phi)$. This conditional expectation is a function of the information set ε^{t-1}, is computed at a specific price p_t and is a function of a specific prior, ϕ_0. Thus, for a given history ε^{t-1}, the minimization selects the admissible prior that yields the lowest expectation $\widehat{x}(p_t|\varepsilon^{t-1};\phi)$. In other words, at each point in time, the firm looks at the historical data and is concerned that, conditional on posting a price, demand at that price is the lowest possible (subject to the constraint on prior distributions). The firm then maximizes over P_t under the belief $\widehat{x}(p_t|\varepsilon^{t-1};\phi_0^{\min})$ evaluated at the worst-case prior $\phi_0^{\min}$.

The minimization step in (4) is relatively easy to solve. We conjecture that the minimizing ϕ_0 is such that, for a given price P_t, it implies that the worst-case expected demand realization $\widehat{x}(p_t|\varepsilon^{t-1};\phi_0)$ is low. Having solved for the optimal policy rule, including the value function, we can then verify that the conjecture is verified. 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 formula (2). The latter needs to be verified, but it is also intuitive: higher demand $x(p_t)$ increases not only current profits but also future expected profits because demand is weakly increasing.

Finally, the true DGP is a line parallel to the prior boundaries of the intervals defined in (3), but with a different intercept, and is given by

$$x(p) = \overline{\gamma} - bp \tag{5}$$

with $\gamma_l < \overline{\gamma} < \gamma_h$. Specifying this distribution is only required here for characterizing the dynamics of the optimal pricing policy functions, derived under the ambiguity expressed in (3), but evaluated under the true DGP.

4.3 Updating and re-evaluation

We now describe the learning process of the firm. We follow Epstein and Schneider (2007) in modeling how the set of priors about the conditional demand is updated as new, but noisy, data on quantities is realized.[11] Let $\Upsilon_{t-1}(\varepsilon^{t-1})$, to be described below, denote the set of posterior beliefs about the unknown demand function $x(p)$, given information ε^{t-1}.

The dynamics of learning can be summarized by a process of one-step-ahead conditional beliefs. However, in contrast to the Bayesian case, there is now a typically non-degenerate set assigned to every history. We are interested in modeling an environment where the decision maker starts with a possibly large set of prior theories on the demand schedules and then uses the observed information to discard some prior theories that seem statistically unlikely.

Epstein and Schneider (2007) propose that this re-evaluation takes the form of a likelihood ratio test. In particular, the decision-maker discards all priors ϕ_0 that do not pass a likelihood-ratio test against an alternative prior that puts maximum likelihood on the sample. Posteriors are formed only for priors that pass the test. This re-evaluation of prior theories is significantly simplified when each prior is a Dirac measure on a particular demand schedule and the LR test is performed pointwise. In particular, take the observed past quantities, and

[11] Technically, our approach is an extension of the Epstein and Schneider (2007) framework to infinite dimensions, and learning about functions rather than single parameters.

only consider the likelihood ratio test done at the price points where the quantities have been observed. Let P^{t-1} denote the vector of observed prices in the past. The likelihood ratio test can be broken into two operations: first, compute the posterior $\phi_{t-1}\left(x(p_{t-1}); \varepsilon^{t-1}, \phi_0\right)$ of having observed a particular value $x(p_{t-1})$. The Dirac measure makes this updating step trivial. The posterior probability is always equal to one or zero, depending on the prior:

$$\phi_{t-1}\left(x(p_{t-1}); \varepsilon^{t-1}, \phi_0\right) = \left\{ \begin{array}{l} 1 \text{ if } \phi_0\left(x(p_{t-1})\right) = 1 \\ 0 \text{ if } \phi_0\left(x(p_{t-1})\right) = 0 \end{array} \right\}$$

The second element of the update is the likelihood ratio computation. As defined in (3), $\phi_0\left(x(p_j)\right) = 1$ if $x(p_j) \in [\gamma_l - bp_j, \gamma_h - bp_j]$ and equals zero otherwise. Thus, here we can compute the data density at each price P_j observed in the vector of past P^{t-1}, conditional on a particular value of such $x(p_j)$:

$$L_{N_{t-1}(P_j)}(x(p_j)) \equiv \prod_{i=1}^{N_{t-1}(P_j)} f\left(q_i(p_j)|x(p_j)\right)$$

where $f\left(\cdot|x(p)\right)$ denotes the density of a normal distribution with mean $x(p)$ and variance σ_z^2, $N_{t-1}(P_j)$ is the number of times that price P_j has been observed up until time $t-1$ and $q_i(p_j)$ records the demand realizations at that price.

The statistical re-evaluation of prior theories takes the following form. Let $\widetilde{x}(p_j)$ be the particular function $x(.)$ that takes probability one under a given prior $\widetilde{\phi}_0$. Then, for each element P_j, the decision-maker restricts attention to priors $\widetilde{\phi}_0$ such that the hypothesis $x(p_j) = \widetilde{x}(p_j)$ is not rejected by an asymptotic likelihood-ratio test performed on the given sample, where the critical value of the $\chi^2(1)$ distribution is $-2\log\alpha$. Specifically, the likelihood ratio test does not eliminate priors that satisfy:

$$L_{N_{t-1}(P_j)}(\widetilde{x}(p_j)) \geq \alpha \max_{x(p_j)} L_{N_{t-1}(P_j)}(x(p_j)) \tag{6}$$

Eliminating a prior $\widetilde{\phi}$ means that the firm determines a particular demand function is inadmissible given the observed data.

For example, consider significance levels of 5% and 2.5%, which leads to $\alpha = 0.15$ or a critical value $\delta_\alpha = 1.96$ and $\alpha = 0.08$ or a critical value $\delta_\alpha = 2.23$, respectively. If $\alpha = 1$ then we only remain with the maximum likelihood estimate and this becomes a standard Bayesian problem.It is useful to note that this re-evaluation of theories has to be done after each history of events.

Thus, the set of posterior beliefs consists of the demand schedules associated with priors

ϕ_0 that have survived elimination, both through the statistical or economic steps:

$$\Upsilon_{t-1}(\varepsilon^{t-1}) = \left\{ \begin{array}{c} \widehat{x}(p) : \widehat{x}(p_j) \in [\gamma_l - bp_j, \gamma_h - bp_j];\ L_{N_{t-1}(P_j)}(\widehat{x}(p_j)) \geq \alpha \max_{x(p_j)} L_{N_{t-1}(P_j)}(x(p_j)) \\ \text{and } x(p_j) \leq x(p_k) \text{ for } \forall\ p_j \geq p_k \end{array} \right\}$$

The set of one-step-ahead conditional beliefs, is thus a set $\mathcal{P}_{t-1}\left(\varepsilon^{t-1}\right)$ of normal distributions

$$\mathcal{P}_{t-1}\left(\varepsilon^{t-1}\right) = \left\{ f\left(\cdot|\widehat{x}(p)\right) : \widehat{x}(p) \in \Upsilon_{t-1}(\varepsilon^{t-1}) \right\}$$

When the decision-maker entertains posting at time t a price P', he considers a set of normal distributions for demand at P', which differ in their mean. The mean can take any of the posterior values $\widehat{x}(p')$ that are associated with priors ϕ that have not been eliminated by the decision maker as unlikely, given the observed history of quantity realizations, i.e. those $\widehat{x}(p') \in \Upsilon_{t-1}(\varepsilon^{t-1})$.

4.4 Updating with one repeatedly posted price

To build intuition for the updating formulas suppose the demand history is such that it only contains observations of demand at some P_0, for N number of times. The sample average demand $\widehat{q}_N(p_0) \equiv \sum_{i=1}^{N} \frac{q_i(p_0)}{N}$ serves as the maximum likelihood estimate of $x(p_0)$. The sample mean has a sample standard deviation of $\widehat{\sigma}_N = \sigma_z/\sqrt{N}$.

We find it analytically useful to describe the lower and upper bound of the prior tunnel, compared to the true DGP in (5), as

$$\gamma_l = \overline{\gamma} - \nu\sigma_z;\ \gamma_h = \overline{\gamma} + \nu\sigma_z \tag{7}$$

and the realized demand as the average demand under the true DGP, shifted by a multiple ψ of its sample standard deviation

$$\widehat{q}_N(p_0) = \overline{\gamma} - bp_0 + \psi\widehat{\sigma}_N$$

The learning step in (6) amounts to keeping the priors ϕ_0 that satisfy

$$x(p_0) \in [\widehat{q}_N(p_0) - \delta_\alpha\widehat{\sigma}_N, \widehat{q}_N(p_0) + \delta_\alpha\widehat{\sigma}_N]$$

where δ_α is the desired critical value. The resulting allowed set of demands at p_0 is the

intersection

$$[\overline{\gamma} - bp_0 - \nu\sigma_z, \overline{\gamma} - bp_0 + \nu\sigma_z] \cap [\overline{\gamma} - bp_0 + (\psi - \delta_\alpha)\,\widehat{\sigma}_N, \overline{\gamma} - bp_0 + (\psi + \delta_\alpha)\,\widehat{\sigma}_N] \tag{8}$$

The intersection in (8) results in a non-empty set if and only if

$$|\psi| \leq \nu\sqrt{N} + \delta_\alpha \tag{9}$$

Clearly the restriction is most binding for $N = 1$, which says that the sample mean should not to be too large so that even the lower bound of the desired confidence interval becomes larger than the prior upper bound, and, reversely that it's not too low. For example, if $\delta_\alpha = \nu = 1.96$ then $|\psi| < 2 * 1.96$. The restriction is more likely to be satisfied as N is larger.

To complete the description of the learning process, consider the case in which the intersection in (8) results in an empty set. This is a situation in which the confidence interval around the observed average demand is too narrow to intersect the prior tunnel. Since the decision-maker only considers demand schedules in the latter, he treats the observed demand as unlikely until it intersects at least in one point the prior tunnel. This means that the critical value δ_α is increased until it reaches $|\psi| - \nu\sqrt{N}$, so that condition (9) is satisfied.

The worst-case demand $x^*(p_0)$ is the minimum of the demands that survive the re-evaluation step:

$$x^*(p_0)|\,(\widehat{q}_N(p_0)) = \max\{\overline{\gamma} - bp_0 - \nu\sigma_z, \overline{\gamma} - bp_0 + (\psi - \delta_\alpha)\,\widehat{\sigma}_N\} \tag{10}$$

Intuitively, the worst-case demand can be the lower bound of the confidence interval if the lower bound of the confidence interval is above the lower bound of the prior tunnel, a condition summarized by:

$$\nu > (\delta_\alpha - \psi)\,/\sqrt{N} \tag{11}$$

This is more likely to happen if the confidence interval is narrower, which is determined by a larger N and a smaller critical value δ_α, if the average demand is larger, through a higher ψ, and if the prior tunnel is wider, controlled by a larger ν. The worst-case demand instead can remain to be the initial one, $\overline{\gamma} - bp_0 - \nu\sigma_z$, if the opposite condition holds.

Having determined the worst-case $x^*(p_0)$, we can find the solution to the rest of the demand curve. In particular, for prices higher than p_0 the worst-case posterior is the worst-case prior

$$x^*(p')|\,(\widehat{q}_N(p_0)) = \overline{\gamma} - bp' - \nu\sigma_z \text{ for } \forall\ p' > p_0$$

For prices lower than p_0, there is a threshold $p_2(\psi, N)$, characterized by

$$\overline{\gamma} - \nu\sigma_z - bp_2 = \overline{\gamma} - bp_0 + (\psi - \delta_\alpha)\,\widehat{\sigma}_N$$

at which demand under the initial worst-case demand, given by the left hand side, equals the lower bound of the demand at the observed price p_0. For prices between $p_2(\psi, N)$ and p_0 the worst-case posterior is higher than the worst-case prior because of the downward sloping curve restriction. In fact, in that case the lowest demand that satisfies the weak monotonicity is the worst-case demand at $x^*(p_0)$. For prices below $p_2(\psi, N)$, the worst-case is restricted now by the worst-case prior.

To summarize, having observed $\widehat{q}_N(p_0)$, the worst-case demand is:

$$x^*(p') \equiv \min x(p')|\,(\widehat{q}_N(p_0)) = \left\{ \begin{array}{c} \max\left\{\overline{\gamma} - bp' - \nu\sigma_z, x^*(p_0)|\,(\widehat{q}_N(p_0))\right\} \text{ for } p' \leq p_0 \\ \overline{\gamma} - bp' - \nu\sigma_z \text{ for } p' > p_0 \end{array} \right\} \quad (12)$$

where $x^*(p_0)|\,(\widehat{q}_N(p_0))$ is given by (10).

4.4.1 Kinked expected demand

The important property of the learning process is that it can generate kinks at the observed prices. Indeed, the worst-case expected demand in (12) has a kink at p_0, as long as (11) is satisfied. Figure 2 is an example of a plot for $x^*(p')$ for the above situation, where $P_0 = 1$, $\widehat{q}_N(p_0)$ is given by the demand under the true DGP, i.e. $\psi = 0$, and condition (11) is satisfied, so that the lower bound of the confidence interval is above the lower bound of the prior tunnel. The function has in fact two kinks, at P_0 and $P_2(\psi, N)$.

The kink generated at the observed P_0 can obviously create price stickiness. If the firm considers increasing the price, it will act as if the expected demand is given by the lower bound of the prior tunnel, which is characterized by a discrete jump down from P_0. If conversely, it considers decreasing the price, on the interval $(p_2(\psi, N), p_0)$, then the firm acts as if there is no gain in demand and thus it is not optimal to lower there the price.

The intuition behind the kinked expected demand is the following. The firm does not restrict demand to be part of a particular parametric family of functions, hence observations are useful mostly in updating expected demand locally, not globally. As the firm gathers information at one price, it is becoming increasingly confident about the demand there. Specifically, as the number of those observations increases, the confidence interval shrinks to the point the firm is convinced by the observed data that the demand is very likely to be above its initial worst-case belief. However, due to the non-parametric stance on the demand

schedules, having observed demand at that price puts only a few restriction on the possible values demand takes away from that point.

The firm updates its view on the demand at the rest of the price support by considering bounds on what the new information implies. In particular, demand cannot decrease to the left of the observed price and it can fall up to the initial worst-case bound for higher prices. By considering the whole set of prior demand schedules that are consistent with the observed data, the firm acts as if there is a kink at the observed price. At this price the firm looks, from the perspective of an econometrician, as being more optimistic about demand than at other prices. Once the expected demand has a kink, it is then clear that for a range of small enough cost shocks it is optimal for the firm not to change its price.

To showcase the model counterpart of expected utility, we can consider several comparisons. The starkest one is that where the agent knows the true DGP in (5). In this case the expected demand is smooth everywhere and the optimal price is the solution to

$$\max_{P_t} e^{0.5\sigma_z^2}(P_t - C_t)e^{\overline{\gamma}-b\log P_t} \tag{13}$$

which results in the standard markup over real marginal cost:

$$P_t^{RE} = \frac{b}{b-1}C_t \tag{14}$$

There is clearly no price stickiness in this case.

A more complex environment is to bring in learning but consider the firm as taking a parametric view on the shape. For example, if we model the uncertainty it faces as

$$x(p_t) = \gamma - bp_t;\ \gamma \sim N(\overline{\gamma}, \sigma_\gamma^2);\ b \sim N(\overline{b}, \sigma_b^2)$$

we impose the knowledge of a linear demand curve, just with unknown constant and slope. In this case, once the agent has an estimate of the coefficients it simply draws a line between two points using that mean estimate. Even if there is uncertainty, this uncertainty is not point by point but it is simply the estimation risk of the whole function. Thus, this alternative setup cannot generate kinks.[12]

[12]Balvers and Cosimano (1990), Bachmann and Moscarini (2011) and Willems (2011) are some examples of models with active learning about a parametric demand curve.

4.5 Key modeling ingredients

There are two key modeling ingredients for our mechanism of rigidity. The first one is the non-parametric nature of learning. The role of this ingredient is to make uncertainty reduction local in nature. As described above, a simple parametric view on demand, such as learning about a linear demand curve, does not generate kinks. Here, instead, our mechanism emphasizes the plausible feature that the strongest reduction in uncertainty occurs at the prices that have been actually posted. The second ingredient is that this uncertainty should ultimately matter for decision so the mechanism requires some uncertainty aversion. The objective is to have a lower certainty equivalent of the price associated with the higher uncertainty.

These two ingredients can be potentially implemented in different environments of uncertainty. The first is within the expected utility framework, where uncertainty is limited to risk in the form of a unique prior. There, one needs to characterize the entire posterior distribution over functions, a challenging task even for high-level non-parametric econometrics.[13] Importantly, in terms of the economics behind the mechanism, to generate the local reduction in uncertainty, the initial prior over functions needs to include *some* a-priori demand non-differentiability. Finally, the latter non-differentiability can generate a kink in the demand variance that would need to be accompanied by risk aversion to have an effect on the pricing decisions.

In this paper we have taken a different approach, namely to use a model of learning under ambiguity. The difference, and in many aspects the advantage, compared to the expected utility case, is that the firm needs to characterize only the worst-case demand, and not the whole set of posterior beliefs. In addition, the firm does not need a-priori demand non-differentiability in the prior set of entertained functions in order to generate kinks. Instead, the non-differentiability comes entirely from the ambiguity aversion, which generates a kink in the expected demand from the switch in the worst-case beliefs.

It is useful to note in this context that the prior set Υ_0 that we describe contains functions that do not have restrictions on their derivatives. That is the reason why the worst-case demand can range from being locally flat or vertical, as long as it belongs to the prior tunnel. This lack of additional restrictions is done here for simplicity. We could impose limits on the derivatives of the demand functions but that would come at the cost of a more convoluted characterization of the updated set of likely demands. Importantly, imposing limits on the derivatives would still lead to a non-differentiability in the worst-case demand. Intuitively, when the firm entertains setting a higher price than the one for which demand is known, it

[13] See Ichimura and Todd (2007) for a survey of semi- and non-parametric estimators and Blundell et al. (2008) for a recent contribution on non-parametric estimation of demand curves.

is worried about an elastic demand, with potentially some bounds on those elasticities. This belief switches, and creates a kink in the perceived derivative of the demand function, when the firm considers setting a lower price, for which the worst-case is a more inelastic demand.

5 Optimal pricing

5.1 A static optimization problem

In this subsection we describe a static version of the profit maximization. In particular ,at the beginning of each period t, the firm chooses the price P_t to maximizes the end-of-period profits under the worst-case conditional expectation of demand:

$$\max_{p_t} \min_{\pi \in \mathcal{P}_{t-1}(\varepsilon^{t-1})} E^{\pi} \left(P_t - C_t\right) e^{x(p_t)+z_t}$$

Recall that the learning process delivers a set of Dirac measures on the elements $x(p)$. Thus, under the worst-case posterior, the demand $x(p)$ equals $x^*(p)$ with probability 1. There is no risk around that estimated mean, but rather only that coming from z_t. Thus, the problem becomes similar to that in (13), except that there is a set of possible conditional demands:

$$\max_{p_t} \min_{x(p_t)} e^{0.5\sigma_z^2} \left(P_t - C_t\right) e^{x(p_t)}$$

To highlight the solution through analytical representations, we present a case where the firm has observed one price P_0, for N number of times at an average quantity $\widehat{q}_N(p_0)$. We have described how learning about the demand schedule works in this case in section 4.4. Suppose that condition (11) holds so that there is a kink at p_0. Denote the lower bound of the confidence interval:

$$\underline{q}_N(p_0) \equiv \widehat{q}_N(p_0) - \delta_\alpha \sigma_z / \sqrt{N}$$

Further denote the difference between $\underline{q}_N(p_0)$ and the lower bound of the prior tunnel as

$$\Delta(p_0) \equiv \underline{q}_N(p_0) - [\overline{\gamma} - bp_0 - \nu\sigma_z]$$

The worst-case demand is given by

$$x^*(p') \equiv= \left\{ \begin{array}{c} \underline{q}_N(p_0), \text{ for } p' \in [p_2(\psi, N), p_0] \\ \overline{\gamma} - bp' - \nu\sigma_z \text{ for } p' < p_2(\psi, N) \text{ and } p' > p_0 \end{array} \right\}$$

To characterize the optimal price, notice that the first order condition at an interior solution

is the rational expectations price, as in (14): $P_t^{RE} = \mu C_t$, where $\mu \equiv b/(b-1)$ is the markup. The reason is that, even if the firm prices under the lower bound of the prior tunnel, we have assumed that there is the same elasticity as under the true DGP. Moreover, the optimal price will not be $p \in [p_2(\psi, N), p_0)$ as the demand is the same at that interval but the price is highest at p_0. So, we only need to compute the profit at P_0 and compare it to that arising from setting the RE one. The former is:

$$\upsilon(P_t = P_0) = e^{0.5\sigma_z^2} (P_0 - C_t) e^{\underline{q}_N(p_0)}$$

For ease of exposition, define a hypothetical value of cost $C_0 \equiv P_0/\mu$, for which the price P_0 would be the optimal RE price. The profit can be rewritten as

$$\upsilon(P_t = P_0) = e^{0.5\sigma_z^2 + \overline{\gamma} - \nu\sigma_z} \left(\frac{C_0}{C_t}\mu - 1 \right) C_t (\mu C_t)^{-b} \left(\mu \frac{C_0}{C_t} \right)^{-b} e^{\Delta(p_0)} \tag{15}$$

The profit at a RE price simply sets $C_0 = C_t$ and $\Delta(p_0) = 0$ in (15), so that

$$\upsilon(P_t = P_t^{RE}) = e^{0.5\sigma_z^2 + \overline{\gamma} - \nu\sigma_z} (\mu - 1) C_t (\mu C_t)^{-b} \tag{16}$$

which is the standard profit of a flexibly chosen optimal price, except that the log demand is lower by $\nu\sigma_z$ compared to the valuation under the true DGP in (13).

We can now ask the question of what is the range of cost shocks C_t for which it is optimal to stick with P_0. For this, we write the difference in profits from sticking or moving from P_0 as the difference between the profits in (15) and (16) and treat it as a function of $r_t^c \equiv C_t/C_0$:

$$h(r_t^c) \equiv \left(\frac{\mu}{r_t^c} - 1 \right) (r_t^c)^b e^{\Delta(p_0)} - (\mu - 1) \tag{17}$$

The function is above zero for $r_t^c = 1$, which obviously comes from the condition that there is a kink at p_0, i.e. that $\Delta(p_0) > 0$. Moreover, $h'(r_t^c)_{|r_{t=1}} = 0$ and it is concave so it reaches a maximum at $r_t^c = 1$. We are interested in finding the values r_t^c for which it equals zero. To resort to an analytical solution, we take a second order approximation of h around $r_t^c = 1$ and find the roots of that quadratic function as

$$r_1^c = 1 - \sqrt{\frac{2(1 - e^{-\Delta(p_0)})}{b(b-1)}};\ r_2^c = 1 + \sqrt{\frac{2(1 - e^{-\Delta(p_0)})}{b(b-1)}}$$

The result is that in a range of cost shocks $[C_0 r_1^c, C_0 r_2^c]$ it is optimal to set the price P_0. As the shocks become larger in absolute value the firm sets the RE price.

Because the above conclusion is drawn on a second order approximation of $h(r_t^c)$, the range for which there is stickiness is symmetric. We can check the third derivative of h and find that it adds the term $-\frac{2}{3}b\,(b-2)\,(r_t^c-1)^3$. Thus, if $b>2$ (a condition easily satisfied by empirically reasonable values), the function is lower (higher) for the higher (smaller) root r_2^c (r_1^c). So the nonlinear function $h(r_t^c)$ will intersect zero at values that are both smaller than the corresponding $r_{1,2}^c$. This shows that there is asymmetry: the inaction region is longer to the left than to the right so that there is a more likely pass through for positive cost shocks than negative ones. The intuition is that the profit function is more sensitive to higher cost shocks: if the firm does not change its price it suffers more from the loss in markup than if it considers symmetric lower cost shocks.

5.2 Dynamics: a three-period model

In our model the observations in the information set depend on actions. Indeed, posting different prices leads to noisy signals about different parts of the unknown demand schedule. Thus, this becomes a dynamic problem in that choosing a price not only leads to static profits but to future benefits in the form of learning demand. This influence goes through two effects: one is deterministic, by increasing the number of times at which the posted price is observed. The second is through the random innovation that will be observed at the end of the period. The former effect arises in this model from the presence of ambiguity. The second is more general, appearing also in dynamic problems with experimentation as in the multi-arm bandit problems.

Solving fully optimal learning problems while allowing for experimentation is a difficult numerical task. The main computational problem here is that the state space explodes as the number of posted prices increases with time. For this reason we take the approach of studying a three-period model, described below, such that in the last period there are only static profits to be gained and no continuation utility. We believe that even the three-period model with learning is rich enough to capture most of the important effects of the many, possibly infinite, periods version of the dynamic model.

There are three periods. Let us start from the beginning of the second period, when the firm starts with the information from previously observed demand realizations, denoted below by ε^1. At this point in time, the firm also knows the cost c_2.

The dynamic problem of the firm is to choose the optimal price P_2 that maximizes the

worst-case expectation of the discounted sum of the second and third period profits:

$$\max_{P_2} \min_{\pi \in \mathcal{P}_1(\varepsilon^1)} E^{\pi} \left[(P_2 - C_2)\, e^{x(p_2|\varepsilon^1)+z_2} + \beta v(P_3^*) \right] \tag{18}$$

where P_3^* denotes the optimal price set in period 3, conditional on the ε^1 and the new demand signal $q(p_2)$ realized at the end of period 2 at the price P_2; E^{π} denotes the worst-case conditional expectation, which uses the estimate of demand $x^*(p_2|\varepsilon^1)$.

The third period problem is a static maximization, characterized in section 5.1:

$$v(P_3^*) = \max_{P_3} \min_{\pi \in \mathcal{P}_2(\varepsilon^1, q(p_2))} E^{\pi} \left(P_3 - C_3 \right) e^{x(p_3|\varepsilon^1, q(p_2))+z_3} \tag{19}$$

The first period consists of letting the firm choose the prices that act as the initial state variable in the problem described above. This allows us to study what would be the price in which the firm would mostly invest knowing about.

5.2.1 Parametrization

Here we are interested in illustrating the main mechanisms of the model. It is important to note that we do not have a discrete space for the cost as that may mechanically generate discreteness in prices even in a standard model. The Markov process 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. The benchmark parametrization is in Table 1. We set $b = 6$, the constant $\overline{\gamma} = 0$ and the critical value $\delta_\alpha = 1.96$, which corresponds to a 95% confidence interval. We set the cost shock parameters ρ_c and σ_c to values calculated by Eichenbaum et al. (2011), where they observe marginal costs. We normalize $\overline{c} = (b-1)/b$ so that $P^{RE} = 1$. We set the discount factor $\beta = 0.99[1 - (1 - e^{-1/30})]$, where the second part of the discounting models that a 'pricing regime' lasts on average 30 weeks in the data, as documented by Stevens (2014). We are left with setting the width of the worst-case prior tunnel. Here we set $\nu = 2$, which is argued in Ilut and Schneider (2014) as a reasonable upper bound on ambiguity, and explore with setting the standard deviation of demand shocks σ_z.

Table 1: Calibrated parameters

b	$\overline{\gamma}$	δ_α	ρ_c	σ_c	ν	σ_z	β
6	0	1.96	0.14	0.11	2	0.1	0.9575

5.3 Results

Worst-case expected demand

We first plot the worst-case expected average demand for the case where the firm has observed only one price, namely $p_1 = 0$, in Figure 2. The blue solid lines represent the bounds on the prior tunnel, the blue dotted line is the true DGP, the red cross is the average demand observed at p_1, and the red vertical line denotes the 95% confidence interval around it. The black line plots the worst-case demand, having observed that information, which forms an obvious kink at p_1.

To illustrate the role of certainty, in figure 3, we increase the number of times for which the same average demand has been observed. This corresponds to increasing N in our previous calculations. As the 95% confidence interval shrinks, the decision maker is more confident that the observed average demand reflects the true one, and the kink becomes stronger. This effect is at the heart of decreasing hazard function in this model: as the firm accumulates knowledge at the observed price, it becomes more confident and thus more unlikely to change the price, a feature that is documented in the data.

When the firms observes two prices, as for example it happens in the third period of our simple model, the worst-case posterior demand can have two kinks. Figure 4 shows this case, which adds an average demand observation for a higher p_2. In a similar fashion, more information at p_2 strengthens the kink, as shown in Figure 5.

Pricing policy functions

In section 5.1 we developed analytical solution and intuition for a static pricing optimization. We develop that here through the use of graphical representation. Figure 6 plots such an optimal pricing policy under RE, in brown, against ambiguity, in blue. There is a clear area of inaction, for which the firm finds it optimal not to change its price. As shown in section 5.1, this inaction region is stronger to the left, implying that cost increases are more likely to be pass-through than similar cost decreases.

Figure 7 plots the optimal price for the case where a higher p_2 has been observed, which lead to a kink in expected demand as shown in Figure 4. The two kinks in expected demand manifest themselves as areas of inaction around the two past observed prices. This captures the discreteness of the policy function: previously observed prices become 'focal points'. For example, suppose there is an increase in the cost that, according to the blue policy function with just one observed price, would have otherwise resulted in setting P^{RE}. In the case of two observed prices, the dark solid line shows that the optimal price does not need to adjust monotonically, but rather jump at p_2 and stick there for a range of cost realizations.

Experimentation

Figure 8 plots the pricing policy of period 2, where the firm has observed the price p_1 and takes into account the effect of its pricing decision on the future valuation. This is marked by the dark solid line. 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 hand, by moving to another price it can expect to learn something new and potentially valuable. Which force dominates depends on state variables. Figure 9 is an example of the former effect being stronger, which leads to more stickiness than the static policy function. To further explore this, we compute the policy function in the case where firm will repeat the static last period problem forever, without ever updating its information set again. The continuation value in this case is the present discounted value of the stream of expected profits from the third period, but all this changes is the discount factor, increasing it to $\widetilde{\beta} = \beta/(1-\beta)$. The policy function is shown in Figure 9. Not surprisingly, this only increases stickiness. In these cases gaining more information about where the firm currently stands is important for the future problem, and outweighs any experimentation incentives. This is because the observed price is right at the median and would be close to the optimal price if the firm receives a cost shock close to the mean - that is where the bulk of future realizations of cost are likely to be anyway, hence learning about this part of the demand curve is useful.

To showcase experimentation and the role of the state variables, we now assume that the observed price p_1 is not the median price but corresponds to the 25th percentile of the cost distribution. Figure 10 plots this case and illustrates that this is not a very useful price to learn at. The firm is not likely to choose again to pick such a low price unless it gets very low cost shock realizations and thus it finds it optimal to move earlier away from it so that there is less stickiness to the right of p_1. This is a case where the incentive to experiment rather than to learn more at the same price wins out. This effect is magnified in the case where the relevant discount factor is increased to $\widetilde{\beta}$, as shown in Figure 11.

These pricing policies are taking as given an initial price p_1. Due to our three-period setup we can ask what is the initial price, or prices, that the firm would like most to know about. In particular, we consider the following cases.

In the first case we let the firm choose one price that it can know extremely well. The optimal such price is $p_1 = 0$, which is also the average flexible price. This is intuitive - the firm wants to get information around where it expects future shocks to realize. This also means that the optimal policy figures we already have discussed above are drawn exactly for that optimal p_1.

In the second case we let the firm choose two price p_1 signals, rather than just one, and we assume that each signal has been observed just once. Then we find that the firm will choose both signals to be at $p_1 = 0$. Thus, instead of getting noisy signals at two different price points, the firm prefers to get a better signal at that high-probability price point p_1. This speaks again to the tension between experimentation and acquiring further information at a point the firm has already seen in the past.

In the third case we let the firm choose two p_1 signals, but we make them perfectly revealing. Then the firm will choose two different signals, at $p_1^h = 0.07$ and $p_1^l = -0.10$. Figure 12 plots the optimal pricing policy in period 2, where the firm has observed perfectly demand at these two prices. In this case the kinks are so deep that the two flat spots almost take up all of the price space. The black line plots the policy function of the forward looking firm, while the blue line is the static maximization. Because of the deep kinks, the two functions are essentially on top of each other.

A final experiment is to compute the policy function of a firm that has seen the two prices $p_1^h = 0.07$ and $p_1^l = -0.10$ only once each. Notice that this is not necessarily ex-ante optimal, since the firm picks those two prices at time 1 only if the associated number of time of having observed those prices would be large, and otherwise it will pick $p_1 = 0$ twice. The policy function, shown in Figure 13, has two flat spots, and also a lot of action in between. Now there are also interesting differences in the functions between the static and the forward-looking firm. We see that the firm finds it optimal to acquire more information about an already established kink. This means that the forward looking firm is more likely to either (i) stay at one of its kinks or (ii) move back to one of its kinks.

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. We do not provide general characterization of these tradeoffs but explore them in numerical experiments of the type we reported above. While the literature on multi-arm bandit problem is extensive, it usually provides some analytical characterizations only in the case of risk and the arms being independent. In our model neither of these conditions are met. First, the arms are correlated, since observing demand realizations at one price is partially informative about the rest of the demand schedule. Second, there is ambiguity over the payoffs. A recent advance in the theoretical literature on ambiguity and bandit problems is Li (2013), which finds that in a context of independent ambiguous arms, the incentive to experiment is weaker compared to the risk case. 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. Our numerical results, with the further addition of correlated arms, are consistent with a similar conclusion.

6 Nominal Rigidity

The model presented so far was one of real rigidity, in which P is interpreted as a real price. In particular, there was nothing that prevented nominal adjustments. For example if the firm knew that the aggregate price level had shifted, it could exactly change its nominal price to achieve the same real price and stay at the "safe" place.

We structure this section as follows. First, we enrich the model so as to make a distinction between real and nominal prices. We 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 monopolistically competitive firms that sell to a final good industry. The firm's demand is thus a function of the 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 does not know the technology of its industry and is ambiguous about it. This leads to ambiguous beliefs about the relevant industry price level, and thus about the demand-relevant relative price. As a result, in addition to not knowing the demand curve, the firm is uncertain about its appropriate relative price argument. Thus, the firm faces two dimensions of ambiguity – the demand function itself is ambiguous, and its argument is ambiguous. The firm sets an optimal nominal pricing action that is robust to both. We show that this turns the real rigidity generated in the previous section into nominal rigidity.

Second, we provide empirical evidence based on US data for the time-variation of the relationship between aggregate and industry prices. Here we discuss the lack of statistical confidence that an econometrician has, when estimating this time-variation at different horizons, in rejecting the null hypothesis that aggregate prices are not informative about industry prices. Thus, consistent with the approach that resulted in real rigidity, the firm is now also put on equal footing to the econometrician that cannot easily reject the fact that aggregate prices are typically not a useful signal about the relevant relative price.

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:

$$C_t = \left(\int C_{jt}^{\frac{b-1}{b}} dj \right)^{\frac{b}{b-1}} \tag{20}$$

This final good demand defines the aggregate price index P_t

$$P_t = \left(\int \bar{P}_{jt}^{1-b} dj \right)^{\frac{1}{1-b}} \tag{21}$$

where $\bar{P}_{jt}$ are the price indices of the separate industries.

Our preferred interpretation of this setup is that the final household consumes different types of final goods that are produced by industries with potentially different structures.[14] Each industry j has a representative final goods firm, which produces its good by aggregating over intermediate goods i with the technology

$$C_{jt} = f_j^{-1} \left(\int f_j(C_{ijt}) v_j(z_{it}) di \right) \tag{22}$$

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 price index $\bar{P}_{jt}$ such that

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

where C_{ijt} is the amount purchased of good variety i by industry j. Solving the cost minimization problem of the representative firm in industry j yields

$$C_{ijt} = f_j'^{-1} \left(\frac{P_{it}}{\bar{P}_{jt}} \frac{f' \left(\frac{C_{jt}}{\xi_{jt}} \right)}{v(z_{it})} \right) \equiv H_j \left(\frac{P_{it}}{\bar{P}_{jt}}, C_{jt}, z_{it} \right) \tag{23}$$

The demand of industry j for a given intermediate good i is a function of the relevant relative price, $\frac{P_{it}}{\bar{P}_{jt}}$, 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 $Y_{it} = C_{ijt}$, and uses labor L_{it} in the production function:

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

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

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. Moreover, 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 periods. They observe 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

Thus, a firm does not know the specific functional form of the demand it faces, but rather needs to estimate it using its observables. To make the problem more tractable, we assume that firm i understands that the aggregate industry demand C_{jt} and the demand shocks z_{it} enter multiplicatively so that[15]

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

The firm can also use the known structure of aggregate demand

$$C_{jt} = \left(\frac{\bar{P}_{jt}}{P_t}\right)^{-b} C_t \tag{24}$$

to write its demand as

$$C_{ijt} = H_j\left(\frac{P_{it}}{\bar{P}_{jt}}\right)\left(\frac{\bar{P}_{jt}}{P_t}\right)^{-b} C_t \exp(z_{it}) \tag{25}$$

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 $H_j(.)$, which captures the effect of changing its own price vis-a-vis its direct competitors. Using the convention that lower case letters denote the logs of upper-case variables, the firm obtains a linear expression in an unknown function, h_j, an unknown variable, p_{jt}, known effects, c_t

[15]This assumption does not affect the overall structure of the information problem the firm faces. The learning framework we describe easily extends to the case of estimating demad as a function of multiple variables, without any significant conceptual differences.

and bp_t, and an unobserved shock, z_{it}:

$$y_{it} = h_j(p_{it} - p_{jt}) + c_t - b(p_{jt} - p_t) + z_{it}. \tag{26}$$

We assume that the firm has a prior belief that the function h_j is such that

$$h(x) \in [-\gamma - bx, \gamma - bx],$$

and hence lies in the type of a prior tunnel studied previously. Combining the two unknown effects in demand, we obtain an unknown function of two variables

$$x(p_{it}, p_{jt}) \equiv h(p_{it} - p_{jt}) - bp_{jt} \tag{27}$$

Knowing that h_j falls in a tunnel with a slope of $-b$ tells the firm that the *systematic* effect of the unobserved industry price p_{jt} cancels out, so that for any given p_{jt}

$$\begin{aligned} x(p_{it}, p_{jt}) &\leq \gamma - b(p_{it} - p_{jt}) - bp_{jt} = \gamma - bp_{it} \\ x(p_{it}, p_{jt}) &\geq -\gamma - b(p_{it} - p_{jt}) - bp_{jt} = -\gamma - bp_{it} \end{aligned} \tag{28}$$

On average, changes in p_{jt} do not have an effect on the demand of an individual firm i. The intuition is that industry prices have two effects. On the one hand, a fall in p_{jt}, keeping everything else constant, makes the output of industry j relatively cheaper compared to the rest of the economy, and this increases the average industry demand, and hence the demand of firm i. On the other hand, a fall in p_{jt} makes good i more expensive relative to the average good in its industry, and thus decreases demand. On average, the two effects cancel out. An intuitive way to describe is that a fall in p_{jt} leads to an increase in the overall pie of industry j, but ceteris paribus, the firm expects to receive a smaller share of that pie.

However, the effects of p_{jt} cancel out only *on average*, and industry prices can have important local effects. Locally, the industry demand function $h(.)$ can be more or less elastic than aggregate demand, and the overall effect of an increase in p_{jt} could be either positive or negative. Let us take the case of a fall in p_{jt} as an example again. If $h(.)$ happens to be more elastic, locally, than aggregate demand, this fall in p_{jt} will actually reduce firm i's demand. The overall industry pie increases just the same, but now the fall in p_{jt}, keeping p_{it} constant, leads to a greater reduction in the share of the pie going to firm i. The prior knowledge on the tunnel in which $h(.)$ lays only tells the firm that the average elasticity of h is equal to the elasticity of aggregate demand, but local elasticities could be quite different.

The firm has two sources of information on p_{jt}. First, every T periods, the firm conducts

marketing reviews that reveal the current industry price. 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.

6.3 Uncertainty about the relationship with aggregate prices

Even if the firm knows that the relevant relative price is not given by the observed aggregate price, the firm can use the latter to extract information about the industry price 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 production functions f_j), it does not know the exact functional form of that relationship.[16] 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

$$p_{jt} = p_{js} + \phi(p_t - p_{js}) + \nu_{jt}, \tag{29}$$

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 about ν_{jt} are Gaussian white noise processes, but with different, possibly time-varying variances.[17] We model the uncertainty about the cointegrating function in a similar fashion to the way we deal with uncertainty about the demand function. 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, Υ_ϕ, around the true DGP $\phi(p_t - p_{js}) = p_t - p_{js}$. The set of possible functions intentionally allows for a weak relationship between industry and aggregate inflation in the short-run. We model this by specifying that for small values of $|p_t - p_{jt}|$, i.e. small inflationary pressure, the function $\phi(.)$

[16] 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.

[17] The assumption of White Noise ν_{jt} is not crucial and can be justified through a worst-case scenario as well. For simplicity, we assume it outright.

lives in an interval around 0,

$$\phi(p_t - p_{js}) \in [-\gamma_p, \gamma_p], \text{ for } |p_t - p_{js}| \leq K. \tag{30}$$

This allows for functions that imply no relationship between the aggregate inflation and industry inflation, at least in the short. In the long-run, however, there is clearly a relationship and the set of possible functions outside of this region grows linearly with the true DGP $p_t - p_{jt}$:

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

The particular boundaries of the interval of possibilities are chosen to be continuous and to define a set of priors that is analytically tractable. This is done purely for convenience, however, and has no bearing on the rest of the argument. The magnitude of K is chosen to be high enough so that in between reviews the function $\phi(.)$ belongs to the set described by equation (30). 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 that can last even 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 complex, non-linear relationships locally. Intuitively, this means that 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, our set of priors explicitly allows for the possibility that the current short-run relationship is weak, even though in the long-run the firm expects prices to rise in lock-step.

For tractability, we again focus on the limiting case where the variance function of the GP distributions goes to zero, so conditional on a prior, one function $\phi(.)$ has probability 1 and all others probability zero. Therefore, the unknown portion of the firm's demand can be written as

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

and we can define the function

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

where $\hat{r}_{it} = p_{it} - p_{js}$. Thus, demand is a function of the relative price $p_{it} - p_{js}$, relative to the last observed review signal p_{js}, and the ambiguous component of the current industry price: $\phi(p_t - p_{js})$.

The firm understands that its demand is ambiguous in two dimensions. First, the function $x(.)$ itself is ambiguous. Second, one of the arguments of the function is ambiguous. The firm looks for an optimal pricing action, p_{it}, that is robust to this ambiguity. As instructed by the recursive multiple priors utility shown in equation (4), this amounts to maximizing the expected end-of-period profits and continuation utility under the worst-case belief about demand. We have discussed the role of demand uncertainty in the simpler model of section 4. The second channel is here the ambiguity represented by $\phi(.)$. The worst-case belief is a function $\phi(.)$ chosen so to imply the lowest possible demand at any given relative price $\hat{r}_{it}$. This additional ambiguity dimension, now through the worst-case $\phi(.)$, defines a new, worst-case demand function

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

In essence, the firm acts *as if* nature draws the true DGP to be the worst possible relationship between aggregate prices and industry prices out of the entertained set Υ_ϕ of such functions. This leaves the firm with a worst-case demand that is only a function of the observed $\hat{r}_{it}$ and the unobserved noise term ν_{jt}. This demand is the object that the firm can learn about through its past prices and quantities, since according to the optimal behavior under ambiguity, it believes that nature has minimized demand in this same fashion at any point in time.

6.3.1 Signals on relevant relative price

Finally, we assume that the firm performs reviews on a fixed schedule, with a new signal arriving every T periods. The idea is that reviews are costly and time consuming and cannot be done every period, but since they are useful, they are done on a regular basis. We do not model the microfoundations of the review selection process, but rather view the assumption of a new review every T periods as a convenient way to model the salient features of what happens in practice. We realize that in reality the review decision is most likely state-dependent, but as long as the reviews do not happen every period, introducing state-dependent review would not change our analysis and conclusions. Since this is not central to

the main argument, 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).

Given this structure of signal arrival, the beliefs of the firm about future signals evolve as follows. Every T periods the firm's beliefs get recentered at the true value of the industry price, hence if there is a review at time t, then $E_t(p_{jt}) = p_{jt}$. The firm expects that the signal at the next review is given by

$$E_t(p_{j,t+T}) = p_{js} + \min_{\phi \in \Upsilon_\phi} E_t(\phi(p_{t+T} - p_{js})),$$

which only serves to shift the expected nominal price needed to achieve some desired relative prices from period $t + T$ onwards.

6.3.2 Nominal rigidity from real rigidity

One last aspect to account for is that the shock ν_{jt} is unobserved, and hence the estimation target, $x^*(.)$, is in fact a time-varying function of the one observed argument $\hat{r}_{it}$,

$$x^*(\hat{r}_{it}, \nu_{jt}) = x_t(\hat{r}_{it}), \tag{31}$$

where the unobserved noise term serves as a random disturbance to $x_t(\hat{r}_{it})$. For tractability, we assume that the induced time-variation is well approximated by

$$x_t(r) = x_{t-1}(r) + \varepsilon_t;\ \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{32}$$

Putting together (26), (27) and (31) the demand facing the firm is

$$y_{it} = x_t(\hat{r}_{it}) + c_t + b(p_t - p_{js}) + z_{it} \tag{33}$$

which is a known function of the observed aggregates, namely price p_t and quantity c_t, an unknown function $x_t(.)$ 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, with the added wrinkle that since the learning target is time-varying, the firm will optimally discount past observations.

The kinks are formed in the space of relative prices $\hat{r}_{it}$. However, the base of the relevant relative price $\hat{r}_{it}$, the last review signal p_{js}, does not change often. To keep this relative price constant, in order to 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, and given the structure of this economy, 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. We illustrate this behavior in a quantitative model in section 7.

We have used the argument that the firm is concerned that aggregate prices do not reveal much information about industry prices. In the next sub-section we are interested in evaluating how reasonable is that view using empirical evidence for US data.

6.4 Empirical evidence on the link between aggregate and industry prices

Here we investigate the empirical relationship between aggregate and industry prices at different horizons. Our analysis uses the Bureau of Labor Statistics' most disaggregated 130 CPI indices as well as aggregate CPI inflation.

In our first exercise, we use a simple regression method to look into the statistical significance of the relationship. 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 then run rolling regressions of the form:

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

over three-year windows starting in 1995 and ending in 2010.[18] 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 aggretate 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.

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

This can be seen in Figure 14 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.

We next explore further this characteristic by estimating a hidden state model where the connection between the aggregate and industry is allowed to be time-varying. Consider the following state space representation:

$$\pi_{j,t,k} = \alpha_{j,k,t}\pi^a_{t,k} + \sigma_{u,j,k}u_t \tag{34}$$

$$\alpha_{j,k,t} = \alpha_{j,k,t-1} + \sigma_{v,j,k}v_t \tag{35}$$

where u_t and v_t are white noise and the other variables have been defined previously. Equation (34) represents the measurement equation, while equation (35) is the transition evolution of the unobserved estimate $\alpha_{j,k,t}$. The subscripts (j, k) show that this estimation is done for each pair of industry j and horizon k.

The hidden state $\alpha_{j,k,t}$ can be estimated using a standard Kalman filter. The filter generates the Kalman smoothed estimates $\widehat{\alpha}_{j,k,t|N}$, which is the conditional expectation of $\alpha_{j,k,t}$, based on the information of the whole sample of size N.[19] Similarly, it produces the smoothed estimate of the uncertainty $\Sigma_{j,k,t|N}$ around that estimate. Thus, the conditional time-t distribution is given by

$$\alpha_{j,k,t} \sim N(\widehat{\alpha}_{j,k,t|N}, \Sigma_{j,k,t|N})$$

For a pair (j, k), we analyze the sample path of the estimate $\widehat{\alpha}_{j,k,t|N}$ and its uncertainty $\Sigma_{j,k,t|N}$. We analyze how often we cannot reject the null that $\widehat{\alpha}_{j,k,t|N}$ equals zero at some confidence value. Define that fraction of times, out of the whole sample, to be $n_{j,k}$. For a given horizon k, we vary the industries j and denote the average over $n_{j,k}$ as n_k. Finally, we vary the horizon k and collect the resulting n_k. We interpret the measure n_k as the strength of statistical evidence for the firm to consider it reasonable to believe that within the horizon given by k, the relation of aggregate inflation to industry inflation is typically zero.

We first plot the estimated distribution of $\alpha_{j,k,t}$ for the Carbonated drinks industry, which turns out to be a typical industry for our results. In particular, Figures 15 to 22 plot the estimate $\widehat{\alpha}_{j,k,t|N}$ and the 95% confidence interval around it based on the estimated uncertainty $\Sigma_{j,k,t|N}$, for various inflation horizons, ranging from 1 to 3, 6, 12, 24, 36, 48 and

[19] Alternatively, we could have reported the conditional moments based only on information up to time t. Instead, we give the econometrician the most available data, in the form of information on the whole sample, and report the results based on the Kalman smoother.

60 months.

The pictures show that for short horizons it is typical that we cannot reject the null of no predictive content from aggregate inflation for the evolution of industry-level inflation. Not surprisingly, the relationship becomes more significant as the horizon lengthens. In addition, there is a lot of time-variation in the estimated effect. This type of evidence supports the idea that the firm considers a wide set of beliefs about the short-run relationship between the two measures of inflation. As in our model, this set shrinks for the longer-run relationship. Figure 23 plots the value of $n_{j,k}$, defined above, and shows that the fraction of times that we cannot reject the null of $\widehat{\alpha}_{j,k,t|N} = 0$ is indeed high at most horizons and decreases with the horizon.[20]

We find similar patterns when we repeat this analysis over industries. Figure 24 plots the value n_k, defined above, and shows that on average an econometrician cannot reject the null of zero effect of aggregate inflation for a fraction of times that decreases with the horizon, from about 75% at 1 month to about 20% even for 5 years.

7 Quantitative model

We build a quantitative model with nominal prices. The model uses the same layers of production as in the description of the nominal rigidity section but it now expands by endogenizing marginal costs and introducing a law of motion for the aggregate price level. The model is intended for studying, through a more quantitative lens, the individual decision problem of an ambiguity averse firm that faces demand uncertainty. Because we focus on the individual behavior, a more precise way to view the setup analyzed 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: First, there is a unit interval of continuum of intermediate monopolistic firms indexed by i, where each firm sells a differentiated product. They sell to industries, indexed by j. Second, an industry buys from monopolistic firms and sells to a final good producer. The industries are competitive firms. Third, there is a firm producing a final good to be sold to the representative consumer.

[20]The 14 horizons correspond to 1, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 42, 48 and 60 months.

7.1.1 Agents and shocks

Representative agent

There is a representative household that consumes and works, whose problem is

$$\max \sum \beta^t \left[\log C_t - \chi \int L_{i,t} di \right]$$

subject to the budget constraint

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

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 industries j with a CES aggregator with elasticity b as shown in (20). The solution to the cost minimization problem of the representative agent is to demand from each industry the amount given by (24).

The $j-th$ industry

The technology and resulting cost minimization solution of the $j-th$ industry are described by equations (22) and (23). The industries are competitive in producing the j good. They do not exploit the demand for their variety j by the representative consumer, and make zero profits.

The $i-th$ monopolistic firm

The demand for the monopolistic firm i comes from the industry j in the form of (23) which we have further restricted to be described in (25) as

$$Y_{i,t} = H_j \left(\frac{P_{it}}{\bar{P}_{jt}} \right) \left(\frac{\bar{P}_{jt}}{P_t} \right)^{-b} C_t \exp(z_{it})$$

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 $N(0, \sigma^2_a)$. The real flow profits are therefore:

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

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

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

Nominal aggregate spending

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 $N(0, \sigma^2_s)$.

In turn, we can use the optimal hours decision of the household to substitute out for W_t:

$$\frac{W_t}{P_t} = \chi C_t$$

so that the real flow profits can be written as

$$\upsilon_{i,t} = \left(\frac{P_{it}}{P_t} - \frac{\chi S_t}{\omega_{it} A_t P_t} \right) Y_{i,t} \tag{36}$$

7.1.2 Demand uncertainty

True data generating process

We first characterize the determination of demand under the true data generating process (DGP). We use a simple true DGP: each industry type j has the same CES function f_j and

v_j in (22) of the form

$$f_j(C_{ijt}) = C_{ijt}^{\frac{b-1}{b}};\ v_j(z_{it}) = z_{it}^{1/b}$$

These standard CES aggregators imply the following demand for intermediate good i :

$$C_{j,i,t} = C_{jt}\varepsilon_{it}\left(\frac{P_{i,t}}{\overline{P}_{j,t}}\right)^{-b}$$

Thus, under the true DGP, the demand function is simply

$$y_{i,t} = -bp_{i,t} + c_t + bp_t + z_{it} \tag{37}$$

Notice that the whole layer of industry demand has dissapeared in this DGP. 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 $\overline{p}_{j,t}$.

Monopolistic firm's information

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 exactly the same as described previously, where equation (33) gives the demand to be estimated as

$$y_{it} = x_t(\hat{r}_{it}) + c_t + bp_t + z_{it} \tag{38}$$

and

$$\hat{r}_{it} = p_{it} - E_t(p_{jt})$$

is the price relative to the expectations of the industry level price. The expectations are based on the information set of the firm, and the worst-case beliefs described in Section 6.

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 $\bar{P}_{j,t}$, 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 that the underlying demand is $x(p_{i,t}) = -bp_{i,t}$.

This is the flexible price, rational expectations general equilibrium version of our economy.[21] Its solution, up to a log-linear term, is:

$$p_t^{flex} = \log \frac{b\chi}{b-1} + \log S_t - \log A_t \tag{39}$$

and the optimal solution for the rational expectations firms' price is to simply subtract $\log \omega_{it}$ from the aggregate level.

The ambiguity averse firm has all the knowledge about aggregate equilibrium relationships of a rational expectations 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 $\upsilon_{i,t}$

$$\max_{P_{it}} \min_{\widehat{x}(P_{it}|\varepsilon^{t-1})} E^{X(P_{it}|\varepsilon^{t-1})} \upsilon_{i,t} \tag{40}$$

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.

We find it useful to compare this economy with one where the only difference is that the measure zero firms knows the true demand function but are subject to physical menu costs when changing their price. For a better comparison, these firms still only maximize the static profit, which in this case is

$$\upsilon_{i,t}^{MC} = \left(\frac{P_{it}}{P_t} - \frac{\chi S_t}{\omega_{it} A_t P_t} \right) Y_{i,t} - f W_t I_{P_{it} \neq P_{it-1}} \tag{41}$$

where the latter term reflect the menu cost expressed in wages paid. The objective of this comparison is to help us understand what does the new type of cost of not changing the price proposed in this paper brings compared to the standard menu cost.

Optimization problem and equilibrium

The individual ambiguity-averse firm takes as given the only relevant aggregate state variable, namely the aggregate price level, given by (39), and maximizes the objective given by (40), where profits are defined in (36), subject to the demand uncertainty in (38) and the assumed information structure. Finally, we compare the individual pricing behavior of this type of firm with its menu cost version, where profits are determined in (41).

[21] A similar approach of a flexible aggregate price level is taken 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. In that case, we need to employ usual Krusell-Smith type conjectures on the equilibrium aggregate price level.

GLS: Discounting past observations

Due to the unobserved noise ν_{jt}, there are random disturbances to the unknown demand function, and hence $x_t(\hat{r}_{it})$ is time varying as in (32). This time-variation essentially implies that there is added error in signals further into the past, and hence those signal have lower signal-to-noise ratio than more current signals. The firm understands this, and accounts for it in the optimal way, given its understanding that the induced time-variation is given by (32), by placing a progressively smaller weight on past observations. The exact weighting of observations is computed by maximizing the likelihood function of the data. Let the vector $\mathbf{y} = [y_{it}, y_{i,t-1}, \ldots, y_{i,0}]'$ denote the set of signals at a given point $\hat{r}_0$ the firm has seen in the past. They have a joint-Normal distribution and we denote their variance as Σ_y, which accounts for the extra observational error in past signals. Lastly, let $e_1 = [1, \ldots, 1]$, and then the optimal weighting of observations is given by

$$\hat{x}_t(p) = \frac{e_1 \Sigma_y^{-1} \mathbf{y}}{e_1 \Sigma_y^{-1} e_1}$$

7.2 Results

7.2.1 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 $\mu = 0.00046$ to match is set 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 $\rho_a = 0.975$ 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 7 parameters that refer to the firm's problem. We start by choosing an elasticity of substitution of $b = 6$, implying a markup of 20%. We set the critical value used in the statistical evaluation step $\delta_\alpha = 1.96$, corresponding to a 95% confidence interval.

There are thus 5 parameters left. For this 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 using empirical evidence on how difficult is it to predict the one-period-ahead quantity. In particular, using our 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 do this across all items within a category/market and also for the item with

most sales in its category. We compute the absolute in-sample prediction error $(Q - X\widehat{\beta})/\overline{Q}$, where $\widehat{\beta}$ are the regression coefficients based on the regression and $\overline{Q}$ is the mean quantity.

Table 4 reports the results for the moments of the prediction error of these types of regression. We find that the median absolute ranges from 18% to 48% of the average quantity. We calibrate σ_z to generate a similar median error for the prediction of quantity under the true DGP of our model. For the benchmark model we use $\sigma_z = 0.5$, which corresponds to a median forecast error of $0.50 * 0.675 = 0.3375$, matching our sample average.[22]

The persistence and standard deviation of the idiosyncratic productivity are parameters that are standard in menu cost models. That literature suggests using pricing moments such as the fraction of price increases and the average size of price changes to calibrate them (see for example Vavra (2014)). There are 2 parameters that are novel to the learning model proposed here. The first one is the width of prior tunnel, controlled by ν, which is the multiple of standard deviations that the firm uses to form the initial set of possible demand. The second one is the time-variation of the underlying demand function, given by σ_ε^2.

For the two learning parameters we find it informative to use the following two pricing moments: the frequency of posted price changes and the frequency of 'reference price' changes. As in Gagnon et al. (2012), we define a 'reference price' the modal price within a rolling window of 13-weeks. The ambiguity parameter comes out at making the width of the prior tunnel equal to plus-minus two standard deviations of the demand shock, a bound argued as reasonable in Ilut and Schneider (2014). The parameter for σ_ε^2 means that each past observation receives a weight in the estimation of the function that declines by a factor of 95%. Finally, for the menu-cost model, where the information parameters ν and σ_x^2 do not matter, we calibrate f to the same frequency of posted price changes, conditional on the rest of the structural parameters being the same as in the ambiguity model. Table 2 presents the whole set of parameters.

Table 2: Parameters

β	μ	σ_s	ρ_a	σ_a	δ_α	σ_z	ρ_ω	σ_ω	ν	σ_x	f
$0.97^{(1/52)}$	0.00046	0.0015	0.975	0.003	1.96	0.5	0.95	0.09	2	0.23	0.023

7.2.2 Pricing behavior

Pricing moments

Table 3 presents pricing moments generated by the models against their empirical counterpart. Only the first four moments are targeted by the calibration.

[22] Here we used that $\Phi(-0.6745) = 0.25$, with $\Phi(.)$ denoting the standard normal cdf.

Table 3: Pricing moments

Moment	Data	Model w ambiguity	Model w menu cost
(1) Fraction of price increases	51%	52.14%	53.64%
(2) Average size of price changes	25.4%	25.02%	16.94%
(3) Frequency of posted price changes	22.85%	21.73%	21.96%
(4) Frequency of 'reference price' changes	5.96%	5.29%	9.31%
(5) Fraction of time at modal price	85.4%	72.8%	59.4%
(6) Probability modal price is the max price	75.96%	46.07%	52.06%
(7) Probability of revisiting a price	62.1%	66.07%	0%
(8) Average number of unique prices (13 weeks)	2.62	2.5	3.62

The baseline model does a good job at matching the targeted moments. It is interesting to note that by keeping the rest of the structural parameters the same, we can find a menu cost value so that the model with rational expectations matches the same frequency of posted price changes. However, this comes at a significant loss in matching moments (2) and (4). Thus, the ambiguity model, in the process of matching the frequency of reference price changes, also generates a larger average size of changes. The reason is that most of the changes in the baseline model come from large switches between reference prices, rather than the smaller changes implied by the menu cost. These large changes come in the ambiguity model together with sticky reference prices.

The additional pricing moments show the main mechanism of the model. Moment (5) indicates that the modal price accounts for a large fraction of the posted prices, similar to the data. Movements from the modal price are relatively symmetric, as suggested by moment (6). In the data the modal price is instead more likely to be the maximal price. The model generates strong memory in prices, so that conditional on a price change, the probability of selecting the same price in the last 26 weeks is about 66%, very similar to the empirical moment. A window of 13 weeks experiences a relatively small number of unique prices, as in the data. In comparison, in the model with menu cost, prices spend much less time at the modal price, have no price memory and are less discrete within a specific time-window.

Figure 25 plots the histogram of the price changes implied by, respectively, the ambiguity and menu cost models. The latter produces a bimodal distribution, typical for menu cost models. Compared to this, the ambiguity model produces both a bigger mass of small price changes and a bigger mass of large price changes. Figure 26 plots the distribution of price changes for a 'typical' category/market in our dataset, namely salted snacks in New York. While some of the larger spikes can be attributed to 'sales', the data indicates a high frequency of both large and small price changes.[23]

[23]See Alvarez et al. (2014) for an analysis of small and large price changes in the French and US data.

The reason for generating larger price changes is the existence of kinks in the policy function and the resulting potential for frequent, large price changes as the firm switches between the prices at those kinks. Small price changes are generated because the policy function resembles the flexible price policy in some situations. On the one hand, this can happen because the history of shocks may be such that the 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 usually tracks the flexible price outside the kinks. Thus, because of the endogeneity of what appears as a cost of changing a price in the ambiguity model, there large and small price changes co-exist.[24]

Figure 27 plots a typical sample of shocks (demand, productivity and money supply) and the implied pricing behavior in the top left panel. Figure 28 shows the implied flexible price path for a firm hit with the same shocks. Figure 29 contains the path of the ambiguity averse firm. As indicated by the population moments, this path is characterized by discreteness and strong memory: prices return very often to the same values. Figure 30 shows in more depth that the menu cost model behaves differently than the ambiguity model. In particular, as suggested by table 3 there is no price memory for the former. Moreover, the conditional behavior of the two paths is different in the sense that the price flexibility is not greatly synchronized, although by construction the two paths have very similar unconditional frequency of price changes.

Policy functions

Figure 31 plots pricing policies, as functions of the idiosyncratic productivity, in the case that two previous prices have been observed once each. The kinks are very small and thus the policy function resembles the flexible price one, and is characterized by large flexibility and likely small price changes. However, as the number of observations at those prices increases, in the case plotted in Figure 32 to ten, the kinks become deeper. In this situation we will mostly observe few and large (discrete) price changes, with switches to and from 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, from the perspective of monetary non-neutrality, is the optimal pricing behavior as a function of monetary policy shocks. We define here the degree of monetary neutrality 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. Figure 33 plots pricing policies in the case of one previously observed price. Compared to the menu cost

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

version, the implied inaction is smaller and thus the monetary neutrality stronger. As we increase the number of times that this price has been observed, the inaction region becomes wider, to the extent that it generates more stickiness than the menu cost version, as shown in Figure 34, which plots the case of ten such previous observations.

Having multiple observed prices leads to different effects of monetary shocks. Figure 35 plots the case in which two previous prices have been observed once each. We see that there are two flat areas in the pricing function corresponding to those two previously experienced prices. Compared to the one observed price case, from Figure 33, there is more inaction since now there is a second kink and thus there is a more significant effect of the monetary policy shock. Compared to the menu cost version, the implied inaction is still smaller and thus the monetary neutrality stronger.

As the number of observations at those prices increases, in the case plotted in Figure 36 to ten, the kinks at the two prices become deeper. Compared to the one observed price case, from Figure 34, the monetary shocks in this case have smaller overall effects. The reason is that with one deep kink the price would essentially not respond at all to the shock. However, with two kinks, while the price would stay fixed for a range of policy shocks, it would also switch drastically to move to the other kink. When it switches, the price gets closer to the flexible version.

Indeed, in this example, if we start from the mean and analyze negative monetary policy shocks, we see that the price is first above the flexible one, then switches to the other kink and is thus below the flexible line. As the shock becomes even more negative, the optimal price policy intersects with the flexible version. For this range of monetary policy shocks, the fact that there are two kinks implies that the optimal markup is not so far from the flexible one compared to menu cost or the one kink version. In addition, this behavior also means that the sign of the effect on the average quantity sold changes with the size of the monetary policy shock. Consider again a monetary policy shock that is initially negative until the optimal price switches from the high to its low value. For that range, the price is too high compared to the flexible price and the firm sells on average less. However, after the switch to the low value, the firm prices at a markup below the flexible one and in the process sells more on average. As the monetary shock becomes increasingly negative the firm sticks to this price which eventually will lower again its quantity below its flexible version. As the inaction region extends further to the left than the menu cost version, this negative effect on quantity will be significant even for large negative shocks.

To summarize, monetary policy shocks lead to effects on optimal prices 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 lead the firm to optimally select prices that implies larger kinks. Following such a history, 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.

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 a posted price. The limited knowledge allows the firm to only characterize likely bounds on the possible non-parametric 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 drastic 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 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 in order 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 parsimonious mechanism naturally predicts that prices should be sticky, unless shocks are sufficiently large. In addition, it can explain many pricing facts that have proven challenging for standard models to match: prices exhibit 'memory' as firms find it optimal to stick to a discrete distribution of prices; retailers often alternate between a regular and a sale price, both sticky; the probability of observing a price change is decreasing in the time

since the last price movement; small and large price changes coexist in the data.

While our mechanism appears promising, more work remains to test empirically its predictions but also its implications for the aggregate properties of DSGE models. For example, the fact that firms are more reluctant to change prices after a long period of inaction may have interesting implications, such as history dependence or even asymmetry in an economy's response to shocks. This could, in turn, imply some novel policy implications. We plan on exploring these avenues further in future research.

References

Alvarez, F., H. L. Bihan, and F. Lippi (2014): "Small and large price changes and the propagation of monetary shocks," NBER Working Paper.

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*, 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.

Blundell, R., M. Browning, and I. Crawford (2008): "Best nonparametric bounds on demand responses," *Econometrica*, 1227–1262.

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

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

Burstein, A. T. (2006): "Inflation and output dynamics with state-dependent pricing decisions," *Journal of Monetary Economics*, 53, 1235–1257.

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

Carlton, D. W. (1986): "The Rigidity of Prices," *American Economic Review*, 76, 637658.

Cecchetti, S. G. (1986): "The Frequency of Price Adjustment: A Study of the Newsstand Prices of Magazines," *Journal of Econometrics*, 31, 255274.

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.

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.

——— (2007): "Learning Under Ambiguity," *Review of Economic Studies*, 74, 1275–1303.

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.

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.

Ichimura, H. and P. E. Todd (2007): "Implementing nonparametric and semiparametric estimators," *Handbook of Econometrics*, 6, 5369–5468.

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

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.

L'Huillier, J.-P. (2011): "Consumers imperfect information and price rigidities," Manuscript, Einaudi Institute for Economics and Finance.

Li, J. (2013): "The K-armed bandit with multiple-priors," Manuscript, McGill University.

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.

——— (2015): "Rigid pricing and rationally inattentive consumer," *Journal of Economic Theory*, forthcoming.

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.

——— (2011): "Price setting in forward-looking customer markets," *Journal of Monetary Economics*, 58, 220–233.

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

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.

——— (2005): "Customer anger at price increases, changes in the frequency of price adjustment and monetary policy," *Journal of Monetary Economics*, 52, 829–852.

——— (2011): "Fair pricing," *Journal of the European Economic Association*, 9, 952–981.

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

SHESHINSKI, E. AND Y. WEISS (1977): "Inflation and costs of price adjustment," *The Review of Economic Studies*, 287–303.

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.

Table 4: Predicting demand

		(1) Across all items				
		Median	p10	p25	p75	p90
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: $(Q - X\widehat{\beta})/\overline{Q}$.

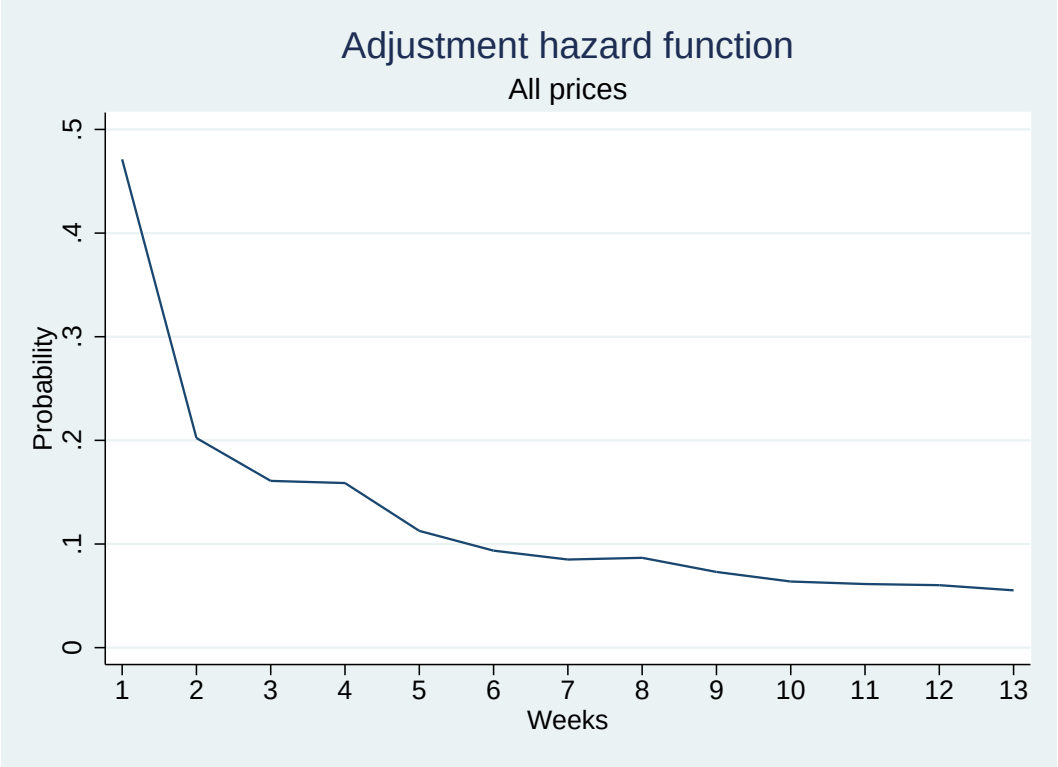


Figure 1: Figure

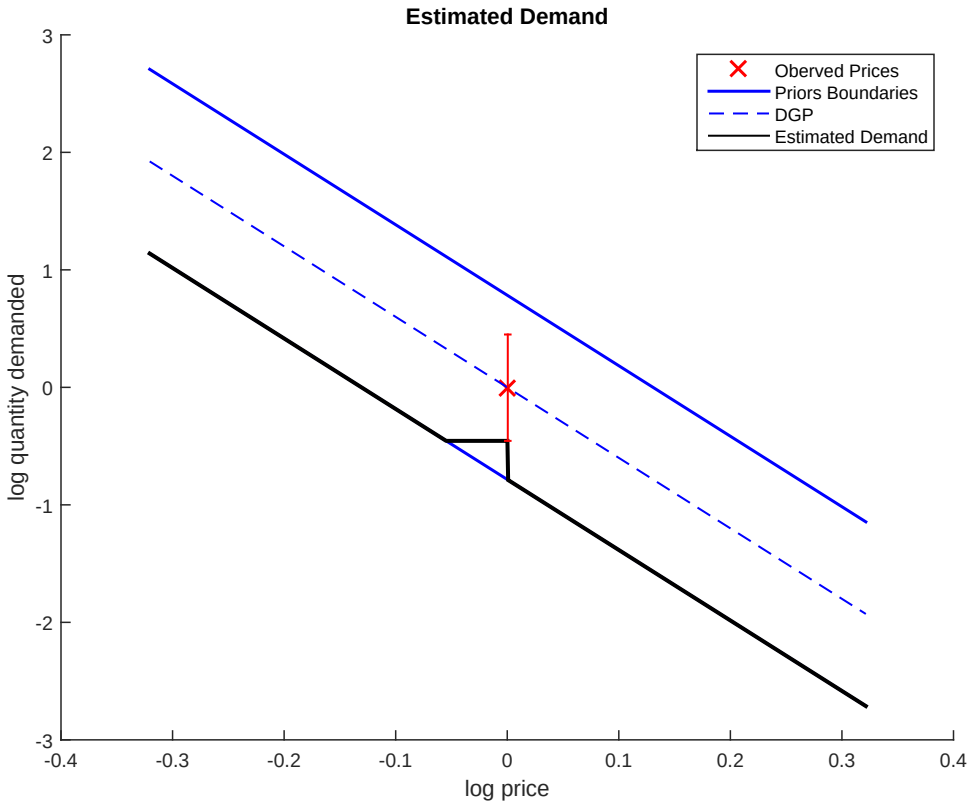


Figure 2: Figure

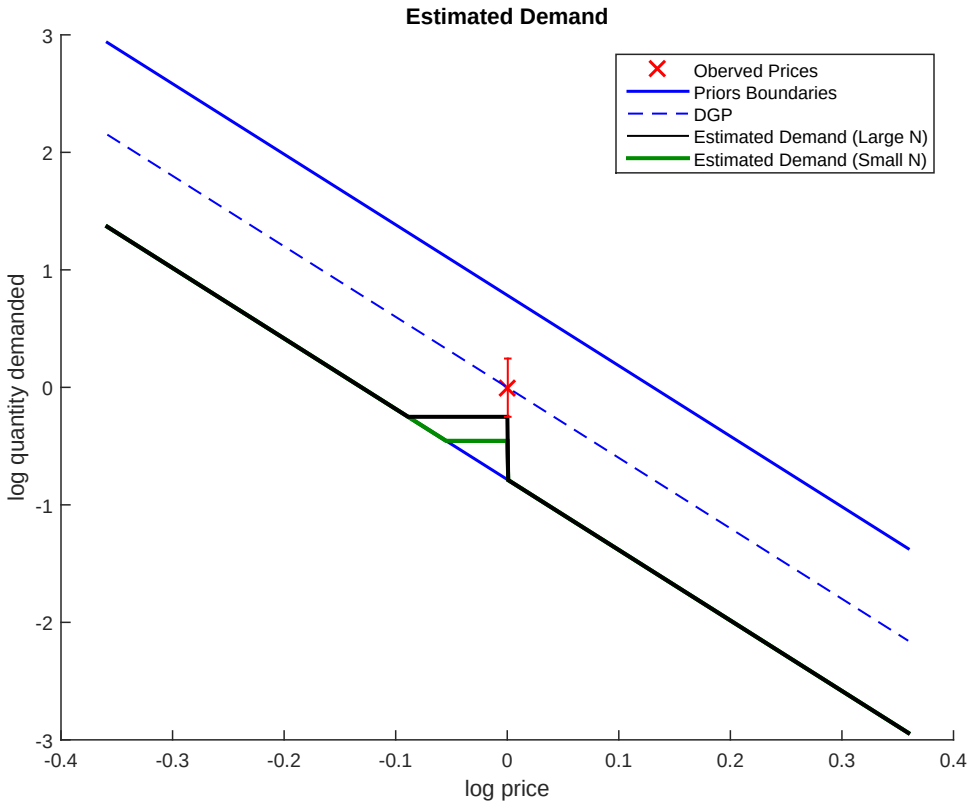


Figure 3: Figure

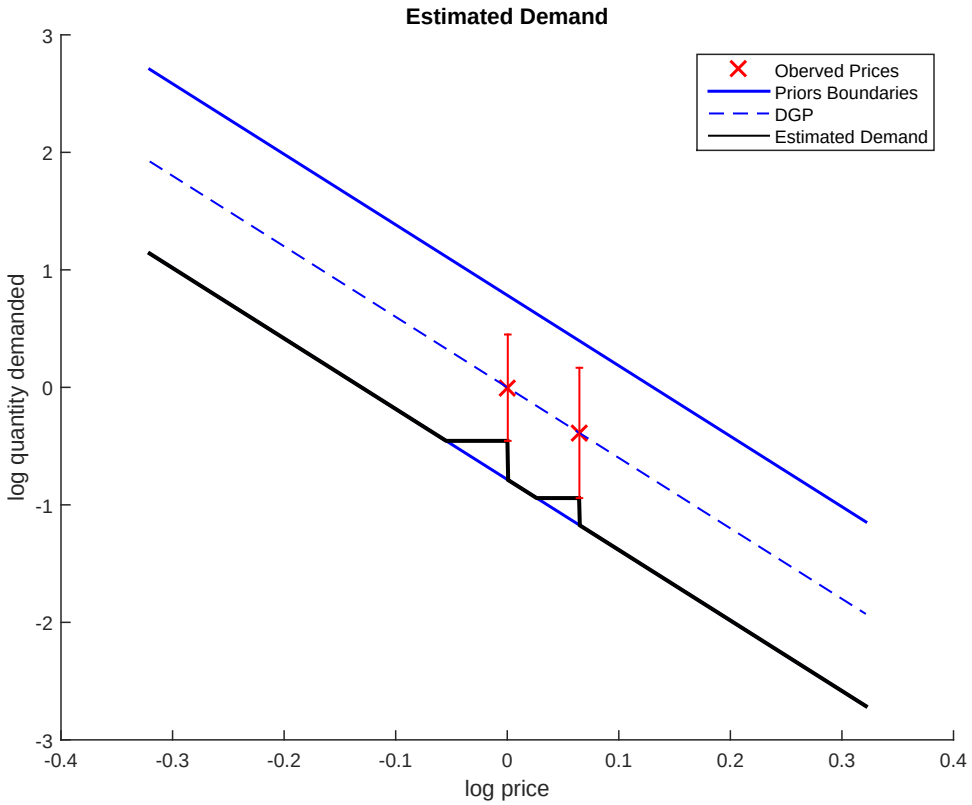


Figure 4: Figure

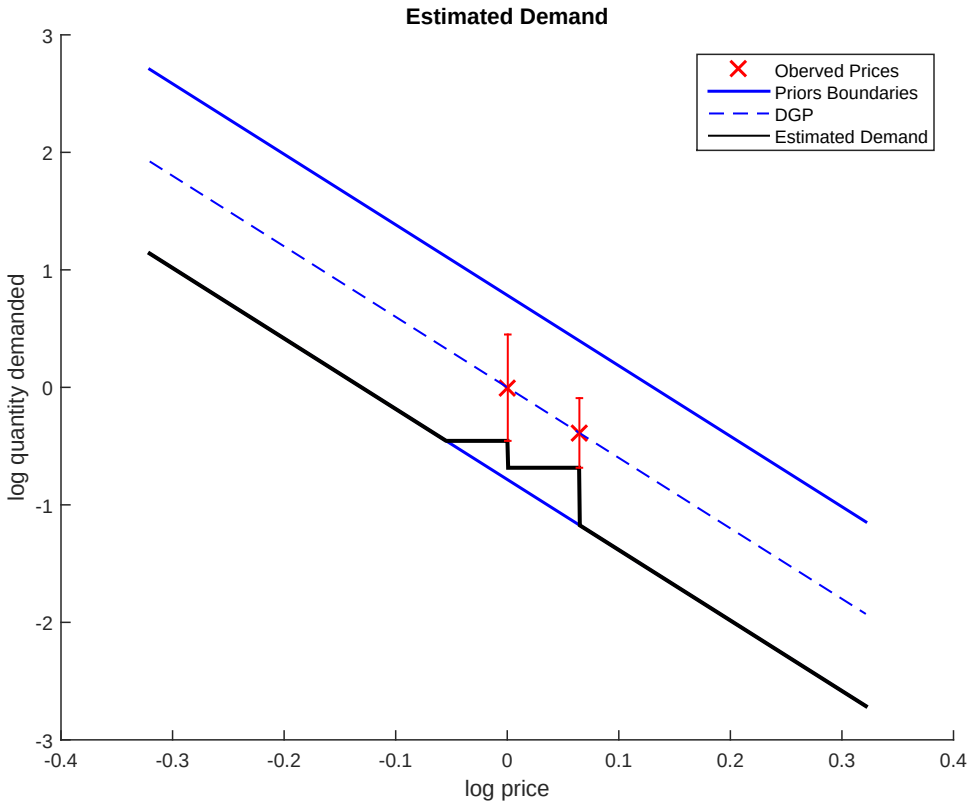


Figure 5: Figure

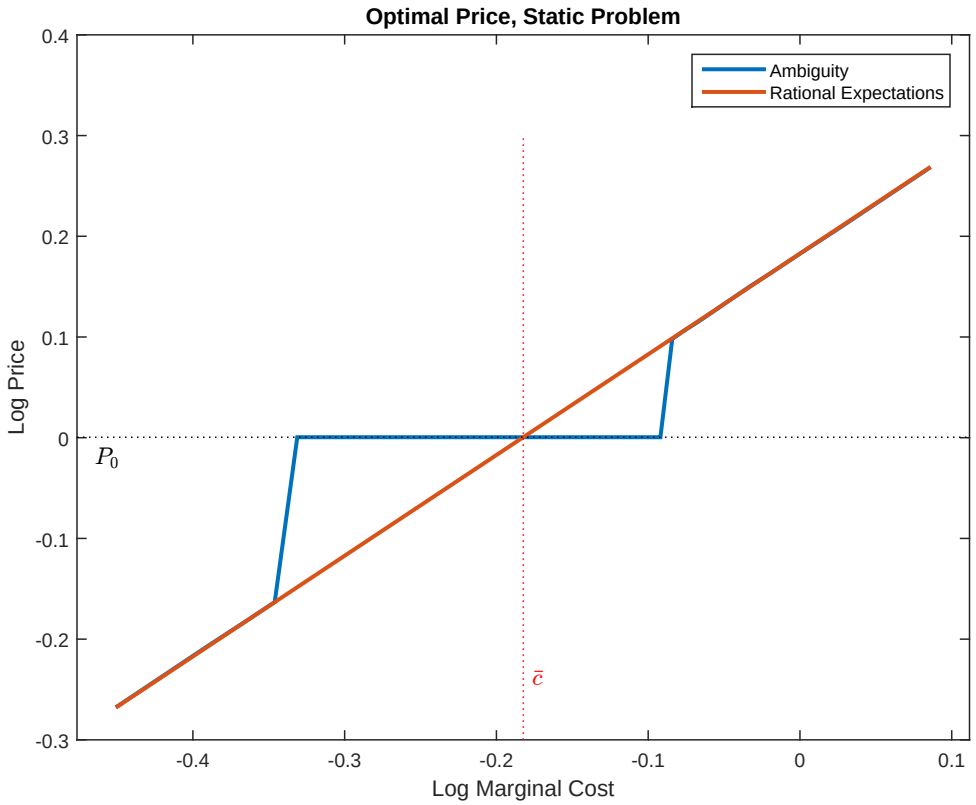


Figure 6: Figure

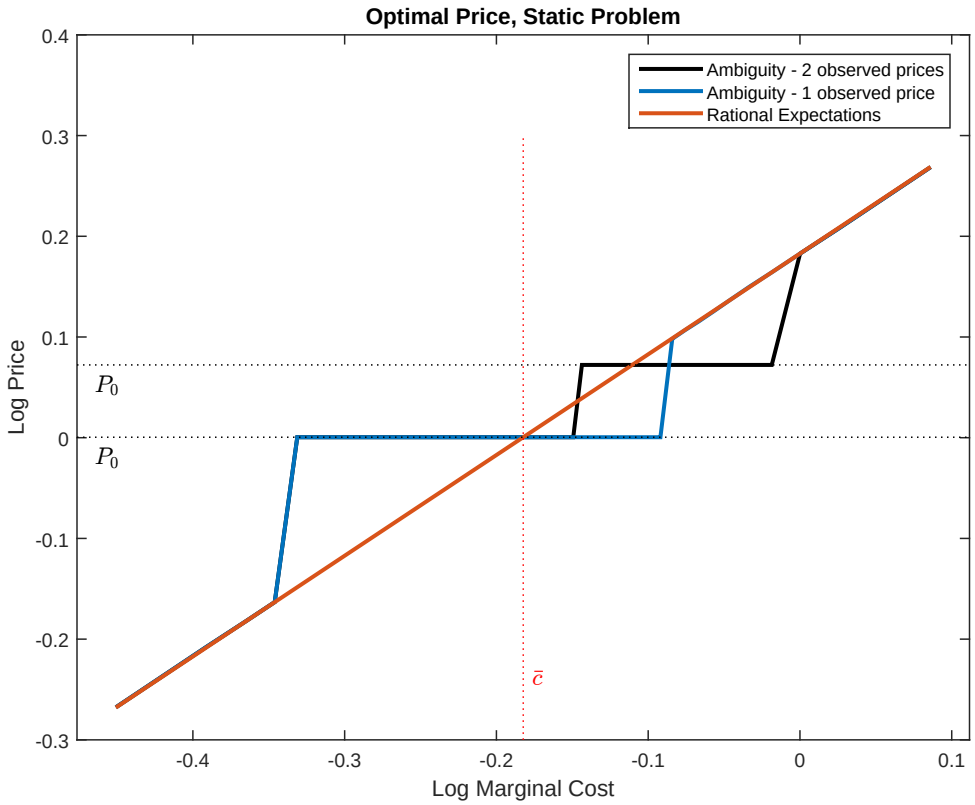


Figure 7: Figure

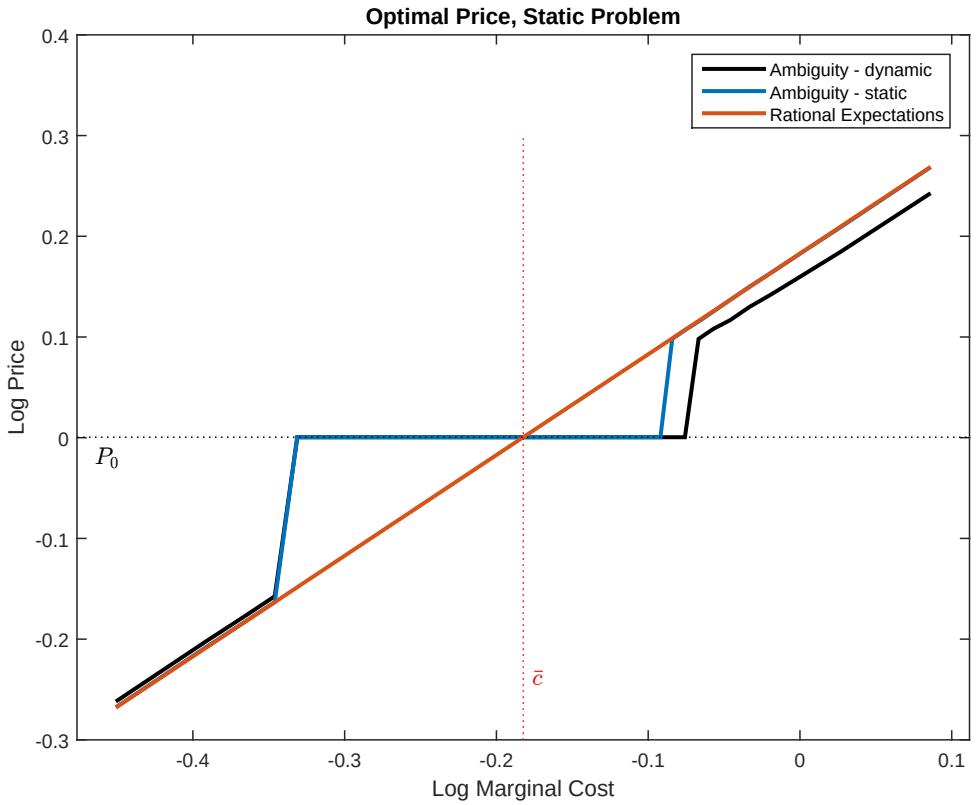


Figure 8: Figure

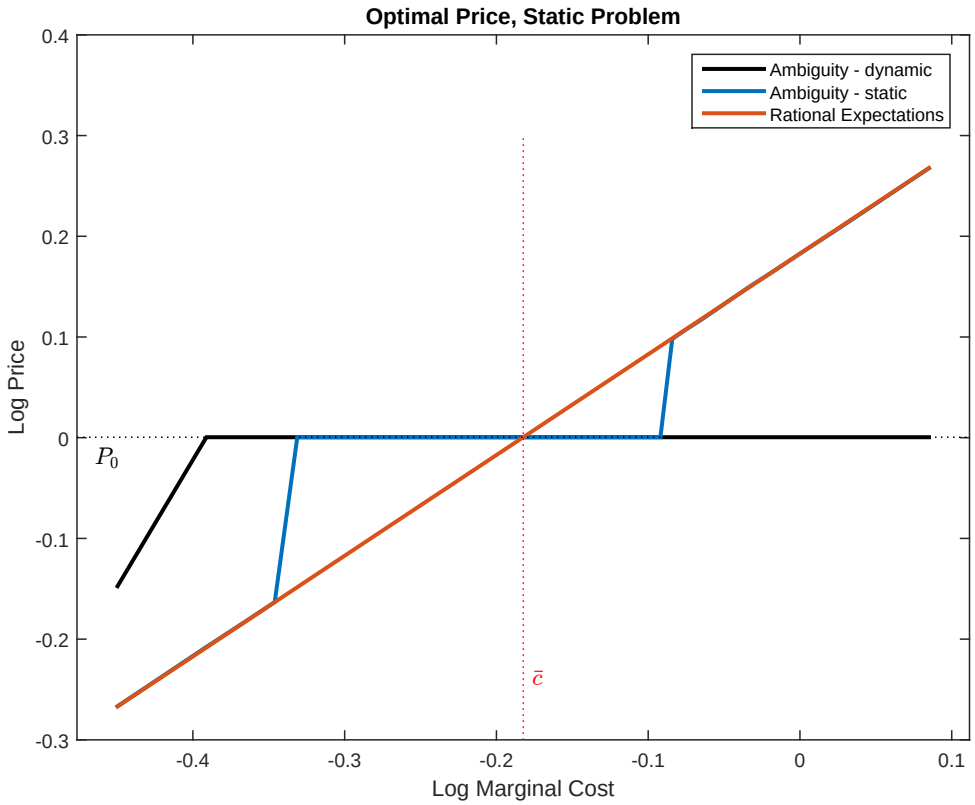


Figure 9: Figure

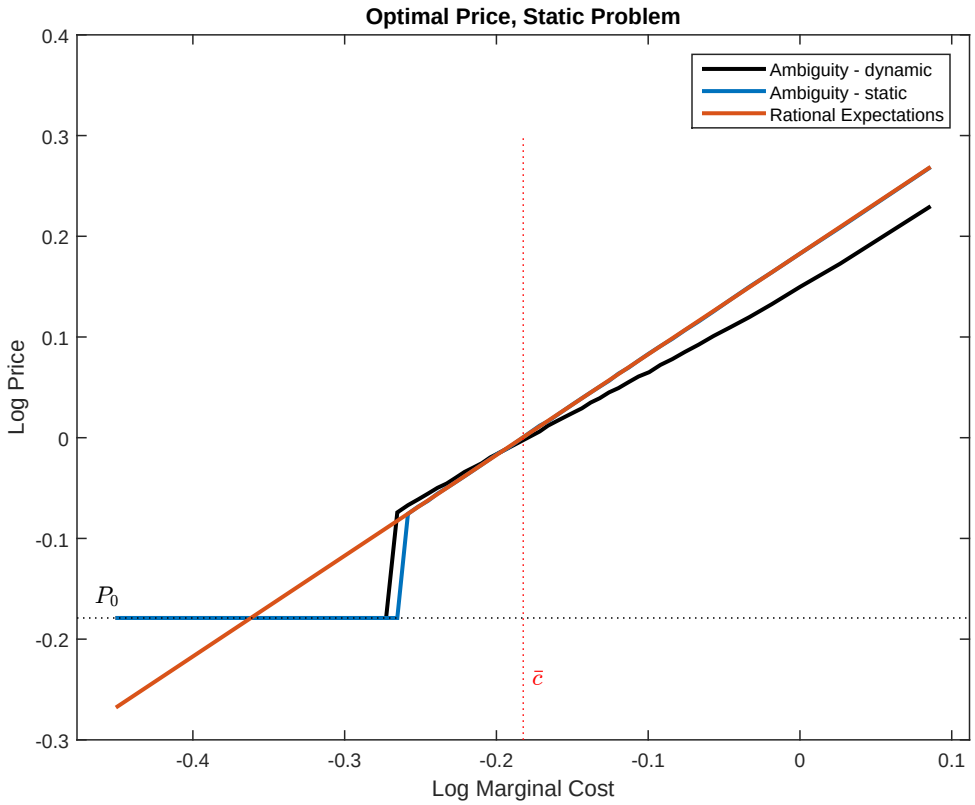


Figure 10: Figure

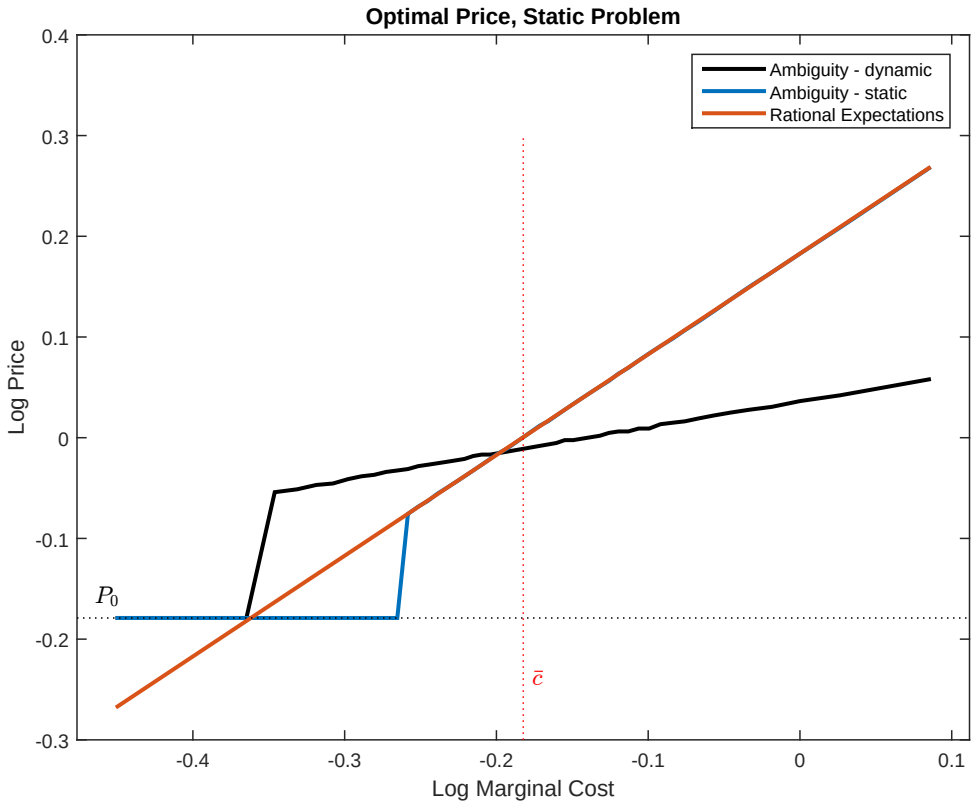


Figure 11: Figure

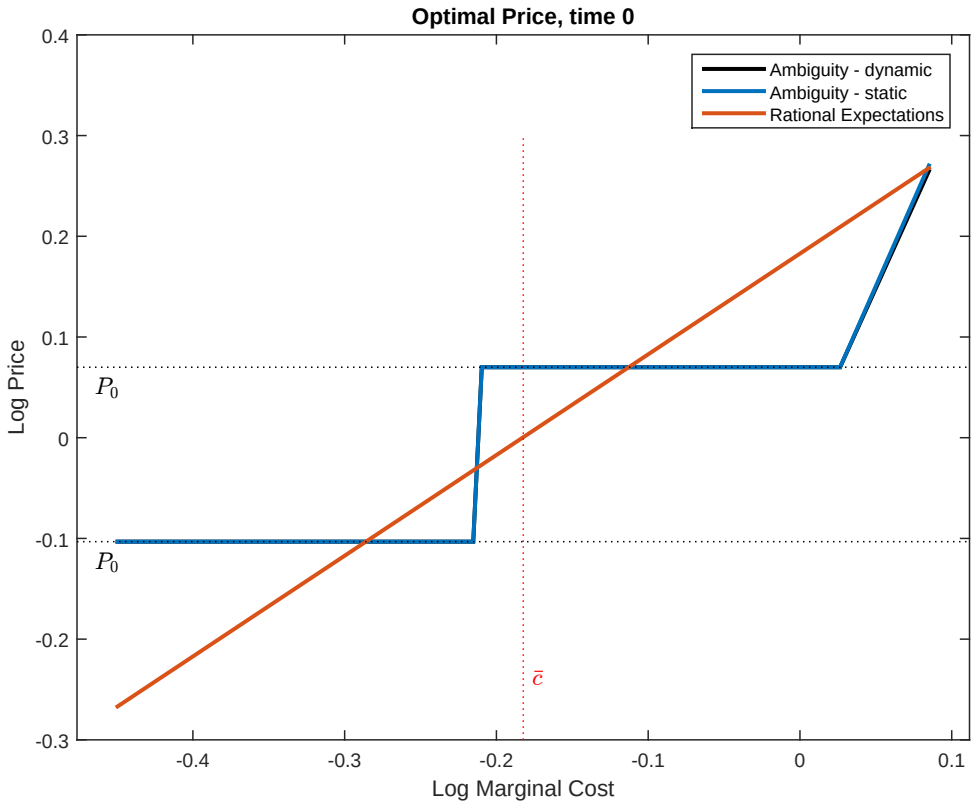


Figure 12: Figure

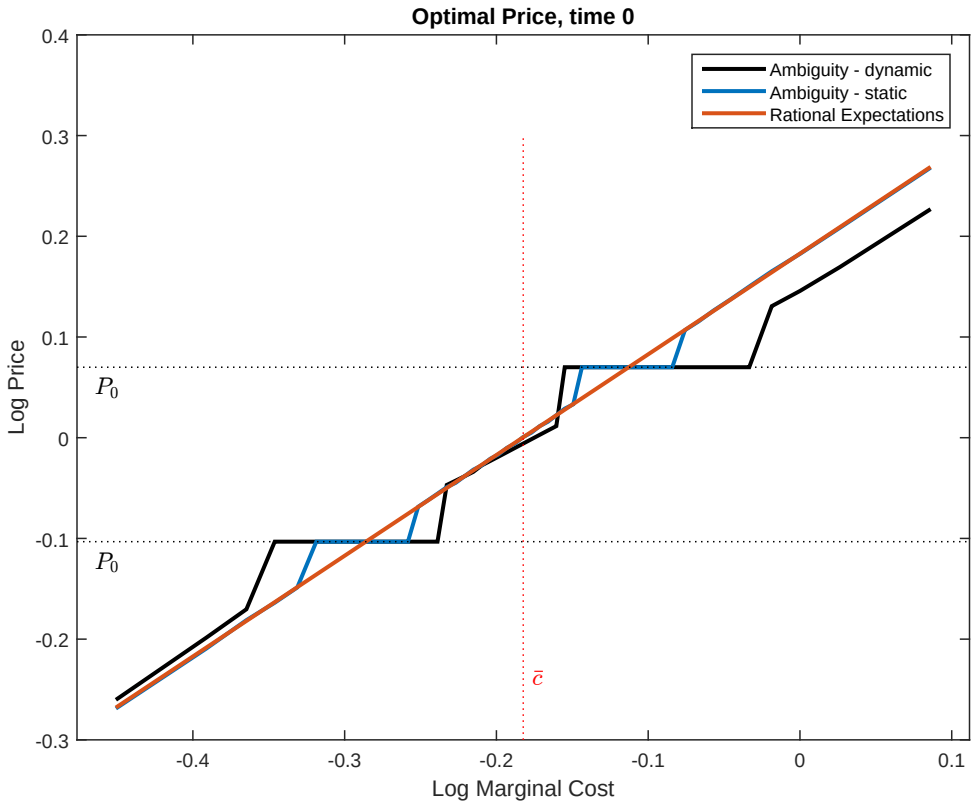


Figure 13: Figure

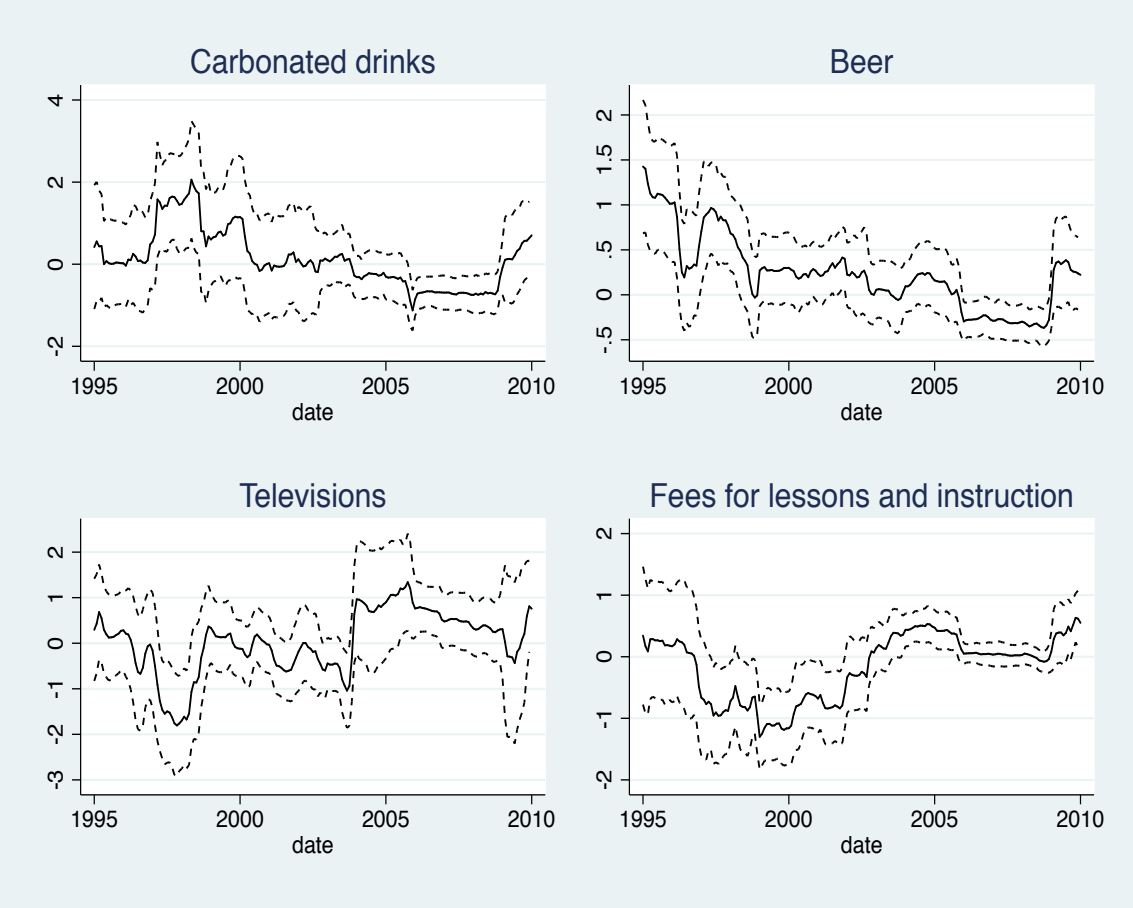


Figure 14: 3-year rolling regressions of 3-month industry inflation on 3-month aggregate inflation for four categories. The solid line plots the point estimate of regression coefficient on aggregate inflation. The dotted lines plot the 95% confidence intervals.

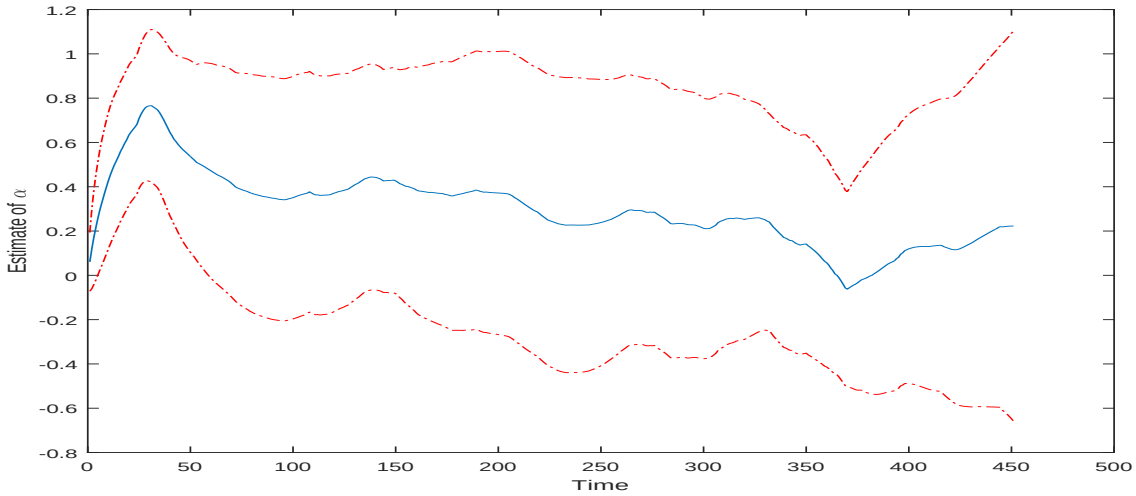


Figure 15: Effect of current aggregate inflation on Carbonated drinks industry inflation, both measured over a 1 month horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% confidence interval using the smoothed estimated uncertainty.

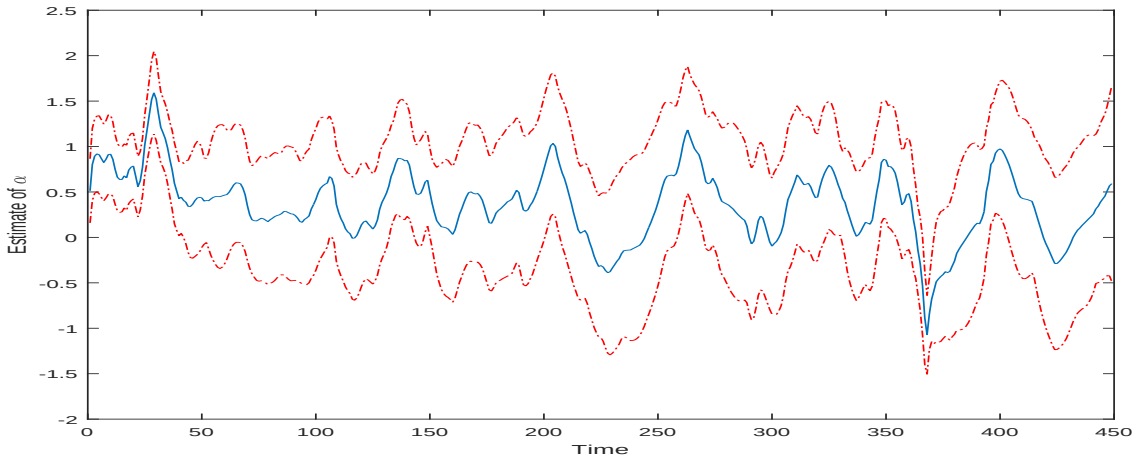


Figure 16: Effect of current aggregate inflation on Carbonated drinks industry inflation, both measured over a 3 months horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% confidence interval using the smoothed estimated uncertainty.

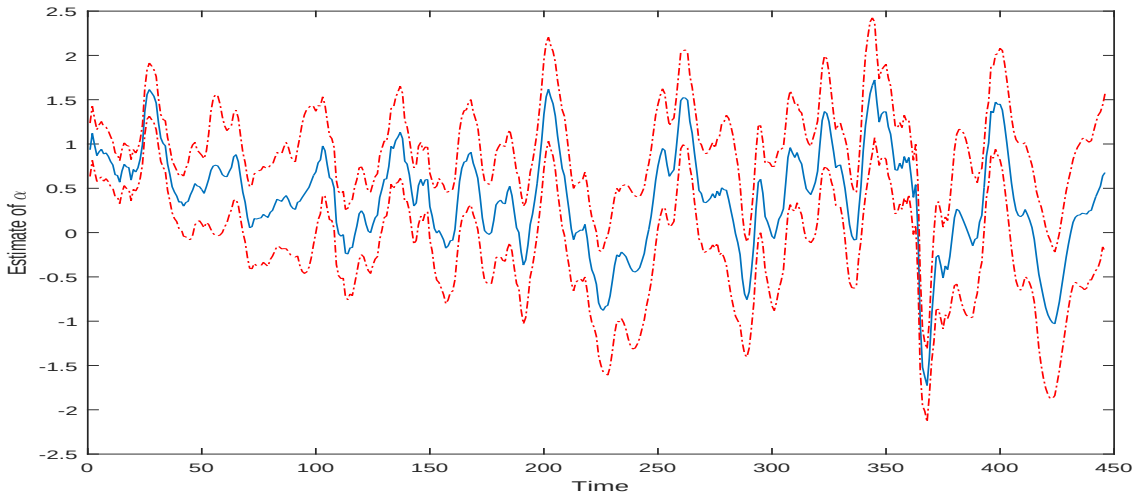


Figure 17: Effect of current aggregate inflation on Carbonated drinks industry inflation, both measured over a 6 months horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% confidence interval using the smoothed estimated uncertainty.

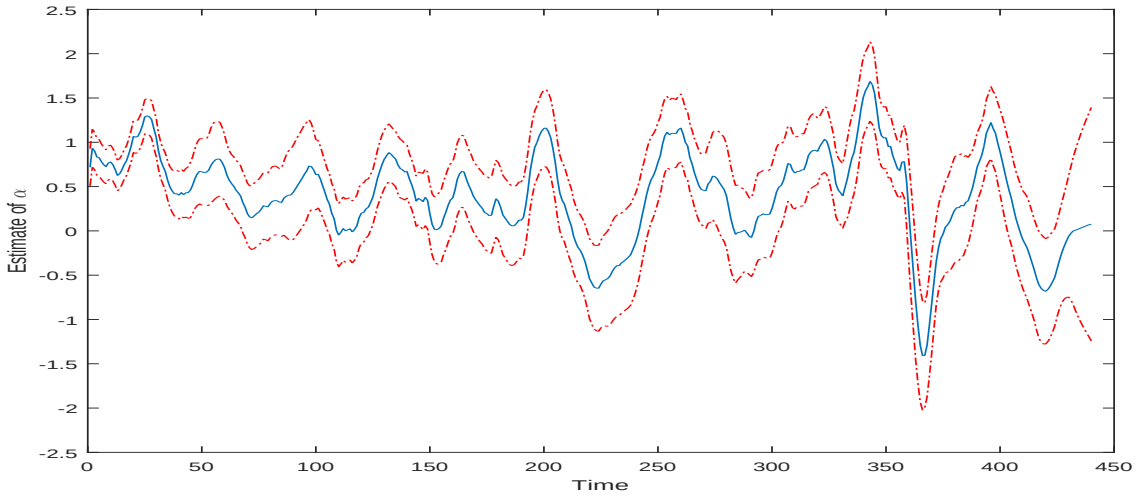


Figure 18: Effect of current aggregate inflation on Carbonated drinks industry inflation, both measured over a 1 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% confidence interval using the smoothed estimated uncertainty.

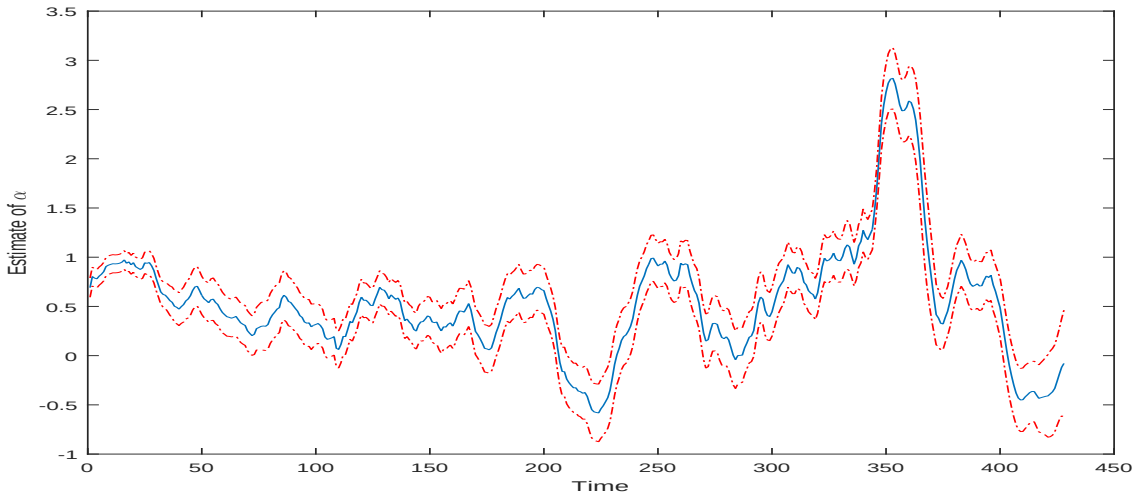


Figure 19: Effect of current aggregate inflation on Carbonated drinks industry inflation, both measured over a 2 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% confidence interval using the smoothed estimated uncertainty.

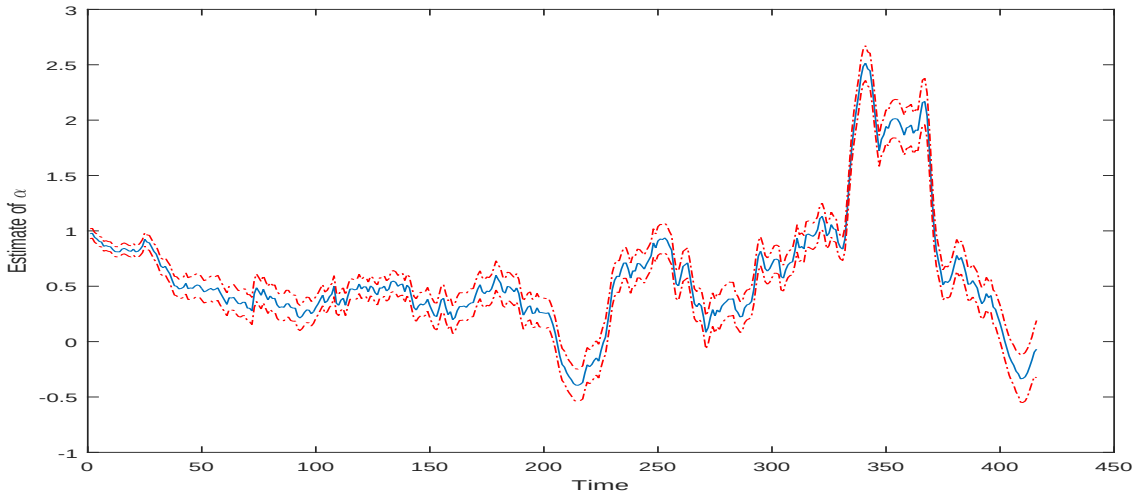


Figure 20: Effect of current aggregate inflation on Carbonated drinks industry inflation, both measured over a 3 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% confidence interval using the smoothed estimated uncertainty.

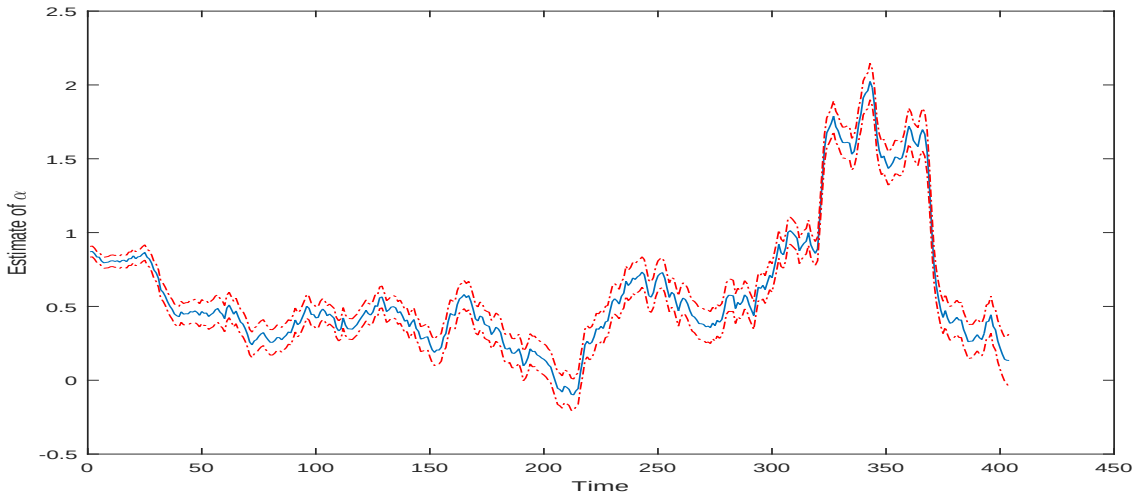


Figure 21: Effect of current aggregate inflation on Carbonated drinks industry inflation, both measured over a 4 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% confidence interval using the smoothed estimated uncertainty.

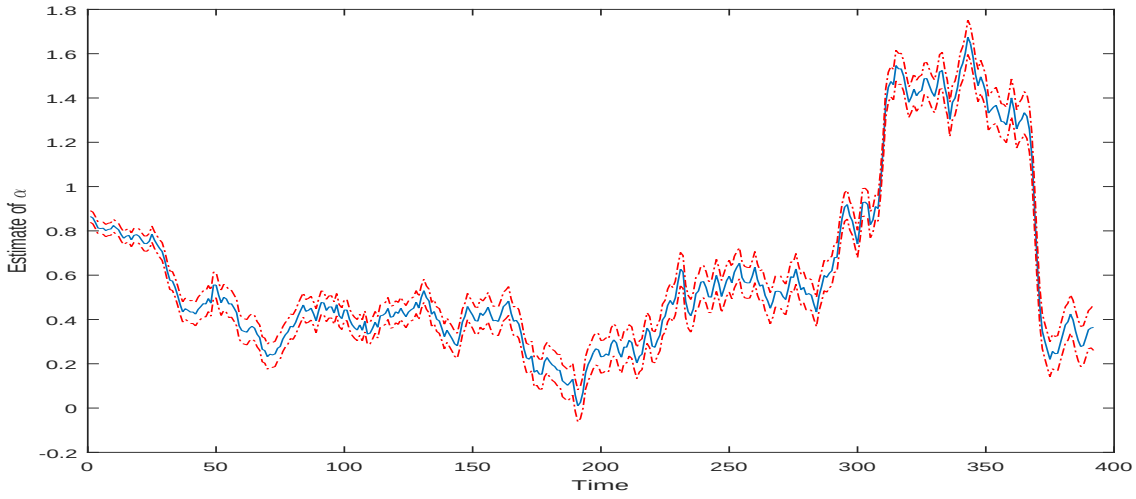


Figure 22: Effect of current aggregate inflation on Carbonated drinks industry inflation, both measured over a 5 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% confidence interval using the smoothed estimated uncertainty.

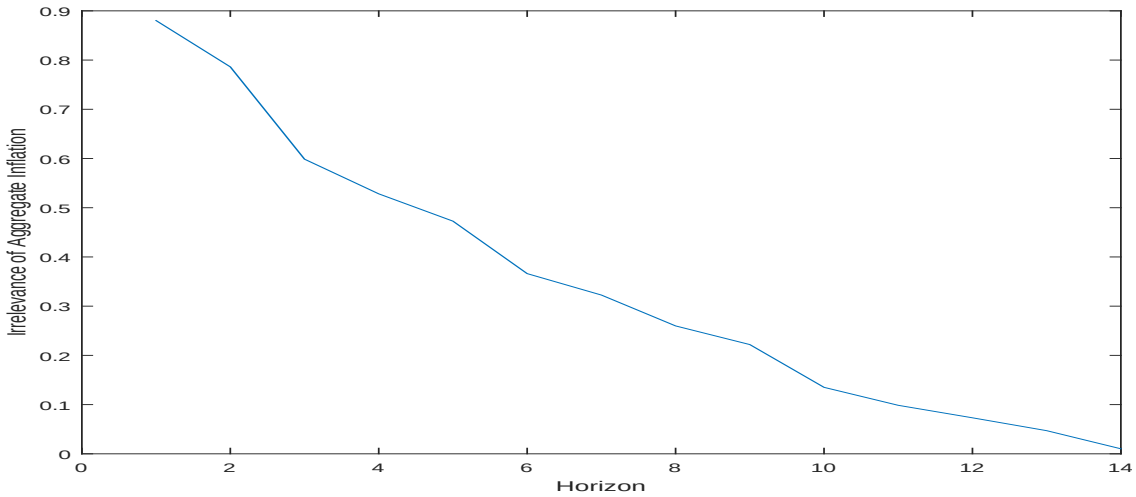


Figure 23: Fraction of times that the current aggregate inflation has an estimated effect on Carbonated drinks industry inflation that is not statistically different from zero, at 95% confidence interval. Inflations are computed over various horizons, ranging from 1 month to 5 years.

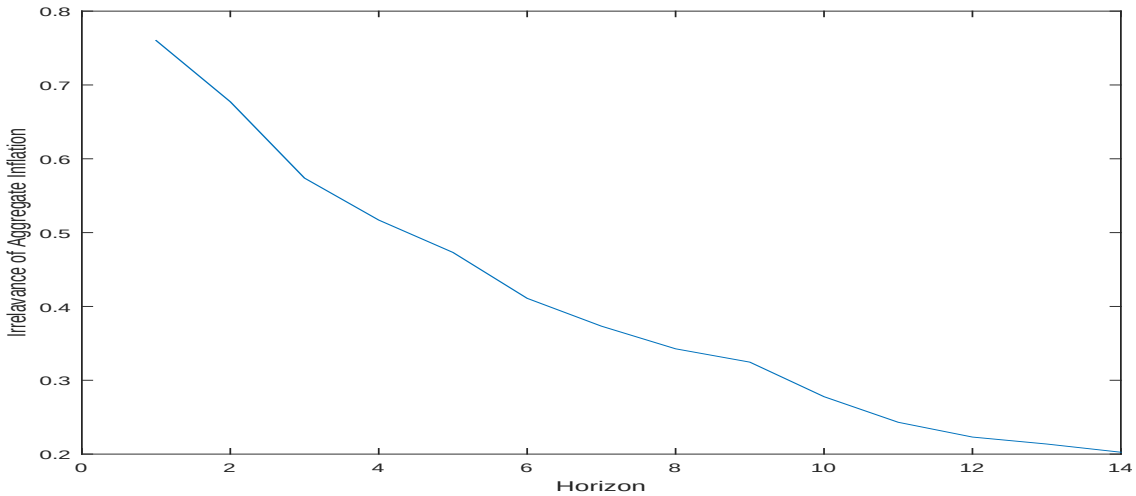


Figure 24: Average, over industries, of the fraction of times that the current aggregate inflation has an estimated effect on industry inflation that is not statistically different from zero, at 95% confidence interval. Inflations are computed over various horizons, ranging from 1 month to 5 years.

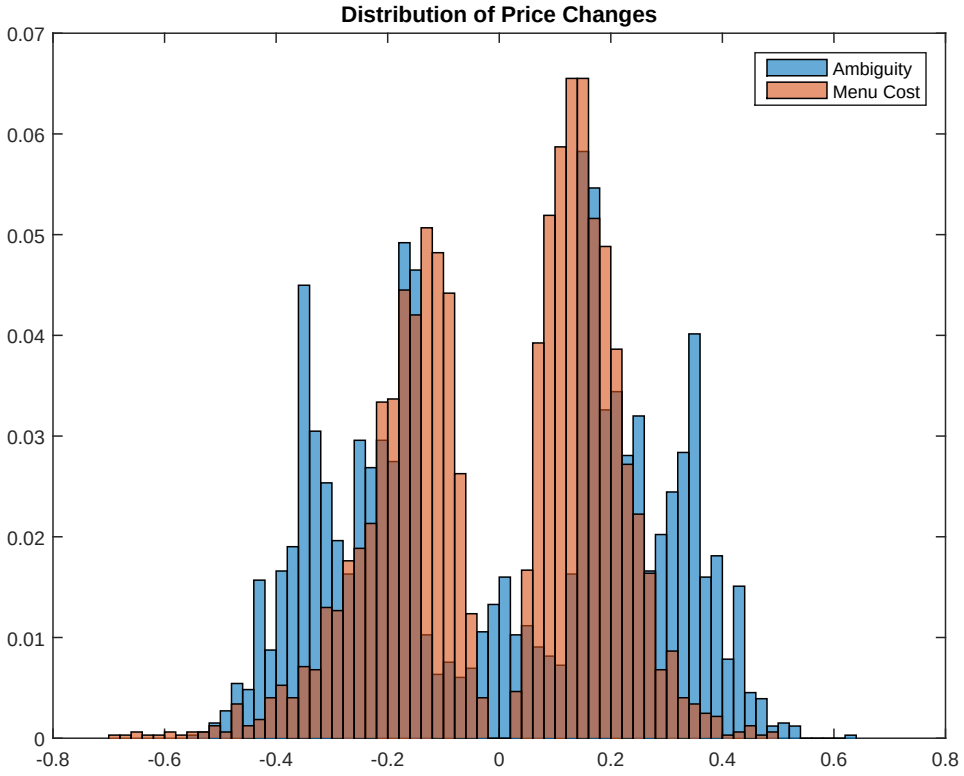


Figure 25: Figure

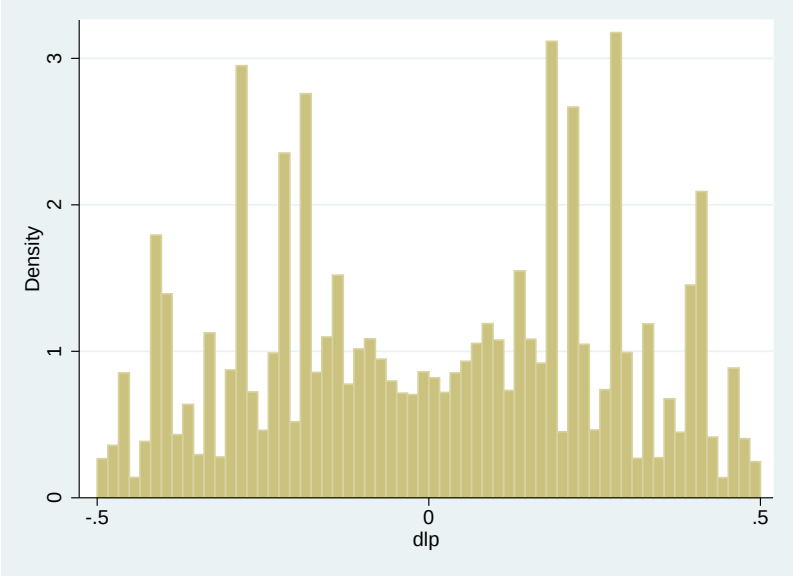


Figure 26: Figure

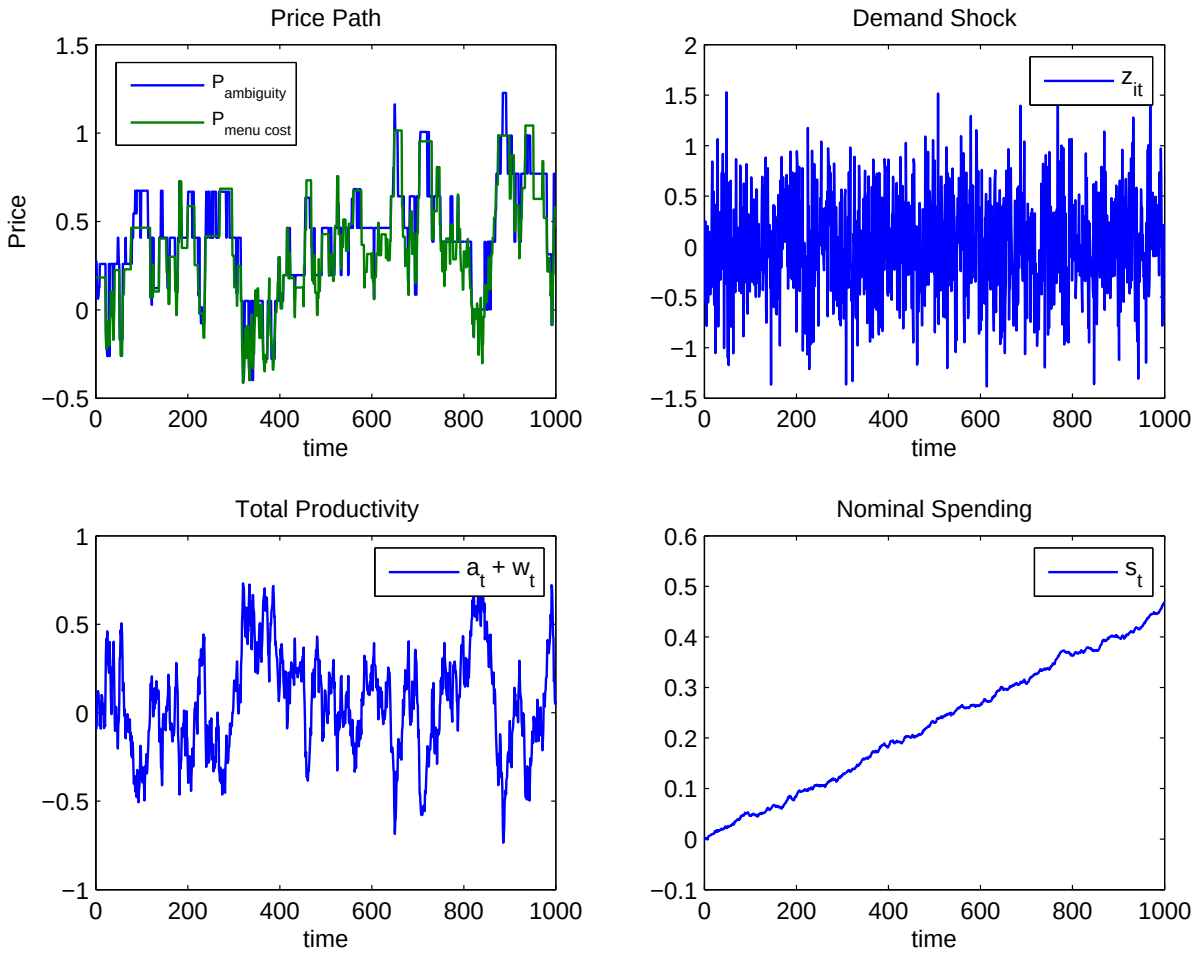


Figure 27: Figure

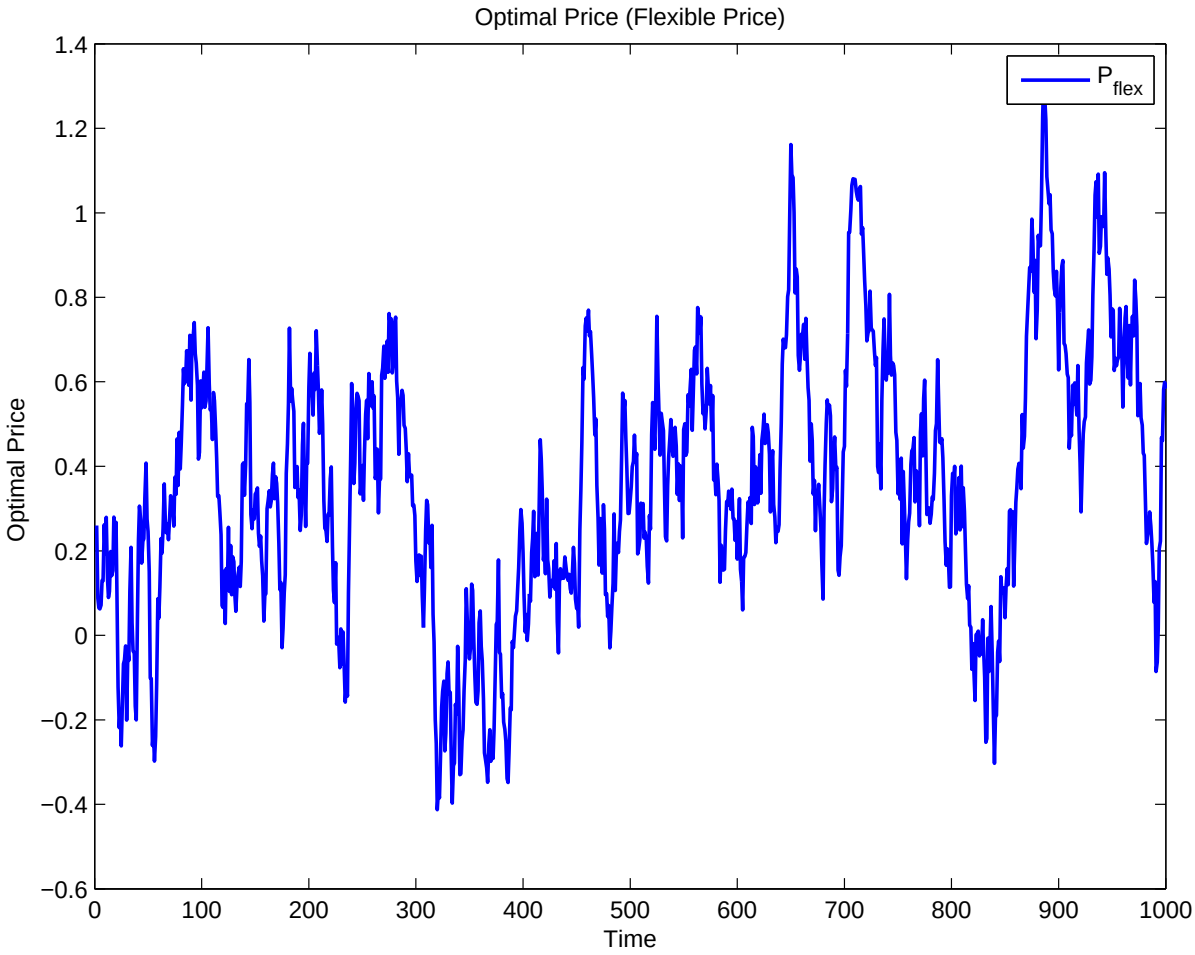


Figure 28: Figure

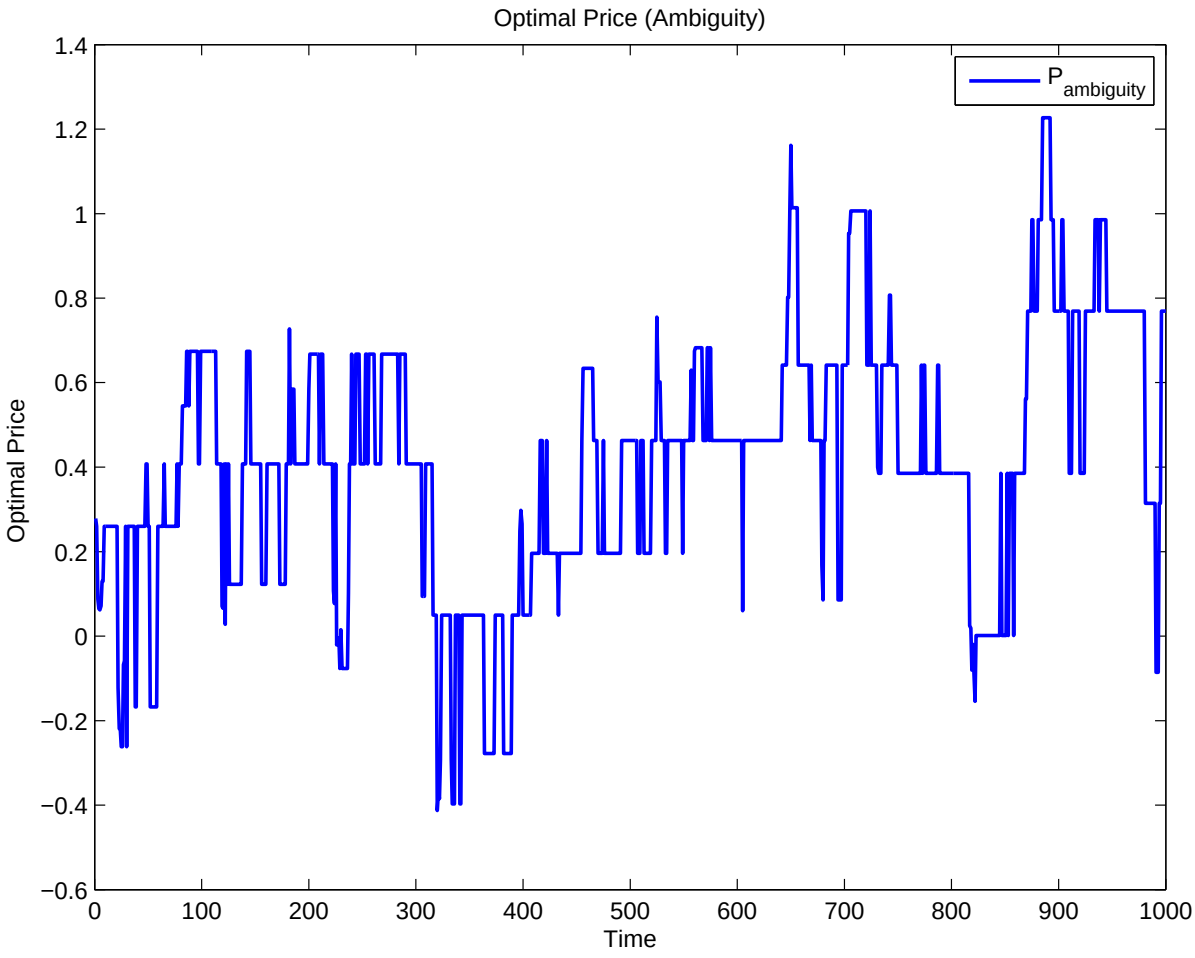


Figure 29: Figure

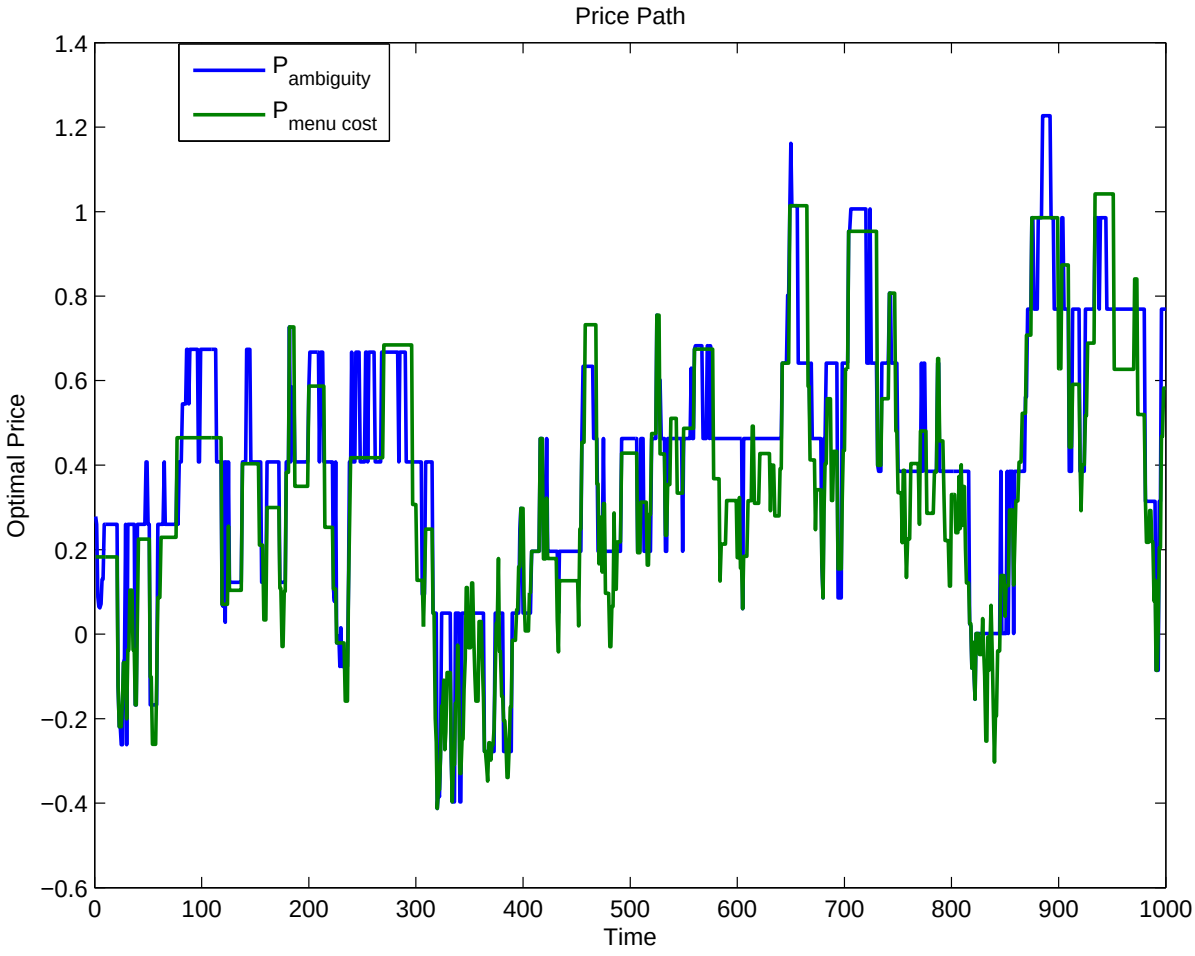


Figure 30: Figure

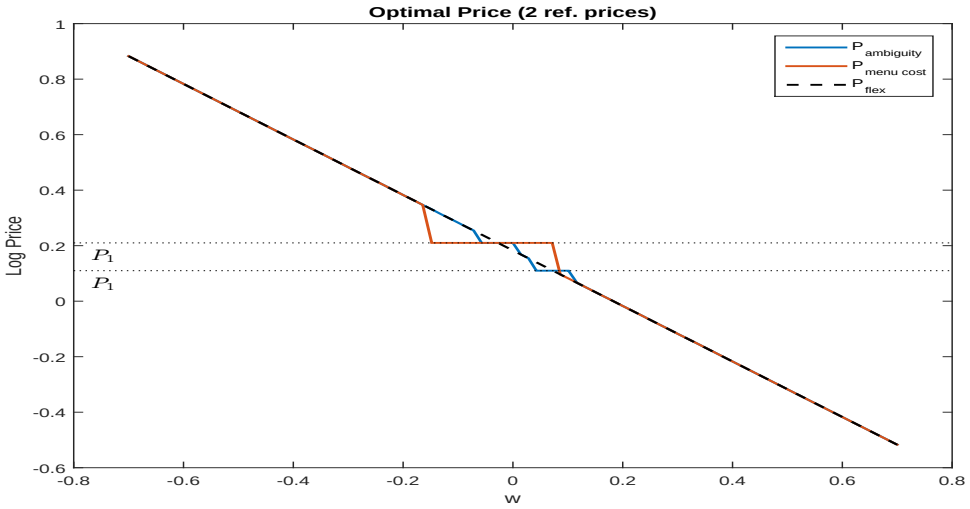


Figure 31: Figure

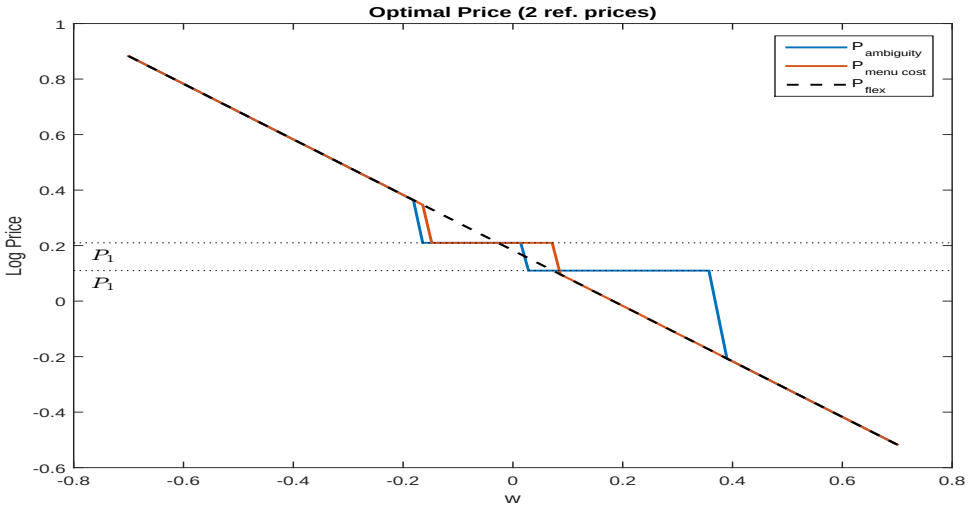


Figure 32: Figure

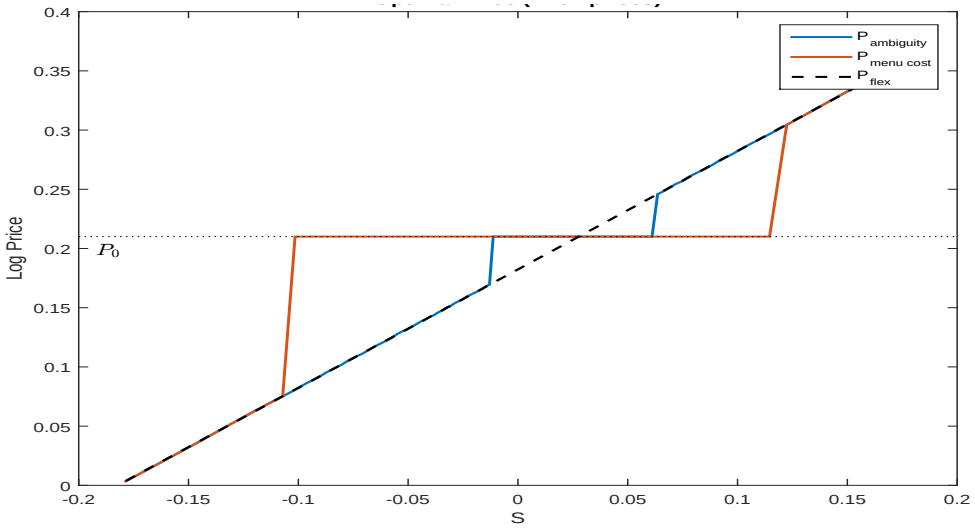


Figure 33: Figure

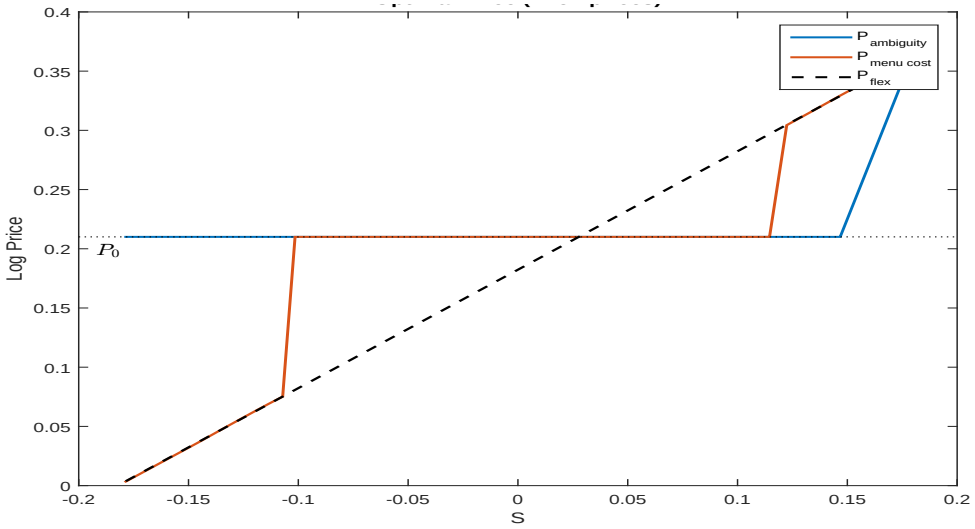


Figure 34: Figure

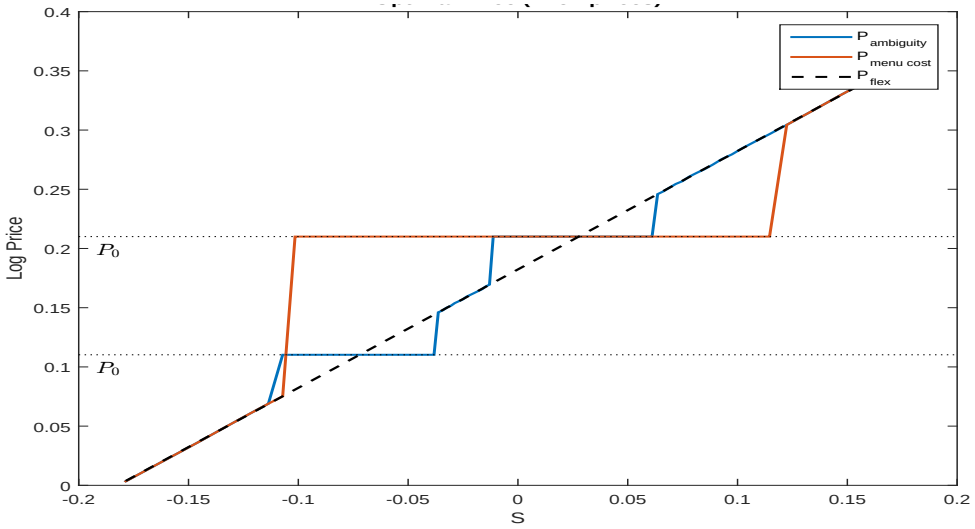


Figure 35: Figure

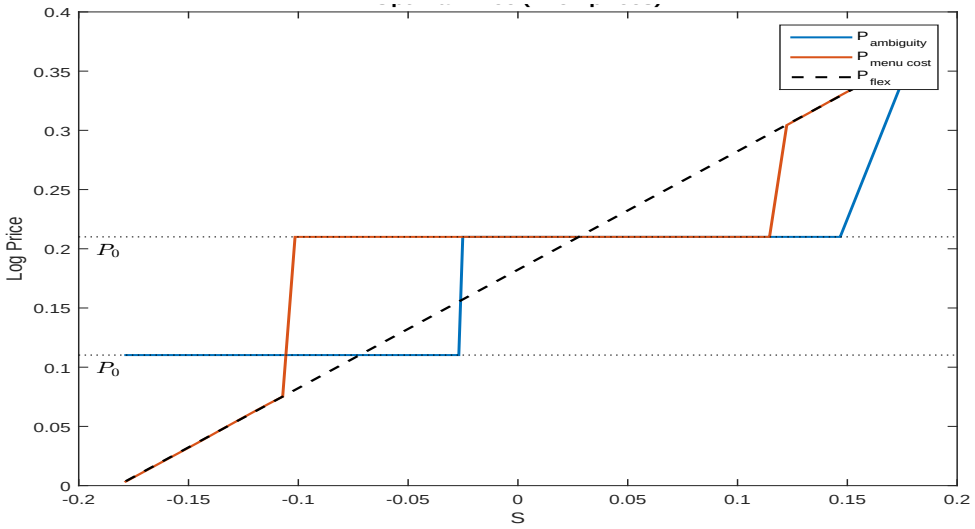


Figure 36: Figure