Paralyzed by Fear: Rigid and Discrete Pricing under Demand Uncertainty

Cosmin Ilut	Rosen Valchev	Nicolas Vincent
Duke University & NBER	Boston College	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 consistent with a number of salient micro facts. In the modelics575(the)-527(monop)-27(olistic)-527(rm)-527(fa ambiguous signal of the prices of its direct competitors in the short runicsour rigidity becomes nominal in nature.

Email addresses: Ilut cosmin.ilut@duke.edu, Vhev: vhev@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.

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 resulting characteristic of the worst-case relationship is to make the aggregate price not informative about the unobserved industry price. The nature's reaction defines a worst-case demand schedule that is a function of this non-ambiguous relative price. 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

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 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

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 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:

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

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

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

$$m(p)\ 2\ [\ _l \quad bp;\ _h \quad bp]: \tag{3}$$

This set is motivated as a limit on the ambiguity the rm faces, and its size will be calibrated based on what the rm could reject at standard 5% levels with a small sample of observations. Intuitively, the interpretation is that while the rm's marketing department provides it with some possible DGPs, it is not con dent enough to restrict itself to probabilistically weighing di erent 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 $_0$ with observed signals.

We further specify the set of $_0$ by studying the limiting case when the covariance function K goes to zero almost surely. In that case, $_0$ consists entirely of Dirac measures, on the space of measurable, downward sloping functions. So that for any given prior $_0\ 2\ \ _0$, there is a unique function $x(p)$ which has probability one, and all other d-374(53 27(y)27(ard)-258(alopi 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

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

where $\pi(\varepsilon_t; c_t)$ 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, μ_0, from the set of admissible priors $\mathcal{P}_0$. 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, $\mathcal{P}_0$, of the demand function $x(p_t)$, and hence over the conditional expectation of demand, further denoted by $\hat{x}(p_t|\varepsilon^{t-1}; \mu)$. 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, μ_0. Thus, for a given history $\varepsilon^{t-1};$ the minimization selects the admissible prior that yields the lowest expectation $\hat{x}(p_t|\varepsilon^{t-1}; \mu):$ 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 $\hat{x}(p_t|\varepsilon^{t-1}; \mu_0^{\min})$ evaluated at the worst-case prior $\mu_0^{\min}:$

The minimization step in (4) is relatively easy to solve. We conjecture that the minimizing $_0$ is such that, for a given price P_t; it implies that the worst-case expected demand realization $b x(p_t | "^{t-1}; _0$

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

$_0$ that have survived elimination, both through the statistical or economic steps:

$$_{t\ 1}("^{t\ 1}) = \left(\hat{x}(p) : \hat{x}(p_j) \in [\ _l \quad \hat{b}p_j;\ _h \quad \hat{b}p_j];\ L_{N_{t\ 1}(P_j)}(\hat{x}(p_j)) \qquad \max_{x(p_j)} L_{N_{t\ 1}(P_j)}(x(p_j)) \text{ and } x(p_j) \quad x(p_k) \text{ for } \forall\, p_j \quad p_k \right)$$

The set of one-step-ahead conditional beliefs, is thus a set $P_{t\ 1}("^{t\ 1})$ of normal distributions

$$P_{t\ 1}\ "^{t\ 1} =_{1\,1}$$

intersection

$$[-\quad bp_0 \quad _z; - \quad bp_0 + \quad _z] \setminus [- \quad bp_0 + (\quad) b_N; - \quad bp_0 + (\quad + \quad) b_N] \qquad (8)$$

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

$$j \quad j \quad \frac{p}{N} + \qquad (9)$$

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 con dence interval becomes larger than the prior upper bound, and, reversely that it's not too low. For example, if $= \quad = 1{:}96$ then $j \quad j < 2 \quad 1{:}96{:}$ The restriction is more likely to be satis ed 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 con dence 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 is increased until it reaches $j \quad j \quad \frac{p}{N};$ so that condition (9) is satis ed.

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

$$x\ (p_0) j (b_N (p_0)) = \max f- \quad bp_0 \quad _z; - \quad bp_0 + (\quad) b_N g \qquad (10)$$

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

$$> (\quad) = \frac{p}{N} \qquad (11)$$

This is more likely to happen if the con dence interval is narrower, which is determined by a larger N and a smaller critical value $;$ if the average demand is larger, through a higher $;$ and if the prior tunnel is wider, controlled by a larger . The worst-case demand instead can remain to be the initial one, $- \quad bp_0 \quad _z;$ if the opposite condition holds.

Having determined the worst-case $x\ (p_0);$ we can nd 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

$$x\ (p^0) j (b_N (p_0)) = - \quad bp^0 \quad _z \text{ for } 8\ p^0 > p_0$$

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

$$- \quad z \quad bp_2 = - \quad bp_0 + (\quad) \, b_N$$

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 $p_2(\;;N)$ 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 $p_2(\;;N);$ the worst-case is restricted now by the worst-case prior.

To summarize, having observed $b\phi_N(p_0);$ the worst-case demand is:

$$x(p^0) \quad \min x(p^0) j (b\phi_N(p_0)) = \begin{cases} \max f - \quad bp^0 \quad z; x(p_0) j (b\phi_N(p_0)) g & \text{for } p^0 \quad p_0 \\ - \quad bp^0 \quad z & \text{for } p^0 > p_0 \end{cases} \tag{12}$$

where $x(p_0) j (b\phi_N(p_0))$ 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 $x(p^0)$ for the above situation, where $P_0 = 1;$ $b\phi_N(p_0)$ is given by the demand under the true DGP, i.e. $\quad = 0;$ 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 $P_2(\;;N)$:

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 $(p_2(\;;N); p_0);$ 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 rm 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 rm acts as if there is a kink at the observed price. At this price the rm 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 rm 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

$$\max$$

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 $_0$ 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 rm 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 pro t maximization. In particular ,at the beginning of each period *t;*

is the rational expectations price, as in (14): $P_t^{RE} = \quad C_t$, where $\quad b/(b \quad 1)$ is the markup. The reason is that, even if the rm 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 $p\ 2\ [p_2(\ ; N); p_0)$ as the demand is the same at that interval but the price is highest at p_0: So, we only need to compute the pro t at P_0 and compare it to that arising from setting the RE one. The former is:

$$(P_t = P_0) = e^{0:5 \ {}^2_z} (P_0 \quad C_t)\, e^{\underline{q}_N (p_0)}$$

For ease of exposition, de ne a hypothetical value of cost $C_0 \quad P_0 = \ ;$ for which the price P_0 would be the optimal RE price. The pro t can be rewritten as

$$(P_t = P_0) = e^{0:5 \ {}^2_z + - \quad z} \quad \frac{C_0}{C_t} \quad 1 \quad C_t (\ C_t) \ ^b \quad \frac{C_0}{C_t} \ ^b \ e^{\ (p_0)} \tag{15}$$

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

$$(P_t = P_t^{RE}) = e^{0:5 \ {}^2_z + - \quad z} (\quad 1)$$

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 $''^{1}$. 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 pro ts:

$$\max_{P_2} \min_{2P_1("^1)} E \left[(P_2 \quad C_2)\, e^{x(p_2 j "^1)+z_2} + \quad (P_3) \right] \tag{18}$$

where P_3 denotes the optimal price set in period 3, conditional on the $"^1$ and the new demand signal $q(p_2)$ realized at the end of period 2 at the price P_2; E denotes the worst-case conditional expectation, which uses the estimate of demand $x\ (p_2 j "^1)$:

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

$$(P_3) = \max_{P_3} \min_{2P_2("^1;q(p_2))} E \ (P_3 \quad C_3)\, e^{x(p_3 j "^1;q(p_2))+z_3} \tag{19}$$

The rst period consists of letting the rm 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 rm 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

$$c_t \quad \overline{c} = \ _c (c_{t \ 1} \quad \overline{c}) + \ _c \ ^c_t$$

where ^c_t is white noise. The benchmark parametrization is in Table 1. We set $b = 6$, the constant $^-$ = 0 and the critical value $= 1{:}96;$ which corresponds to a 95% con dence interval. We set the cost shock parameters $_c$ and $_c$ to values calculated by Eichenbaum et al. (2011), where they observe marginal costs. We normalize $\overline{c} = (b \quad 1){=}b$ so that $P^{RE} = 1:$ We set the discount factor $= 0{:}99[1 \quad (1 \quad e^{\ 1=30})];$ 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 $= 2;$ 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 $_z$:

Table 1: Calibrated parameters

b	$^-$		$_c$	$_c$		$_z$

5.3 Results

Worst-case expected demand

We rst plot the worst-case expected average demand for the case where the rm 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% con dence 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 gure 3, we increase the number of times for which the fqr2(o)-22e-293(a)01thebsermofdel:eres]927 Td [(forms)-326knoer4roorm255num55ned thaeum55n3(namely

1 i [(=) -464 (0)]d (927 11c227 (99een27 27 (ed)

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 $\tilde{\beta} = \beta/(1-\beta).$ 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 $\tilde{\beta};$ 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.

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:

$$C_t = \int C^{b \; 1}$$

This final good demand defines the aggregate price index P_t

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

where 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 P_{jt} such that

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

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

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

The demand of industry j for a given intermediate good i is a function of the relevant relative price, $\frac{P_{it}}{P_{jt}}$, overall industry output C_{jt}, and demand shocks z_{it}. We denote this function by H_j and note that it is a transformation of the functions f_j and, *distributed as N*

P4jA equiarInti walermntive ntermroe-1(t)tiv1(o)n-246(its-3745(hat)-346(ith1(oe-146(ieonsom28(ev-246(

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

$$h(r) 2 [\quad br; \quad br];$$

and hence lies in the type of a prior tunnel studied previously. The function is ambiguous, and we will again focus on the limiting case of Delta priors where each prior awards one possible function probability 1, and all others probability 0. The admissible functions are all weakly decreasing functions that fall in the tunnel above.

There are two sources of uncertainty in demand { uncertainty about the shape of demand, $h(:)$, and uncertainty about the relevant price index p_{jt}. Uncertainty about demand is going to be handled in a manner very similar to the previous discussion on real rigidity, hence next we turn to the uncertainty about p_{jt}.

6.3 Uncertainty about the relationship with aggregate prices

The rm has twotun277ditfunct3((function)]TJ/F21 11.9230.6 Tf 6.739 0 Td [(p)]TJ/F48 7.9701 Tf 5.875

p_{jh}

where p_{js}

confident in extrapolating this long-run relation to short-run fluctuations, and entertains functions $\ (.)$ 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 focus on the limiting case where the variance function of the GP distributions goes to zero, so conditional on a prior, one function $\ (.)$ has probability 1 and all others probability zero.

6.3.1 Worst-case beliefs

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

$$h(\hat{r}_{it} \quad (p_t \quad p_{js}) \quad {}_{jt}) \quad b(\ (p_t \quad p_{js}) + {}_{jt});$$

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

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

$$h \ (\hat{r}_{it}; \ {}_{jt}) = \min h(\hat{r}_{it} \quad (p_t \quad p_{js}) \quad {}_{jt}) \quad b(\ (p_t \quad p_{js}) + {}_{jt}) \tag{29}$$

This is the demand level that would prevail if nature draws the worst possible $\ (.)$, conditional on a particular choice of $h(.)$ and price $\hat{r}_{it}$. Note that in the short run $\ (p_t \quad p_{js})$ 2 $[\ {}_p; \ {}_p]$, and hence variation in p_t does not change the set of possible numerical values that could be realized through $\ (p_t \quad p_{js})$. Hence the minimization can equivalently be recast in terms of minimizing over a parameter, $\ $ 2 $[\ {}_t; \ {}_p]$, which represents the conditional expectation of p_{jt}. Since movements in p_t do not affect the minimization problem, the solution is given by

$$(p_t \quad p_{js}) =$$

Intuitively, the worst-case cointegrating relationship implies that movements in the aggregate price are not informative about the industry prices in the short-run. This is because when there is no such informative relationship, nature is free to choose the worst possible expectation of p_{jt}, given a demand function $h(.)$ and price choice $\hat{p}_{it}$.

Since the transitory shocks $_{jt}$ are not observed, we can also take an expectation over them and define the expected h :

$$x(\hat{p}_{it}) = E_t(h \ (\hat{p}_{it}; \ _{jt}))$$

This 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. For tractability, we assume that the implied expectational errors follow a normal distribution,

$$h \ (\hat{p}_{it}; \ _{jt}) = x(\hat{p}_{it}) + \varepsilon_{it}; \ \varepsilon_{it} \quad N(0; \ ^2_\varepsilon); \tag{30}$$

6.3.2 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.[18]

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

18

at the next review is given by

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

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

6.3.3 Nominal rigidity from real rigidity

The firm uses past signals to learn about the worst-case demand. Putting together (26) and (30), the demand facing the firm is

$$y_{it} = x(\hat{r}_{it}) + c_t + b(p_t - p_{js}) + \varepsilon_{it} + z_{it} \tag{31}$$

which is a known function of the observed aggregates, namely price p_t and quantity c_t, an unknown function $x(.)$ of its perceived relevant relative price and Gaussian noise. This forms

$\beta_{j,k,t}$, based on the information of the whole sample of size N.[20] 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

$$\beta_{j,k,t} \sim N(\hat{\beta}_{j,k,t|N}, \Sigma_{j,k,t|N})$$

For a pair (j,k), we analyze the sample path of the estimate $\hat{\beta}_{j,k,t|N}$ and its uncertainty $\Sigma_{j,k,t|N}$. We analyze how often we cannot reject the null that $\hat{\beta}_{j,k,t|N}$ 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 $\beta_{j,k,t}$ 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 $\hat{\beta}_{j,k,t|N}$ 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 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 $\hat{\beta}_{j,k,t|N} = 0$ is indeed high at most horizons and decreases with the horizon.[21]

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.

[20] 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.

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

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.

7.1.1 Agents and shocks

Representative agent

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

$$\max \sum \beta^{t} \left(\log C_t - \int L_{i;t} di \right)$$

subject to the budget constraint

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

where q_{t+1} is the stochastic discount factor, b_{t+1} is state contingent claims on aggregate shock, $\Pi_{i;t}$ 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

$$Y_{i;t} = H_j \left(\frac{P_{it}}{P_{jt}}\right) \left(\frac{P_{jt}}{P_t}\right)^{b} C_t \exp(z_{it})$$

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 $N(0; \sigma^2)$ and ε^{a}_{t} is iid $N(0; \sigma^2_a)$. The real flow profits are therefore:

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

Monopolistic firms are owned by the representative agent, and thus they discount pTf 10.623eeg39552 where ^s_t is iid $N(0; \, ^2_s):$

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

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

so that the real ow pro ts can be written as

$$_{i;t} = \quad \frac{P_{it}}{P_t} \quad \frac{S_t}{!_{it}A_tP_t} \quad Y_{i;t} \tag{34}$$

7.1.2 Demand uncertainty

True data generating process

We rst 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 :

$$C_{j;i;t} = C_{jt}"_{it} \quad \frac{P_{i;t}}{\overline{P}_{j;t}} \quad ^{b}$$

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

$$y_{it} = \quad bp_{i;t} + c_t + bp_t + z_{it} \tag{35}$$

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 rm retains all the uncertainty about the direct competitors, re ected in the unknown, relevant price $p_{j;t}$:

Monopolistic rm's information

As in section 6, we assume that the rm observes the aggregate P_t and C_t, but not its demand function. The learning process is exactly the same as described previously in section 6, where equation (31) gives the demand to be estimated as

$y_{it} = x($^.9552 Tf 3.556 1.794 Td[(9552 Tf 3.556 1.9552 Tf 3.556 1.9552 Tf -2

$x(\hat{r}_{it})$ is:

$$x(\hat{r}_{it}) \; 2 \; [$$

the static pro t, which in this case is

$$ {}^{MC}_{i;t} = \quad \frac{P_{it}}{P_t} \quad \frac{S_t}{!_{it} A_t P_t} \quad Y_{i;t} \quad f W_t I_{P_{it} \neq P_{it} \ 1} \tag{39}$$

where the latter term re ect 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.

Reset shocks

Because we have modeled so far that the price-sensitive component of demand $x(\hat{r}_{it})$ is constant through time, the rm can in principle learn it perfectly as it accumulates new information. However, it is plausible that the rm is concerned that the underlying demand shifts and thus it has to start learning it again. We model the decay in the informational content of observation by introducing shocks to this learning capital, which we call 'reset shocks'. The interpretation of this shock is that there are events that change the competitive landscape of the rm, such as for example the entrance/exit of competitors, the in ow/out ow of customers. The rm nds these situations as resetting the information it has accumulated.

A reset event happens with a constant probability and for all prices it increases the con dence interval for the expected demand. The reset shock brings the posterior estimates closer to the prior, i.e. it makes the past learning less useful. In particular, for each relative price r_n that has been observed, the reset shock expands its con dence interval. For example, if before the shock

$$x(r_0) \ 2 \ [b_N(r_0) \quad \frac{z}{\sqrt{N(r_0)}}; b_N(r_0) + \quad \frac{z}{\sqrt{N(r_0)}}]$$

with the reset shock being realized the new true demand is shifted around each element of $x(r_0)$ by

$$e(r_0) = x(r_0) \quad \frac{z}{\sqrt{N(r_0)}}$$

where is a parameter. At the moment of the shock, this is equivalent to nding a fraction of the $N(r_0)$ which is used in computing the 95% con dence interval:

$$(\quad + \quad) \frac{z}{\sqrt{N(r_0)}} = \quad \frac{z}{\sqrt{N(r_0)}}; \quad = \quad \frac{}{+}^{2} < 1$$

So, conditional on a reset shock, we can reparametrize by modeling the state variables $N(r_n)$ of the rm as becoming $N(r_n)$; where

For example, can be equal to 0 such that the rm discounts all the past information. Everything else in the analysis proceeds as before.

We assume that the rm does not observe the reset shock. This means that past information $N(r_n)$ decays deterministically at rate $'$ $(1$ $+$ $)$.[24] The rst component is the probability that a shock has not hit and the second is the amount of loss in information conditional on a shock. In this case, the state variable entering period t that captures the 'information relevant' number of times, $\tilde{N}_{t-1}(r)$; for which a rm has observed a price r is computed recursively as

$$\tilde{N}_t(r) = ' \quad \tilde{N}_{t-1}(r) + I_{r_t}$$

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 $= 1.96;$ 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 $(Q - X\hat{b})/\overline{Q};$ where $\hat{b}$ 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 $0.50 \times 0.675 = 0.3375;$ matching our sample average.[25]

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 specific to the learning model proposed here. The first one is the width of prior tunnel, controlled by ; which is the multiple of standard deviations that the firm uses to form the initial set of possible demand. The second one is the rate of information decay, $';$ necessary for the model to not collapse to full information about the true DGP.

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). Finally, for the menu-cost model, where the information parameters and $'$ 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.

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

Table 2: Parameters

the modal price, have no price memory and are less discrete within a speci c 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.[26]

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 rm switches between the prices at those kinks. Small price changes are generated because the policy function resembles the exible price policy in some situations. On the one hand, this can happen because the history of shocks macy 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 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 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 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 pro t function. In addition, we show that if publicly available indicators

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.

Gagnon, E., D. L

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.

Matejka, 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 rms, and aggregate uctuations," *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): \In ation and costs of price adjustment," *The Review of Economic Studies*

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: $(Q \quad X\, \mathrm{b})=\overline{Q}$.

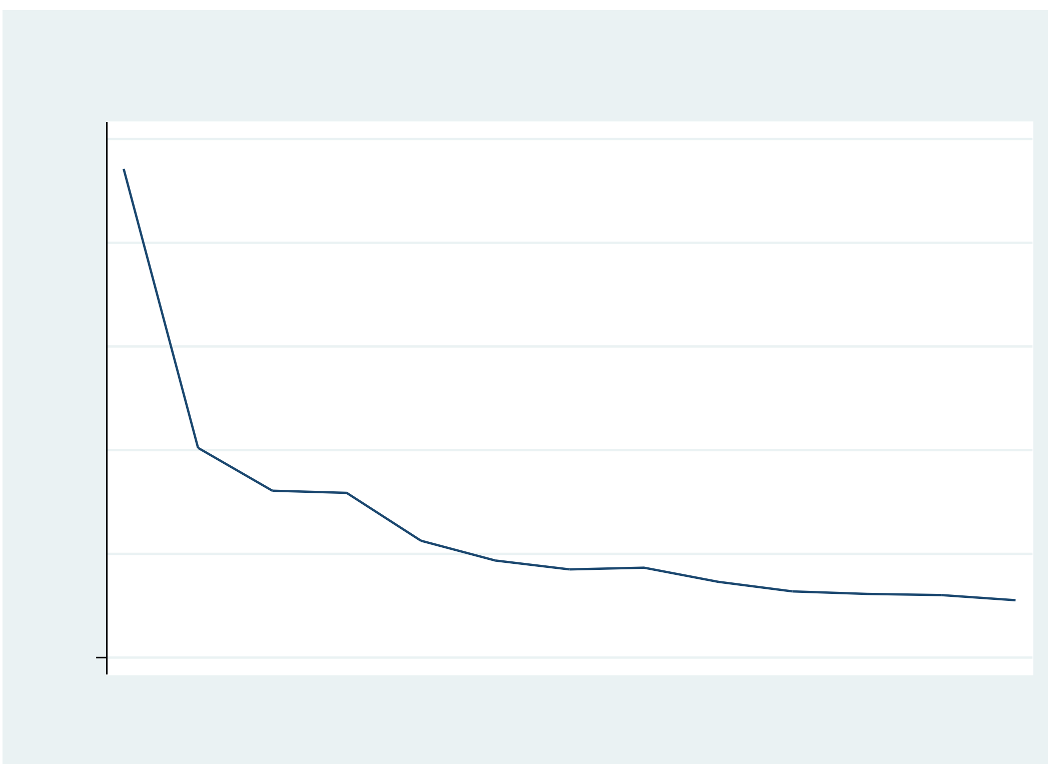


Figure 1: Figure

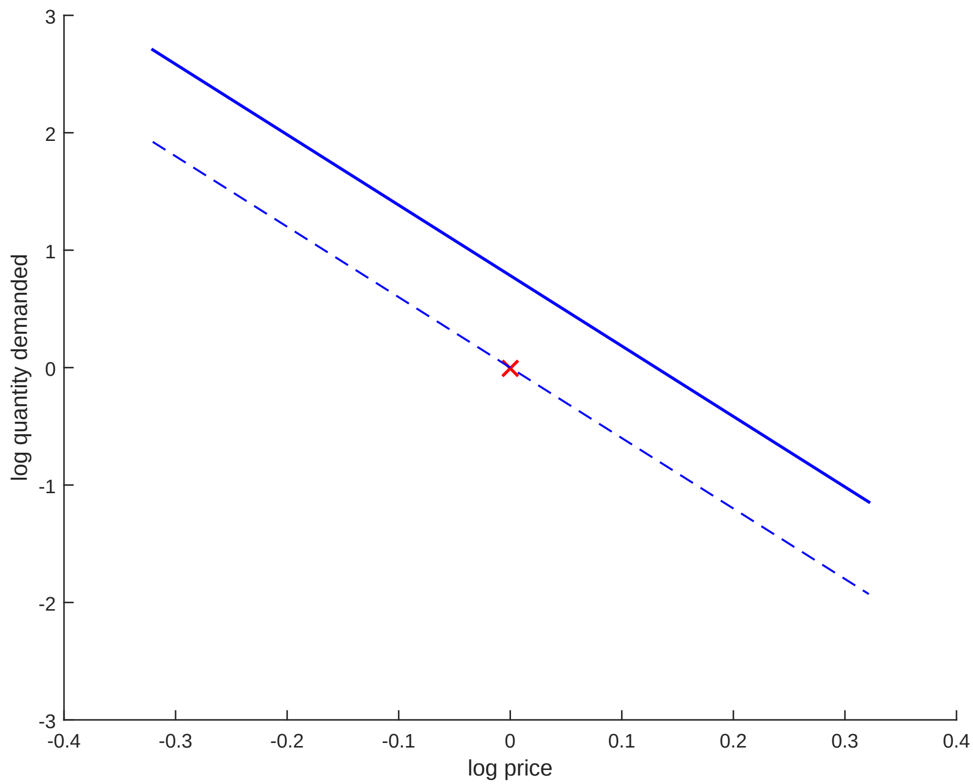


Figure 2: Figure

Figure 3: Figure

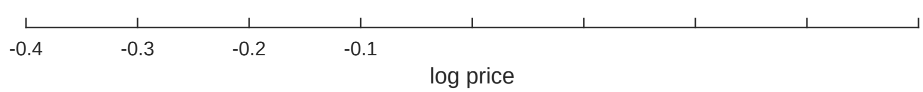


Figure 4: Figure

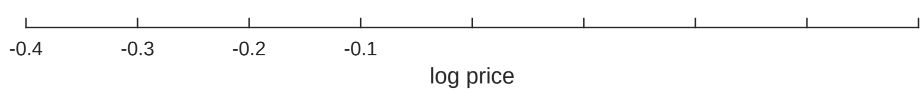


Figure 5: Figure

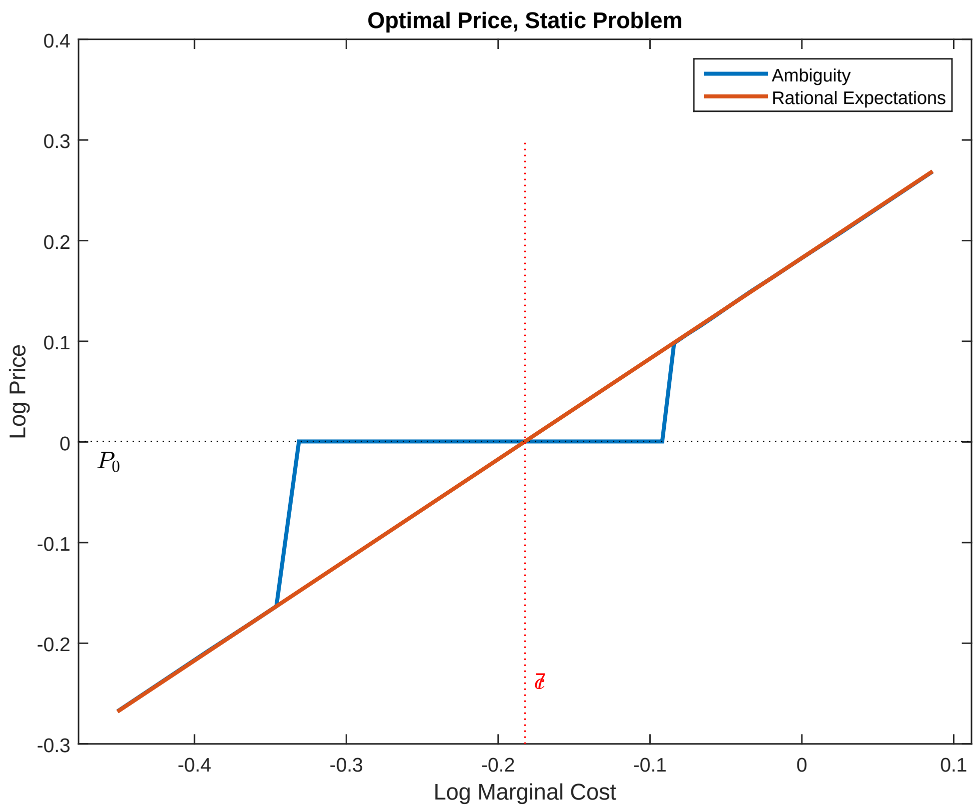


Figure 6: Figure

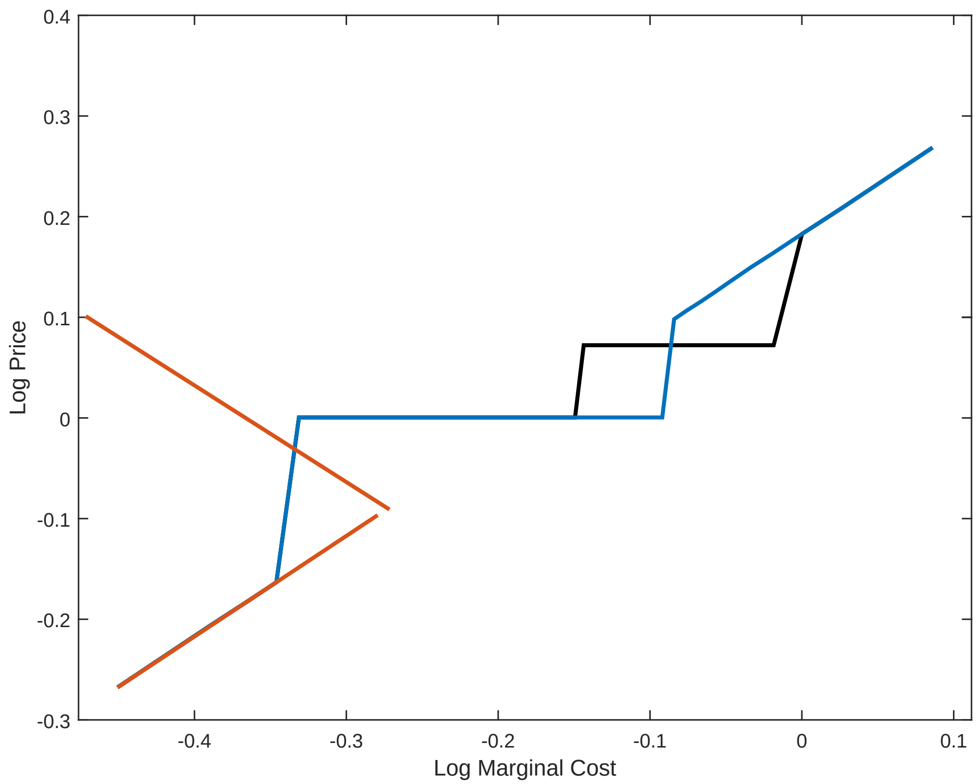


Figure 7: Figure

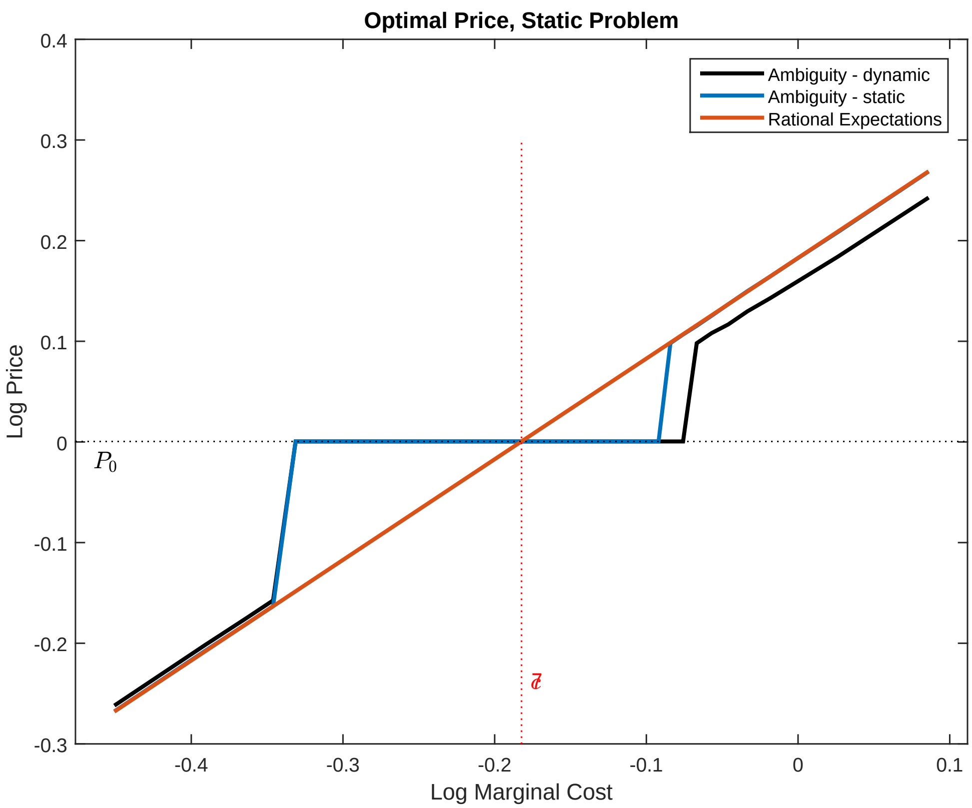


Figure 8: Figure

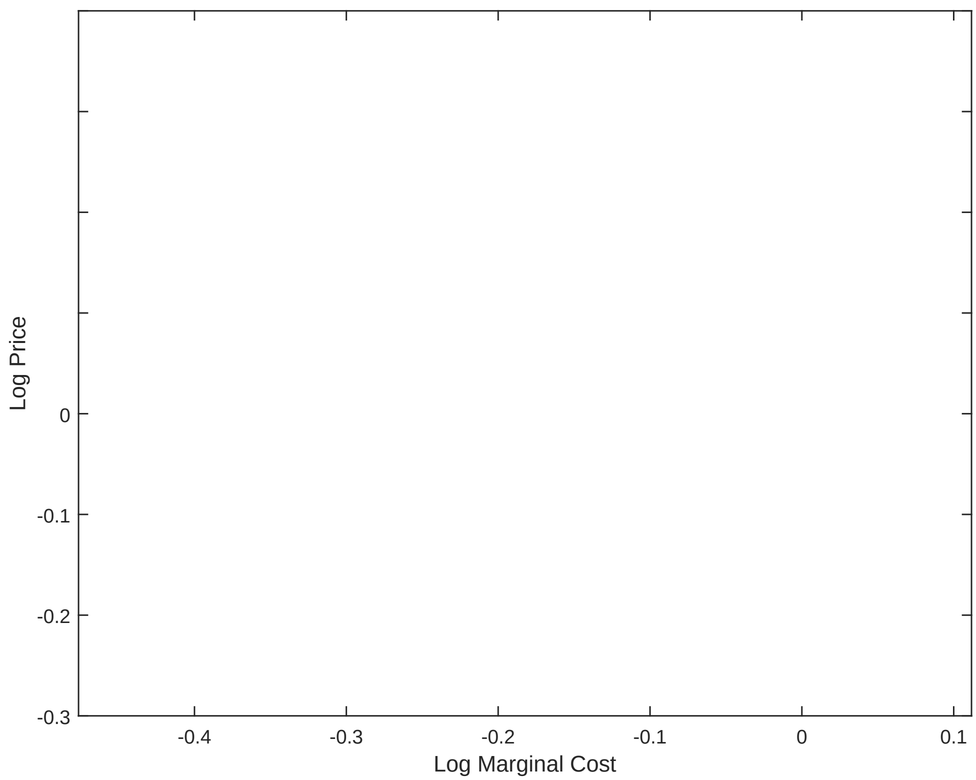


Figure 9: Figure

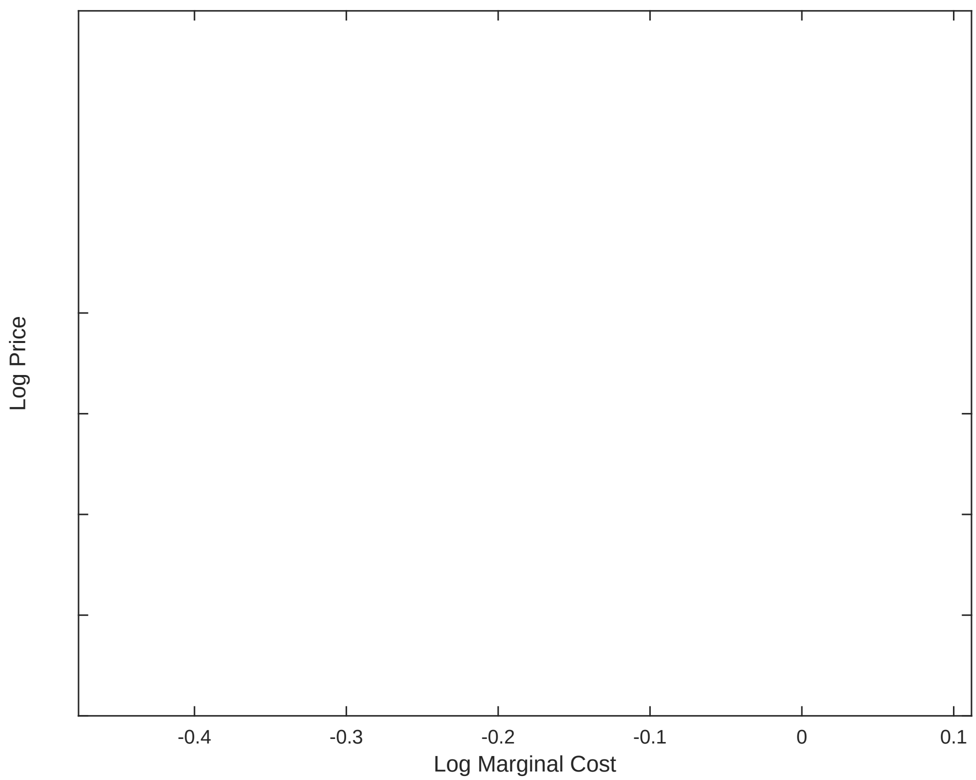


Figure 10: Figure

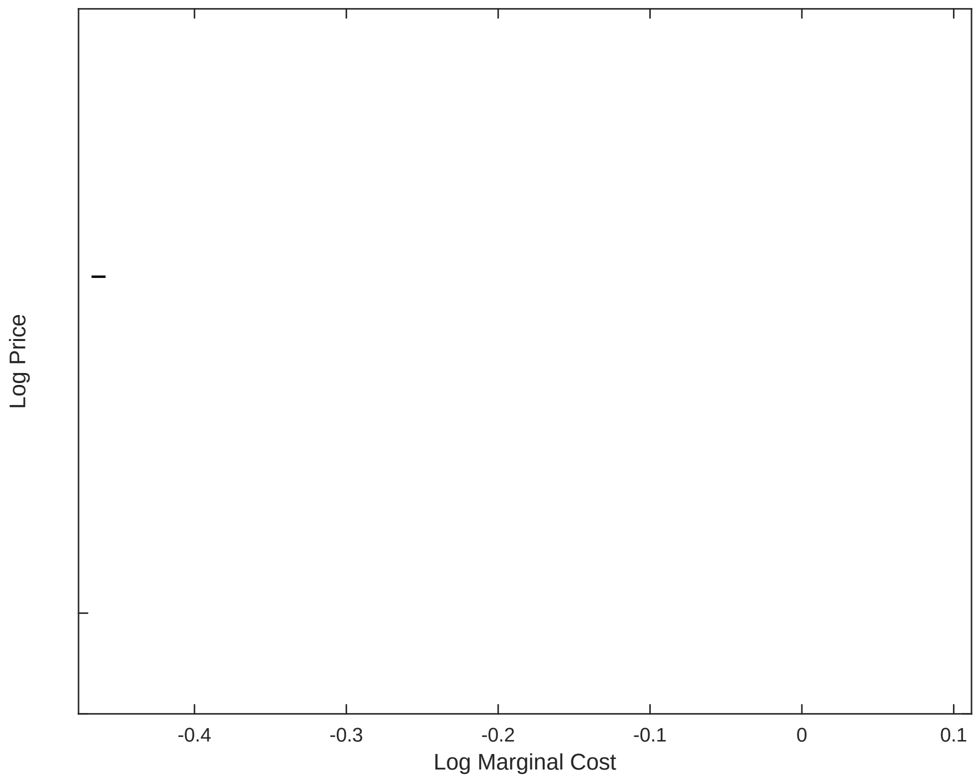
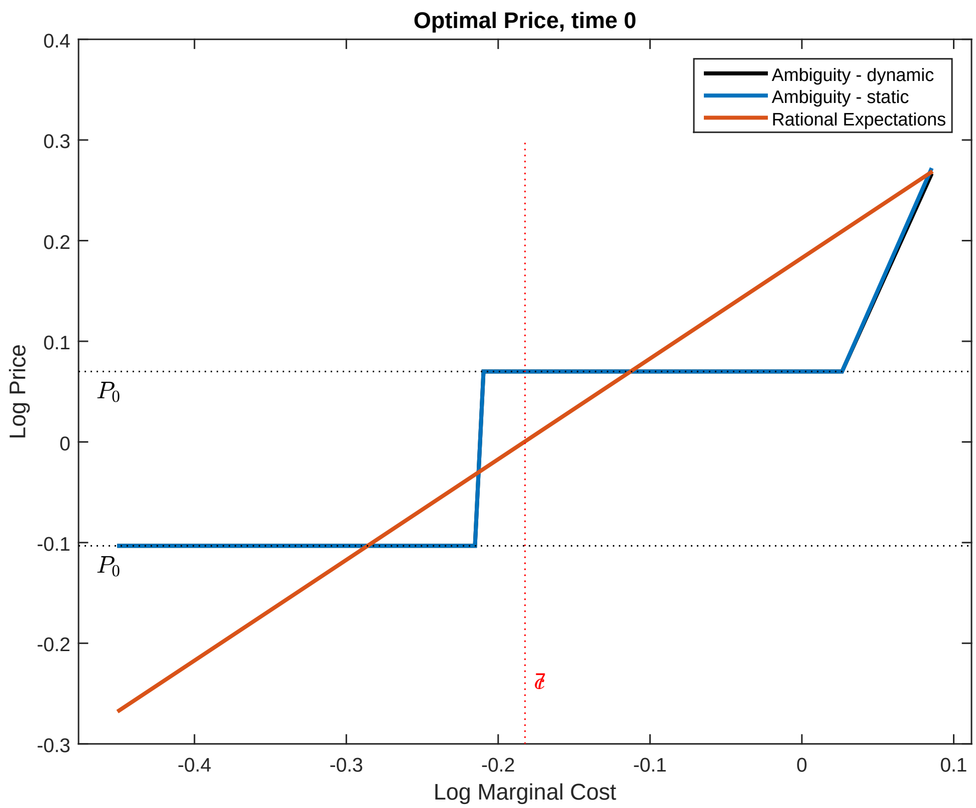


Figure 11: Figure


Optimal Price, time 0
Ambiguity - dynamic
Ambiguity - static
Rational Expectations
Log Price
Log Marginal Cost
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.3
-0.2
-0.1
0
0.1
P_0
P_0

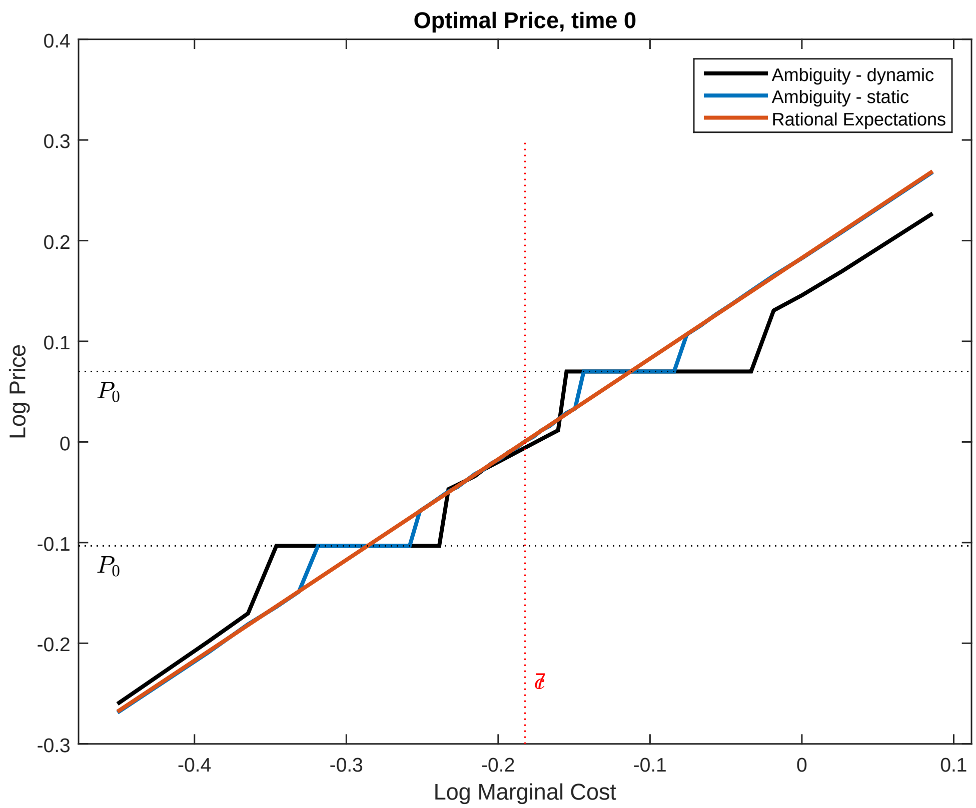


Figure 13: Figure

Figure 14: 3-year rolling regressions of 3-month industry in ation on 3-month aggregate in ation for four categories. The solid line plots the point estimate of regression coe cient on aggregate in ation. The dotted lines plot the 95% con dence intervals.

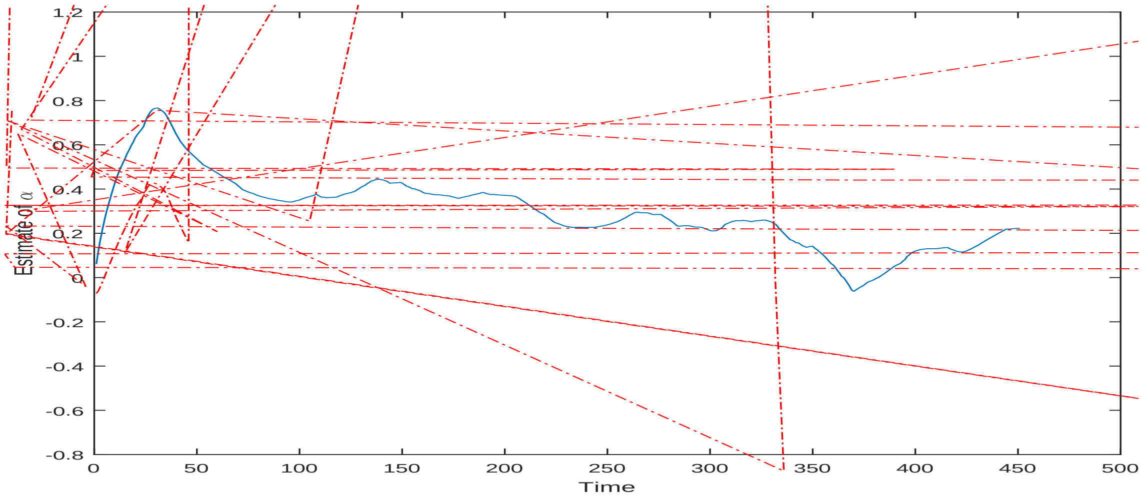


Figure 15: E ect of current aggregate in ation on Carbonated drinks industry in ation, both measured over a 1 month horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% con dence interval using the smoothed estimated uncertainty.

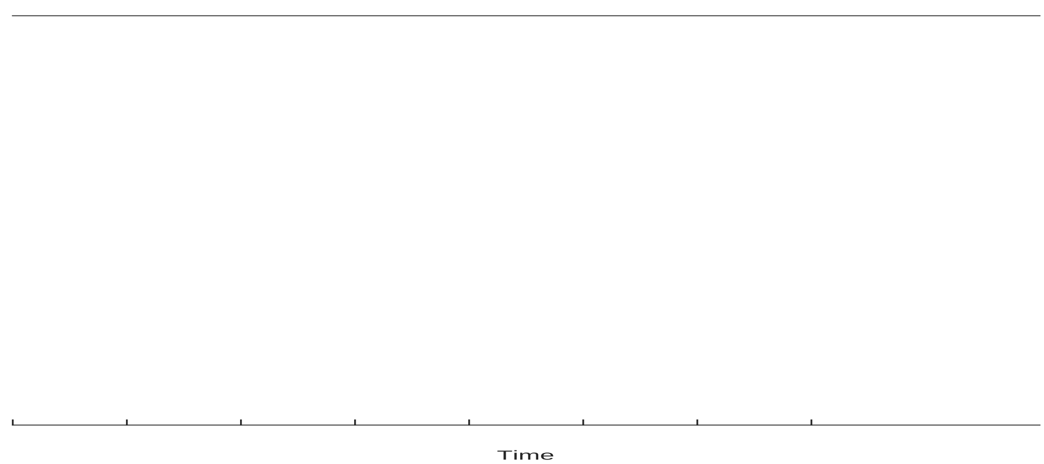


Figure 16: E ect of current aggregate in ation on Carbonated drinks industry in ation, both measured over a 3 months horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% con dence interval using the smoothed estimated uncertainty.

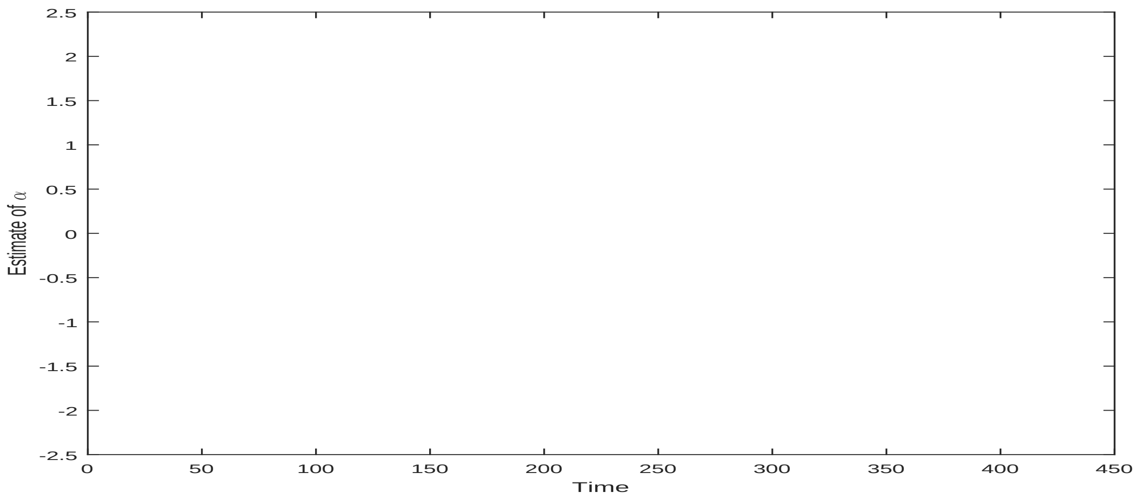


Figure 17: E ect of current aggregate in ation on Carbonated drinks industry in ation, both measured over a 6 months horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% con dence interval using the smoothed estimated uncertainty.

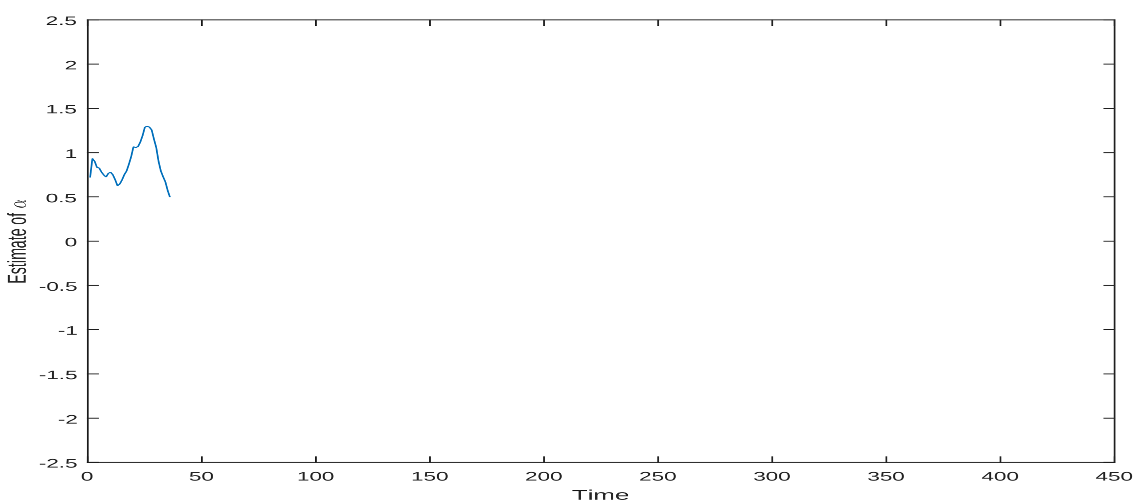


Figure 18: E ect of current aggregate in ation on Carbonated drinks industry in ation, both measured over a 1 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% con dence interval using the smoothed estimated uncertainty.

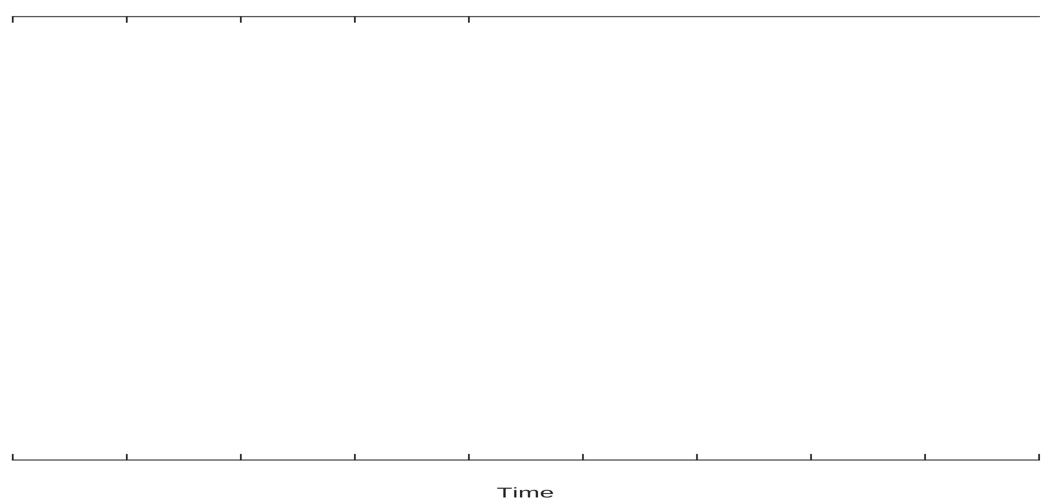


Figure 19: E ect of current aggregate in ation on Carbonated drinks industry in ation, both measured over a 2 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% con dence interval using the smoothed estimated uncertainty.

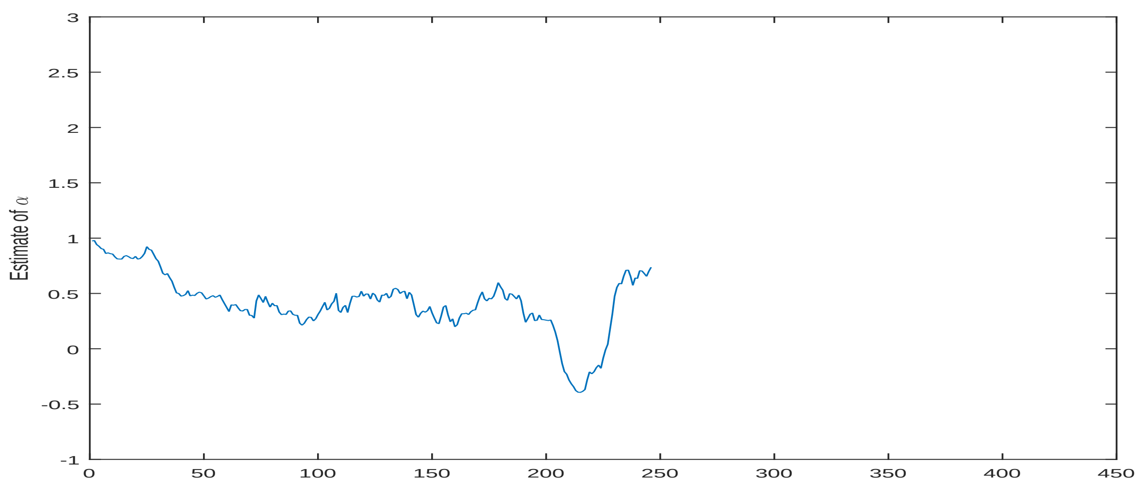


Figure 20: E ect of current aggregate in ation on Carbonated drinks industry in ation, both measured over a 3 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% con dence interval using the smoothed estimated uncertainty.

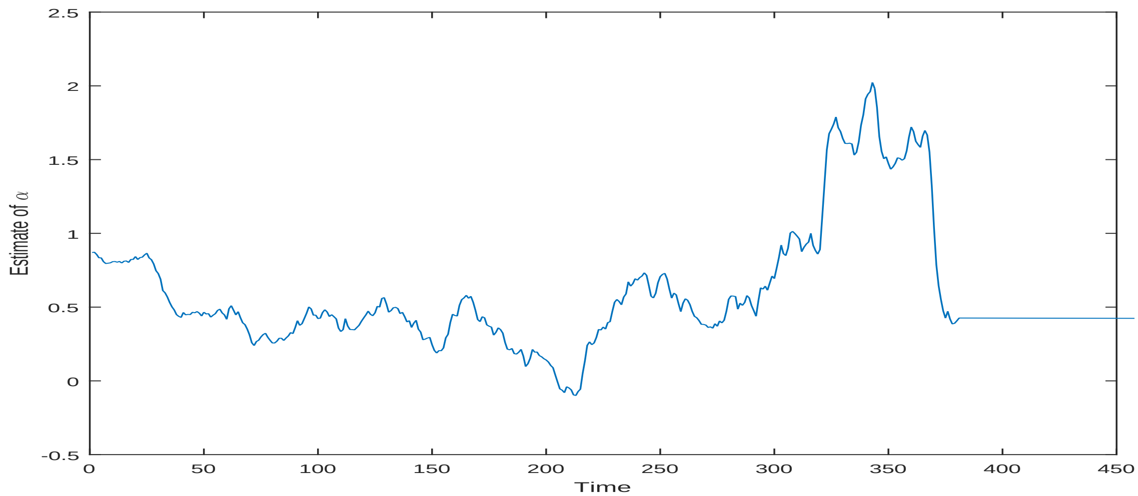


Figure 21: E ect of current aggregate in ation on Carbonated drinks industry in ation, both measured over a 4 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% con dence interval using the smoothed estimated uncertainty.

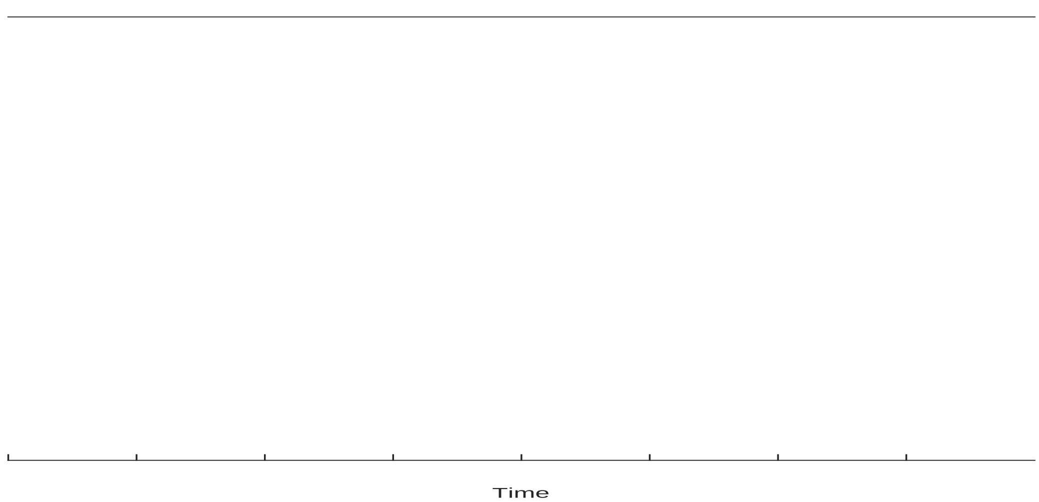


Figure 22: E ect of current aggregate in ation on Carbonated drinks industry in ation, both measured over a 5 year horizon. The solid line plots the Kalman smoothed estimate. The dotted line plots the 95% con dence interval using the smoothed estimated uncertainty.

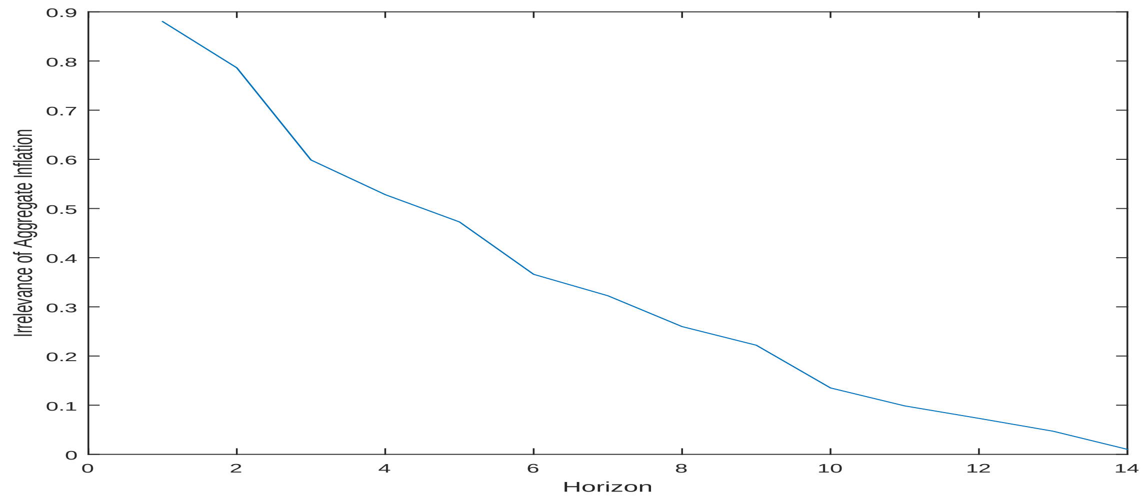


Figure 23: Fraction of times that the current aggregate in ation has an estimated e ect on Carbonated drinks industry in ation that is not statistically di erent from zero, at 95% con dence interval. In ations are computed over various horizons, ranging from 1 month to 5 years.

Figure 24: Average, over industries, of the fraction of times that the current aggregate in ation has an estimated e ect on industry in ation that is not statistically di erent from zero, at 95% con dence interval. In ations are computed over various horizons, ranging from 1 month to 5 years.

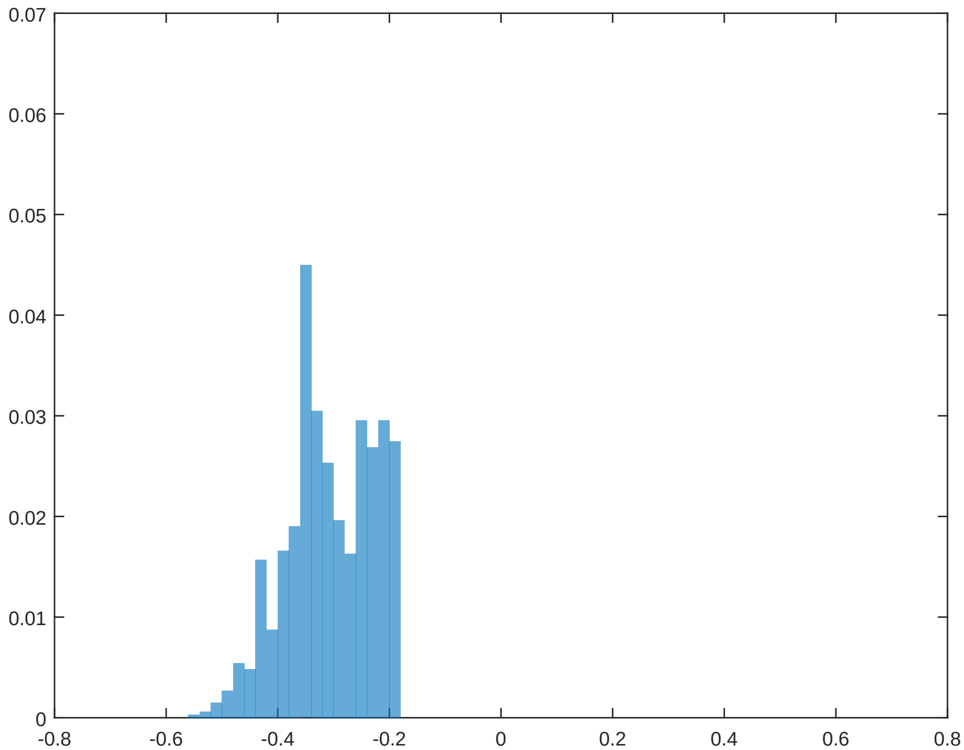


Figure 25: Figure

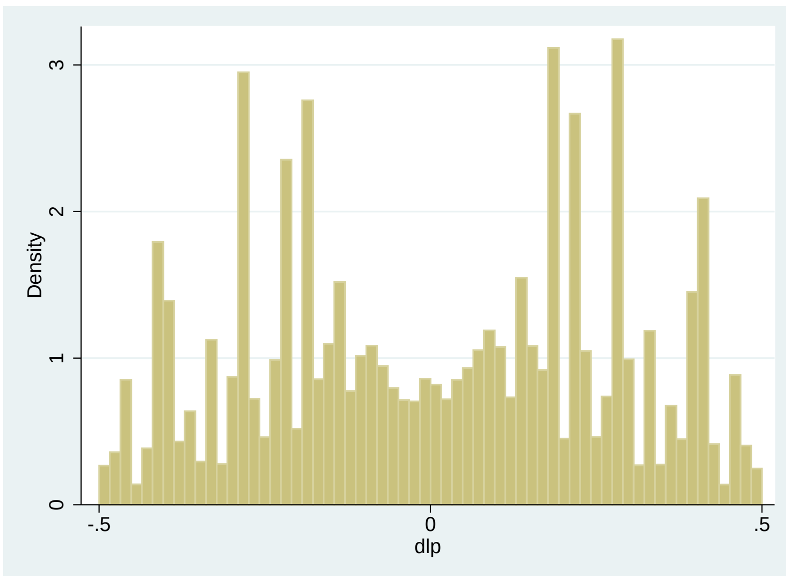


Figure 26: Figure

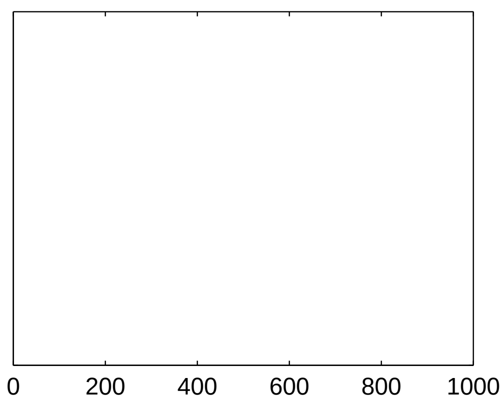


Figure 27: Figure

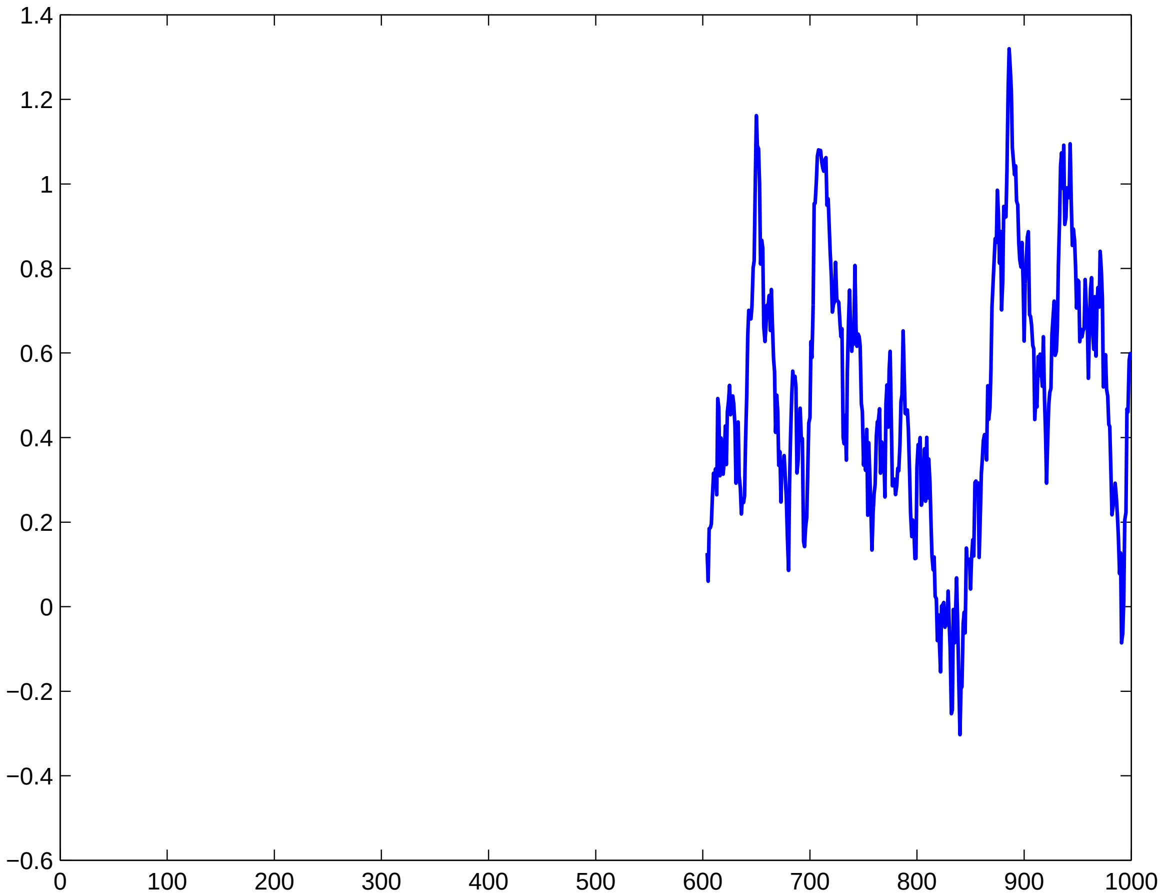


Figure 28: Figure

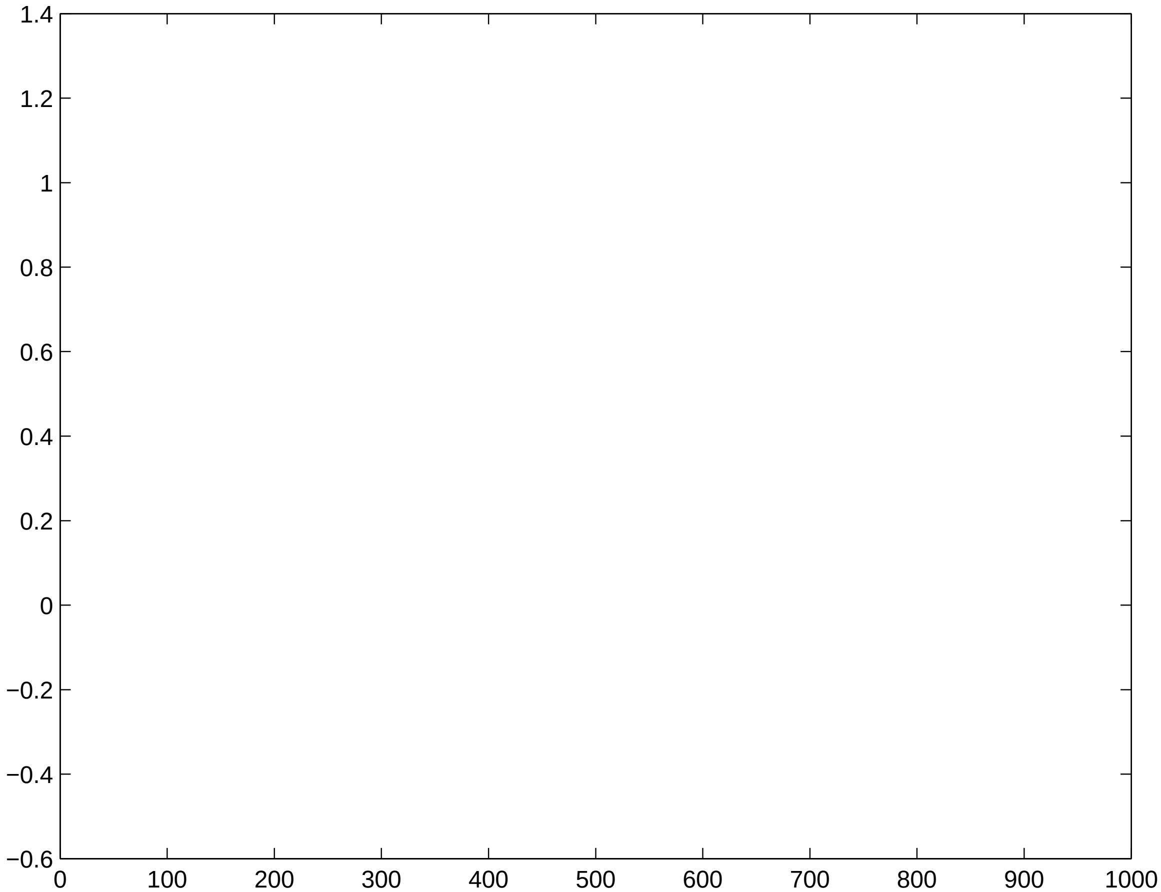
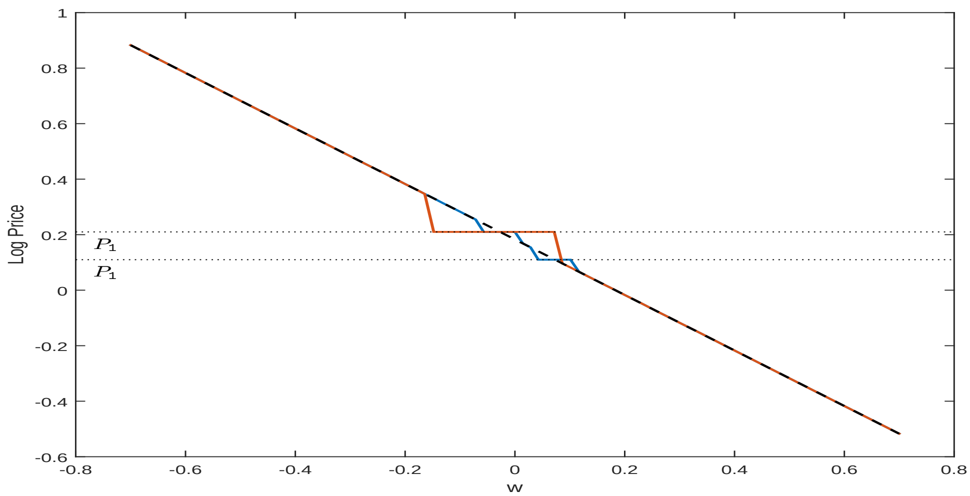


Figure 29: Figure

Figure 30: Figure



1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Log Price
P_1
P_1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
w

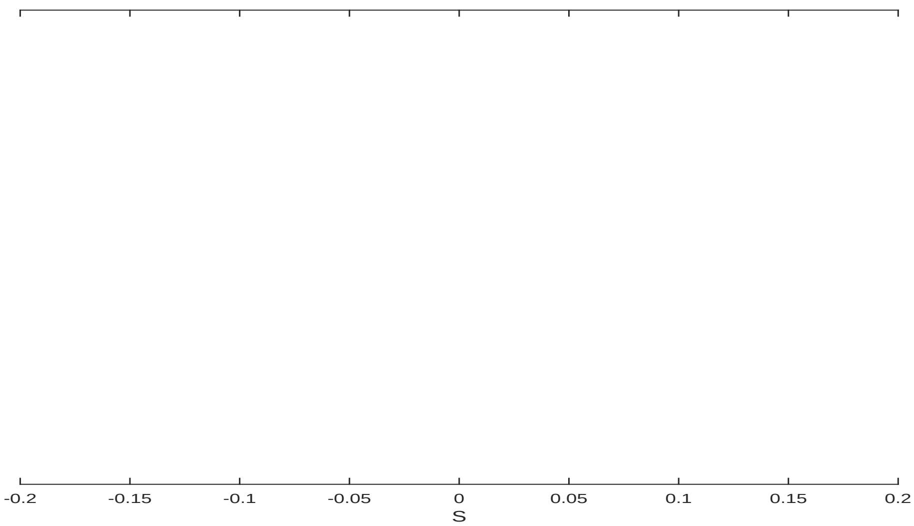


Figure 33: Figure

Figure 34: Figure

Figure 35: Figure

s

Figure 36: Figure