

# Product Life Cycle, Learning, and Nominal Shocks\*

David Argente<sup>†</sup>

University of Chicago

Chen Yeh

UIUC

This version: November 2017

First version: September 2015

[\[Link to the latest version\]](#)

## Abstract

In this paper we study the role of product entry and exit in propagating nominal shocks to the real economy. Toward that goal, we show that product turnover has an extensive role in the aggregate economy and that the frequency and size of price adjustments are negatively related to a product's age. We exploit information from product-level characteristics and the timing of products' launches to provide empirical support that these stylized facts can be rationalized by an active learning motive: firms with new products are faced with demand uncertainty and can optimally obtain valuable information by varying their prices. Building on the empirical findings, we construct a menu cost model with active learning and quantify the importance of age-dependent pricing moments for the propagation of nominal shocks. In the calibrated version of our model, the cumulative response of output to a nominal shock more than doubles compared to the standard menu cost model and this response is higher during economic booms.

*JEL Classification Numbers: D4, E3, E5.*

*Key words: menu cost, firm learning, optimal control, fixed costs, nominal shocks.*

---

\*We thank Fernando Alvarez, Erik Hurst, Francesco Lippi, Robert Shimer, and Joseph Vavra for their advice and support. We want to thank Bong Geun Choi, Steve Davis, Elisa Giannone, Mikhail Golosov, Veronica Guerrieri, Greg Kaplan, Oleksiy Kryvtsov, Munseob Lee, Robert Lucas Jr., Sara Moreira, Jón Steinsson, Nancy Stokey, Harald Uhlig as well as seminar participants at the University of Chicago, Bank of Mexico, IEA (2015), EGSC at Washington University in St. Louis (2015), Midwest Macro (2015), Econometric Society European Winter Meeting (2015), Econometric Society North American Summer Meeting (2016), SED (2016), EIEF, and the Bank of Italy. David Argente gratefully acknowledges the hospitality of the Bank of Mexico and the Einaudi Institute for Economics and Finance where part of this paper was completed.

<sup>†</sup>Main author. Email: [dargente@uchicago.edu](mailto:dargente@uchicago.edu). Address: 1126 E. 59th Street, Chicago, IL 60637.

# 1 Introduction

A recent surge of research on product innovation has shown that product creation and destruction is extensive in the U.S. and has emphasized its importance for macroeconomic performance (e.g. [Broda and Weinstein, 2010](#)). In particular, new products are prevalent. [Argente, Lee and Moreira \(2017\)](#) find that one in three products are either created or destroyed in a given year and that more than 20 percent of the products are less than one year old. While turnover rates are clearly high, whether product creation and destruction are quantitatively important for the propagation of nominal shocks to the real economy is not clear.<sup>1</sup>

In this paper, we study the real effects of a nominal stimulus in an economy featuring high rates of product turnover and age-dependent pricing moments. To do so, we start by using a large panel of US products to document new facts about the distribution of both the duration and size of price changes over the product’s life cycle; a dimension which price-setting models ignore. We show that pricing moments at the product level strongly depend on their age: entering products change their prices twice as often as average products and the average size of these adjustments is 50 percent larger compared to the average price change.<sup>2</sup>

We then explore the underlying reason why new products display such different pricing dynamics. The empirical evidence from a select number of industries shows that the introduction of a new product comes with a significant amount of demand uncertainty from the firm’s perspective. These firms vary their prices to obtain valuable information about their demand curves.<sup>3</sup> We provide further evidence of this and extend the set of previous empirical findings in the literature to a broad line of product categories. Exploiting variation in the timing of product launch within retailers, we find that retailers carry forward information obtained during the first launch to any subsequent launch of the same product at a different location. As a result, retailers adjust prices less frequently and less aggressively after the first launch of their product. Furthermore, the negative relation between a product’s pricing moments and its age should be more pronounced for more novel products. To confirm this relation, we construct a unique *newness* index that measures the novelty of a product. We indeed find that more novel, entering products adjust their prices more often and by larger amounts.

---

<sup>1</sup>Nominal shocks, as discussed in [Lucas \(2003\)](#), can affect real variables and relative prices in the short run but not in the long run. [Shapiro and Watson \(1988\)](#) refers to these shocks as “demand” shocks.

<sup>2</sup>The patterns at exit are quite different as the frequency and absolute size of the price adjustments stay mostly constant before exit at the product level. Nonetheless, the frequency and depth of sales increase significantly at exit which indicates that firms attempt to liquidate their inventories before phasing out their products. These findings are described in more detail in the appendix.

<sup>3</sup>[Gaur and Fisher \(2005\)](#), for example, surveys 32 US retailers and finds that 90 percent conduct price experiments to learn about their demand.

Our set of stylized facts is the first in the price-setting literature that focuses on age-dependent pricing moments. While previous contributions have found heterogeneity in pricing moments along other dimensions (e.g. [Nakamura and Steinsson, 2009](#)), none of them has focused on a product’s life cycle. Our empirical findings provide insights on the underlying reason why firms adjust prices. This is relevant because the theoretical literature on price-setting shows that different types of price changes have substantially different macroeconomic implications.<sup>4</sup> We argue that the age distribution of products at a given point in time is important for the propagation of a nominal shock to real output.

To support our argument, we build a quantitative menu cost model in the spirit of [Golosov and Lucas \(2007\)](#). The model encompasses demand uncertainty and active learning similar to that in [Bachmann and Moscarini \(2012\)](#).<sup>5</sup> Whenever a firm’s product enters the market, it is unaware about its elasticity of demand. However, the firm does know that its elasticity is constant, and that it is either high or low. Firm-specific demand shocks prevent a firm from directly inferring its type: a firm that sets a relatively high price and observes a low amount of sales cannot distinguish between the fact that its product has a high elasticity of substitution or whether the realization of its demand shock was simply low. To deal with this uncertainty, firms engage in a Bayesian learning process and form beliefs about their elasticity.<sup>6</sup> Changes in its price alter the speed at which the firm learns about its elasticity of demand. Thus, the firm balances its incentives between maximizing static profits and deviating in order to affect its posterior beliefs in the most efficient way so that it acquires more valuable information about its type. This is also known as the trade-off between current control and estimation. As a firm ages and learns more about its elasticity, incentives for active learning decline and the dispersion of price changes decreases. This is consistent with evidence from other markets such as the newly deregulated market of frequency response in the UK. [Doraszelki, Lewis and Pakes \(2016\)](#) find that in response to uncertainty, firms experiment with their bids by adjusting them more frequently and by larger amounts. As firms acquire more information about demand, the adjustment of the bids became less frequent and smaller.

Then, we calibrate this framework to standard pricing moments and our newly found

---

<sup>4</sup>Especially if the timing of the price change is endogenous (e.g. [Caplin and Spulber \(1987\)](#)) or exogenous (e.g. [Calvo \(1983\)](#)).

<sup>5</sup>Their focus is very different as they study how negative first moment aggregate shocks induce risky behavior. In their model, when firms observe a string of poor sales, they become pessimistic about their own market power and contemplate exit. At that point, the returns to price experimentation increase as firms “gamble for resurrection.” However, we observe no evidence for differential pricing moments at a product’s exit in the data.

<sup>6</sup>Our work is related to the literature in optimal control problems with active learning that have been studied in many areas of economics since [Prescott \(1972\)](#). Its application to the theory of imperfect competition consists of relaxing the assumption that the monopolist knows the demand curve it faces. The first application of this concept can be found in [Rothschild \(1974\)](#). More recent examples can be found in [Wieland \(2000b\)](#), [Ilut, Valchev and Vincent \(2015\)](#) and [Willems \(2016\)](#).

pricing moments that vary over a product’s life cycle, and quantify the cumulative real output effect of a nominal spending shock. Our findings indicate that the output response is 2.3 times larger than the benchmark model with no demand uncertainty. The reasoning behind this result is twofold. First, the learning incentives dampen the selection effect; that is, pricing with active learning motives pushes firms away from the margin of price adjustment and lessens the mass of firms that adjust their prices due to a nominal shock.<sup>7</sup> Since firms have an additional motive to change prices, they are less sensitive to changes in their costs.<sup>8</sup>

Second, the concept of a product’s life cycle introduces an additional form of cross-sectional heterogeneity in the adjustment frequency. In an environment with active learning incentives combined with menu costs, uncertain firms are willing to adjust their prices more often to acquire information on their demand. These firms will most likely adjust their price several times before firms with sharper beliefs adjust their price once after a nominal shock. However, all price changes by more uncertain firms that occur after the one in response to the nominal shock have no effect on real output because these firms have already adjusted to the shock. Given that the model is calibrated to match the average frequency of the price changes, firms that are more certain about their type significantly delay the adjustment of the aggregate price level after a nominal shock. These firms tend to be older and have lower frequency of price adjustment on average. As a result, this delay reduces the selection in timing of price changes.<sup>9</sup>

The remainder of this paper is organized as follows: In Section 2, we present the data and the main empirical findings. In section 3, we set up a quantitative model of menu costs that is able to explain our stylized facts. In addition, we develop the relevant conditions for active learning and describe what learning regimes might occur. In Section 4, we discuss our results on the propagation of nominal shocks and compare our results with other models used in the literature. We also discuss the sensitivity of our results to matching other moments that vary over the product’s life cycle such as sales profiles and exit rates. We also extend the model to include endogenous entry to quantify the response of output in periods of high and

---

<sup>7</sup>The term “selection” was introduced by [Golosov and Lucas \(2007\)](#) to indicate that firms that change prices after a nominal shock are those whose prices are in greatest need of adjustment. Given that the distribution of the size of price changes fully encodes this type of selection, a wide range of papers in recent years have taken advantage of micro data to match the size distribution of price changes, such as [Nakamura and Steinsson \(2008\)](#), [Midrigan \(2011\)](#), [Karadi and Reiff \(2016\)](#) and [Vavra \(2014\)](#).

<sup>8</sup>Although this logic is similar to the one described in [Alvarez, Bihan and Lippi \(2014\)](#) in relation to the response of real output to nominal shocks, our framework does not fit the class of models for which the kurtosis of the distribution of price changes is a sufficient statistic. This is because firms might optimally choose to deviate from the static profit-maximizing price in absence of a menu cost in order to learn about their demand elasticity.

<sup>9</sup>Recent contributions highlighting the importance of the selection in timing are [Kiley \(2002\)](#), [Nakamura and Steinsson \(2009\)](#), [Sheedy \(2010\)](#), [Alvarez, Lippi and Paciello \(2011\)](#) and [Carvalho and Schwartzman \(2015\)](#).

low product entry. Section 5 concludes. The appendix provides proofs to the propositions, additional empirical findings, and extensions of the model.

## 2 Stylized Facts of Product’s Life Cycle

In this section, we use a large scanner data set to show a new set of stylized facts on a product’s life cycle in the US economy. We begin by showing the importance of new products both in terms of their count and revenue relative to the aggregate. Then, we develop a set of facts that clearly show that pricing moments at the product level are considerably different across a product’s age; in particular near entry. At entry, the frequency of regular price changes, the absolute size of regular price adjustments, and the cross-sectional standard deviation of regular price changes are higher. All of these moments approximately settle to their respective averages as the product matures. Furthermore, the fraction of large price changes, defined as those changes larger than two standard deviations, is considerably larger at the beginning of the product’s life cycle. We hypothesize that an active learning motive can rationalize these facts. The last part of this section provides additional evidence consistent with this interpretation by exploiting the variation in products’ entry time across different stores and their characteristics across different product categories.

### 2.1 Data

The life cycle patterns of products’ prices have typically not been studied much as the requirements on the data are quite stringent. Doing so requires a large panel of products with information about their entry date and prices at a high sampling frequency. The Consumer Price Index (CPI) Research Dataset, for example, is only available at a monthly frequency and the age of products is unknown. All Entry-Level Items (ELIs) are added to the CPI basket long after its first appearance in the national market.

For this reason, we rely on the IRI Marketing data set instead that provides more than ten years of data at the store-product-week level. The data is generated by point-of-sale systems: each retailer reports the total dollar value of its weekly sales and total units sold for each product. A product is identified by its Universal Product Code (UPC), a code consisting of 12 numerical digits that is uniquely assigned to each specific product and represents the finest level of disaggregation at the product level.

The data contains approximately 2.4 billion transactions from January 2001 to December 2011 that represents roughly 15 percent of household spending in the Consumer Expenditure Survey (CEX). Our sample contains approximately 170,000 products and 3,000 distinct stores

across 43 metropolitan areas (MSA). The data covers 31 product categories and includes detailed information about each product such as its brand, volume, color, flavor, and size.<sup>10</sup>

Given the properties of the data, we can identify the first appearance of a UPC in a certain store by using the retail and product identifiers. We assume that if a UPC changes, some noticeable characteristic of the product has also changed. This is because it is rare that a meaningful quality change occurs without a change to its UPC. Considering each UPC as a product is, in fact, a very broad definition since it includes classically innovative products, which are “breakthrough” products that deliver innovation to an existing or new product category; line extension products, which are new products within an already existing category; and temporary products, which have a short life cycle and are typically seasonal. We find that product line extensions, such as flavor or form upgrades or novelty and /seasonal items, are much more prevalent than the introduction of new brands.

Using UPC and retailer identifiers, we are able to determine at what week and store each product first appears. We define entering products as those that enter the US market after January 2002. Our data starts from January 2001, thus an entering product is one that has no observable transactions in any store across the US for at least one year. This assumption avoids the inclusion of products with a left-censored age. In addition, we only consider products that entered the market before the first week of 2007. We impose this restriction for two reasons. First, the prices of products born during downturns can have substantially different patterns than those of products born in normal times.<sup>11</sup> More importantly, IRI Symphony undertook a substantial reorganization of its product categories and expanded their scope at the beginning of 2007. Thus, the data after this specific date might include some entering products that might not correspond to actual product introductions.<sup>12</sup> By restricting our sample of entering products between January 2002 and January 2007, we

---

<sup>10</sup>The product categories include Beer, Carbonated Beverages, Coffee, Cold Cereal, Deodorant, Diapers, Facial Tissue, Photography Supplies, Frankfurters, Frozen Dinners, Frozen Pizza, Household Cleaners, Cigarettes, Mustard & Ketchup, Mayonnaise, Laundry Detergent, Margarine & Butter, Milk, Paper Towels, Peanut Butter, Razors, Blades, Salty Snacks, Shampoo, Soup, Spaghetti Sauce, Sugar Substitutes, Toilet Tissue, Toothbrushes, Toothpaste, and Yogurt. The dataset is discussed in more detail in [Bronnenberg, Kruger and Mela \(2008\)](#). See also [Alvarez, Bihan and Lippi \(2014\)](#), [Chevalier and Kashyap \(2014\)](#), [Gagnon and Lopez-Salido \(2014\)](#), [Stroebel and Vavra \(2014\)](#) and [Coibion, Gorodnichenko and Hong \(2015\)](#) for applications of the data to related questions.

<sup>11</sup>[Moreira \(2015\)](#), for example, provides evidence that the average business size across cohorts is significantly affected by aggregate economic conditions at inception.

<sup>12</sup>More specifically, IRI undertook the following actions: i) reorganization of private-label items (i.e. organic private labels are broken out for some categories), ii) dropping of UPCs that have not moved in past years, iii) collapse of UPCs into a main UPC to avoid clutter (i.e. products that came to a store as part of a special promotional code rather than with a standard UPC code), iv) reorganization of categories (i.e. a category might have increased in scope and as a consequence experienced an increase in items), and v) addition of UPCs that were introduced at the beginning of each stub. All of these are consistent with changes in the number of entering and exiting UPCs due to changes in the product stub rather than new product introductions or products being phased out.

avoid this reclassification bias.

Further, in order to minimize concerns of potential measurement error in the calculation of product-level entry and exit, our baseline sample excludes private label products and only considers products that last at least two years in the market. We exclude private label items from the analysis because all private-label UPCs have the same brand identification so that the identity of the retailer cannot be recovered from the labeling information. We exclude short-lived products in order to minimize the problem that some UPCs might get discontinued only to have the same product appear with a new UPC as noted by [Chevalier, Kashyap and Rossi \(2003\)](#). We also drop promotional items or products with very little revenue to minimize biases due to measurement error.

## 2.2 The Importance of New Products

[Broda and Weinstein \(2010\)](#) emphasize the importance of entering and exiting products for the aggregate performance of the US economy. [Argente, Lee and Moreira \(2017\)](#) show that product turnover in the US is substantial as one third of all products are either created or destroyed in a given year and more than 20 percent of US products are aged less than one year. In this subsection, we sketch an identical picture in our sample to highlight the importance of entering products. We begin by using information on the number of new products, exiting products, and the total number of products in each category  $c$  to define aggregate entry and exit rates at the product level:

$$n(t, s) = \frac{\sum_i N_i(t, s)}{\sum_i T_i(t)}$$

$$x(t, s) = \frac{\sum_i X_i(t, s)}{\sum_i T_i(s)}$$

where  $N_i(t, s)$ ,  $X_i(t, s)$ , and  $T_i(t)$  are number of entrant products, exit products, and total products in period  $t$  relative to period  $s$ . We define the entry rate in period  $t$  relative to  $s$  as the number of new products in period  $t$  relative to period  $s$  as a share of the total number of products with strictly positive sales in period  $t$ . A new product is one that records at least one transaction in any store in period  $t$  and that was not sold in any store in period  $s$ . Creation and destruction are the revenue weighted analogues of the entry and exit rates.

Using a scanner dataset collected at the store level offers the advantage of observing, for the categories available, the entire universe of products for which a transaction is recorded in a given week. For this reason, we can distinguish between products entering the market and products being launched at each store, where our unit of observation is every UPC×Store



pair. We find a substantial degree of entry of products at both levels.

Table I reports the entry and exit rates for the case in which  $t$  and  $s$  are one and five years apart. It shows that 15 percent of the UPCs in the market and on average 27 percent of the products in each store entered in the last year. Approximately 45 percent of the products in the market entered in the last five years accounting for 30 percent of total expenditures. At the store level, 66 percent of all products sold were first introduced by the store in the last five years and they account for more than half of the total revenue of the store. Although the exit rate is very similar to the entry rate of products, destruction is lower than creation. This lower rate means that consumers spend more on new products than on products that are about to exit. The rate of product turnover indicates that at any point in time, there is a large amount of products being launched or being phased out.

Panel A in table II shows that the median duration of a product in a given store is only slightly above three years.<sup>13</sup> The remarkably large rates of product creation and destruction, both at the store and at the market level, along with a short product life cycle indicate that the pricing of entering products are relevant for determining the dynamics of aggregate prices.

## 2.3 Empirical Strategy

To study the price dynamics of products along their life cycle, we begin by computing the average retail price in a given week:

$$P_{mcjst} = \frac{\text{sales}_{mcjst}}{\text{units}_{mcjst}}$$

where  $m$ ,  $c$ ,  $j$ ,  $s$ , and  $t$  index markets (at the MSA level) are product categories, UPCs, stores and time respectively. A considerable advantage of the IRI Symphony is that it provides information on whether and when a product was on sale in a certain store (the so-called “sales flag”) that is absent in other scanner data sets. Since our goal is to study the speed of price adjustment following a nominal shock, we focus on studying the life cycle patterns of regular price changes given that retailers’ use of sales in our data do not vary with macroeconomic conditions.<sup>14</sup> Nonetheless, the main stylized facts discussed below are

---

<sup>13</sup>Since our dataset ends the last week of 2011 and we are considering products that entered the last week of 2006 at the latest, right censoring is only an issue for products that last more than 261 weeks in the market.

<sup>14</sup>Using the same data set, Coibion, Gorodnichenko and Hong (2015) find that retailers’ use of sales does not vary with the unemployment rate. Anderson, Malin, Nakamura, Simester and Steinsson (2013) argue that sales prices are governed by sticky plans and they are planned in advance according to a “trade promotion calendar”. They also find that retailers do not respond to macroeconomics shocks by adjusting the size or frequency of sales. In addition, we do not find evidence that retailers use sales to actively learn at entry as neither the frequency or the size of sales are larger. See figure A.XV in the online appendix for a more



robust to the inclusion of sales in the analysis.

We adopt the same conventions as [Coibion, Gorodnichenko and Hong \(2015\)](#) to distinguish between regular price changes and sales. A *regular* price change is defined as any change in price that is larger than one cent or 1 percent in absolute value. For prices larger than 5 dollars in value, this cut-off is 0.5 percent. To identify sales, we use the sales flag provided in the data, but our results are robust to applying the sales filter introduced by [Nakamura and Steinsson \(2008\)](#).<sup>15</sup> The size of a price change is calculated as the log difference between the price levels in the current and the previous week. Thus, we get:

$$\Delta P_{mcjst} = \ln(P_{mcjst}) - \ln(P_{mcjst-1})$$

Let  $a = 1, \dots, A$  denote the number of weeks since entry (which we will define as the *age* of the product) where  $a = 1$  denotes entry. To assess the movements of the pricing moments over the life cycle of a product, we adopt the following empirical specification:

$$Y_{jsct} = \alpha + \sum_{a=1}^A \phi_a D_{js}^a + \theta_{js} + \tau_t + \gamma_c + \varepsilon_{jsct} \quad (1)$$

where  $j, s, t$  and  $c$  are the UPC, store, time period and cohort  $c = t - a$  the product belongs to respectively.  $Y_{jsct}$  is the variable of interest (e.g. the price change indicator or the size of the price change).  $D_{js}^a$  is a dummy variable that takes the value of one if the product is in its  $a^{\text{th}}$  week since entry.  $\theta_{js}$  captures the fixed effects at the UPC-store level whereas  $\tau_t$  and  $\gamma_c$  denote time and cohort fixed effects respectively. We are interested in the regression coefficients  $\{\phi_a\}_{a=1}^A$  which capture age heterogeneity on the pricing moment of interest.

In our empirical specification, it is not possible to identify the heterogeneous effects of age conditional on a product’s cohort and time period due to perfect collinearity. To resolve this, we estimate equation 1 using two different normalizations. The first assumes that trends appear only in cohort effects. In this case, we replace the time fixed effects with the seasonally adjusted unemployment rate at the level of Metropolitan Statistical Area (MSA) to control for local cyclical economic variation.<sup>16</sup> The second assumes trends only appear in the period

---

elaborate discussion on this issue.

<sup>15</sup>Specifically, under this approach, a good is on sale if a price is reduced but returns to its same previous level within four weeks. [Coibion, Gorodnichenko and Hong \(2015\)](#) use two approaches to identify a price spell. The first treats missing values as interrupting price spells. In the second approach, missing values do not interrupt price spells if the price is the same before and after the periods with missing values. Since the incidence of sales from applying these two approaches does not significantly differ from the one identified by the sales flag provided in the IRI Marketing dataset, we use the union of sales flags obtained from applying these two approaches and the flag provided in the IRI Marketing data to identify the incidence of sales. Our results are not sensitive to any of these choices.

<sup>16</sup>Some examples of studies that use this proxy approach are [Deaton and Paxson \(1994\)](#), [Gourinchas and](#)

effects. In this case, the time fixed effects are included in the estimation of equation 1 and we use the local unemployment rate to represent the cohort fixed effects. Our baseline results use the latter approach, but our results are not sensitive to the chosen normalization.

## 2.4 Pricing Moments over the Life Cycle of US Products

We first use regression specification 1 with a price change indicator as the dependent variable. Figure 1 plots the frequency of regular price changes over the life cycle of a product. The dots represent the estimates of the age fixed effects  $\{\phi_\alpha\}_{\alpha=1}^A$  computed with equation 1.<sup>17</sup> The newer products clearly have their prices changed more often. We state our first stylized fact:

**EMPIRICAL FACT 1.** The average frequency of price adjustment declines with the product’s age. The decline is most pronounced at the early stage of the product’s life cycle as entering products change their prices twice as often as the average product.

The frequency of price adjustment is almost 4 percentage points higher at entry and takes approximately 20 weeks to settle to its average value of 5 percent. The magnitude of this significantly higher frequency is best reflected in the expected amount of time it takes for a product to change its price.<sup>18</sup> If we maintain the frequency of adjustment at entry, then a price should change approximately every 12 weeks. This is *twice* as often relative to the average of 24 weeks that we observe in the data.

Our findings are consistent with those in Alvarez, Borovicková and Shimer (2015). They test whether hazard rates of price changes depend on the age of the product once unobserved heterogeneity is taken into account. They find this is the case and argue that this statistical model is a reasonable representation of the data. Figure A.I in the appendix decomposes these price changes into increases and decreases. We find similar results for both the frequency of price increases and decreases.

In order to study whether the magnitude of these price adjustments also changes over the life cycle of the product we use a similar approach but instead use the absolute size of price changes as the dependent variable. Figure 2 depicts our results. During the first few months, the absolute value of price changes is much larger and almost 5 percentage points higher than the average which amounts to approximately 9 percent. Further, the dispersion of price changes as measured by the weekly cross-sectional standard deviation is almost 40 percent

---

Parker (2002), Aguiar and Hurst (2013), and De Nardi, French and Jones (2010).

<sup>17</sup>Specifically, we plot  $\hat{\alpha} + \hat{\phi}_\alpha$  for every  $\alpha \in \{1, \dots, A = 50\}$ .  $\hat{\alpha}$  is the unconditional average of the frequency of price changes of a product that has been in the market for 50 weeks.

<sup>18</sup>This is equal to  $-1/\ln(1 - f)$  where  $f$  denotes the frequency of price adjustment.

larger during the first four months after entry with respect to its level 12 months after the product is launched. Importantly, this fact holds for both price increases and decreases as shown in figures A.II and A.III in the appendix.<sup>19</sup> This leads to our second stylized fact:

**EMPIRICAL FACT 2.** The absolute size of price adjustment declines monotonically with product’s age. The decline is most pronounced at the early stage of the product’s life cycle as the average absolute size of entering products is almost twice as large as the average change.

Our baseline specification uses a non-parametric specification for the age of a product. Our stylized facts remain robust whenever we use a linear specification in the age of a product. This is summarized in table III. Thus, we conclude that firms not only price more often but also in a more extreme fashion during the early stages of their products’ life cycles.

Next we investigate whether very large price changes are more or less frequent as products get older. To do so, we follow the approach by Alvarez, Bihan and Lippi (2014) to minimize the issue of heterogeneity across products and stores. We define “cells” at the UPC-store level, say  $(j, s) = i$ , and standardize each price change at this level through  $z_{it} = (\Delta p_{it} - \mu_i) / \sigma_i$  where  $\mu_i$  and  $\sigma_i$  are the mean and standard deviation of price changes in cell  $i$  across time. Price changes equal to zero are disregarded.

Figure 3 shows the distribution of regular price changes larger than two standard deviations as a function of the age of the product. We observe a sizeable share of large price changes close to entry particularly during the first 20 weeks. About 40 percent of the price changes larger than two standard deviations of the product life cycle occur during these weeks. The distribution of large price changes is roughly uniform after that.<sup>20</sup> Figure A.V in the appendix shows that this occurs for price changes in both directions.<sup>21</sup> Our third stylized fact is then summarized by:

**EMPIRICAL FACT 3.** About 40 percent of large price changes occur during the first 20

---

<sup>19</sup>Bachmann and Moscarini (2012) focus on the pricing moments of products at exit. In their work, at the end of the product life cycle, the returns to price experimentation increase as firms “gamble for resurrection”. We found little support for this mechanism in the IRI Marketing data. Figure A.XIV in the appendix shows that both the frequency and size of regular price changes stay mostly constant at exit. Nonetheless, figure A.XIV shows that there is an increase in the frequency of sales during the last weeks of the life cycle of the product. We interpret this pattern as an increase in “clearance” sales.

<sup>20</sup>This finding shows that idiosyncratic shocks from fat-tailed distributions need to be used with caution when used to generate large price changes. In standard menu cost models it is often assumed that these shocks arrive at a constant rate. This assumption includes the family of Poisson shocks used in Midrigan (2011) and Karadi and Reiff (2016) where the distribution of large price changes is independent of age. This is relevant as fat-tailed shocks have drastic implications on the degree of non-neutrality. Their presence reduces the selection effect after a monetary shock as the mass of firms responding to it is smaller.

<sup>21</sup>Figure A.VI in the appendix shows that our results are not sensitive to our standardization since the fraction of non-standardized price changes larger than 30 percent shows the same pattern.

weeks of the product’s life cycle. A quarter of them are observed in the first three weeks of a product’s entry.

## 2.5 The Case for Active Learning

Our empirical findings so far have not shed any light on the economic mechanism that could rationalize them. In this section, we argue that active learning can rationalize our empirical findings, and we provide additional empirical evidence to support this claim. There is already a substantial body of literature that shows that firms actively use their prices to obtain information about their demand. This is widespread across the US economy according to empirical and anecdotal evidence from [Gaur and Fisher \(2005, US retailers\)](#), [Einav, Kuchler, Levin and Sundaresan \(2015, online listings on eBay\)](#), and [Campbell and Eden \(2005, grocers\)](#).

Whenever firms face uncertainty over the demand of their product due to some aggregate (e.g., industry- or location-specific shocks) or idiosyncratic (e.g., consumer taste shocks) factors, they can gain information from large movements in a product’s price. This is because these factors are less likely to be important whenever a firm’s sales move due to large, deliberate variations in their price. At entry, a firm might not be aware of some of its demand characteristics. As a result, entering products should change their prices more often and by larger amounts.<sup>22</sup>

Further, active learning has two additional testable implications. First, if the same product is launched at different times or different locations by the same retailer, we should observe that the patterns of learning attenuate after the introduction. This is true when assuming demand characteristics are persistent across time and space. Under this scenario, retailers incorporate attained information from the introduction to later rounds in which the incentives for learning are dampened. Second, if firms are actively learning about their demand, then we should observe that active learning increases when the demand for the product being launched is more uncertain. This increase can occur if the product in question is more

---

<sup>22</sup>An alternative explanation could be penetration pricing. Under this scenario, firms increase long-run profits by launching a low-priced product to secure market share or a solid customer base. This tactic then results in higher future profits as the firm is able to benefit from the consumer’s higher willingness to pay. In the appendix, we provide evidence that this strategy does not seem to be consistent with our empirical findings. We do not find evidence of lower entry prices (figure [A.IX](#)) or rapid price increases at the beginning of the product’s life cycle (figure [A.VIII](#)). In fact, figure [A.VII](#) shows that price increases and decreases are mostly balanced after entry. Conditional on having at least two price changes during the first 6 months, less than 15 percent of the products record only price increases and the probability of two consecutive price increases is less than half. To provide further evidence, in appendix [B.3](#) we extend the standard price-setting model to include customer base concerns and show that the implications of the model are not consistent with several of the stylized facts we have documented.

novel or innovative. The nature of the IRI Marketing data allows us to test both of these conjectures.

### 2.5.1 Timing of Product Launch

The first conjecture consists of studying whether retailers carry forward any information obtained during the first launch to any subsequent launch of the same product at a different location. For this purpose, we take advantage of the variation in products' launch dates across different periods in time or locations. If information is held at the retailer level and they are uncertain about some demand characteristics, then retailers should adjust prices less frequently and less aggressively after the first launch of their product.<sup>23</sup>

To show this, we first divide every UPC-store pair into two different "waves." A UPC-store pair belongs to the first wave if it was launched by a retailer before the completion of the first year since the UPC was first introduced in the national market. Then, a UPC-store pair belongs to the second wave if it was introduced by the same retailer at least one year after the product was first launched at the national level.

Figure 4 shows that after following the baseline regression specification 1 for each wave, products in the first wave have a higher frequency of adjustment at entry than those in the second wave. The figure also shows the same patterns for the absolute size of price changes. On average, at entry, the size of price changes is 7 percent larger than the mean for products in the first wave and only 4 percent larger for products in the second wave. The size of absolute price changes then converges back to the mean of its respective wave.

In the appendix, we show that these findings also hold if we condition on the fact that waves should occur across different cities and, importantly, these patterns occur for both price increases and decreases. These figures show that retailers obtain relevant information about the demand their product faces in the first wave. As a result, their incentives to learn at the time of entry in the second wave are then lower. Thus, there is less need for active learning, and we see that the age-dependence of products' pricing moments become attenuated.

### 2.5.2 Newness Index

To strengthen our findings on learning, we confirm the hypothesis that the age-dependence of pricing moments is more pronounced for products that are more novel. Our definition of what constitutes a "new" product in our earlier empirical exercises is relatively broad. However, it is very well possible that product introductions vary significantly in their degree of novelty.

---

<sup>23</sup>We assume that markets across space are not completely independent. This assumption means that any information obtained on the local demand of some product in Chicago is at least somewhat informative for another city such as New York.

A new product could be an entirely new brand, exist within an incumbent brand, or simply be an improvement or variation on an existing product (e.g., new color or flavor). In order to quantify the novelty of a product, we first compare the pricing behavior of retailers when launching a new brand relative to the rest of the products. The reasoning for this comparison is that brand extensions are more rare and constitute larger innovations relative to products already in the market. We do so using the following regression specification:

$$Y_{jsct} = \alpha + \gamma \text{age}_{jsct} + \phi \text{age}_{jsct} \times \text{NewBrand}_{jsct} + \theta_{js} + \tau_t + \gamma_c + \varepsilon_{jsct}. \quad (2)$$

where as before  $j$ ,  $s$ ,  $c$ , and  $t$  are the UPC, store, cohort and time period, respectively. The dependent variable  $Y_{jsct}$  either represents the price change indicator of the absolute size of price changes.  $\text{NewBrand}_{jsct}$  is an indicator that equals one if the product introduction is a new brand with a new volume.  $\gamma$  captures the heterogeneity of pricing moments in age, and  $\phi$  measures the strength of this heterogeneity with respect to the novelty of the product’s brand. Not surprisingly, the effect of age is negative and strongly significant as  $\gamma$  only summarizes the patterns described in section 2.4 with a linear function. More importantly, table IV shows that the coefficient for the interaction term  $\phi$  is also negative and significant and shows that newer products tend to adjust their prices more often and by larger amounts during the first six months of their life cycle.

In order to provide a more comprehensive measure of the novelty of each product, we construct a *newness index* that uses detailed information about the characteristics of each UPC provided in the IRI Marketing data set. The index counts the number of new and unique attributes a product has at the time of its introduction relative to all of the other products ever sold by a store within the same category. Our measure assigns a higher value to products with more unknown features to the store. Our aim is to capture the novelty of a product from the store’s perspective in order to study whether its pricing patterns differ when it sells a product whose demand parameters are more uncertain (i.e., novel).

We define a product  $j$  in category  $k$  as a vector of characteristics  $V_{kj} = [v_{j1}, v_{j2}, \dots, v_{jN_k}]$  where  $N_k$  denotes the number of attributes we observe in category  $k$  in our data.<sup>24</sup> Then, if  $\Omega_{kst}$  contain the set of product characteristics for each product ever sold in category  $k$  at store  $s$  at time  $t$ , then the *newness index* of a product  $j$  in category  $k$ , launched at time  $t$ ,

---

<sup>24</sup>For example, the product category beer consists of  $N_{beer} = 9$  attributes for each barcode: vendor, brand, volume, type (e.g., ale or lager), package (e.g., can or keg), flavor, size (e.g., bottle or six pack), calorie level (e.g., light or regular) and color.

and in store  $s$  is defined as follows:

$$\text{NI}_{jst}^k = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbb{1}[v_{ji} \notin \Omega_{kst}]. \quad (3)$$

For example, if a new product within the beer category enters with a flavor and a volume that has never been sold at the store before, its newness index is  $2/9$ . We assume that each attribute is equally weighted in order to remain agnostic about the relative importance of each attribute to the degree of newness of a product.<sup>25</sup> To understand if stores launching more novel products are actively learning, we estimate:

$$Y_{jst}^k = \alpha + \gamma \text{age}_{jst} + \phi \text{age}_{jst} \times \text{NI}_{jst}^k + \theta_{js} + \tau_t + \gamma_c + \varepsilon_{jst}. \quad (4)$$

where  $\phi$  is our coefficient of interest because it summarizes the degree of heterogeneity in a product’s age depending on its novelty. Table V shows that our index has substantial power to explain the price-setting patterns we observe at entry. The index confirms the second testable conjecture on learning by showing that more novel products change prices more actively at the beginning of their life cycle. This evidence provides additional support to the active learning hypothesis. The remainder of the paper takes these empirical facts as given and extends the standard price-setting models to include the concept of the product’s life cycle through an active learning mechanism.

### 3 Quantitative Model

Our framework is a discrete time menu cost model in the tradition of [Goloso and Lucas \(2007\)](#) in which firms face uncertainty on their demand curves. A firm’s type, that is, its elasticity of demand, is either high or low, but the firm does not know its type. As a result, a firm forms beliefs on its type. It can adjust its price to change the speed at which its beliefs get updated. Thus, a firm faces a trade-off between maximizing its static profits and gaining more information about its type when choosing its price. The active learning mechanism is based on [Mirman, Samuelson and Urbano \(1993\)](#) and its implementation in general equilibrium is closely related to [Bachmann and Moscarini \(2012\)](#). However, we deviate from their framework by removing the firm’s fixed costs of production. This removal eliminates the “gambling for resurrection” effect for which we cannot find evidence in our micro-level

---

<sup>25</sup>Our index should only be considered as an approximation of the novelty of an item given that it relies heavily on the number of attributes provided by the data that might not describe a product in its entirety. On average, we observe ten product characteristics in each category.



data. Furthermore, the moments in the data indicate that the frequency and absolute size of price changes converge from above to a fixed level over time. This convergence indicates that the incentives for price changes are *not* driven by active learning alone. We deal with these empirical regularities by adding a menu cost and firm-level idiosyncratic shocks. We choose to model active learning in the most transparent manner that is still consistent with the data. Nonetheless, we explore several extensions to our benchmark model to show its robustness (section 4.4) and to highlight further implications (section 4.3).

### 3.1 Households

Households in the economy maximize the expected, discounted utility over aggregate consumption  $C_t$  and labor supply  $L_t$  that is characterized by:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[ \frac{C_t^{1-\theta} - 1}{1-\theta} - \frac{\omega}{1+\chi} L_t^{1+\chi} \right]$$

where  $\mathbb{E}_t$  denotes the expectations operator conditional on information available to the household at time  $t$ . Households have CRRA preferences over an aggregate consumption good with risk aversion parameter  $\theta$  and the level of disutility is denoted by  $\omega$ . The inverse Frisch elasticity is given by  $\chi$  and households discount by a factor  $\beta$  per period.

The aggregate consumption good  $C_t$  is a Cobb-Douglas composite of two Dixit-Stiglitz indices of differentiated goods:

$$C_t = C_{1t}^\eta C_{2t}^{1-\eta} \text{ with } C_{it} = \left[ \int_{k \in J_i} (\alpha_t^i(k))^{\frac{1}{\sigma_i}} (c_t^i(k))^{\frac{\sigma_i-1}{\sigma_i}} dk \right]^{\frac{\sigma_i}{\sigma_i-1}}$$

There are two continua of differentiated goods consisting of perishable consumption units or services. A good is indexed by a pair  $(i, k)$  where the first index  $i \in \{1, 2\}$  denotes the good's *type*. Its *variety* within a group type is denoted by  $k \in J_i$ . Goods come in two types or baskets:  $J_1$  and  $J_2$  are the groups of *specialties* and *generics* respectively. A group  $J_i$  is characterized by its Lebesgue measure  $\varphi_i$ . Varieties within the first basket are hard to substitute with each other whereas generic varieties are mutually substitutable with a relatively high elasticity of substitution, thus we have  $\sigma_2 > \sigma_1$ . Each good is assumed to be produced by a single monopolistically competitive producer.<sup>26</sup> Each good is identified through the index pair  $(i, k)$ . We assume that  $i$  is time-invariant whereas the consumer's variety-specific preference shocks  $\alpha_t^i(k)$  are drawn every period, independently over time, and within and across group types. Draws are the same for all consumers. Households'

<sup>26</sup>As a result, we will use the term "firm" and "product" interchangeably.

decisions are taken *after* observing these taste shocks.

Within each period, households choose how much to consume of each differentiated good to maximize the level of the aggregate consumption good  $C_t$ . For a given level of spending  $S_t$ , we obtain the following downward-sloping demand curve for each differentiated good  $(i, k)$ :

$$c_t^i(k) = \alpha_t^i(k) \left( \frac{p_t^i(k)}{P_{it}} \right)^{-\sigma_i} \frac{\eta_i S_t}{P_{it}} \quad (5)$$

where, with some abuse of notation, the income shares for each basket are given by  $\eta_1 = \eta$  and  $\eta_2 = 1 - \eta$ .  $p_t^i(k)$  denotes the price of a good  $(i, k)$  in period  $t$  and the price index of group  $i$  is denoted by:

$$P_{it} = \left( \int_{k \in J_i} \alpha_t^i(k) (p_t^i(k))^{1-\sigma_i} dk \right)^{1/1-\sigma_i} \quad (6)$$

These price indices satisfy the following equalities:

$$\begin{aligned} P_{1t}C_{1t} + P_{2t}C_{2t} &= S_t \\ P_{1t}^\eta P_{2t}^{1-\eta} &= P_t \\ P_t C_t &= S_t \end{aligned}$$

Thus, the aggregate price level  $P_t$  is such that  $P_t C_t$  is the minimum amount of expenditure necessary to obtain  $C_t$  units of the aggregate consumption good. We assumed that the realization of taste shocks are independent across groups. Since there is a continuum of goods within each basket  $i$ , we can use a law of large numbers.<sup>27</sup> This implies:

$$P_{it} = \left( \int_{k \in J_i} p_t^i(k)^{1-\sigma_i} dk \right)^{1/1-\sigma_i}$$

where the law of large numbers gives us that  $\int \alpha_t^i(k) dk \xrightarrow{p} \mathbb{E}(\alpha_t^i(k)) = 1$  since we normalize the expected value of the taste shocks to be equal to unity.

Households have access to a complete set of Arrow-Debreu securities. Therefore, the period  $t$  budget constraint is characterized by:

$$P_t C_t + \mathbb{E}_t(q_{t,t+1} B_{t+1}) \leq B_t + W_t L_t + \sum_i \int_{k \in J_i} \Pi_t^i(k) dk$$

where  $W_t$  denotes the nominal wage rate and  $\Pi_t^i(k)$  is the profit that households receive from

---

<sup>27</sup>In particular, we use the *Glivenko-Cantelli* theorem. The argument is identical to the one in [Bachmann and Moscarini \(2012\)](#).

owning the firm producing good  $(i, k)$ .  $B_{t+1}$  denotes the state-contingent payoffs in period  $t + 1$  from purchasing assets in period  $t$ . These claims are priced in period  $t$  by the unique (stochastic) discount factor  $q_{t,t+1}$ . The first order conditions of a household's intertemporal maximization problem are then:

$$q_{t,T} = \beta^{T-t} \left( \frac{C_T}{C_t} \right)^{-\theta} \frac{P_t}{P_T} \quad (7)$$

$$\frac{W_t}{P_t} = \omega L_t^\chi C_t^\theta \quad (8)$$

These equations describe the determination of asset prices and labor supply.

### 3.2 Firms

In contrast to most state-dependent pricing models, firms in our framework set prices under incomplete information. At any point in time  $t$ , a firm  $(i, k)$  can observe its total amount of sales  $q_t^i(k)$  after setting some price  $p_t^i(k)$ . Due to the Dixit-Stiglitz specification above, these sales are comprised of:

$$q_t^i(k) = \alpha_t^i(k) \left( \frac{p_t^i(k)}{P_{it}} \right)^{-\sigma_i} \frac{\eta_i S_t}{P_{it}}$$

Given that this demand specification is log-linearly separable, we obtain:

$$\begin{aligned} \log(q_t^i(k)) &= \log(\alpha_t^i(k)) - \sigma_i \log\left(\frac{p_t^i(k)}{P_{it}}\right) + \log(\eta_i S_t) - \log(P_{it}) \\ &= -\sigma_i \log(p_t^i(k)) + \log(S_t) + (\sigma_i - 1) \log(P_{it}) + \log(\eta_i) + \log(\alpha_t^i(k)) \end{aligned}$$

where we define  $\log(\alpha_t^i(k)) = \varepsilon_t^i(k)$ . If the time index  $t$  is temporarily dropped and a monopolistically competitive firm producing good  $(i, k)$  is considered, then with some abuse of notation, we can rewrite the above sales equation as:

$$q = -\sigma_i p + s + \mu_i + \varepsilon_k \quad (9)$$

A firm  $(i, k)$  does not know to which group it belongs to, that is, it does not know whether its type satisfies  $\sigma_i = \sigma_1$  or  $\sigma_i = \sigma_2$ . Furthermore, it does not observe the realization of the (log) taste shock  $\varepsilon_k$ . From the firm's point of view, there is no longer a one-to-one mapping between quantities and prices. Whenever a firm sets a price  $p$ , demand  $q$  can be high for three reasons: (1) the variety belongs to a basket within which substitution is hard (and therefore the market power is high), that is,  $\sigma_i = \sigma_1$ , (2) the realization of the taste shock  $\varepsilon_k$  is high, or (3) the basket to which the firm belongs to only has a few competitors (i.e., low  $\varphi_i$ ) so  $\mu_i$  is

high and the consumer spends much of his or her income on goods in basket  $i$ . For example, if a firm knows its elasticity of substitution  $\sigma_i$  but does not know the realization of the taste shock  $\varepsilon_k$  and hence takes expectations over it, then the optimal set price is characterized by  $p^* = \frac{\sigma_i}{\sigma_i-1}MC$  where  $MC$  denotes the firm's marginal cost. Thus, we immediately can deduce  $P_i = \left(\frac{\sigma_i}{\sigma_i-1}MC\right) \varphi_i^{1/1-\sigma_i}$  and  $\log(P_i) = \frac{1}{1-\sigma_i}\log(\varphi_i) + \log\left(\frac{\sigma_i}{\sigma_i-1}MC\right)$ . Since  $\sigma_i > 1$ ,  $\log(P_i)$  strictly decreases with  $\varphi_i$ .

A firm does observe the amount of sold quantities  $q$  of its product. It can use this information to learn about its elasticity of demand and update its type. As a result, a firm might want to deviate from the static profit maximizing price to learn more about the price elasticity of its corresponding basket. Our quantitative results rely on CES preferences that most of the price-setting models use. However, our results on firm-level learning do not rely on these specific type of preferences. The results will hold as long as the demand function is linearly separable after some uniform transformation.

Taste shocks are specific to each variety, but these are unobserved by the firm. Furthermore, a firm is unaware of its type  $i$  but uses observed sales as a informative signal to learn about its type. As a result, its pricing policy is independent of  $(i, k)$  and we can drop this index without loss of generality for determining the optimal pricing strategy. Our setup imposes the following timing on the firm's pricing decisions and the consumer's realized demand shock for each period.

1. A firm decides on its price  $p$  before the realization of the demand shock  $\varepsilon_k$ .
2. The shock  $\varepsilon_k$  is realized and households decide to consume  $\exp(q) = \exp(-\sigma_i p + s + \mu_i + \varepsilon_k)$  conditional on the set price  $p$ .
3. The firm is contractually obliged to supply  $\exp(q)$ .

Let  $\lambda$  denote the firm's prior belief in a low elasticity of demand (i.e.  $\sigma_i = \sigma_1$ ) and  $f$  the probability distribution function of  $\varepsilon_k$ . Whenever a firm observes log sales  $q$  and aggregate income  $s$ , and sets some price  $p$ , it can update its prior beliefs to the posteriors  $\lambda'$  according to Bayes' rule:

$$\begin{aligned} \lambda' &= B(\lambda, p, q, s) \\ &= \frac{\lambda f(q + \sigma_1 p - \mu_1 - s)}{\lambda f(q + \sigma_1 p - \mu_1 - s) + (1 - \lambda) f(q + \sigma_2 p - \mu_2 - s)} \\ &= \left[ 1 + \frac{1 - \lambda}{\lambda} \frac{f(q + \sigma_2 p - \mu_2 - s)}{f(q + \sigma_1 p - \mu_1 - s)} \right]^{-1} \end{aligned}$$

However, dynamic decision-making requires knowing the evolution of beliefs conditional on a given state  $i$ , where the true data generating process for sold quantities is  $q = -\sigma_i p + s + \mu_i + \varepsilon_k$ .

The firm rationally anticipates that the price it sets will affect the informative quantity it will observe the period after. Prices in period  $t$  affect a firm's future beliefs in period  $t + 1$  conditional on the true state being  $i$ . As a result, the firm's motives are not solely rooted in the maximization of its static profits because a firm's pricing strategy can increase the value of its sales' informativeness. The posterior belief, conditional on state  $i$ , is equal to:

$$b_i(\lambda, p, \varepsilon) = B(\lambda, p, -\sigma_i p + s + \mu_i + \varepsilon, s).$$

In the remainder of this paper, we assume that log demand shocks are normally distributed. Thus, we get  $\varepsilon_k \sim N(0, \sigma_\varepsilon^2)$ . After some algebra, we can derive that:

$$b_i(\lambda, p, \varepsilon) = \left( 1 + \frac{1 - \lambda}{\lambda} \exp \left( \frac{1}{2} (-1)^{1(i=2)} \left[ \left( \frac{\varepsilon}{\sigma_\varepsilon} \right)^2 - \left( p \cdot \frac{\Delta\sigma}{\sigma_\varepsilon} - \frac{\Delta\mu}{\sigma_\varepsilon} + \frac{\varepsilon}{\sigma_\varepsilon} \right)^2 \right] \right) \right)^{-1} \quad (10)$$

where  $\Delta\sigma \equiv \sigma_2 - \sigma_1 > 0$ . Expression 10 shows that the speed of learning or the rate at which posterior beliefs change is heavily influenced by the firm's pricing decision. In fact, the expression indicates that posterior beliefs are more responsive to prices whenever the signal-to-noise ratio  $\Delta\sigma/\sigma_\varepsilon$  is high. Further, if the firm sets a price equal  $\Delta\mu/\Delta\sigma$ , then the posterior beliefs do not change regardless of the realization of the taste shock. As a result, at this price, beliefs are self-reinforcing.

This model of active learning with discrete types is parsimonious and is a sufficient ingredient for generating the empirical findings of section 2.4. Alternatively, we could work with a continuum of types such that a firm's beliefs are captured through a probability distribution. In appendix B.1, we show that it is unlikely that our quantitative results will be affected under this model because it shares many of the same features as the model with discrete types.

However, there are additional advantages to using discrete types. First, it simplifies the computational procedure significantly. The simplest case of active learning with a continuum of types that is still tractable features uncertainty about its demand elasticity only and Gaussian conjugates its priors. Even under this case, a firm's beliefs will consist of at least a pair, that is, a mean and a variance, which is more than the single state for the prior belief in our setup. As a result, our computational procedure suffers much less from the curse of dimensionality. Second, Kiefer and Nyarko (1989) show that under a continuum of types there are multiple limit beliefs that are outcomes of optimal policy but that do not coincide with the true parameter values. Under discrete types we show that only one limit belief exists that does not converge to the truth.

### 3.2.1 Two Period Model with Active Learning

To display the active learning mechanism as clearly as possible, we use a model with only two periods. If the firm is type  $i$ , then its profits are conditional on setting a price  $p$  equal to  $\Pi^i(p)$ . We impose that  $\Pi^i(\cdot)$  is strictly concave, which is standard.

By construction, the firm only cares about maximizing myopic profits in the second period. These myopic profits are a linear combination of the concave functions  $\Pi^1(p)$  and  $\Pi^2(p)$  based on the prior belief  $\lambda$ , that is,  $M(p; \lambda) = \lambda\Pi^1(p) + (1 - \lambda)\Pi^2(p)$ . Thus, we must have:

$$V_2(\lambda) \equiv \max_{p \in \mathcal{P}} M(p; \lambda)$$

Therefore, its maximizer  $p^M(\lambda)$  is unique and monotonically increasing in the belief  $\lambda$ .<sup>28</sup> In the first period however, the firm must balance its incentives between obtaining higher myopic profits and sharpening its posterior beliefs to increase its continuation value.

$$V_1(\lambda_0) = \max_{p \in \mathcal{P}} \left\{ M(p; \lambda_0) + \beta \mathbb{E}_\varepsilon \left[ \lambda_0 V_2(b_1(\lambda_0, \log(p), \varepsilon)) + (1 - \lambda_0) V_2(b_2(\lambda_0, \log(p), \varepsilon)) \right] \right\}$$

where  $b_i(\lambda_0, \log(p), \varepsilon)$  is defined as in (10) and we denote its maximizer as  $p^*(\lambda_0)$ . Further, the policy  $p^M(\lambda)$  is the optimal pricing function whenever the firm is unable to affect its posterior beliefs. As a result, a firm actively learns with its price at the belief  $\lambda_0$  if it deviates from this myopic price function. This deviation  $|p^*(\lambda_0) - p^M(\lambda_0)|$  reflects the firm's incentive to gain information to increase the speed of learning at the expense of its current period profits.

*Convexity of the value of information.* A relatively large literature has established that the firm's active learning is formally captured by a continuation value that is convex in a firm's beliefs.<sup>29</sup> The following lemma establishes this feature.

**LEMMA 1.** The value function  $V_2(\cdot)$  is convex and  $\mathcal{C}^2$ .

*Proof.* See appendix A1.1. □

The result is shown explicitly for the two period setup. However, it is generalizable to the infinite period framework. The convexity of  $V_2(\cdot)$  is important because, to establish sufficient conditions for active learning, we follow [Mirman, Samuelson and Urbano \(1993\)](#). In their

<sup>28</sup>A formal argument can be found in proposition 2 in the appendix.

<sup>29</sup>For example, this argument can be found in [Aghion, Bolton, Harris and Jullien \(1991\)](#) and [Mirman, Samuelson and Urbano \(1993\)](#).

proposition 1, the convexity of  $V_2(\cdot)$  is one of their two sufficient conditions. Informally, the second condition states that adjustments in prices must be capable of increasing the informativeness of a firm’s sales.

*Incentives for active learning.* A firm sets its price to identify its elasticity of substitution by “separating” these two possible demand curves as far apart as possible. This separation indicates that the price at which the demand curves cross in expectation results in sales that are completely uninformative. Thus, we deduce that the expected demand curves cross if and only if  $p = \frac{\Delta\mu}{\Delta\sigma}$ . This intersecting price can be defined as the confounding price  $\hat{p}$ . If the firm decides to choose its active learning policy  $p^*(\lambda)$  to be equal to  $\hat{p}$ , then there should be no benefits from active learning. This reasoning is formalized in the following proposition:

**PROPOSITION 1.** Let  $\hat{p} = \frac{\Delta\mu}{\Delta\sigma} \in (p_2^*, p_1^*)$ , then there exists a confounding belief  $\hat{\lambda}$  such that either one of the following two cases hold.

- I.  $p^*(\hat{\lambda}) = p^M(\hat{\lambda}) = \hat{p}$ .
- II.  $p^*(\lambda)$  is discontinuous at  $\lambda = \hat{\lambda}$ .

Furthermore, the confounding belief  $\hat{\lambda}$  is unique up to  $\lambda \in \{0, 1\}$  and strictly increasing (decreasing) in  $\Delta\mu$  ( $\Delta\sigma$ ).

*Proof.* See appendix A.2.2. □

*Numerical example.* In this example, we parameterize the profit function using CES demand curves with elasticities  $\sigma_1$  and  $\sigma_2$  and, for simplicity, we set  $\lambda_0 = \hat{\lambda}$ . Figure 5 plots the firm’s static profits, the continuation value, and the total payoff (which is the sum of the latter two) as a function of the firm’s set price  $p$ . The dotted lines at the extremes of the figure depict the optimal prices  $p_2^*$  and  $p_1^*$  under perfect information. Given that  $M(p; \lambda_0)$  is a weighted sum of strictly concave functions in  $p$ , it is strictly concave in  $p$  itself. By definition, it is maximized at  $p^M(\lambda_0)$ . The concavity of  $M(p; \lambda_0)$  illustrates the costs of active learning as prices far away from  $p^M(\lambda_0)$  represent profit losses in the first period.

Figure 5 shows that  $\mathcal{V}(\cdot; \lambda_0)$  is convex. It also shows that its minimum lies at the confounding price  $\hat{p}$ . The reason is that a firm’s sales become completely informative at the confounding price. In this case, small deviations from the confounding price lead to large gains. Thus, the benefits from active learning are strongly related to the convexity of  $\mathcal{V}(\cdot; \lambda_0)$ . For example, prior beliefs closer to zero and one lead to less convex continuation values. The reason is because the marginal benefit of information decreases for firms that are more certain about their type. The convexity of  $\mathcal{V}(\cdot; \lambda_0)$  is also affected by the signal-to-noise ratio. For



extremely large values of  $\sigma_\varepsilon$ , for instance, the optimal policy converges to the myopic policy. This is because there is no amount of variation in its price that the firm can use to induce an informative signal. As a result, the firm behaves as if its price does not affect its posterior beliefs, which is equivalent to behaving myopically.

A firm bases its pricing strategy by maximizing its total payoff, which is the sum of strictly concave and convex functions. In this example, the total payoff is double-peaked and its global maximum is at  $p^*(\lambda_0)$ .<sup>30</sup> The figure shows that the global maximum lies in the interior of  $\mathcal{P} = [p_2^*, p_1^*]$  and, most importantly, the optimal pricing strategy deviates from its myopic counterpart.<sup>31</sup> As mentioned above, the degree of active learning is captured by the difference between  $p^*(\lambda_0)$  and  $p^M(\lambda_0)$ .<sup>32</sup> In appendix B.1, we show that the active learning mechanics with a continuum of types are identical in a two-period setting. Just as in this section, a firm faces the “current control-estimation” trade-off by maximizing a total payoff that consists of a strictly concave myopic profit function and a convex continuation value.

*Active learning regimes.* The gains from active learning are strongly related to the convexity of  $\mathcal{V}(\cdot; \lambda_0)$ . This, in turn, is determined by the prior belief, the signal-to-noise ratio, and the discount factor. A firm’s prior belief determines how certain it is about its type. A firm has less incentive to engage in active learning as its belief moves closer to zero or one. The signal-to-noise ratio summarizes the sensitivity of a firm’s posterior beliefs to price deviations relative to the confounding price. Thus, firms that face extremely large levels of noise will basically never receive an informative signal through their sales. As a result, they have no incentives to actively learn. Lastly, the discount factor indicates how much a firm values more information in future periods. The convexity of  $\mathcal{V}(\cdot; \lambda_0)$  determines the shape of the total payoff function. In our previous numerical example, the total payoff function was double-peaked because  $\mathcal{V}(\cdot; \lambda_0)$  was sufficiently convex but this might not always be the case.

The shape of the total payoff function determines the active learning regime.<sup>33</sup> In our setup, there are two qualitatively different regimes determined by the shape of  $\mathcal{V}(p; \lambda_0)$ : extreme and moderate active learning. Under extreme active learning, the total payoff function

---

<sup>30</sup>In general, the sum of concave and convex functions can have multiple peaks, however the results of our baseline framework always have either a single or a double-peaked continuation value.

<sup>31</sup>In proposition 3 of the appendix, we derive a set of sufficient conditions to guarantee that  $p^*(\lambda_0) \in [p_2^*, p_1^*]$  for all  $\lambda_0$ .

<sup>32</sup>In our example, total expenditure  $S$  is constant across the two periods. Suppose that  $S_1 \neq S_2$ , then note that the incentives to engage in active learning increase if the firm expects demand to increase in the second period since the instantaneous profits are proportional to aggregate demand. In particular, if  $\frac{S_1}{P_1} < \frac{S_2}{P_2}$ , then the cost of active learning is lower in the first period. In the data, the demand for a new product increases the first few periods after entry which increases the incentives to actively learn early on the life cycle. More than 80% of the variation in the demand for a product can be explained by the extensive margin of demand.

<sup>33</sup>The terminology is borrowed from Keller and Rady (1999) who show that different active learning regimes could arise in a problem of a seller choosing quantities subject to a randomly changing state.

is double-peaked. As a result, the firm never chooses to price at the confounding price, and  $p^*(\lambda)$  displays a discontinuity at  $\lambda = \hat{\lambda}$ . Since the value of information is minimized at the confounding belief  $\hat{\lambda}$ , the firm has the most incentive to change its price at this specific belief and deviates in a discontinuous fashion. But, under moderate active learning, the total payoff function is single-peaked and the policy function  $p^*(\cdot)$  is continuous between  $p_2^*$  and  $p_1^*$ .

Figure 6 depicts the two active learning regimes. The thin gray line shows the myopic policy function  $p^M(\lambda)$  that is monotonically increasing in  $\lambda$  whereas the purple line is the policy function  $p^*(\lambda)$  under active learning. The figure shows that  $p^M(\lambda)$  and  $p^*(\lambda)$  are bounded from below and above by  $p_2^*$  and  $p_1^*$ , respectively, which proposition 3 in the appendix predicts. Under extreme active learning, the policy function shows a discontinuity at the confounding belief. The firm actively learns mostly near  $\hat{\lambda}$  as it tries to keep the informativeness of its observed sales as high as possible. It can only do this to a limited extent as otherwise the firm would lose too many static profits. With moderate active learning, the myopic policy coincides with the active learning policy at the confounding price  $\hat{p}$  as predicted by proposition 1. Once the firm updates its posterior closer to the boundaries (i.e.,  $\lambda \in \{0, 1\}$ ), the incentives for active learning decline again as the firm's information set converges to the complete information case. In this case, the myopic and active learning policies coincide at  $\lambda \in \{0, 1\}$ . Hence, the firm would never pay the opportunity costs (i.e., give up static profits) through active learning whenever its beliefs reach either zero or one.<sup>34</sup>

### 3.2.2 Dynamic Pricing Policies under Incomplete Information

Firms are ex ante identical but can generate heterogeneous ex-post pricing paths as different realizations of the log demand shocks induce differently updated posterior beliefs. A firm has access to a linear production technology in labor. Its production function is given by:

$$y_t^i(k) = z_t^i(k)\ell_t^i(k)$$

where  $y_t^i(k)$  denotes the output of some firm  $(i, k)$  in period  $t$ . Similarly,  $\ell_t^i(k)$  is the quantity of labor the firm uses for production purposes in period  $t$ . Its idiosyncratic productivity is given by  $z_t^i(k)$ . Labor is supplied competitively at the nominal rate  $W_t$ , then a firm's static

---

<sup>34</sup>In the two-period model with menu costs, the firm must decide to either adjust its price or maintain it at the same level. Under perfect information, the firm follows a standard  $(s, S)$  policy and the region of inaction depends on the curvature of the profit functions and the menu cost. But, under demand uncertainty, the width of the inaction band also depends on the firm's prior belief and it is larger close to the confounding belief because the variance in the belief changes is higher. This variance induces a high option value of waiting that is reflected in the larger width of inaction (figure A.XVII). This inaction, in turn, reduces the adjustment frequency. On the other hand, higher uncertainty pushes the firm to adjust for a given region of inaction.

profits, conditional on being type  $i$ , are equal to:

$$\left( p_t^i(k) - \frac{W_t}{z_t^i(k)} \right) \alpha_t^i(k) \left( \frac{p_t^i(k)}{P_{it}} \right)^{-\sigma_i} \frac{\eta_i S_t}{P_{it}}$$

We assume that log productivity follows a mean-reverting process:

$$\log(z_{t+1}^i(k)) = \rho \cdot \log(z_t^i(k)) + \sigma_\zeta \zeta_{t+1}^i(k) \text{ where } \zeta_{t+1}^i(k) \sim N(0, 1)$$

A firm chooses a price to determine the trade-off between maximizing current profits and obtaining more accurate information in the future about its elasticity of demand. Since a firm cannot observe the realization of its demand shock  $\alpha_t^i(k)$  whenever it has to decide on its pricing policy, it has to take expectations over it. Due to our normalization  $\mathbb{E}(\alpha_t^i(k)) = 1$ , we obtain a firm's ex-interim expected profits conditional on type  $i$ , setting some price  $p$ , and having idiosyncratic productivity  $z$ :

$$\begin{aligned} \Pi_t^i(p; z) &= \mathbb{E} \left[ \left( p - \frac{W_t}{z} \right) \alpha_t^i(k) \left( \frac{p}{P_{it}} \right)^{-\sigma_i} \frac{\eta_i S_t}{P_{it}} \right] \\ &= \left( p - \frac{W_t}{z} \right) \left( \frac{p}{P_{it}} \right)^{-\sigma_i} \frac{\eta_i S_t}{P_{it}} \end{aligned}$$

A firm does not know its elasticity of demand, so it takes expectations over these ex-interim profits using their current prior belief  $\lambda_t$ . Furthermore, firms are required to pay a fixed cost of  $\psi$  in units of labor to adjust their nominal price. This results in a firm's (ex-ante) expected profits. Thus, we define:

$$\Pi_t(p; z) \equiv \lambda_t \Pi_t^1(p; z) + (1 - \lambda_t) \Pi_t^2(p; z) - \psi W_t \cdot \mathbf{1}(p \neq p_{t-1}^i(k))$$

where  $\mathbf{1}(A)$  is an indicator function equal to unity whenever the statement  $A$  holds true. Given these constraints, a firm chooses a path of prices  $\{p_t^i(k)\}_{t \geq 0}$  to maximize the expected, discounted profits:

$$\mathbb{E}_0 \sum_{t=0}^{\infty} q_{t,t+1} \Pi_t(p_t^i(k); z_t^i(k))$$

where the expectation is with respect to the path of future beliefs, demand, and productivity shocks. Any firm makes its pricing decisions while taking aggregate prices, spending, and the wage rate as given. These variables are determined in general equilibrium and are summarized

by the aggregate state  $\xi_t \equiv (P_{1t}, P_{2t}, W_t, S_t) \in \Xi$ . In the following, we focus on a stationary equilibrium in which nominal aggregate spending trends are at a constant rate  $\tilde{\pi} \geq 0$ :

$$\log(S_{t+1}) = \log(S_t) + \tilde{\pi}$$

Thus, there is no aggregate uncertainty and the state  $\xi_t$  is constant in this stationary equilibrium. Profits are then discounted at the rate  $\beta$ .

Our framework is a hybrid version of standard state-dependent pricing models and frameworks that feature active learning, which in our setup only adds a state variable. Firms start out with a prior  $\lambda_0$  and an initial productivity draw. They then choose their entry price optimally without paying a menu cost. In our model firms have substantial incentives to learn their type at the beginning of their life cycle by adjusting their prices to obtain information. As the product matures the gains to obtaining additional information are extremely small and they do not offset the menu cost. Given that the frequency and absolute size of price changes at these stages are non-negligible, we capture the incentives for price changes at the later stages of a product's life cycle through standard state-contingent channels: idiosyncratic cost shocks and allowing for positive inflation levels.

Thus, a firm's dynamic programming problem is summarized by the following Bellman equation:

$$V(\lambda, z, p_{-1}) = \max \{V^A(\lambda, z), V^N(\lambda, z, p_{-1})\}$$

where the value of adjusting and not adjusting are respectively given by:

$$\begin{aligned} V^A(\lambda, z) &= \max_{p \geq 0} \left( p - \frac{W}{z} \right) \left[ \lambda \eta \frac{p^{-\sigma_1}}{P_1^{1-\sigma_1}} + (1-\lambda)(1-\eta) \frac{p^{-\sigma_2}}{P_2^{1-\sigma_2}} \right] \frac{S}{P} - \psi \frac{W}{P} \\ &\quad + \beta \lambda \int_{z'} \int_{\varepsilon} V \left( b_1(\lambda, \log(\frac{p}{1+\tilde{\pi}}), \varepsilon), z', \frac{p}{1+\tilde{\pi}} \right) dF(\varepsilon) dG(z', z) \\ &\quad + \beta(1-\lambda) \int_{z'} \int_{\varepsilon} V \left( b_2(\lambda, \log(\frac{p}{1+\tilde{\pi}}), \varepsilon), z', \frac{p}{1+\tilde{\pi}} \right) dF(\varepsilon) dG(z', z) \\ V^{NA}(\lambda, z, p_{-1}) &= \left( p_{-1} - \frac{W}{z} \right) \left[ \lambda \eta \frac{p_{-1}^{-\sigma_1}}{P_1^{1-\sigma_1}} + (1-\lambda)(1-\eta) \frac{p_{-1}^{-\sigma_2}}{P_2^{1-\sigma_2}} \right] \frac{S}{P} \\ &\quad + \beta \lambda \int_{z'} \int_{\varepsilon} V \left( b_1(\lambda, \log(\frac{p_{-1}}{1+\tilde{\pi}}), \varepsilon), z', \frac{p_{-1}}{1+\tilde{\pi}} \right) dF(\varepsilon) dG(z', z) \\ &\quad + \beta(1-\lambda) \int_{z'} \int_{\varepsilon} V \left( b_2(\lambda, \log(\frac{p_{-1}}{1+\tilde{\pi}}), \varepsilon), z', \frac{p_{-1}}{1+\tilde{\pi}} \right) dF(\varepsilon) dG(z', z) \end{aligned}$$

We define the optimal pricing policy  $p^*(\lambda, z)$  as the maximizer associated with the value function  $V^A(\lambda, z)$ . In a menu cost model without active learning, a price-setting firm only considers its static profits and its effect on the continuation value through the fact that changing prices is costly.<sup>35</sup> However, sales are observable and informative. Thus, a firm can also affect its posterior beliefs through its price. This is highlighted by the posterior belief functions  $b_1$  and  $b_2$  in the firm's continuation value. As a result, the policy function  $p^*(\lambda, z)$  reflects the optimal deviation from the myopic policy function that summarizes the balance between sacrificing static profits and increasing the rate at which it learns about its type.<sup>36</sup>

### 3.3 Stationary Equilibrium

#### 3.3.1 Aggregate Price Consistency

We assume that every firm starts out with the prior  $\lambda_0 \in (0, 1)$  in the beginning of its life cycle, thus firms are ex-ante homogeneous in this dimension. However, different realizations of the idiosyncratic taste shocks lead to ex-post heterogeneity of a firm's prior belief  $\lambda$  in the cross-section. Furthermore, firms also become heterogeneous due to different realizations of the idiosyncratic productivity shock in the cross-section. Note there is not only cross-sectional dispersion in beliefs across firms of different types but also within groups. This dispersion in firms' beliefs and their idiosyncratic productivity is captured by the cross-sectional distribution  $\varphi_i(\lambda, z)$  for firms of type  $i$ . We previously defined the aggregate price index as:

$$P_i = \left( \int_{k \in J_i} p^i(k)^{1-\sigma_i} dk \right)^{\frac{1}{1-\sigma_i}}$$

However, the optimal pricing policy  $p^*(\lambda, z)$  is independent from  $i$  and  $k$ . To obtain price consistency in the aggregate, we thus require:

$$P_i = \left( \int p^*(\lambda, z)^{1-\sigma_i} d\varphi_i(\lambda, z) \right)^{\frac{1}{1-\sigma_i}} \quad (11)$$

---

<sup>35</sup>This class of frameworks include standard price-setting models such as Barro (1972), Dixit (1991), Golosov and Lucas (2007) and Alvarez and Lippi (2014) for example.

<sup>36</sup>Note that our framework is fundamentally different from most price-setting models with learning. In the framework by Baley and Blanco (2017), a firm is faced with uncertainty about its productivity. As a result, the problem can be formulated as a Kalman-Bucy filtering problem. Information however evolves exogenously: in their baseline case, these flows are driven by Brownian motions and a Poisson shock. In contrast, our model considers firms that can explicitly affect their set of information. As a result, the flow of information becomes an endogenous object.

### 3.3.2 Labor Market Clearing

The market clearing condition for goods is explicitly incorporated in the firm's problem, thus the only remaining factor market to clear is the labor market. In the remainder of our analysis, we use the log utility in consumption (i.e.,  $\theta = 1$ ) and an inverse Frisch elasticity of zero (i.e.,  $\chi = 0$ ). This restriction with separable, additive utility in consumption and leisure means that wages are proportional to aggregate spending.<sup>37</sup> This gives us:

$$W = \chi S$$

Total nominal spending  $S$  equals  $PC$ , where  $P = P_1^\eta P_2^{1-\eta}$ , and thus gives us an expression for the *real* wage rate:

$$\frac{W}{P} = \chi C$$

Given the linear production technology, labor demand is simply characterized by:

$$L^d = \sum_i \int_{k \in J_i} \frac{c^i(k)}{z^i(k)} dk$$

Labor market clearing means  $L^d = L$  where  $L$  is equal to one third in our calibration exercises.

### 3.3.3 Equilibrium

We focus on a stationary equilibrium in which any dying firm is immediately replaced by a new firm. The latter is a type 1 firm with probability  $\lambda_0$  that serves as its prior belief at entry. We simplify the analysis by normalizing the measure of firms to one and by organizing the industry composition as follows:

$$J_1 = [0, \gamma_1] \text{ and } J_2 = (\gamma_1, 1].$$

Our restrictions on entry then mean that  $\gamma_1 = \lambda_0$ , which guarantees a balanced measure of in- and out-flows at the product level. Whenever nominal total spending grows deterministically at the rate  $\tilde{\pi}$ , there is no aggregate uncertainty. If  $W$  is the economy's numéraire, then we can define a stationary equilibrium.<sup>38</sup>

---

<sup>37</sup>See Golosov and Lucas (2007) who use the same specification for consumer preferences. In their setup, this specification means that wages are proportional to the stock of money. Thus, it grows at the same rate as inflation. We have a similar proportionality rule as wages become proportional to total spending which grows at the rate of inflation  $\tilde{\pi}$ .

<sup>38</sup>In the appendix A.5, we describe the numerical algorithm to solve this framework computationally.

**Definition 1 (Stationary equilibrium)** *A stationary equilibrium is a tuple  $(W, P_1, P_2, S)$  and a pair of invariant distributions  $(\varphi_1(\lambda, z), \varphi_2(\lambda, z))$  such that the real variables are constant. Thus:*

- I. *Consumers maximize utility by consuming varieties  $c^i(k)$ ,  $k \in J_i$ ,  $i \in \{1, 2\}$ .*
- II. *Firms maximize profits by adopting  $p^*(\lambda, z)$  when adjusting prices.*
- III. *Factor markets clear.*
- IV. *Prices are consistently aggregated.*
- V. *Firms die at the rate  $\delta$  and enter the economy as a type 1 firm with probability  $\lambda_0$ .*

## 4 Propagation of Nominal Shocks

### 4.1 Calibration and Results

Because the IRI Symphony data is weekly, we set the model period at one week. As a result, the discount factor is set at  $\beta = 0.96^{1/52}$  that reflects an interest rate of around 3.8% and incorporates the exogenous exit rate of 0.4%. This rate comes directly from the IRI Symphony data at the UPC-store level. The mean yearly growth rates of nominal and real GDP equal  $g_n = 0.04$  and  $g_r = 0.02$  respectively. Since there is no long-run real growth in the model economy, we set  $\tilde{\pi} = (g_n - g_r)^{1/52} = 0.00038$  as the weekly rate of inflation. Furthermore, the standard deviation of the taste shock  $\sigma_\varepsilon$  equals 0.4, which matches the standard deviation of sold quantities conditional on no price change in the IRI Symphony, which is the value of 42 percent that is reported in [Eichenbaum, Jaimovich and Rebelo \(2011\)](#). Lastly, the disutility of labor  $\chi$  is chosen so that the aggregate employment is approximately  $\frac{1}{3}$  because we normalize the amount of time available to the consumer to unity.

The remaining parameters are calibrated to match various micro-data moments. There are seven remaining parameters: two elasticities of substitution ( $\sigma_1$  and  $\sigma_2$ ), the prior belief at entry  $\lambda_0$ , the basket division of income  $\eta$ , the fixed menu cost  $\psi$ , and the persistence and standard deviation of idiosyncratic productivity  $\rho$  and  $\sigma_\zeta$ . These parameters are calibrated jointly and are selected to hit eight moments from the data: the average frequency of adjustment, the average size of increases, the average size of decreases, the fraction of price changes that are increases, the frequency of adjustment on the second and tenth week after entry, and the absolute size of the price changes during the second and tenth week after entry.

As is standard in the class of menu cost models, the fixed cost of adjustment  $\psi$  partially governs the average frequency and size of adjustments. The extent to which active learning is more present at the beginning of a product's life cycle is determined by the amount of



information a firm has at entry summarized by the signal-to-noise ratio  $(\sigma_2 - \sigma_1)/\sigma_\varepsilon$  and the prior belief at entry  $\lambda_0$  that represents the fraction of firms facing the elasticity of substitution  $\sigma_1$ . We assume this parameter to be equal across all entering firms. As the incentives to actively learn decrease, price changes are mainly driven by idiosyncratic cost shocks. Thus, the parameters  $\rho$  and  $\sigma_\zeta$  have a relatively large impact on the pricing moments at the later stage of the product’s life cycle.

Table VI shows the model’s best parameters in terms of fitting moments, and table VII displays the resulting moments from the framework compared to the data. The productivity parameters are in line with previous estimates in the menu cost literature. The model matches the frequency of adjustment and fraction of increasing price changes closely. The specification for the menu cost  $\psi$  means that the total adjustment costs in the economy represent approximately 0.7% of steady-state weekly revenues. The cost conditional on adjustment is around 1.4%, which is in line with the estimates in Zbaracki, Ritson, Levy, Dutta and Bergen (2004). The value of  $\sigma_1$  is in the range of values typically used in the menu cost literature. The model requires a somewhat large  $\sigma_2$  to induce enough active learning. Nonetheless,  $\sigma_2$  is well within the estimates of Broda and Weinstein (2010) who compute elasticities of substitution for a variety of products using data similar to ours.

The model also performs well in replicating the life cycle patterns in the frequency of price adjustments and the absolute size of the adjustments. Figure 7 shows that in our simulations entering products are more likely to adjust prices and they do so by larger amounts. This is driven by 1) the size of the signal-to-noise ratio and 2) the overall level of the elasticities of substitution since firms with lower market power (i.e., high elasticities of substitution) have higher incentives to get their prices “right” as the opportunity costs of active learning (i.e., sacrificing static profits) are higher. The incentives to actively learn also affect the size distribution of price changes by generating large price changes endogenously. Our calibration matches the standard deviation of price changes and price changes in the 75th percentile in absolute value without explicitly targeting them. The model, however, underpredicts the prevalence of price changes in the 90th percentile of the size distribution. <sup>39</sup>

---

<sup>39</sup>The resulting hazard of price changes in our economy is downward-sloping with a small hump at short durations (figure 8). This is not obvious at first glance, and it is the result of several opposing forces. The presence of a menu cost typically results in upward-sloping hazard rates as firms are less likely to adjust after they reset their prices. On the other hand, as firms learn, the probability of consecutive price changes is larger at entry since new information might cause a new adjustment. This force, in addition to the fact that the variance of idiosyncratic shocks is large relative to the rate of inflation, contributes to the decrease in the slope of the hazard at long horizons. This is in contrast to the hazard rate in Bachmann and Moscarini (2012) that is completely flat at zero with a spike at 21 months. This is because of the presence of idiosyncratic cost shocks and due to the fact that in order to match the life cycle moments, the signal to noise ratio in our calibration is larger, which increases the incentives to learning.

## 4.2 Implications

We perform a counterfactual experiment in which the log nominal output increases permanently by a size that is comparable to a one week doubling of the nominal output growth rate. We observe that on impact approximately 70 percent of the nominal shock goes into output. In a baseline menu cost model with full information, this value is around 60 percent.<sup>40</sup>

Alvarez, Lippi and Passadore (2017) show that in a large class of continuous time, price-setting models, both state- and time-dependent, the effect of a small nominal shock is identical across these models. In discrete time, however, the response when the shock occurs conveys information about the aggregate price flexibility. To quantify this flexibility, we follow Caballero and Engel (2007) and compute the "Flexibility Index": an accounting relation that describes how inflation will respond when a small shock occurs and that is fully pinned down by the current distribution of the firms' desired price changes and the adjustment hazard.

This index is valid in all models including our model with active learning. We begin by decomposing the price response when the shock occurs into intensive and extensive margins. If  $x_t(\lambda) \equiv \ln\left(\frac{p_t}{p_t^*(\lambda)}\right)$  is the price gap, then it can be defined as the difference between a firm's current price and its desired price, that is, the price it chooses as a function of its beliefs conditional on adjustment. If the economy-wide distribution of price gaps is given by  $f(x, \lambda)$ , we assume that firms have an adjustment probability  $\Lambda(x, \lambda)$  that is increasing in their price gap.<sup>41</sup> If there is some unexpected, positive shock  $\Delta S > 0$  to firms' desired prices, the price response equals:

$$\lim_{\Delta S \rightarrow 0} \frac{\Delta \pi}{\Delta S} = \underbrace{\int \Lambda(x, \lambda) f(x, \lambda) dx d\lambda}_{= \text{intensive}} + \underbrace{\int x(\lambda) \Lambda_x(x, \lambda) f(x, \lambda) dx d\lambda}_{= \text{extensive}} \quad (12)$$

that can be seen as the sum of two components: intensive and extensive margins. The intensive margin measures the contribution to inflation of the firms' products whose prices would have adjusted without the aggregate shock taking place. These firms adjust to the aggregate shock by changing the size of their adjustment. Equation 12 shows that this margin equals the frequency of adjustment. The extensive margin captures the strength of the selection effect and measures the additional inflation contribution of firms whose decision to

---

<sup>40</sup>The full information model reflects the Golosov-Lucas benchmark with two types of firms (those facing elasticity  $\sigma_1$  or  $\sigma_2$ ), but both groups of firms are fully aware of their type. This model is then calibrated to match the same moments as the model with active learning and features the same fraction of firms of each type. Since we are interested in the effect of active learning on real output, the full information Golosov-Lucas model with two types is the appropriate benchmark.

<sup>41</sup>To simplify the math, we assume here that a positive small shock  $\Delta S$  does not affect firms' beliefs. However, our results do take this effect into consideration as we calculate the extensive and intensive margins numerically.

adjust is either triggered or canceled by the aggregate shock. This margin becomes naturally more relevant as the number of firms near the margin of adjustment increases (i.e., large  $A_x(x, \lambda)$ ) or when the difference between adjusting and not adjusting is large (i.e., large  $|x(\lambda)|$ ). Calibrated to the same frequency of price adjustment, the difference between any two models in quantifying the effect solely reflects the difference in the extensive margin.<sup>42</sup> The desire to actively learn from prices pushes firms away from the margin of adjustment; both lowering the mass of firms at the original bounds of inaction and substantially reducing the importance of the extensive margin. Even though this approach recovers only the price response to the shock, [Berger and Vavra \(2015\)](#) show that it is highly predictive of the overall price stickiness in the economy.

Furthermore, the half-life of the real response more than doubles in our framework with respect to that of the Golosov-Lucas benchmark. This is because, by introducing the product's life cycle, we introduce cross-sectional heterogeneity in the frequency of price adjustments across firms of different ages endogenously. As a result, the coefficient of variation of price spells duration is 30 to 35 percent larger than in the full information benchmark.

Actively learning firms have vastly higher frequencies of price changes. These firms will most likely adjust their price several times before firms with sharper beliefs after a nominal shock. However, all price changes after the first one made by firms actively learning have no consequence on the output because these firms have already adjusted to the shock. Given that the model is calibrated to match the average frequency of price changes, the fact that firms certain about their type have, on average, a lower frequency of price adjustment significantly delays the adjustment of the aggregate price level after a nominal shock. In other words, a higher level of cross-sectional heterogeneity in the duration of price spells reduces selection in timing after an unanticipated monetary shock as pointed out by [Alvarez, Lippi and Paciello \(2011\)](#) and [Carvalho and Schwartzman \(2015\)](#).<sup>43</sup>

Quantitatively, the cumulative effects on real output are 2.3 times larger under demand uncertainty than in the model with full information (as shown in figure 9).<sup>44</sup> In our baseline setup with firms producing only one good ( $n = 1$ ), no fat-tailed shocks and no random menu costs, our model with active learning has cumulative effects on real output that are about 2/5 of that in the Calvo framework. This magnitude is very comparable to the multi-product

---

<sup>42</sup>In the Calvo model, the extensive margin is zero as there is no selection effect.

<sup>43</sup>[Nakamura and Steinsson \(2009\)](#) illustrate this concept within the context of a simple Calvo model. In that framework, the degree of monetary-non-neutrality is convex in the frequency of price changes. If, for example, the overall frequency of price adjustment in the economy is a convex function of the frequency of price changes of firms actively learning and those certain about the elasticity they face, heterogeneity in the cross-sectional distribution of firms will amplify monetary non-neutrality.

<sup>44</sup>The area under the impulse response in the full information model is 1/6 of that in the Calvo model which is also found by [Alvarez, Bihan and Lippi \(2014\)](#).

model by [Alvarez and Lippi \(2014\)](#) for the case of  $n = 10$  or to the single product Golosov-Lucas model with random menu costs (also known as the “CalvoPlus” specification) where the fraction of free price adjustments is 80 percent ( $l = 0.8$ ). Furthermore, the cumulative real effects in our framework are only 20 percent lower than the  $n = \infty$  case in [Alvarez, Bihan and Lippi \(2014\)](#). In contrast, the cumulative effects on real output in our benchmark model with the addition of random menu costs (which are added in order to generate small price changes) are almost 1.3 times as large as in the Golosov-Lucas model with a fraction of free price adjustments of  $l = 0.8$ .<sup>45</sup>

### 4.3 Nominal Shocks in Periods of High Product Entry

Our baseline framework reflects a stationary environment in which the number of entrants is constant over time. To investigate whether cyclical changes in the extensive margin of products play an important role in the amplification of nominal shocks, we construct a dynamic version of the model. There are several reasons why focusing on the business cycle could be important. First, [Argente, Lee and Moreira \(2017\)](#) show that the entry rate of new products is highly procyclical. Second, previous contributions also show that the impact of nominal shocks on real output vary over the business cycle. For example, [Vavra \(2014\)](#) shows that monetary policy is less effective in stimulating real output during downturns. For the sake of brevity, we provide only a summary of the model in this section. A more detailed description can be found in appendix [B.4](#).

Consumers and firms are identical as in the baseline framework. However, a firm’s productivity now consists of two components: an idiosyncratic and an aggregate component denoted by  $z$  and  $Z$  respectively. Furthermore, the entry rate of products is endogenous which allows it to vary with the aggregate state of the economy. In the beginning of each period, a pool of entrants observes the aggregate state and decides if they want to become producers by paying a fixed entry cost (denoted in units of labor). Lastly, we assume that the level of aggregate productivity follows a two-state symmetric Markov chain. A boom and a bust are defined as a one standard deviation increase and decrease, respectively, from average aggregate productivity. The latter is normalized to unity and we calibrate the former by estimating a standard autoregressive process on [Fernald’s \(2014\)](#) utilization-adjusted Total Factor Productivity series. Transition probabilities are then calibrated such that the average length of a cycle is about 35 months.

Despite its simplicity, this setup shows whether nominal shocks are amplified more during booms. In our extension, periods of high aggregate productivity mean periods of high product

---

<sup>45</sup>We calibrate the fraction of free adjustments to match the fraction of small price changes defined as  $|dp_{it}| < \frac{1}{2} \text{mean}|dp|$  which is the data is approximately 40%.

entry. The calibration of this framework shows that the real output effects of a nominal shock are 15 percent larger in booms than during busts. This is because as the entry rate of products increases, the average firm gets younger and a higher proportion of firms then engage in active learning. These firms are less likely to adjust their prices after a nominal shock as their incentives to change their prices in response to idiosyncratic cost shocks are lower.

Although the findings of this section are robust to aggregate shocks of ordinary size, our findings might differ for very large aggregate shocks. In our calibration, the size distribution of price changes plays a large role in determining the degree to which shocks get propagated during booms. Further, the kurtosis of the distribution increases which in turn, weakens the selection effect. However, the possibility exists that with a sufficiently large number of entering firms, the average frequency of price adjustments increases, which can offset this effect. An extreme example of this effect is whenever all firms in the economy are replaced every period. In this case, prices are close to fully flexible and the effects on real output are small.

#### 4.4 Robustness and Extensions

Our benchmark model can capture many features of the data including standard pricing moments and those related to the product’s life cycle as section 2.4 showed. However, there might be other features of the data concerning entering products that could affect our conclusions.

A possible source of concern lies in the fact that entering products do not immediately feature high quantities of sales. In fact, it might require some time to build up sales for new products (e.g., building up a customer base). Our baseline framework does not reflect a gradual buildup of sales for entering products. Thus, we could be overestimating the importance of new products that in turn affects our results on the propagation of nominal shocks. We extend our framework in two different ways to deal with this issue. First, we allow for an exogenous, age trend in the demand shocks similar to [Foster, Haltiwanger and Syverson \(2016\)](#) to incorporate the fact that entering products’ sales grow over time after starting at a relatively low level. Appendix B.3.1 shows the details of our implementation. In this specification, younger products contribute less to aggregate output, but their incentives to actively learn are higher given the prospects of higher sales in the future. These two forces contribute in different directions when measuring the response of real output to a nominal shock, which leaves our results discussed in section 4.2 virtually unchanged.

Second, we extend the canonical price-setting model of [Golosov and Lucas \(2007\)](#) by incorporating customer retention concerns. A firm’s current level of demand depends posi-

tively on (a fraction of) the level of demand in the previous period. As a result, a firm has incentives to set low prices in the beginning of their life cycle to attract customers and build up their customer base. Whenever this base has reached sufficiently high levels, a firm starts exercising its market power by raising its price. These incentives are also known as “investing” and “harvesting”. In appendix B.3, we show that such a framework is not consistent with our stylized facts from section 2.4.

Another possible concern could be our assumption of a constant rate of exit. Younger products are more likely to exit the product market, so our assumption of age-independent exit rates could potentially bias our results on the propagation of nominal shocks. This is because the composition of products is biased toward younger products that experience a higher frequency and absolute size of price adjustment as discussed in section 2.4. In appendix B.2, we show that the product hazard function as a function of age is downward sloping in our data. However, the slope of the hazard function with respect to age is relatively small. Whenever we extend our framework by exogenously incorporating age-dependent exit rates consistent with the data, our conclusions are not affected significantly.

## 5 Conclusion

The increasing availability of micro-level data sets has allowed researchers to delve deeper into the mechanics of a firm’s dynamic pricing behavior. Recent studies have found new insights into firms’ pricing behavior along several dimensions . Although there is substantial anecdotal evidence that firms choose different pricing strategies over the life cycle of their products, the degree of price heterogeneity along this dimension and its aggregate implications have remained largely unexplored. In this paper, we aim to fill this gap by developing the salient facts on the evolution of products’ pricing moments over their life cycle and by providing a structural interpretation for them. We construct a quantitative framework in which firms that face uncertainty about their demand curves can actively learn through their pricing strategies and show that this model can rationalize standard price-setting moments and our set of stylized facts.

In this context, we develop sufficient conditions for active learning to occur and describe the different regimes that could arise in this setup. We then investigate the implications of active learning incentives for the propagation of nominal shocks. The calibration of our model can be interpreted as a hybrid between standard menu cost models and active learning models. It delivers the life cycle facts that support our observations in the data. In our model, relative to the full information benchmark, the real effects of nominal shocks are at least twice as large and persistent when measured by their cumulative effect on real output.

Our quantitative framework contains the minimal amount of ingredients to rationalize the key mechanisms and our empirical findings. Nonetheless, our model could be extended to cover more complicated mechanisms. We have briefly explored several of them, but we leave the full economic implications of these extensions for future research.



## References

- Aghion, Philippe, Patrick Bolton, Christopher Harris, and Bruno Jullien, “Optimal learning by experimentation,” *Review of Economic Studies*, 1991, 58 (4), 621–654.
- Aguiar, Mark and Erik Hurst, “Deconstructing life cycle expenditure,” *Journal of Political Economy*, 2013, 121 (3), 437–492.
- Alvarez, Fernando and Francesco Lippi, “Price setting with menu cost for multiproduct firms,” *Econometrica*, 2014, 82 (1), 89–135.
- Alvarez, Fernando E, Francesco Lippi, and Luigi Paciello, “Optimal Price Setting With Observation and Menu Costs\*,” *The Quarterly journal of economics*, 2011, 126 (4), 1909–1960.
- Alvarez, Fernando, Francesco Lippi, and Juan Passadore, “Are State-and Time-Dependent Models Really Different?,” *NBER Macroeconomics Annual*, 2017, 31 (1), 379–457.
- , Hervé Le Bihan, and Francesco Lippi, “Small and large price changes and the propagation of monetary shocks,” *NBER Working Paper No. 20155*, 2014.
- , Katarina Borovicková, and Robert Shimer, “The proportional hazard model: Estimation and testing using price change and labor market data,” *University of Chicago Mimeo*, 2015.
- Anderson, Eric, Benjamin A Malin, Emi Nakamura, Duncan Simester, and Jón Steinsson, “Informational rigidities and the stickiness of temporary sales,” *NBER Working Paper No. 19350*, 2013.
- Argente, David, Munseob Lee, and Sara Moreira, “Innovation and Product Reallocation in the Great Recession,” 2017.
- Bachmann, Rüdiger and Giuseppe Moscarini, “Business Cycles and Endogenous Uncertainty,” 2012.
- Baley, Isaac and Julio A Blanco, “Firm Uncertainty Cycles and the Propagation of Nominal Shocks,” 2017.
- Barro, Robert J, “A theory of monopolistic price adjustment,” *Review of Economic Studies*, 1972, pp. 17–26.
- Berger, David and Joseph Vavra, “Dynamics of the US price distribution,” 2015.

- Broda, Christian and David E Weinstein, “Product Creation and Destruction: Evidence and Price Implications,” *American Economic Review*, 2010, pp. 691–723.
- Bronnenberg, Bart J, Michael W Kruger, and Carl F Mela, “Database paper-The IRI marketing data set,” *Marketing Science*, 2008, *27* (4), 745–748.
- Caballero, Ricardo J and Eduardo MRA Engel, “Price stickiness in Ss models: New interpretations of old results,” *Journal of Monetary Economics*, 2007, *54*, 100–121.
- Calvo, Guillermo A, “Staggered prices in a utility-maximizing framework,” *Journal of Monetary Economics*, 1983, *12* (3), 383–398.
- Campbell, Jeffrey R and Benjamin Eden, “Rigid prices: evidence from US scanner data,” 2005.
- Caplin, Andrew S and Daniel F Spulber, “Menu Costs and the Neutrality of Money,” *The Quarterly Journal of Economics*, 1987, *102* (4), 703–725.
- Carvalho, Carlos and Felipe Schwartzman, “Selection and monetary non-neutrality in time-dependent pricing models,” *Journal of Monetary Economics*, 2015, *76*, 141–156.
- Caves, Richard E, “Industrial organization and new findings on the turnover and mobility of firms,” *Journal of Economic Literature*, 1998, *36* (4), 1947–1982.
- Chevalier, Judith A and Anil K Kashyap, “Best Prices: Price Discrimination and Consumer Substitution,” 2014.
- , —, and Peter E Rossi, “Why Don’t Prices Rise During Periods of Peak Demand? Evidence from Scanner Data,” *American Economic Review*, 2003, *93* (1), 15–37.
- Coibion, Olivier, Yuriy Gorodnichenko, and Gee Hee Hong, “The cyclicalities of sales, regular and effective prices: Business cycle and policy implications,” *American Economic Review*, 2015.
- Deaton, Angus and Christina Paxson, “Intertemporal Choice and Inequality,” *Journal of Political Economy*, 1994, *102* (3), 437–467.
- Dixit, Avinash, “Analytical approximations in models of hysteresis,” *Review of Economic Studies*, 1991, *58* (1), 141–151.
- Doraszelski, Ulrich, Gregory Lewis, and Ariel Pakes, “Just starting out: Learning and equilibrium in a new market,” 2016.

- Eichenbaum, Martin, Nir Jaimovich, and Sergio Rebelo, “Reference prices, costs, and nominal rigidities,” *American Economic Review*, 2011, *101* (1), 234–262.
- Einav, Liran, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan, “Assessing sale strategies in online markets using matched listings,” *American Economic Journal: Microeconomics*, 2015, *7* (2), 215–247.
- Fernald, John G, “A quarterly, utilization-adjusted series on total factor productivity,” in “in” Federal Reserve Bank of San Francisco 2014.
- Foster, Lucia, John Haltiwanger, and Chad Syverson, “The slow growth of new plants: Learning about demand?,” *Economica*, 2016, *83* (329), 91–129.
- Gagnon, Etienne and David Lopez-Salido, “Small price responses to large demand shocks,” 2014.
- Gaur, Vishal and Marshall L Fisher, “In-store experiments to determine the impact of price on sales,” *Production and Operations Management*, 2005, *14* (4), 377–387.
- Gilchrist, Simon, Raphael Schoenle, Jae Sim, and Egon Zakrajšek, “Inflation dynamics during the financial crisis,” *The American Economic Review*, 2017, *107* (3), 785–823.
- Golosov, Mikhail and Robert E Lucas, “Menu Costs and Phillips Curves,” *Journal of Political Economy*, 2007, *115* (2).
- Gourinchas, Pierre-Olivier and Jonathan A Parker, “Consumption over the life cycle,” *Econometrica*, 2002, *70* (1), 47–89.
- Ilut, Cosmin, Rosen Valchev, and Nicolas Vincent, “Paralyzed by Fear: Rigid and Discrete Pricing under Demand Uncertainty,” 2015.
- Karadi, Peter and Adam Reiff, “Menu Costs, Aggregate Fluctuations, and Large Shocks,” 2016.
- Keller, Godfrey and Sven Rady, “Optimal experimentation in a changing environment,” *Review of Economic Studies*, 1999, *66* (3), 475–507.
- Kiefer, Nicholas M and Yaw Nyarko, “Optimal control of an unknown linear process with learning,” *International Economic Review*, 1989, *30* (3), 571–86.
- Kiley, Michael T, “Partial adjustment and staggered price setting,” *Journal of Money, Credit, and Banking*, 2002, *34* (2), 283–298.

- Lee, Yoonsoo and T Mukoyama, “A Model of Entry, Exit, and Plant-level Dynamics Over the Business Cycle,” 2015.
- Lucas, Robert E., “Macroeconomic Priorities,” *American Economic Review*, 2003, *93*, 114.
- Midrigan, Virgiliu, “Menu costs, multiproduct firms, and aggregate fluctuations,” *Econometrica*, 2011, *79* (4), 1139–1180.
- Mirman, Leonard J, Larry Samuelson, and Amparo Urbano, “Monopoly experimentation,” *International Economic Review*, 1993, *34* (3), 549–563.
- Moreira, Sara, “Firm dynamics, persistent effects of entry conditions, and business cycles,” Technical Report, mimeo 2015.
- Nakamura, Emi and Jón Steinsson, “Five facts about prices: A reevaluation of menu cost models,” *Quarterly Journal of Economics*, 2008, *123* (4), 1415–1464.
- and Jon Steinsson, “Monetary Non-Neutrality in a Multi-Sector Menu Cost Model,” *Quarterly Journal of Economics*, 2009, *125* (3), 961–1013.
- and Jón Steinsson, “Price setting in forward-looking customer markets,” *Journal of Monetary Economics*, 2011, *58* (3), 220–233.
- Nardi, Mariacristina De, Eric French, and John B Jones, “Why Do the Elderly Save? The Role of Medical Expenses,” *Journal of Political Economy*, 2010, *118* (1), 39–75.
- Paciello, Luigi, Andrea Pozzi, and Nicholas Trachter, “Price dynamics with customer markets,” 2014.
- Prescott, Edward C, “The multi-period control problem under uncertainty,” *Econometrica*, 1972, pp. 1043–1058.
- Rothschild, Michael, “A two-armed bandit theory of market pricing,” *Journal of Economic Theory*, 1974, *9* (2), 185–202.
- Shapiro, Matthew D and Mark W Watson, “Sources of business cycle fluctuations,” *NBER Macroeconomics annual*, 1988, *3*, 111–148.
- Sheedy, Kevin D, “Intrinsic inflation persistence,” *Journal of Monetary Economics*, 2010, *57* (8), 1049–1061.
- Stroebel, Johannes and Joseph Vavra, “House prices, local demand, and retail prices,” *NBER Working Paper No. 20710*, 2014.

- Vavra, Joseph, “Inflation Dynamics and Time-Varying Volatility: New Evidence and an Ss Interpretation,” *Quarterly Journal of Economics*, 2014, *129* (1), 215–258.
- Wieland, Volker, “Learning by doing and the value of optimal experimentation,” *Journal of Economic Dynamics and Control*, 2000, *24* (4), 501–534.
- , “Monetary policy, parameter uncertainty and optimal learning,” *Journal of Monetary Economics*, 2000, *46* (1), 199–228.
- Willems, Tim, “Actively learning by pricing: a model of an experimenting seller,” *Economic Journal*, 2016.
- Zbaracki, Mark J, Mark Ritson, Daniel Levy, Shantanu Dutta, and Mark Bergen, “Managerial and customer costs of price adjustment: direct evidence from industrial markets,” *Review of Economics and Statistics*, 2004, *86* (2), 514–533.
- Zellner, Arnold, *An introduction to Bayesian inference in econometrics* 1971.

# Tables and Figures

Table I: Product Entry and Exit

|             | UPC    | UPC    | UPC×Store | UPC×Store |
|-------------|--------|--------|-----------|-----------|
|             | 5-Year | 1-Year | 5-year    | 1-year    |
| Entry       | 0.45   | 0.14   | 0.66      | 0.27      |
| Creation    | 0.29   | 0.07   | 0.47      | 0.15      |
| Exit        | 0.42   | 0.13   | 0.61      | 0.25      |
| Destruction | 0.08   | 0.01   | 0.39      | 0.10      |

Note: The table shows the statistics of the entry rate, exit rate, creation, and destruction for 1-year and 5-year intervals. Columns (1) and (2) show the statistics at the UPC level and columns (3) and (4) at the UPC × Store level.

Table II: Distribution of Duration by UPC × Store

|                             | (1)         | (2)          | (3)              | (4)          |
|-----------------------------|-------------|--------------|------------------|--------------|
|                             | Unweighted  |              | Revenue Weighted |              |
|                             | Weeks since | Observations | Weeks since      | Observations |
|                             | Entry       | since Entry  | Entry            | since Entry  |
| 1 <sup>st</sup> percentile  | 1.0         | 1.0          | 12.4             | 7.7          |
| 25 <sup>th</sup> percentile | 37.4        | 16.3         | 108.4            | 73.5         |
| 50 <sup>th</sup> percentile | 96.3        | 47.1         | 183.7            | 131.8        |
| 75 <sup>th</sup> percentile | 209.5       | 122.0        | 280.6            | 208.3        |
| 99 <sup>th</sup> percentile | 450.7       | 369.7        | 466.3            | 405.1        |
| Mean                        | 134.0       | 83.1         | 198.9            | 148.9        |
| Std. Dev.                   | 118.9       | 90.6         | 117.9            | 100.0        |

Note: The table shows the statistics of the distribution of durations of a UPC × Store pair. In columns (1) and (2) we compute the duration of each UPC × Store pair and aggregate them to the category level using equal weights. Categories are further aggregated using equal weights. In columns (3) and (4) we aggregate to the category level using revenue weights and aggregate across categories using equal weights. Weeks since entry refers to the number of weeks elapsed since the product was first observed. Observations since entry refers to the number of times a product is observed in our data set. A product is observed only if it records a transaction in a given week and store.

Table III: Life cycle Properties of Selected Pricing Moments - First 6 months

| Dependent Variable    | (1)                  | (2)                  | (3)                  | (4)                  |
|-----------------------|----------------------|----------------------|----------------------|----------------------|
|                       | Equal Weights        |                      | Revenue Weights      |                      |
| Frequency             | -0.054***<br>(0.001) | -0.076***<br>(0.002) | -0.072***<br>(0.003) | -0.104***<br>(0.004) |
| Frequency increases   | -0.051***<br>(0.001) | -0.069***<br>(0.002) | -0.066***<br>(0.002) | -0.092***<br>(0.003) |
| Frequency decreases   | -0.002***<br>(0.000) | -0.006***<br>(0.001) | -0.006***<br>(0.001) | -0.012***<br>(0.002) |
| Absolute size         | -0.154***<br>(0.002) | -0.171***<br>(0.005) | -0.154***<br>(0.004) | -0.174***<br>(0.008) |
| Size increases        | -0.194***<br>(0.003) | -0.223***<br>(0.006) | -0.185***<br>(0.005) | -0.223***<br>(0.010) |
| Size decreases        | 0.075***<br>(0.002)  | 0.078***<br>(0.006)  | 0.091***<br>(0.004)  | 0.071***<br>(0.011)  |
| UPC $\times$ Store FE | ✓                    | ✓                    | ✓                    | ✓                    |
| Time FE               |                      | ✓                    |                      | ✓                    |
| Cohort controls       |                      | ✓                    |                      | ✓                    |

Note: The table reports the coefficients (in percent) from the OLS tests. The independent variable is the age of the product and the dependent variables are the moments defined in the table. The sample is the first 6 months (or 26 weeks) after the product was first launched. The controls include UPC  $\times$  store fixed effects, time fixed effects, and cohort controls that are approximated by the local unemployment rate in the city and month the product was launched. Columns (1) and (2) report the coefficients that assume equal weights for each UPC  $\times$  store. Columns (3) and (4) report the results with revenue weights. The standard errors are clustered at the store level. The \*\*\*, \*\*, and \* denote significance at 0.01, 0.05, and 0.10 levels respectively.

Table IV: New Brand - First 6 Months

|                | Equal Weights        |                      | Revenue Weighted     |                      |
|----------------|----------------------|----------------------|----------------------|----------------------|
|                | Frequency<br>(1)     | Size<br>(2)          | Frequency<br>(3)     | Size<br>(4)          |
| Age            | -0.075***<br>(0.002) | -0.171***<br>(0.005) | -0.104***<br>(0.004) | -0.177***<br>(0.008) |
| Age×New Brand  | -0.028***<br>(0.005) | -0.047***<br>(0.014) | -0.020**<br>(0.008)  | -0.085***<br>(0.018) |
| UPC × Store FE | ✓                    | ✓                    | ✓                    | ✓                    |
| Time FE        | ✓                    | ✓                    | ✓                    | ✓                    |
| Cohort         | ✓                    | ✓                    | ✓                    | ✓                    |

Note: The table reports the estimates of equation 2. The independent variable is the age of the product interacted with an indicator that equals one if the brand and the volume of the product are new. The dependent variables are the frequency and absolute size of the price changes. The sample is the first 6 months (or 26 weeks) after the product was first launched. The controls include UPC × store fixed effects, time fixed effects, and cohort controls that are approximated by the local unemployment rate in the city and month the product was launched. Columns (1) and (2) report the coefficients that assume equal weights for each UPC × store. Columns (3) and (4) report the results with revenue weights. The standard errors are clustered at the store level. The \*\*\*, \*\*, and \* denote significance at 0.01, 0.05, and 0.10 levels respectively.

Table V: Newness Index - First 6 months

|                | Equal Weights        |                      | Revenue Weights      |                      |
|----------------|----------------------|----------------------|----------------------|----------------------|
|                | Frequency<br>(1)     | Size<br>(2)          | Frequency<br>(3)     | Size<br>(4)          |
| Age            | -0.076***<br>(0.004) | -0.173***<br>(0.006) | -0.109***<br>(0.010) | -0.177***<br>(0.012) |
| Age×Newness    | -0.053*<br>(0.031)   | -0.228***<br>(0.053) | 0.038<br>(0.058)     | -0.310***<br>(0.101) |
| UPC × Store FE | ✓                    | ✓                    | ✓                    | ✓                    |
| Time FE        | ✓                    | ✓                    | ✓                    | ✓                    |
| Cohort         | ✓                    | ✓                    | ✓                    | ✓                    |

Note: The table reports the estimates of equation 4. The independent variable is the age of the product interacted with the Newness index. The dependent variables are the frequency and absolute size of the price changes. The sample is the first 6 months (or 26 weeks) after the product was first launched. The controls include UPC × store fixed effects, time fixed effects, and cohort controls that are approximated by the local unemployment rate in the city and month the product was launched. Columns (1) and (2) report the coefficients that assume equal weights for each UPC × store. Columns (3) and (4) report the results with revenue weights. The standard errors are clustered at the store level. The \*\*\*, \*\*, and \* denote significance at 0.01, 0.05, and 0.10 levels respectively.



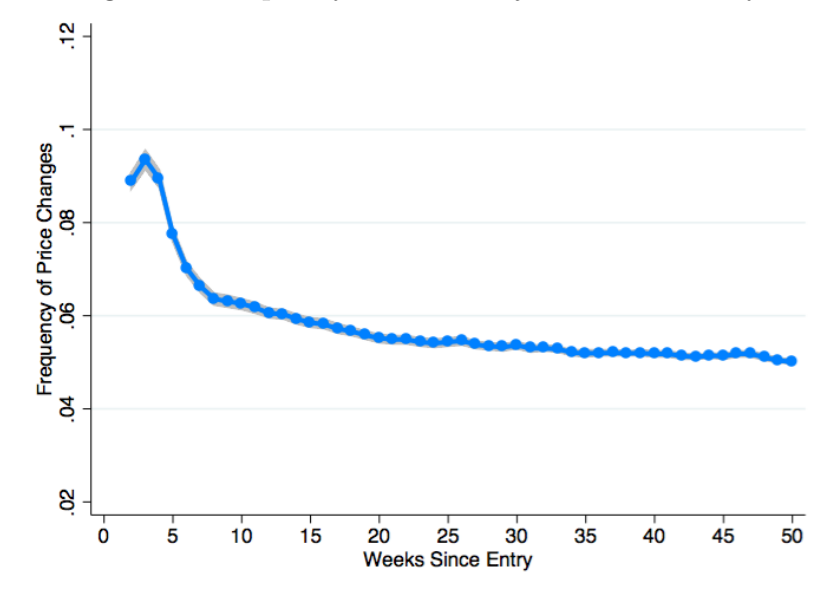
Table VI: Internally Calibrated Values of the Model's Parameters

| Description                     | Parameter      | Value |
|---------------------------------|----------------|-------|
| Elasticity of Substitution 1    | $\sigma_1$     | 4.7   |
| Elasticity of Substitution 2    | $\sigma_2$     | 13.4  |
| Prior Belief at Entry           | $\lambda_0$    | 0.75  |
| Basket Division of Income       | $\eta$         | 0.36  |
| Fixed Cost                      | $\psi$         | 0.02  |
| Productivity Persistence        | $\rho$         | 0.56  |
| Productivity Standard Deviation | $\sigma_\zeta$ | 0.05  |

Table VII: Moments of Price Change Distribution

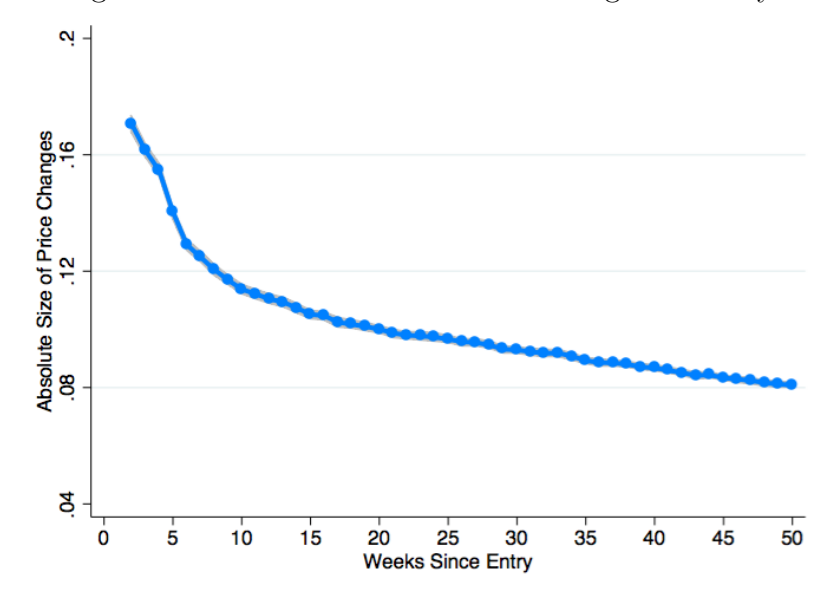
| Moment                      | Data         | Model with Learning | Full Info |
|-----------------------------|--------------|---------------------|-----------|
| Frequency Week 2            | 0.09         | 0.09                | 0.002     |
| Frequency Week 10           | 0.06         | 0.05                | 0.05      |
| Absolute Size Week 2        | 0.17         | 0.17                | 0.05      |
| Absolute Size Week 10       | 0.11         | 0.10                | 0.07      |
| Frequency                   | 0.05         | 0.05                | 0.05      |
| Fraction Up                 | 0.66         | 0.58                | 0.58      |
| Size Up                     | 0.09         | 0.09                | 0.06      |
| Size Down                   | 0.07         | 0.11                | 0.07      |
|                             | Not Targeted |                     |           |
| Std. of Price Changes       | 0.11         | 0.11                | 0.07      |
| 75th Pct Size Price Changes | 0.10         | 0.12                | 0.07      |
| 90th Pct Size Price Changes | 0.18         | 0.14                | 0.08      |

Figure 1: Frequency of Price Adjustment at Entry



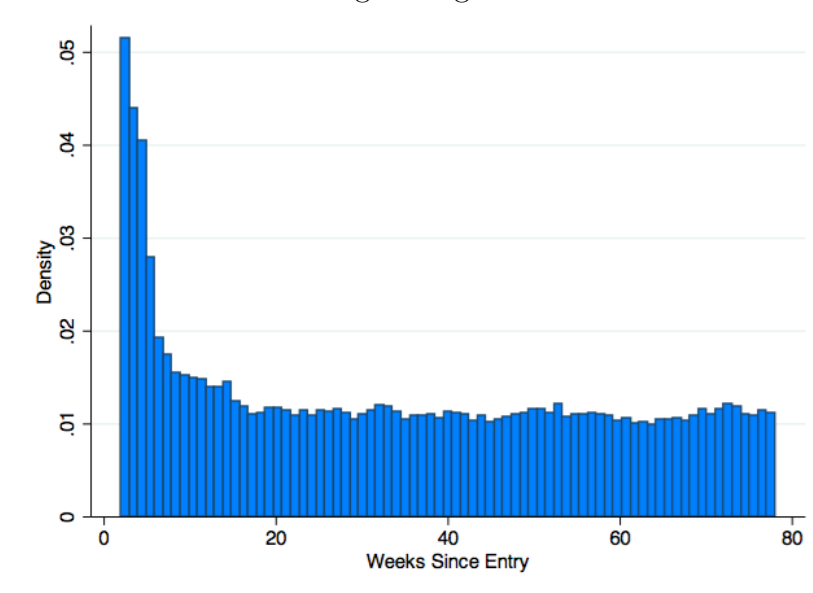
Note: The graph plots the average weekly frequency of price adjustments of entering products. The y-axis denotes the probability that the product adjusts its price in a given week and the x-axis denotes the number of weeks the product has been observed in the data after entry. The graph plots the coefficients for the age fixed effects of equation 1 where we use the regular price change indicator as the dependent variable. Equation 1 is computed by controlling for UPC-store and time fixed effects and the local unemployment rate to represent the cohort fixed effects. The calculation uses approximately 130 million observations and 2.5 million UPC  $\times$  store pairs. The standard errors are clustered at the store level. The underlying data source is the Symphony IRI.

Figure 2: Absolute Value of Price Changes at Entry



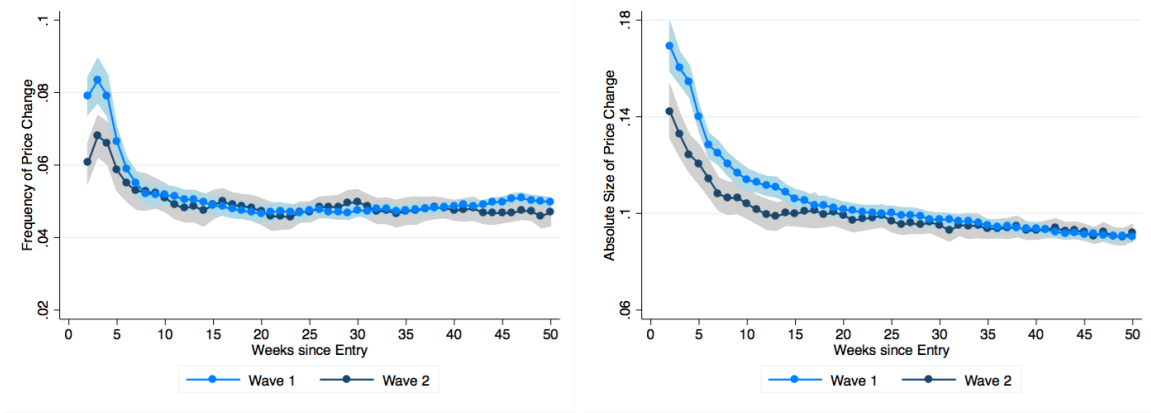
Note: The graph plots the average absolute size of price adjustments of entering products. The y-axis is the absolute value of the log price change in that week, and the x-axis denotes the number of weeks since the product entered. The graph plots the coefficients for the age fixed effects of equation 1 where we use the absolute value of the log price change as the dependent variable. Equation 1 is computed controlling for UPC-store, time fixed effects, and the local unemployment rate to represent the cohort fixed effects. The calculation uses approximately 5.8 million price changes and 2.5 million UPC  $\times$  store pairs. The standard errors are clustered at the store level. The underlying data source is the Symphony IRI.

Figure 3: Fraction of Price Changes Larger than Two Standard Deviations



Note: The figure shows the fraction of price changes larger than two standard deviations from the mean in a given category and store as a function of the age of the product. The products considered are those that last at least two years in the market. Source: IRI Symphony dataset

Figure 4: Pricing Moments by Waves

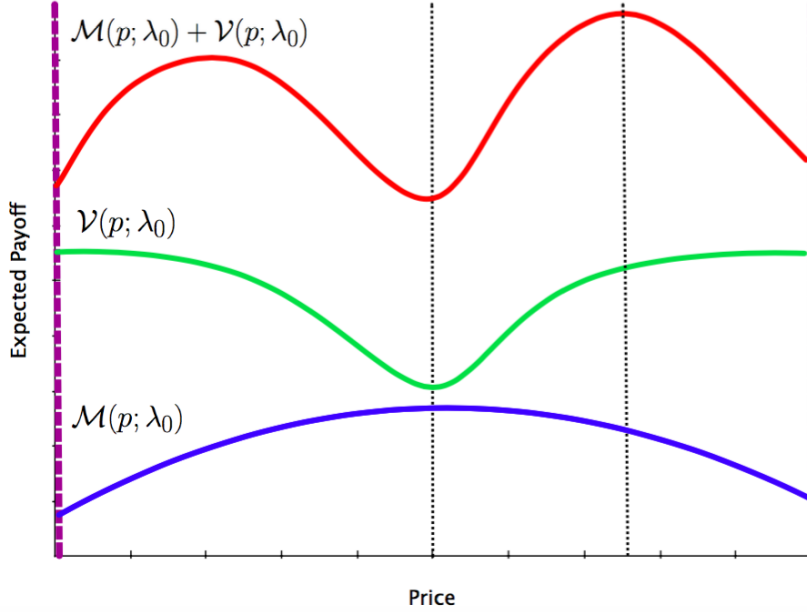


(a) Frequency of Price Changes

(b) Absolute Size of Price Changes

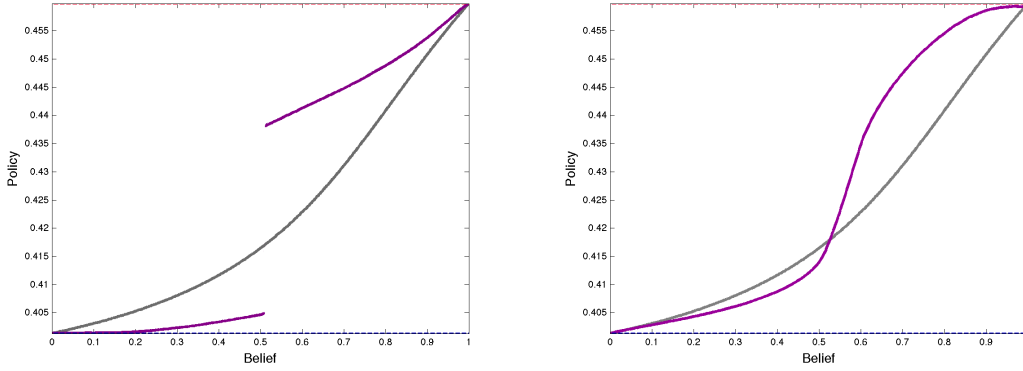
Note: Panel (a) shows the probability of adjusting prices and panel (b) shows the absolute size of the price changes by waves. Wave 1 represents products that were launched at some location during a period in the first year since the product was introduced *nationally*. Wave 2 represents the same products when launched in different stores a year after their national entries. The graphs control for stores, product, and time fixed effects.

Figure 5: Numerical Example of the Two Period Model



Note: Static profits  $\mathcal{M}(p; \lambda_0)$ , continuation value  $\mathcal{V}(p; \lambda_0)$  and total payoff  $\mathcal{M}(p; \lambda_0) + \mathcal{V}(p; \lambda_0)$  at  $\lambda_0 = \hat{\lambda}$ . The dotted purple lines represent the optimal prices  $P_2^*$  and  $P_1^*$ .  $P^M(\lambda_0)$  represents the myopic policy and  $P^*(\lambda_0)$  the policy under active learning.

Figure 6: Active Learning Regimes

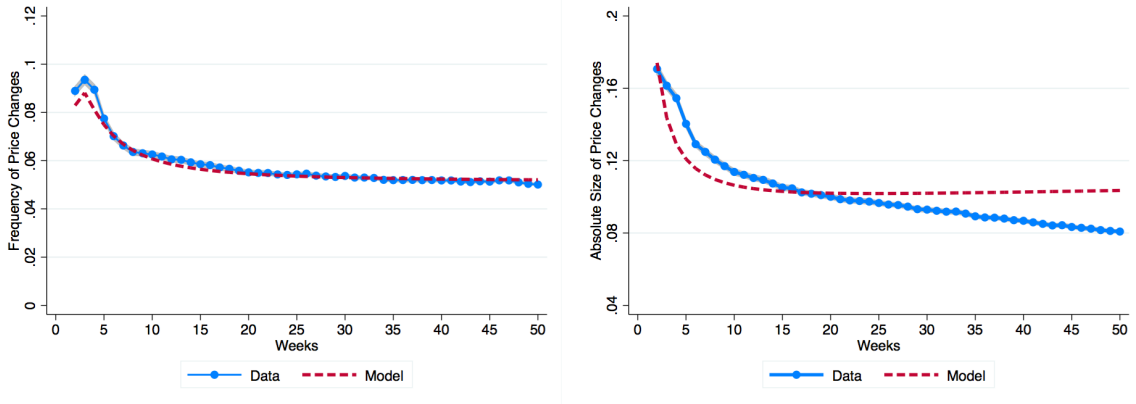


(a) Extreme Active Learning.

(b) Moderate Active Learning.

Note: Panel (a) shows the extreme active learning regime and panel (b) its moderate counterpart. The gray line depicts the myopic policy  $P^M(\lambda)$  and the purple lines the policy under active learning  $P^*(\lambda)$ . The dotted lines at the top and bottom of the panels indicate the optimal prices  $P_1^*$  and  $P_2^*$ .

Figure 7: Model vs Data



(a) Frequency of Price Changes

(b) Absolute Size of Adjustments

Note: The figure shows the results of the model and compares them with the data. We simulate a panel of 1,000 firms over 1,000 periods and compute both the predicted frequency of price adjustments and the absolute size of the price changes over the life cycle of a product. The results of the frequency of price changes are shown in panel (a) and those of the absolute size of price changes are shown in panel (b).

Figure 8: Hazard of Price Change (Model)

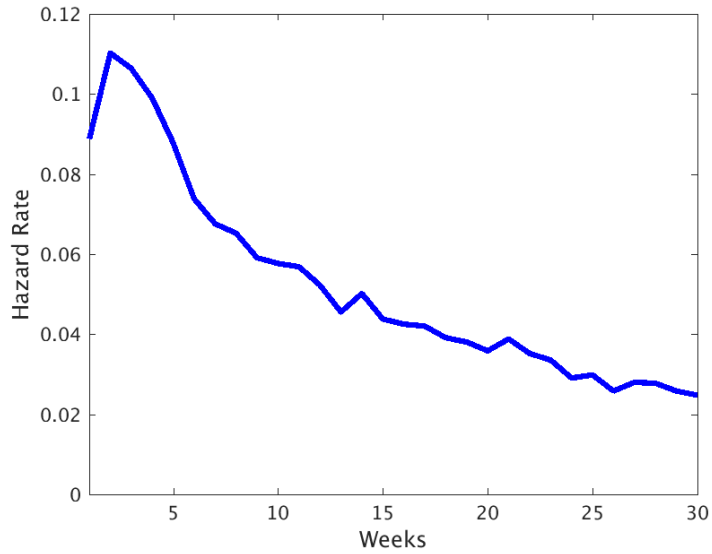
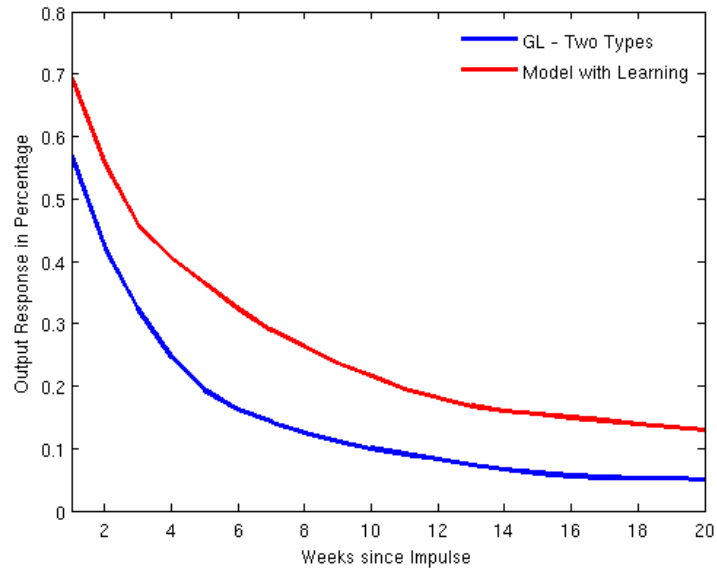


Figure 9: Real Output Response to Nominal Shock



Note: The figure shows the response of the log real output to a 0.00038 increase in the nominal output growth rate. The output response is shown in the graph as a percent of the nominal shock. The red line depicts the output response in Golosov and Lucas (2007) with two different types of firms (i.e.,  $\sigma_1$  and  $\sigma_2$ ) and the blue line is the response in a price-setting model with active learning. Both models are calibrated to match the same moments and feature the same fraction of firms of each type.

# A Appendix

## A.1 Tables and Figures

Table A.I: Life cycle Properties of Selected Pricing Moments - First Year by Category I

| Dependent Variable   | Regular Price Changes |                    | All Price Changes  |                    |
|----------------------|-----------------------|--------------------|--------------------|--------------------|
|                      | Frequency<br>(1)      | Abs. Size<br>(2)   | Frequency<br>(3)   | Abs. Size<br>(4)   |
| Beer                 | -0.070***<br>0.021    | -0.028***<br>0.010 | -0.346***<br>0.038 | -0.067***<br>0.007 |
| Blades               | 0.063***<br>0.017     | -0.089***<br>0.022 | -0.209***<br>0.027 | -0.019<br>0.012    |
| Carbonated Beverages | -0.275***<br>0.017    | -0.119***<br>0.016 | -0.115***<br>0.024 | -0.034***<br>0.012 |
| Cigarettes           | -0.003<br>0.044       | 0.003<br>0.048     | -0.328***<br>0.049 | -0.028<br>0.034    |
| Coffee               | -0.019<br>0.022       | -0.106*<br>0.056   | -0.200***<br>0.028 | -0.016<br>0.014    |
| Cold Cereal          | -0.150***<br>0.010    | -0.446***<br>0.031 | -0.322***<br>0.019 | -0.117***<br>0.015 |
| Deodorant            | -0.051***<br>0.007    | -0.208***<br>0.033 | -0.138***<br>0.017 | -0.014<br>0.012    |
| Diapers              | 0.090***<br>0.026     | -0.049***<br>0.016 | -0.161***<br>0.039 | -0.062***<br>0.014 |
| Facial Tissue        | -0.014<br>0.028       | -0.070**<br>0.029  | 0.072<br>0.049     | -0.010<br>0.002    |
| Frozen Dinner        | -0.161***<br>0.011    | -0.204***<br>0.016 | -0.160***<br>0.015 | -0.077***<br>0.010 |
| Frozen Pizzas        | -0.115***<br>0.015    | -0.168***<br>0.022 | -0.073***<br>0.024 | -0.106***<br>0.014 |
| Household Cleaners   | -0.120***<br>0.027    | -0.269***<br>0.060 | -0.082**<br>0.036  | -0.037<br>0.027    |
| Frankfurters         | -0.132***<br>0.029    | -0.535***<br>0.079 | 0.036***<br>0.054  | 0.019<br>0.039     |
| Laundry Detergent    | -0.063***<br>0.018    | -0.227***<br>0.035 | 0.054***<br>0.046  | -0.040**<br>0.017  |
| Margarine & Butter   | -0.004<br>0.021       | -0.165***<br>0.037 | -0.009<br>0.032    | 0.017***<br>0.026  |

Note: The table reports the coefficients (in percent) from OLS regressions of the 15 categories available in the Symphony IRI. The independent variable is the age of the product, and the dependent variables are the moments defined in the table. The sample is the first year (or 52 weeks) after the product was first launched. The controls include UPC  $\times$  store fixed effects, time fixed effects, and cohort controls that are approximated by the local unemployment rate in the city and month the product was launched. Columns (1) and (2) report the coefficients for regular price changes. Columns (3) and (4) report the results for all price changes (including sales). The standard errors are clustered at the store level. The \*\*\*, \*\*, and \* denote significance at 0.01, 0.05, and 0.10 levels respectively.

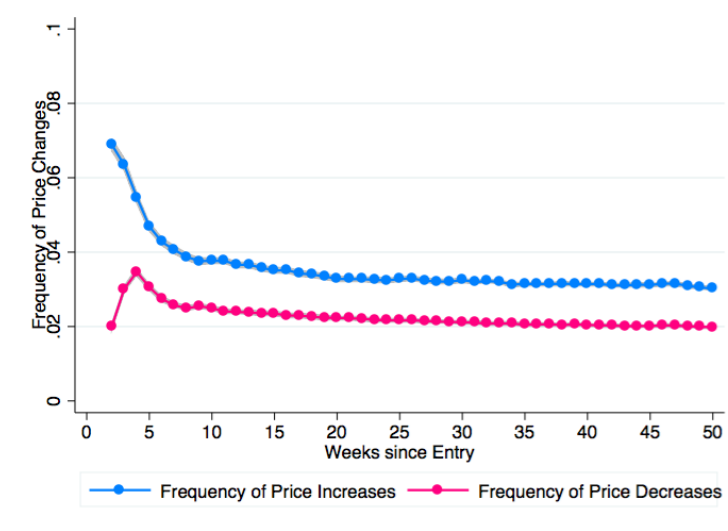
Table A.II: Life cycle Properties of Selected Pricing Moments - First Year by Category II

| Dependent Variable   | Regular Price Changes |                    | All Price Changes  |                    |
|----------------------|-----------------------|--------------------|--------------------|--------------------|
|                      | Frequency<br>(1)      | Abs. Size<br>(2)   | Frequency<br>(3)   | Abs. Size<br>(4)   |
| Mayonnaise           | -0.144***<br>0.041    | -0.086<br>0.080    | 0.677***<br>0.091  | 0.208<br>0.064     |
| Milk                 | -0.052**<br>0.021     | -0.047**<br>0.018  | -0.144***<br>0.031 | -0.027*<br>0.015   |
| Mustard & Ketchup    | -0.017<br>0.020       | -0.118**<br>0.046  | -0.011<br>0.029    | -0.050<br>0.036    |
| Paper Towels         | -0.023<br>0.028       | -0.236***<br>0.053 | -0.339***<br>0.051 | -0.402***<br>0.043 |
| Peanut Butter        | -0.073***<br>0.030    | -0.150***<br>0.029 | -0.242***<br>0.040 | -0.115***<br>0.035 |
| Photography Supplies | 0.502***<br>0.098     | -0.142<br>0.109    | -1.055***<br>0.195 | -0.233**<br>0.095  |
| Razors               | -0.294***<br>0.055    | -0.216***<br>0.072 | -0.849***<br>0.066 | 0.047<br>0.032     |
| Salty Snacks         | -0.368***<br>0.013    | -0.263***<br>0.026 | -0.062***<br>0.018 | -0.111***<br>0.013 |
| Shampoo              | -0.061***<br>0.011    | -0.135***<br>0.026 | -0.275***<br>0.026 | -0.060***<br>0.013 |
| Soup                 | -0.031***<br>0.011    | -0.268***<br>0.032 | -0.214***<br>0.031 | -0.147***<br>0.019 |
| Spaghetti Sauce      | -0.117***<br>0.017    | -0.127**<br>0.056  | -0.050<br>0.035    | 0.017<br>0.022     |
| Sugar Substitutes    | -0.089***<br>0.026    | -0.044<br>0.029    | -0.226***<br>0.041 | -0.005<br>0.043    |
| Toilet Tissue        | 0.046<br>0.034        | -0.107***<br>0.040 | -0.251***<br>0.060 | -0.180***<br>0.037 |
| Toothbrushes         | -0.133***<br>0.012    | -0.233***<br>0.051 | -0.225***<br>0.027 | -0.014<br>0.021    |
| Toothpaste           | -0.166***<br>0.010    | -0.243***<br>0.031 | -0.269***<br>0.016 | -0.055***<br>0.010 |
| Yogurt               | -0.047***             | -0.126***          | -0.033             | -0.014             |

Note: The table reports the coefficients (in percent) from OLS regressions for the 16 categories available in the Symphony IRI. The independent variable is the age of the product, and the dependent variables are the moments defined in the table. The sample is the first year (or 52 weeks) after the product was first launched. The controls include UPC  $\times$  store fixed effects, time fixed effects, and cohort controls that are approximated by the local unemployment rate in the city and month the product was launched. Columns (1) and (2) report the coefficients for regular price changes. Columns (3) and (4) report the results for all price changes (including sales). The standard errors are clustered at the store level. The \*\*\*, \*\*, and \* denote significance at 0.01, 0.05, and 0.10 levels respectively.

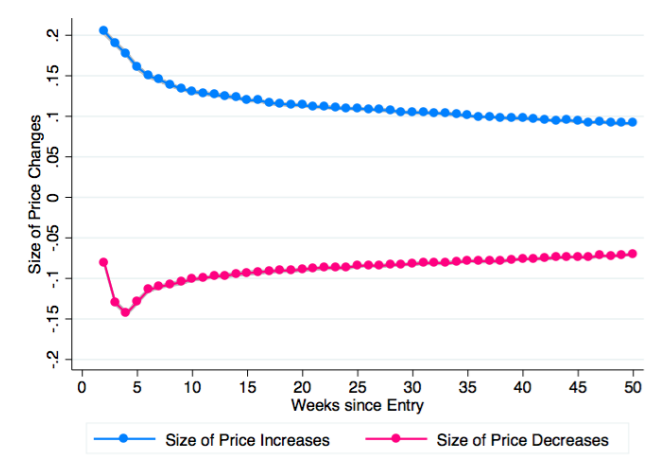


Figure A.I: Frequency of Price Increases and Decreases at Entry



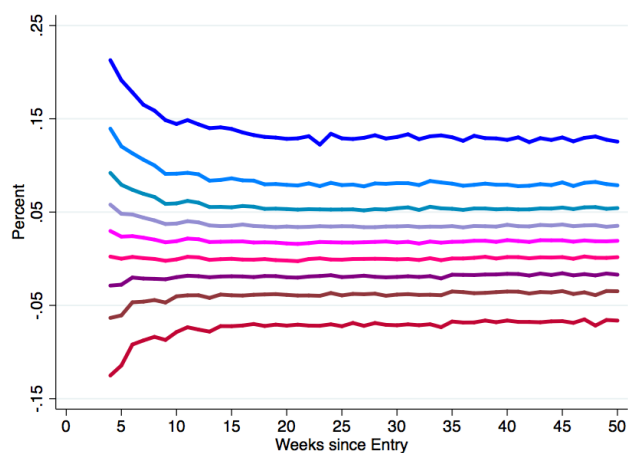
Note: The graph plots the average weekly frequency of price adjustments of products entering the market. The y-axis denotes the probability that the product adjusts prices in a given week, and the x-axis denotes the number of weeks the product has been observed in the data after it entered the market. The graph plots the age fixed effects where we use a regular price change indicator as the dependent variable that controls for the store, UPC, time fixed effects, and the local unemployment rate represents the cohort fixed effects. The blue line indicates the frequency of positive price adjustments and the red line the frequency of negative price adjustments. The calculation uses approximately 130 million observations and 2.5 million stores×UPC pairs. The data source is the Symphony IRI data set.

Figure A.II: Absolute Value of Price Increases and Decreases at Entry



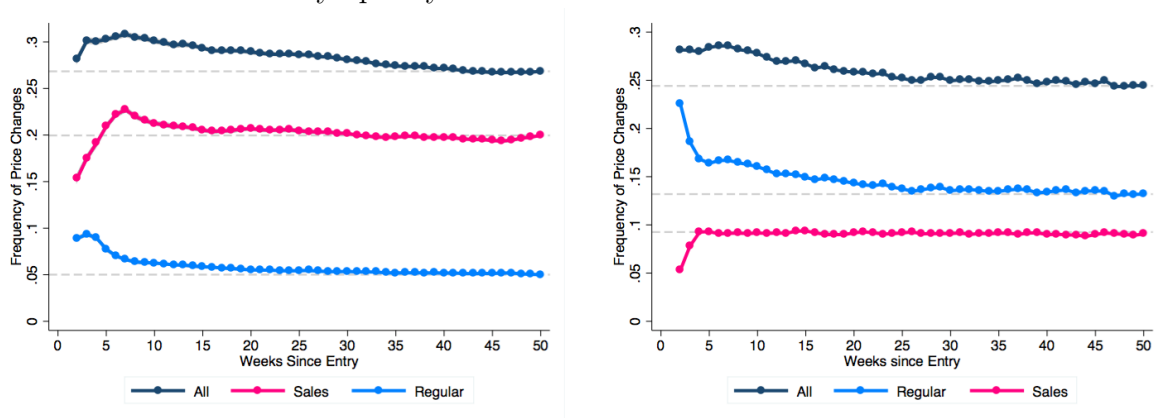
Note: The graph plots the average size of the price adjustments of products that enter the market. The y-axis is the value of the log price change in that week, and the x-axis denotes the number of weeks the product has been observed in the data after it entered the market. The graph plots the age fixed effects where we use the log price change as dependent variable that controls for the store, UPC, time fixed effects, and the local unemployment rate represents the cohort fixed effects. The blue line indicates the average size of positive price adjustments and the red line the average size of negative price adjustments. The calculation uses approximately 5.8 million price changes and 2.5 million stores×UPC pairs. The data source is the Symphony IRI data set.

Figure A.III: Distributions of Price Changes



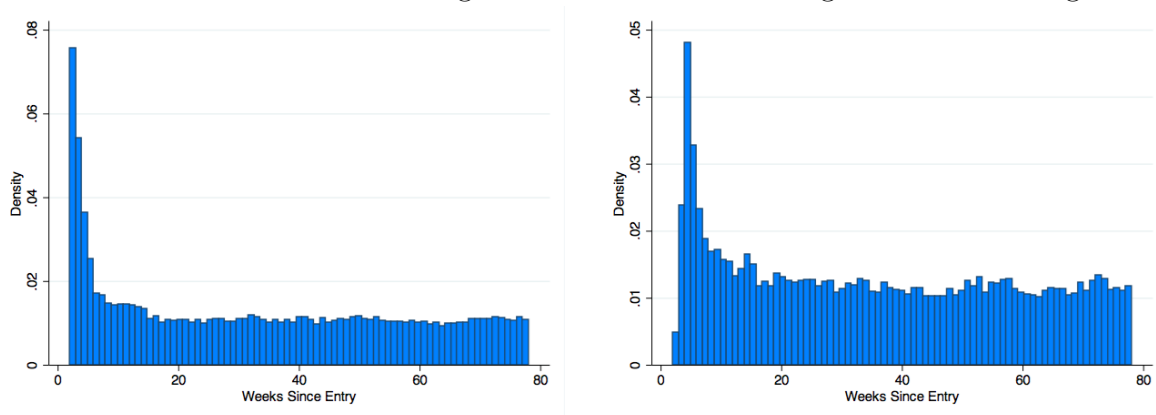
Note: The graph shows the distribution of regular price changes. The percentiles plotted are the 10th, 25th, 33rd, 50th, 66th, 75th, and 90th respectively. The y-axis is the value of the log price change in that week, and the x-axis denotes the number of weeks the product has been observed in the data after it entered the market. The calculation uses approximately 5.8 million price changes and 2.5 million stores $\times$ UPC pairs. The data source is the Symphony IRI data set.

Figure A.IV: Frequency of All Price Changes  
 Panel A: IRI Symphony      Panel B: Nielsen RMS



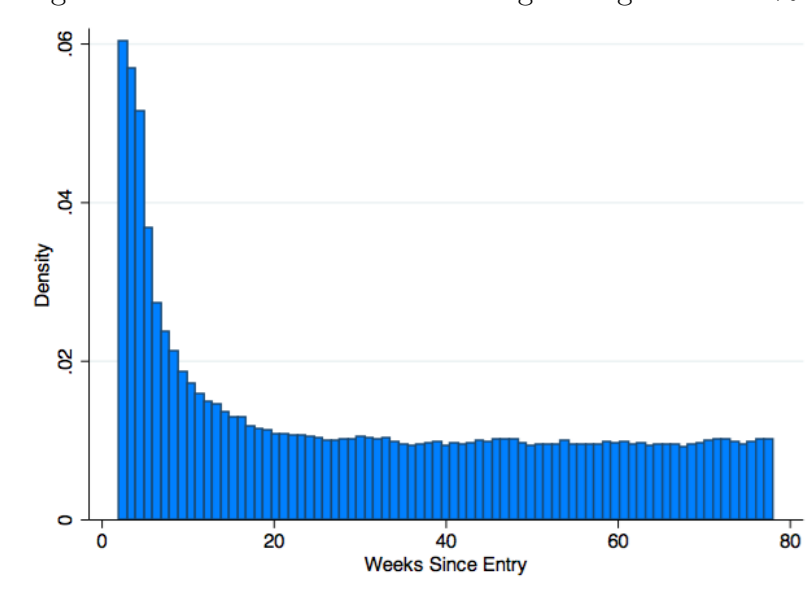
Note: The figure shows the average weekly frequency of all price changes as a function of the number of weeks after a product enters. Panel A shows the frequency of regular prices changes, the frequency of sales, and the frequency of all price changes in the IRI Symphony data. Panel B shows the same variables computed using the Nielsen RMS data for the city of Chicago. Since the Nielsen RMS data do not provide a sales flag, we use the sales filters developed in [Nakamura and Steinsson \(2008\)](#). The graph plots the fixed effects' coefficients for equation 1 where we use the price change indicator as the dependent variable. Equation 1 is computed by controlling for the store, UPC and time fixed effects, and the local unemployment rate represents the cohort fixed effects.

Figure A.V: Fraction of Price Changes Larger than Two Std. (Positive and Negative)  
 Panel A: Positive Price Changes      Panel B: Negative Price Changes



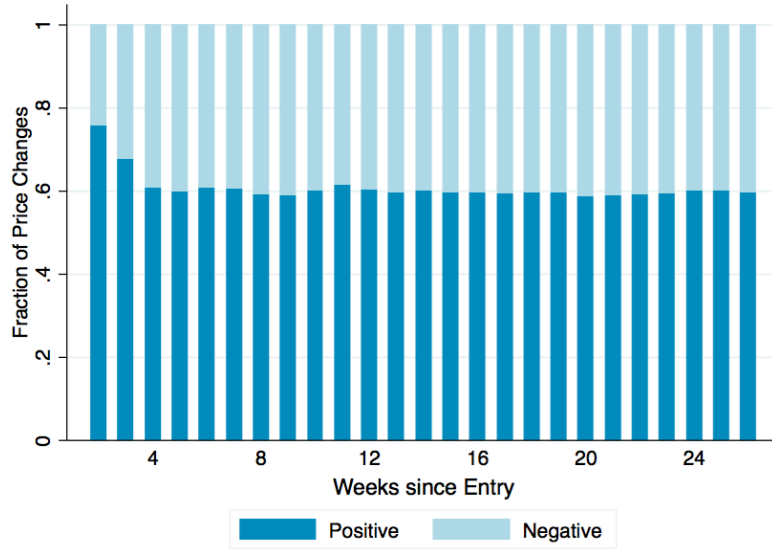
Note: The figure shows the fraction of price changes larger than two standard deviations from the mean in a given category and store as a function of the age of the product. Panel A shows the distribution of large price increases and Panel B the distribution of price decreases. The products considered are those that last at least two years in the market. Source: IRI Symphony dataset.

Figure A.VI: Fraction of Price Changes Larger than 30%



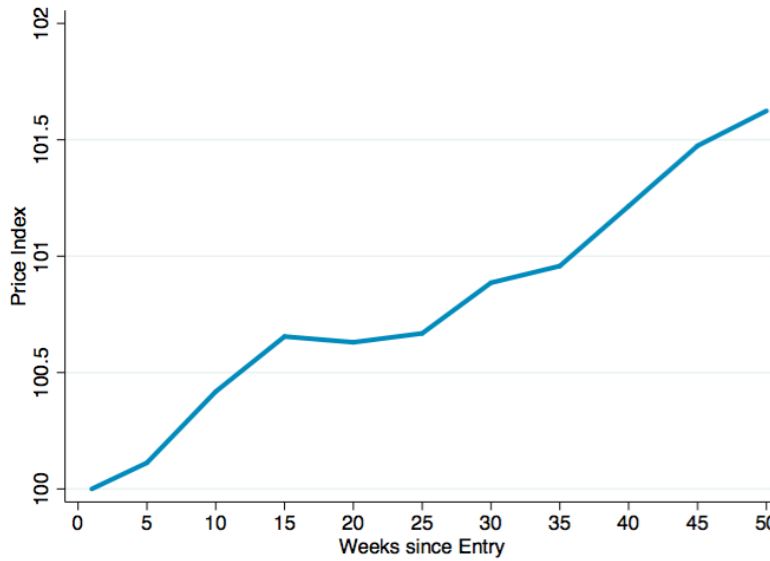
Note: The figure shows the fraction of price changes larger than 30% in a given category and city as a function of the age of the product. The products considered are those that last at least two years in the market. Source: IRI Symphony data set

Figure A.VII: Direction of Price Changes Conditional on Adjustment



Note: The graphs plots the share of price increases and price decreases conditional on adjustment. It considers the first six months after entry. The data source is the Symphony IRI data set.

Figure A.VIII: Price Index for New Products



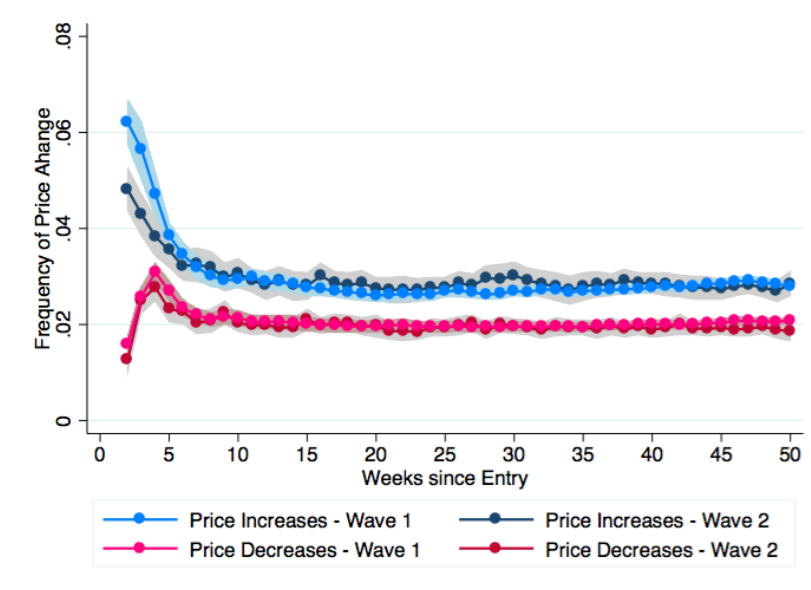
Note: The graph plots a geometric price index for new products. It considers the first year after entry. The expenditure weights are at the UPC level and based on the first year of sales of each product. The data source is the Symphony IRI data set.

Figure A.IX: Distribution of Entry Prices



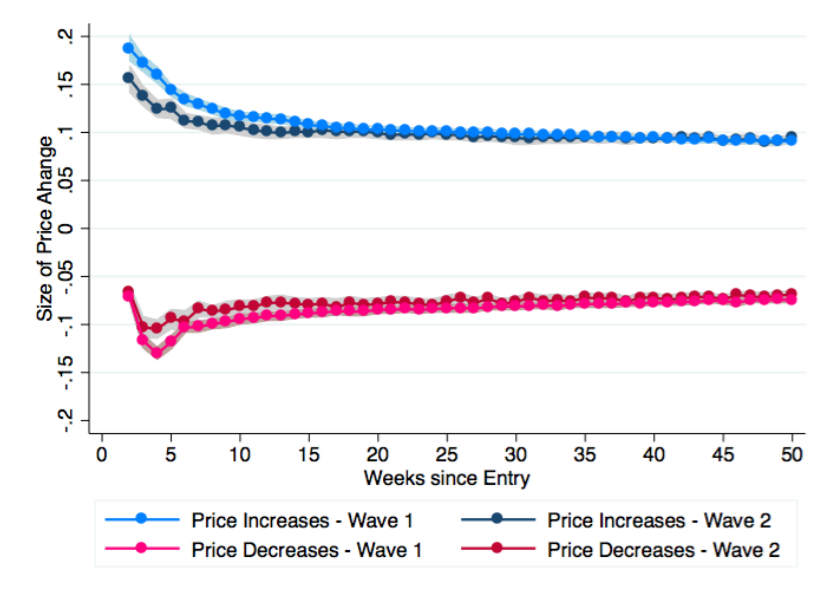
Note: The graph plots the percent difference between the entry price of all new products in our sample with respect to products of the same size, within the same category, at the store in which they were launched. The data source is the Symphony IRI dataset.

Figure A.X: Frequency of Price Adjustment (Positive and Negative)



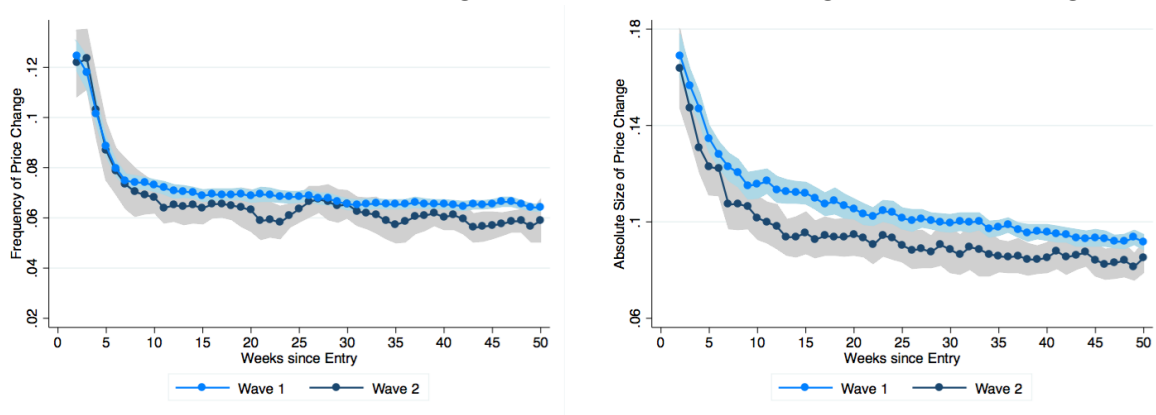
Note: The figure shows the probability of a price adjustment with respect to the mean for both price increases and decreases. Wave 1 represents products that were launched during the first year after the product was introduced. Wave 2 represents the same products when launched in different stores a year later. The graphs control for stores, time, and products fixed effects.

Figure A.XI: Size of Price Adjustments (Positive and Negative)



Note: The figure shows the size of price changes with respect to the mean for both price increases and decreases. Wave 1 represents products that were launched during the first year after the product was introduced. Wave 2 represents the same products when launched in different stores a year later. The graphs control for stores, time, and products fixed effects.

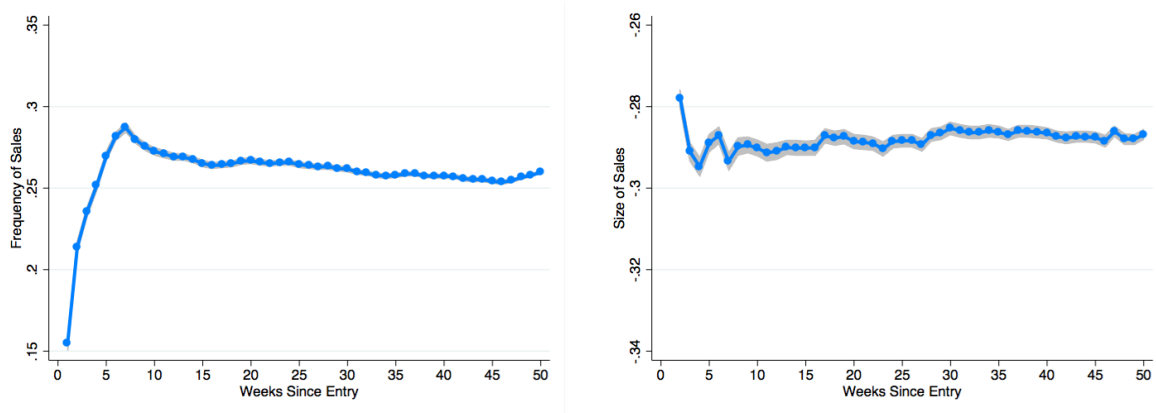
Figure A.XII: Pricing Moments by Waves in Different Cities  
 Panel A: Positive Price Changes  
 Panel B: Negative Price Changes



Note: The figure shows the probability of adjusting prices and the sizes of the adjustments by waves. Wave 1 represents products that were launched during the first year after the product was introduced. Wave 2 represents the same products when launched in different stores (located in different cities) a year later. Panel A shows the frequency of price adjustments and Panel B the absolute size of the price changes. The graphs control for stores, time, and products fixed effects.

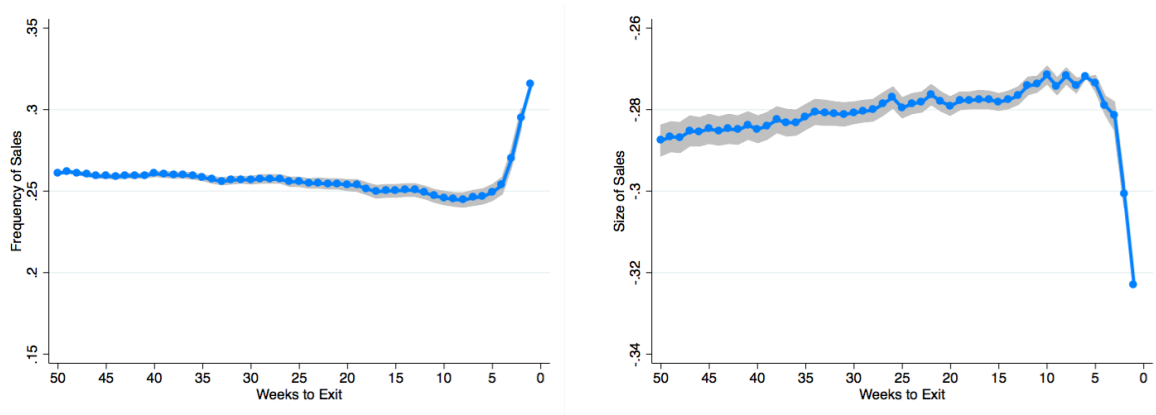


Figure A.XV: Frequency and Size of Sales at Entry  
 Panel A: Frequency  
 Panel B: Size



Note: Panel A plots the frequency of sales changes at entry. Panel B plots the size of the sales at entry. The x-axis denotes the number of weeks a product has been on the market. The graph plots the age fixed effects coefficients for the regression where we use the sales indicator (provided by the data) and the size of the sales (in logs) as dependent variables. The estimates control for store, UPC, time fixed effects, and the local unemployment rate represents the cohort fixed effects. Panel A shows that the probability that a product is on sale is lower at entry. Similarly, the size of sales stays mostly constant during the first year after the product is launched. The standard errors are clustered at the store level. The data source is the Symphony IRI data set.

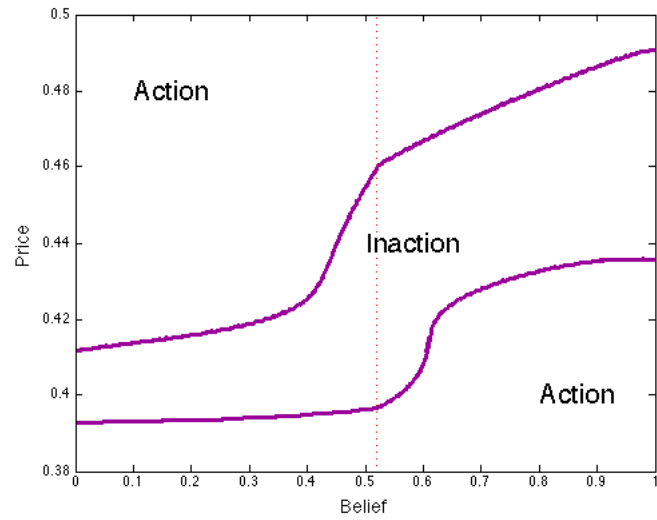
Figure A.XVI: Frequency and Size of Sales at Exit  
 Panel A: Frequency  
 Panel B: Size



Note: Panel A plots the frequency of sales at exit. Panel B plots the absolute size of the sales at exit. The x-axis denotes the number of weeks a product has left in the market before exiting. The graph plots the coefficients for the age fixed effects in the regression where we use the the sales indicator (provided by the data) and the size of the sales (in logs) as dependent variables. The estimates control for store, UPC, time fixed effects, and the local unemployment rate represents the cohort fixed-effects. The figure shows that at exit, products are more likely to be on sale and the size of these discounts are larger. This finding indicates that firms might be attempting to liquidate their inventory before phasing out their product permanently, and they do so by offering extra discounts ("clearance" sales"). The standard errors are clustered at the store level. The data source is the Symphony IRI data set.



Figure A.XVII: Inaction Region: Two-Period Model with Menu Costs



Note: The figure shows the region of inaction for the two-period model with menu costs. The y-axis denotes the previous price and the x-axis the belief. The region inside the purple lines is the region of inaction, and the dotted red line indicates the confounding belief.

## A.2 Proof of Lemmas

### A.2.1 Proof of Lemma 1

*Proof.* Recall that the value function  $V_2(\lambda)$  is given by:

$$V_2(\lambda) = \max_{p \in \mathcal{P}} \left\{ \left( p - \frac{W}{z} \right) \left( \lambda \eta \frac{p^{-\sigma_1}}{\bar{P}_1^{1-\sigma_1}} + (1-\lambda)(1-\eta) \frac{p^{-\sigma_2}}{\bar{P}_2^{1-\sigma_2}} \right) \frac{S}{P} \right\}$$

The function  $f(\lambda, p) \equiv \left( p - \frac{W}{z} \right) \left( \lambda \eta \frac{p^{-\sigma_1}}{\bar{P}_1^{1-\sigma_1}} + (1-\lambda)(1-\eta) \frac{p^{-\sigma_2}}{\bar{P}_2^{1-\sigma_2}} \right) \frac{S}{P}$  is continuous in  $(\lambda, p) \in [0, 1] \times \mathcal{P}$ . The set  $\mathcal{P} = [P_2^*, P_1^*]$  is furthermore compact. Then, the *Theorem of the Maximum* states that  $V_2(\cdot)$  is continuous on  $[0, 1]$ .

Convexity in  $\lambda$  follows almost directly. Fix an arbitrary  $\alpha \in [0, 1]$  and  $\lambda, \lambda' \in [0, 1]$ . Let the convex combination  $\tilde{\lambda}$  be defined as  $\alpha\lambda + (1-\alpha)\lambda'$  and define the myopic policy function:

$$P^M(\lambda) = \arg \max_{p \in \mathcal{P}} \left\{ \left( p - \frac{W}{z} \right) \left( \lambda \eta \frac{p^{-\sigma_1}}{\bar{P}_1^{1-\sigma_1}} + (1-\lambda)(1-\eta) \frac{p^{-\sigma_2}}{\bar{P}_2^{1-\sigma_2}} \right) \frac{S}{P} \right\}$$

Then, we get:

$$\begin{aligned} V_2(\tilde{\lambda}) &= \alpha \left( P^M(\tilde{\lambda}) - \frac{W}{z} \right) \left( \lambda \eta \frac{P^M(\tilde{\lambda})^{-\sigma_1}}{\bar{P}_1^{1-\sigma_1}} + (1-\lambda)(1-\eta) \frac{P^M(\tilde{\lambda})^{-\sigma_2}}{\bar{P}_2^{1-\sigma_2}} \right) \frac{S}{P} \\ &\quad + (1-\alpha) \left( P^M(\tilde{\lambda}) - \frac{W}{z} \right) \left( \lambda' \eta \frac{P^M(\tilde{\lambda})^{-\sigma_1}}{\bar{P}_1^{1-\sigma_1}} + (1-\lambda')(1-\eta) \frac{P^M(\tilde{\lambda})^{-\sigma_2}}{\bar{P}_2^{1-\sigma_2}} \right) \frac{S}{P} \\ &\leq \alpha V_2(\lambda) + (1-\alpha) V_2(\lambda') \end{aligned}$$

Therefore, we showed  $V_2(\alpha\lambda + (1-\alpha)\lambda') \leq \alpha V_2(\lambda) + (1-\alpha) V_2(\lambda')$  which is equivalent to  $V_2(\cdot)$  being convex.  $\square$

## A.3 Proof of Propositions

### A.3.1 Proof of Proposition 1

*Proof.* Note that this proposition holds for the infinite period model as well. Suppose it is optimal for the firm to choose  $P^*(\hat{\lambda}) = \hat{P} \in \text{int}(\mathcal{P})$  for some  $\hat{\lambda} \in (0, 1)$ . We show that a firm's continuation value is equal to zero whenever it chooses its price equal to  $\hat{P}$ . Given some price  $P$  and prior belief  $\lambda_0$ , a firm's continuation value is defined as:

$$\beta \mathcal{V}(P; \lambda_0) \equiv \beta \left( \lambda_0 \mathbb{E}_\varepsilon [V(b_1(\lambda_0, \log(P), \varepsilon))] + (1-\lambda_0) \mathbb{E}_\varepsilon [V(b_2(\lambda_0, \log(P), \varepsilon))] \right)$$

Recall that a firm faces a trade-off between maximizing current period expected profits and the value of information (through sharpening its posterior belief). The latter is captured by  $\mathcal{V}(P; \lambda_0)$ . As a result, a firm's marginal benefits are defined as:

$$\lambda_0 \mathbb{E}_\varepsilon \left[ V'(b_1(\lambda_0, \log(P), \varepsilon)) \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \frac{1}{P} \right] \\ + (1 - \lambda_0) \mathbb{E}_\varepsilon \left[ V'(b_2(\lambda_0, \log(P), \varepsilon)) \frac{\partial b_2(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \frac{1}{P} \right]$$

Therefore, a firm's posterior belief at the confounding price  $\hat{P}$  equals its prior belief, that is, we have:

$$b_1(\lambda_0, \log(P), \varepsilon) \Big|_{P=\hat{P}} = b_2(\lambda_0, \log(P), \varepsilon) \Big|_{P=\hat{P}} \\ = \left( 1 + \frac{1 - \lambda_0}{\lambda_0} \right)^{-1} \\ = \lambda_0$$

for all  $\varepsilon \in \mathbb{R}$ . Also, the expected change in a firm's posterior belief at  $P = \hat{P}$  is exactly equal to zero as:

$$\mathbb{E}_\varepsilon \left[ \frac{\partial b_i(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=\hat{P}} \right] = \mathbb{E}_\varepsilon \left[ \frac{\Delta \sigma}{\sigma_\varepsilon^2} (1 - \lambda_0) \lambda_0 (-1)^{\mathbf{1}(i=2)} \varepsilon \right] \\ = 0$$

for  $i \in \{1, 2\}$  as  $\mathbb{E}_\varepsilon[\varepsilon] = 0$ . Therefore, a firm's expected marginal benefit at  $P = \hat{P}$  reduces to:

$$V'(\lambda_0) \hat{P}^{-1} \mathbb{E}_\varepsilon \left[ \lambda_0 \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=\hat{P}} + (1 - \lambda_0) \frac{\partial b_2(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=\hat{P}} \right] = 0$$

If it is optimal for a firm to choose  $P^*(\hat{\lambda}) = \hat{P}$ , then it must be equal to  $P^M(\hat{\lambda})$  as there are no gains from active learning. Recall that  $P^M(0) = P_2^*$ ,  $P^M(1) = P_1^*$  and  $P^M(\cdot)$  is strictly increasing and continuous in proposition 1. Therefore, the confounding price  $\hat{P} \in \mathcal{P}$  is guaranteed to exist. Furthermore, proposition 1 and the *Intermediate Value Theorem* state that there must be some  $\hat{\lambda}$  such that  $P^M(\hat{\lambda}) = \hat{P}$ .

By construction, we have  $\hat{P} \equiv \frac{\Delta \mu}{\Delta \sigma}$ . Proposition 1 states that  $P^M(\cdot)$  is strictly increasing. As a result, we derive that the confounding belief is strictly increasing (decreasing) in  $\Delta \mu$  ( $\Delta \sigma$ ) as we must have  $P^M(\hat{\lambda}) = \frac{\Delta \mu}{\Delta \sigma}$ .  $\square$

### A.3.2 Proof of Proposition 2

**PROPOSITION 2.** The myopic policy function  $P^M(\cdot)$  is strictly increasing and  $\mathcal{C}$ .

*Proof.* The *Theorem of the Maximum* states that  $P^M(\lambda)$  is a non-empty, compact-valued, and upper hemi-continuous correspondence. However, the objective function is a weighted average of strictly concave functions, thus it is strictly concave itself. As a result,  $P^M(\lambda)$  must be single-valued. This value means that  $P^M(\lambda)$  is not only upper hemi-continuous but continuous.

Appendix A.1 of [Bachmann and Moscarini \(2012\)](#) shows that  $\frac{dP^M(\lambda)}{d\lambda} > 0$  if and only if  $P^M(\lambda) > P^M(0) = P_2^*$  for  $\lambda > 0$ . By construction, this holds for  $\lambda = 1$  as  $P_1^* > P_2^*$  as  $\sigma_2 > \sigma_1$ . Thus, the inequality must hold as well for large enough  $\lambda$  through continuity of  $P^M(\cdot)$ .

Suppose by way of contradiction that for some  $\lambda' > 0$ , we have  $P^M(\lambda') = P^M(0)$  instead. Then for some small  $\Delta > 0$ , we must either have  $P^M(\lambda' - \Delta) > P^M(0)$ ,  $P^M(\lambda' - \Delta) = P^M(0)$  or  $P^M(\lambda' - \Delta) < P^M(0)$ . The first case indicates that  $\frac{dP^M(\lambda)}{d\lambda} < 0$ , which contradicts the equivalence from [Bachmann and Moscarini \(2012\)](#). The second case states that  $P^M(\lambda' - \Delta) = P^M(0)$  over an open interval of small strictly positive values of  $\Delta$ . However, this value cannot be true as the expected profit function is strictly concave. Whenever  $P^M(\lambda' - \Delta) < P^M(0)$ , then we must have  $\frac{dP^M(\lambda)}{d\lambda} \Big|_{\lambda=\ell} < 0$  for all  $\ell \in (0, \lambda')$ . But, this means that for all  $\ell \in (0, \lambda')$ , we have  $P^M(\ell) < P^M(0)$  but we assumed that  $\lim_{\lambda \downarrow 0} P^M(\lambda) = P^M(\lambda')$ . Therefore,  $P^M(\lambda)$  must display a discontinuity at  $\lambda = 0$ . This is the desired contradiction as we showed that  $P^M(\cdot)$  is continuous. Thus,  $P^M(\lambda') > P^M(0)$  must hold for all  $\lambda' > 0$  and  $\frac{dP^M(\lambda)}{d\lambda} > 0$  follows.  $\square$

## A.4 Additional Theoretical Results

**LEMMA 2.** The marginal expected change in a firm's posterior belief is bounded by its absolute value, that is,

$$\left| \mathbb{E}_\varepsilon \left[ \frac{\partial b_i(\lambda_0, \log(p), \varepsilon)}{\partial \log(p)} \mid \varepsilon \in \mathcal{F} \right] \right| \leq \frac{\Delta \sigma}{\sigma_\varepsilon^2} \left( \int_{\varepsilon \in \mathcal{F}} |\log(p) \Delta \sigma - \Delta \mu + \varepsilon| dF(\varepsilon) \right)$$

where the sign of  $\log(p) \Delta \sigma - \Delta \mu + \varepsilon$  is constant for all  $\varepsilon \in \mathcal{F} \subseteq \mathbb{R}$ .

*Proof.* Let  $x \equiv \log(P) \Delta \sigma - \Delta \mu$  and  $\varepsilon$  is contained in some set  $\mathcal{F} \subseteq \mathbb{R}$ . By construction of the ex post belief function  $b_i(\lambda, \log(P), \varepsilon)$ , we obtain:

$$\left| \mathbb{E}_\varepsilon \left[ \frac{\partial b_i(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \mid \varepsilon \in \mathcal{F} \right] \right| = \left| \int_{\varepsilon \in \mathcal{F}} \frac{\exp\left(\frac{(\varepsilon+x)^2 + \varepsilon^2}{2\sigma_\varepsilon^2}\right) \Delta \sigma (x + \varepsilon) (1 - \lambda_0) \lambda_0}{\left( \exp\left(\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right) (1 - \lambda_0) \sigma_\varepsilon + \exp\left(\frac{(x+\varepsilon)^2}{2\sigma_\varepsilon^2}\right) \lambda_0 \sigma_\varepsilon \right)^2} dF(\varepsilon) \right|$$

$$\begin{aligned}
&= \frac{\Delta\sigma}{\sigma_\varepsilon^2} \left| \int_{\varepsilon \in \mathcal{F}} \left( \frac{\exp\left(\frac{(\varepsilon+x)^2}{2\sigma_\varepsilon^2}\right) \lambda_0}{\exp\left(\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right) (1-\lambda_0) + \exp\left(\frac{(\varepsilon+x)^2}{2\sigma_\varepsilon^2}\right) \lambda_0} \right) \times \right. \\
&\quad \left. \left( \frac{\exp\left(\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right) (1-\lambda_0)}{\exp\left(\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right) (1-\lambda_0) + \exp\left(\frac{(\varepsilon+x)^2}{2\sigma_\varepsilon^2}\right) \lambda_0} \right) (x+\varepsilon) dF(\varepsilon) \right| \\
&\leq \frac{\Delta\sigma}{\sigma_\varepsilon^2} \left( \int_{\varepsilon \in \mathcal{F}} |x+\varepsilon| dF(\varepsilon) \right)
\end{aligned}$$

where the last inequality follows as the bracketed terms in the second equality are bounded by  $[0, 1]$  and the sign of  $x + \varepsilon$  remains constant on the set  $\mathcal{F}$  by assumption. This is exactly what we wanted to show.  $\square$

**PROPOSITION 3.** Whenever  $V_2'(1)$  is small enough, then the firm's active learning policy has an interior solution, that is,  $P^*(\lambda_0) \in (P_2^*, P_1^*)$  for all  $\lambda_0 \in (0, 1)$ .

*Proof.* We show the case for  $\lambda_0 \geq \frac{1}{2}$ . The case for  $\lambda_0 < \frac{1}{2}$  follows a very similar process. We derive sufficient conditions such that  $P^*(\lambda_0) \in \text{int}(\mathcal{P})$  for all  $\lambda_0 \in (0, 1)$ . This is equivalent to finding sufficient conditions such that a firm's expected marginal benefits strictly dominate its cost counterpart for  $P = P_2^*$  and vice versa for  $P = P_1^*$ .

By construction of the ex post belief functions  $b_i(\lambda_0, \log(P), \varepsilon)$ , we can derive the following equality:

$$\frac{\partial b_2(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} = -\frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \tilde{\beta}(\varepsilon, \lambda_0, x)$$

where the function  $\tilde{\beta}(\cdot)$  is characterized by:

$$\tilde{\beta}(\varepsilon, \lambda_0, x) = \left( \frac{\lambda_0 \exp\left(\frac{(x+\varepsilon)^2}{2\sigma_\varepsilon^2}\right) + (1-\lambda_0) \exp\left(\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right)}{(1-\lambda_0) \exp\left(\frac{(x+\varepsilon)^2}{2\sigma_\varepsilon^2}\right) + \lambda_0 \exp\left(\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right)} \right)^2$$

We show that  $\tilde{\beta}'(\cdot, \lambda_0, x)$  is strictly increasing if and only if  $x(2\lambda_0 - 1)$  is strictly positive. Furthermore, it satisfies  $\lim_{\varepsilon \rightarrow +\infty} \tilde{\beta}(\varepsilon, \lambda_0, x) = \left(\frac{1-\lambda_0}{\lambda_0}\right)^2$  and  $\lim_{\varepsilon \rightarrow -\infty} \tilde{\beta}(\varepsilon, \lambda_0, x) = \left(\frac{\lambda_0}{1-\lambda_0}\right)^2$ . Let  $x_{\min} = \log(P_2^*)\Delta\sigma - \Delta\mu < 0$  and  $x_{\max} = \log(P_1^*)\Delta\sigma - \Delta\mu > 0$ , then we need to show that:

$$\begin{aligned}
&\eta\lambda_0\Pi_1'(P_2^*) + \frac{\beta}{P_2^*} \mathbb{E}_\varepsilon \left[ \left( \lambda_0 V_2'(b_1(\lambda_0, \log(P_2^*), \varepsilon)) - \right. \right. \\
&\quad \left. \left. (1-\lambda_0) V_2'(b_2(\lambda_0, \log(P_2^*), \varepsilon)) \tilde{\beta}(\varepsilon, \lambda_0, x_{\min}) \right) \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=P_2^*} \right] > 0 \quad (\mathbf{A1})
\end{aligned}$$

$$(1 - \eta)(1 - \lambda_0)\Pi'_2(P_1^*) + \frac{\beta}{P_1^*}\mathbb{E}_\varepsilon \left[ \left( \lambda_0 V'_2(b_1(\lambda_0, \log(P_1^*), \varepsilon)) - (1 - \lambda_0)V'_2(b_2(\lambda_0, \log(P_1^*), \varepsilon))\tilde{\beta}(\varepsilon, \lambda_0, x_{\max}) \right) \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=P_1^*} \right] < 0 \quad (\mathbf{A2})$$

that are the first order conditions with respect to  $P$  in period 1 evaluate at  $P = P_2^*$  and  $P = P_1^*$ . We start by finding a sufficient condition for the first inequality **A1**. To do this, we define the function  $g(\varepsilon, \lambda_0) = \lambda_0 V'_2(b_1(\lambda_0, P_2^*, \varepsilon)) - (1 - \lambda_0)V'_2(b_2(\lambda_0, P_2^*, \varepsilon))\tilde{\beta}(\varepsilon, \lambda_0, x_{\min})$ . For  $x = x_{\min} < 0$ , and we show that  $g(\cdot, \lambda_0)$  is monotonically decreasing. Furthermore, it satisfies  $g(-x, \frac{1}{2}) < 0$  and  $g(-x, 1) > 0$ .

Whenever  $\lambda_0$  is relatively close to  $\frac{1}{2}$ , we show that  $\exists \varepsilon(\lambda_0) < -x_{\min}$  such that  $g(\varepsilon, \lambda_0) > 0$  for all  $\varepsilon < \varepsilon(\lambda_0)$  as  $g(-x_{\min}, \frac{1}{2}) < 0$ .<sup>46</sup> Furthermore, we derive that  $\frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=P_2^*} > 0$  if and only if  $\varepsilon > -x_{\min}$ . Now, denote  $E_1 \equiv (-\infty, \varepsilon(\lambda_0))$ ,  $E_2 \equiv (\varepsilon(\lambda_0), -x_{\min})$  and  $E_3 \equiv (-x_{\min}, +\infty)$ . By construction, it must be that  $E_1 \cup E_2 \cup E_3 = \mathbb{R}$ .

The observations above show that:

$$g(\varepsilon, \lambda_0) \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=P_2^*} > 0$$

for  $\varepsilon \in E_2$ . Thus, it is sufficient to show:

$$\eta \lambda_0 \Pi'_1(P_2^*) + \frac{\beta}{P_2^*}\mathbb{E}_\varepsilon \left[ g(\varepsilon, \lambda_0) \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=P_2^*} \Big| \varepsilon \in E_1 \cup E_3 \right] > 0$$

---

<sup>46</sup>Whenever  $\lambda_0$  is close to one, then the steps of the proof are similar. Instead, we have that  $\varepsilon(\lambda_0)$  is greater than  $-x_{\min}$  though.

Let  $\xi_2 \equiv \max_{\varepsilon \in E_3} \tilde{\beta}(\varepsilon) = \tilde{\beta}(-x_{\min})$ , then observe the following strain of inequalities:

$$\begin{aligned}
& \eta\lambda_0\Pi'_1(P_2^*) + \frac{\beta}{P_2^*}V'_2(1)\frac{\Delta\sigma}{\sigma_\varepsilon^2}(x_{\min} + \mathbb{E}_\varepsilon[\varepsilon|\varepsilon \leq \varepsilon(\lambda_0)] - \xi_2\mathbb{E}_\varepsilon[\varepsilon|\varepsilon \geq -x_{\min}]) < \\
& \quad \eta\lambda_0\Pi'_1(P_2^*) + \frac{\beta}{P_2^*}V'_2(1)\frac{\Delta\sigma}{\sigma_\varepsilon^2}(x_{\min}F(\varepsilon(\lambda_0)) + \mathbb{E}_\varepsilon[\varepsilon|\varepsilon \leq \varepsilon(\lambda_0)] \\
& \quad \quad - \xi_2x_{\min}(1 - F(-x_{\min})) - \xi_2\mathbb{E}_\varepsilon[\varepsilon|\varepsilon \geq -x_{\min}]) = \\
& \quad \eta\lambda_0\Pi'_1(P_2^*) + \frac{\beta}{P_2^*}V'_2(1)\frac{\Delta\sigma}{\sigma_\varepsilon^2}(\mathbb{E}_\varepsilon[x + \varepsilon|\varepsilon \in E_1] - \xi_2\mathbb{E}_\varepsilon[x + \varepsilon|\varepsilon \in E_3]) \leq \\
& \quad \eta\lambda_0\Pi'_1(P_2^*) + \frac{\beta}{P_2^*}V'_2(1)\left(\mathbb{E}_\varepsilon\left[\left.\frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(p)}\right|_{P=P_2^*}\right] \Big| \varepsilon \in E_1\right) \\
& \quad \quad - \xi_2\mathbb{E}_\varepsilon\left[\left.\frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(p)}\right|_{P=P_2^*}\right] \Big| \varepsilon \in E_3\right) < \\
& \quad \eta\lambda_0\Pi'_1(P_2^*) + \frac{\beta}{P_2^*}\mathbb{E}_\varepsilon\left[g(\varepsilon, \lambda_0)\frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)}\right]_{P=P_2^*} \Big| \varepsilon \in E_1 \cup E_3\right]
\end{aligned}$$

where the weak inequality follows from lemma 2 and the last strict inequality from the fact that  $V'_2(1) > V'_2(\lambda')$  for any  $\lambda' < 1$ . This means that we are done whenever we can show:

$$\eta\lambda_0\Pi'_1(P_2^*) + \frac{\beta}{P_2^*}V'_2(1)\frac{\Delta\sigma}{\sigma_\varepsilon^2}(x_{\min} + \mathbb{E}_\varepsilon[\varepsilon|\varepsilon \leq \varepsilon(\lambda_0)] - \mathbb{E}_\varepsilon[\varepsilon|\varepsilon \geq -x_{\min}]) > 0$$

Recall that  $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ . Therefore, we can use standard truncation formulas for our conditional expectations. These formulas give:

$$\begin{aligned}
\mathbb{E}_\varepsilon[\varepsilon|\varepsilon \leq \varepsilon(\lambda_0)] - \xi_2\mathbb{E}_\varepsilon[\varepsilon|\varepsilon \geq -x_{\min}] &= -\frac{\varphi\left(\frac{\varepsilon(\lambda_0)}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{\varepsilon(\lambda_0)}{\sigma_\varepsilon}\right)} - \xi_2\frac{\varphi\left(\frac{-x_{\min}}{\sigma_\varepsilon}\right)}{1 - \Phi\left(\frac{-x_{\min}}{\sigma_\varepsilon}\right)} \\
&> -\varphi(0)\left[\frac{1}{\Phi\left(\frac{\varepsilon(\lambda_0)}{\sigma_\varepsilon}\right)} + \frac{\xi_2}{1 - \Phi\left(\frac{-x_{\min}}{\sigma_\varepsilon}\right)}\right] \\
&> -\varphi(0)\left[\frac{1 + \xi_2}{1 - \Phi\left(\frac{-x_{\min}}{\sigma_\varepsilon}\right)}\right]
\end{aligned}$$

Then, we can frame our first sufficient condition as:

$$\eta\lambda_0\Pi'_1(P_2^*) + \frac{\beta}{P_2^*}V'_2(1)\frac{\Delta\sigma}{\sigma_\varepsilon^2}\left[x_{\min} - \varphi(0)\frac{1+\xi_2}{1-\Phi\left(\frac{-x}{\sigma_\varepsilon}\right)}\right] \quad (\mathbf{B1})$$

In a similar fashion, we derive a sufficient condition for **A2**. Define  $h(\varepsilon, \lambda_0) = \lambda_0V'_2(b_1(\lambda_0, P_1^*, \varepsilon)) - (1 - \lambda_0)V'_2(b_2(\lambda_0, P_1^*, \varepsilon))\tilde{\beta}(\varepsilon, \lambda_0, x_{\max})$ , then we have  $h(\cdot, \lambda_0)$  that is monotonically increasing. Once again, we can show that  $\exists \varepsilon(\lambda_0) > 0$  such that  $h(\varepsilon, \lambda_0) > 0$  if and only if  $\varepsilon > \varepsilon(\lambda_0)$ . By straightforward algebra, we can deduce that  $\frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)}\Big|_{P=P_1^*} > 0$  if and only if  $\varepsilon > -x_{\max}$ . Then, denote

$E_1 \equiv (-\infty, -x_{\max})$ ,  $E_2 \equiv (-x_{\max}, \varepsilon(\lambda_0))$  and  $E_3 \equiv (\varepsilon(\lambda_0), +\infty)$  which satisfies  $E_1 \cup E_2 \cup E_3 = \mathbb{R}$ .

With similar reasoning as before, the following condition is sufficient for **A2** to hold:

$$(1 - \eta)(1 - \lambda_0)\Pi_2'(P_1^*) + \frac{\beta}{P_1^*} \mathbb{E}_\varepsilon \left[ \left( \lambda_0 V_2'(b_1(\lambda_0, \log(P_1^*), \varepsilon)) - (1 - \lambda_0) V_2'(b_2(\lambda_0, \log(P_1^*), \varepsilon)) \tilde{\beta}(\varepsilon, \lambda_0, x_{\max}) \right) \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=P_1^*} \Big| \varepsilon \in E_1 \cup E_3 \right] > 0$$

Let  $\xi_1 \equiv \max_{\varepsilon \in E_1} \tilde{\beta}(\varepsilon) = \left( \frac{\lambda_0}{1 - \lambda_0} \right)^2$ , then we derive a similar chain of inequalities as before:

$$\begin{aligned} & (1 - \eta)(1 - \lambda_0)\Pi_2'(P_1^*) + \frac{\beta}{P_1^*} V_2'(1) \frac{\Delta\sigma}{\sigma_\varepsilon^2} (\xi_1 x_{\max} + \xi_1 \mathbb{E}_\varepsilon [\varepsilon | \varepsilon \geq \varepsilon(\lambda_0)] - \mathbb{E}_\varepsilon [\varepsilon | \varepsilon \leq -x_{\max}]) > \\ & (1 - \eta)(1 - \lambda_0)\Pi_2'(P_1^*) + \frac{\beta}{P_1^*} V_2'(1) \frac{\Delta\sigma}{\sigma_\varepsilon^2} (\xi_1 x_{\max} [1 - F(\varepsilon(\lambda_0))] + \xi_1 \mathbb{E}_\varepsilon [\varepsilon | \varepsilon \geq \varepsilon(\lambda_0)] \\ & \quad - x_{\max} F(-x_{\max}) - \mathbb{E}_\varepsilon [\varepsilon | \varepsilon \leq -x_{\max}]) \\ & (1 - \eta)(1 - \lambda_0)\Pi_2'(P_1^*) + \frac{\beta}{P_1^*} V_2'(1) \frac{\Delta\sigma}{\sigma_\varepsilon^2} (\xi_1 \mathbb{E}_\varepsilon [x_{\max} + \varepsilon | \varepsilon \in E_3] - \mathbb{E}_\varepsilon [x_{\max} + \varepsilon | \varepsilon \in E_1]) \geq \end{aligned}$$

$$\begin{aligned} & (1 - \eta)(1 - \lambda_0)\Pi_2'(P_1^*) + \frac{\beta}{P_1^*} V_2'(1) \left( \xi_1 \mathbb{E}_\varepsilon \left[ \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(p)} \Big|_{P=P_1^*} \Big| \varepsilon \in E_3 \right] \right. \\ & \quad \left. - \mathbb{E}_\varepsilon \left[ \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(p)} \Big|_{P=P_1^*} \Big| \varepsilon \in E_1 \right] \right) > \\ & (1 - \eta)(1 - \lambda_0)\Pi_2'(P_1^*) + \frac{\beta}{P_1^*} \mathbb{E}_\varepsilon \left[ h(\varepsilon, \lambda_0) \frac{\partial b_1(\lambda_0, \log(P), \varepsilon)}{\partial \log(P)} \Big|_{P=P_1^*} \Big| \varepsilon \in E_1 \cup E_3 \right] \end{aligned}$$

where the weak inequality follows from lemma 2 and the last strict inequality from the fact that  $V_2'(1) > V_2'(\lambda')$  for any  $\lambda' < 1$ . This means that we are done whenever we can show:

$$(1 - \eta)(1 - \lambda_0)\Pi_2'(P_1^*) + \frac{\beta}{P_1^*} V_2'(1) \frac{\Delta\sigma}{\sigma_\varepsilon^2} (\xi_1 x_{\max} + \xi_1 \mathbb{E}_\varepsilon [\varepsilon | \varepsilon \geq \varepsilon(\lambda_0)] - \mathbb{E}_\varepsilon [\varepsilon | \varepsilon \leq -x_{\max}]) < 0$$

Using the previous finding on expectations of truncated standard normal random variables, the latter inequality is satisfied whenever the following condition holds:

$$(1 - \eta)(1 - \lambda_0)\Pi_2'(P_1^*) + \frac{\beta}{P_1^*} V_2'(1) \frac{\Delta\sigma}{\sigma_\varepsilon^2} \left( \xi_1 x_{\max} + \varphi(0) \frac{1 + \xi_1}{\Phi\left(\frac{-x_{\max}}{\sigma_\varepsilon}\right)} \right) < 0 \quad (\mathbf{B2})$$



Whenever we define  $\tilde{x}_{\min}$  and  $\tilde{x}_{\max}$  as:

$$\begin{aligned}\tilde{x}_{\min} &\equiv \log(P_2^*)\Delta\sigma - \Delta\mu - \varphi(0)\frac{1+\xi_2}{1-\Phi\left(-\frac{\log(P_2^*)\Delta\sigma-\Delta\mu}{\sigma_\varepsilon}\right)} < 0, \\ \tilde{x}_{\max} &\equiv \xi_1(\log(P_1^*)\Delta\sigma - \Delta\mu) + \varphi(0)\frac{1+\xi_1}{\Phi\left(-\frac{\log(P_1^*)\Delta\sigma-\Delta\mu}{\sigma_\varepsilon}\right)} > 0.\end{aligned}$$

then, it is clear that **B1** and **B2** are satisfied whenever  $V_2'(1)$  is bounded from above. More precisely, we get:

$$V_2'(1) < \bar{V} \equiv \frac{\sigma_\varepsilon^2}{\beta\Delta\sigma} \min \left\{ \frac{\eta\lambda_0\Pi_1'(P_2^*)}{-\tilde{x}_{\min}}, \frac{(1-\eta)(1-\lambda_0)(-\Pi_2'(P_1^*))}{\tilde{x}_{\max}} \right\}. \quad (\mathbf{B})$$

Thus, we have shown  $\mathbf{B} \implies (\mathbf{B1} \text{ and } \mathbf{B2}) \implies (\mathbf{A1} \text{ and } \mathbf{A2})$ . However, we concluded in the beginning of the proposition that  $P^*(\lambda_0) \in \text{int}(\mathcal{P})$  whenever **A1** and **A2** hold. This is exactly what we wanted to show.  $\square$

## A.5 Numerical Algorithm

PARAMETERS.  $\beta, \delta, \sigma_1, \sigma_2, \lambda_0, m, s, \sigma_\varepsilon, \rho$  and  $\sigma_\zeta$ .

NUMÉRAIRE.  $W = 1$ .

ALGORITHM: PSEUDO-CODE.

We assume consumer taste shocks to satisfy  $\varepsilon_k \sim N(m, s^2)$ . As a result, integrals over consumer taste shocks are approximated using Gaussian quadrature methods. With some abuse of notation, let the weights and nodes be denoted by  $\{\omega_j^{GH}\}_{j=1}^M$  and  $\{\zeta_j^{GH}\}_{j=1}^M$  respectively.<sup>47</sup> The quadrature weights and nodes are chosen ‘‘optimally’’. The nodes  $\{\zeta_j^{GH}\}_{j=1}^M$  are the roots of the Hermite polynomial  $H_M(\zeta)$  that is defined as  $H_M(\zeta) = \oint \frac{M!}{2\pi i} \exp(-t^2 + 2t\zeta) t^{-(M+1)} dt$  and the weights are equal to:

$$\omega_i^{GH} = \frac{2^{M-1}M!\sqrt{\pi}}{M^2 H_{M-1}(\zeta_j^{GH})^2}$$

We approximate the continuous AR(1) process for idiosyncratic productivity with a finite state Markov process by following the Tauchen (1986) procedure. The number of Markov states is denoted by  $N_T$ . Let the Markov transition density be denoted by  $\mathcal{M}(z_i, z_j)$  for  $i, j \in \{1, 2, \dots, N_T\} \times \{1, 2, \dots, N_T\}$ .

**I (initialization).** Set  $\bar{P}_1^0, \bar{P}_2^0, \chi^0$  and convergence criteria  $\Delta_\varepsilon, \Delta_\varphi > 0$ . Let the counter  $k$  equal zero.

---

<sup>47</sup>The superscript stands for ‘‘Gaussian-Hermite’’ quadrature. This is useful to approximate functions of the form  $f(x) = \exp(-x^2)$  which includes the family of normal distributions.

**II (out).** Given  $k$ , set  $\bar{P}_1^k$ ,  $\bar{P}_2^k$  and  $\chi^k$ .

**III (in).** Set  $H = \frac{1}{3}$  and calculate  $\Pi^k$  using:

$$H^\eta = \frac{1}{\chi^k} \frac{1}{H + \Pi^k}$$

Then, set  $S^k = H + \Pi^k$ . Define the aggregate state as  $\omega^k = (\bar{P}_1^k, \bar{P}_2^k, S^k)$ .

**IV.** Solve the firm's problem by obtaining  $V(\lambda, z, p_{-1})$ :

$V(\lambda, z, p_{-1}) = \max \{V^A(\lambda, z), V^N(\lambda, z, p_{-1})\}$  where

$$\begin{aligned} V^A(\lambda, z) = & \max_{p \geq 0} \left( p - \frac{1}{z} \right) \left[ \lambda \eta \frac{p^{-\sigma_1}}{\bar{P}_1^{1-\sigma_1}} + (1-\lambda)(1-\eta) \frac{p^{-\sigma_2}}{\bar{P}_2^{1-\sigma_2}} \right] \frac{1}{\bar{P}_1^\eta \bar{P}_2^{1-\eta} \chi} - \frac{\psi}{\bar{P}_1^\eta \bar{P}_2^{1-\eta}} \\ & + \beta \lambda \sum_{i=1}^{N_T} \sum_{j=1}^M \mathcal{M}(z_i, z) \frac{1}{\sqrt{\pi}} \omega_j^{GH} V \left( b_1(\lambda, \log(\frac{p}{1+\bar{\pi}}), \sqrt{2}\sigma_\alpha \zeta_j^{GH} + \mu_\alpha), z_i, \frac{p}{1+\bar{\pi}} \right) \end{aligned}$$

$$+ \beta(1-\lambda) \sum_{i=1}^{N_T} \sum_{j=1}^M \mathcal{M}(z_i, z) \frac{1}{\sqrt{\pi}} \omega_j^{GH} V \left( b_2(\lambda, \log(\frac{p}{1+\bar{\pi}}), \sqrt{2}\sigma_\alpha \zeta_j^{GH} + \mu_\alpha), z_i, \frac{p}{1+\bar{\pi}} \right)$$

$$\begin{aligned} V^N(\lambda, z, p_{-1}) = & \left( p_{-1} - \frac{1}{z} \right) \left[ \lambda \eta \frac{p_{-1}^{-\sigma_1}}{\bar{P}_1^{1-\sigma_1}} + (1-\lambda)(1-\eta) \frac{p_{-1}^{-\sigma_2}}{\bar{P}_2^{1-\sigma_2}} \right] \frac{1}{\bar{P}_1^\eta \bar{P}_2^{1-\eta} \chi} \\ & + \beta \lambda \sum_{i=1}^{N_T} \sum_{j=1}^M \mathcal{M}(z_i, z) \frac{1}{\sqrt{\pi}} \omega_j^{GH} V \left( b_1(\lambda, \log(\frac{p_{-1}}{1+\bar{\pi}}), \sqrt{2}\sigma_\alpha \zeta_j^{GH} + \mu_\alpha), z_i, \frac{p_{-1}}{1+\bar{\pi}} \right) \\ & + \beta(1-\lambda) \sum_{i=1}^{N_T} \sum_{j=1}^M \mathcal{M}(z_i, z) \frac{1}{\sqrt{\pi}} \omega_j^{GH} V \left( b_2(\lambda, \log(\frac{p_{-1}}{1+\bar{\pi}}), \sqrt{2}\sigma_\alpha \zeta_j^{GH} + \mu_\alpha), z_i, \frac{p_{-1}}{1+\bar{\pi}} \right) \end{aligned}$$

$$\text{with } b_i(\lambda, \log(p), \epsilon) = \left[ 1 + \frac{1-\lambda}{\lambda} \frac{F'(\mu_i - \mu_2 + (\sigma_2 - \sigma_i)\log(p) + \epsilon)}{F'(\mu_i - \mu_1 + (\sigma_1 - \sigma_i)\log(p) + \epsilon)} \right]^{-1}$$

$$\text{and } \mu_i = (\sigma_i - 1)\log(\bar{P}_i) + \log(\eta_i)$$

**IV.** Store the optimal pricing policy function  $P^*(\lambda, z)$  for every  $(\lambda, z) \in [0, 1] \times Z$ .

**V. SIMULATION.** Simulate a panel of  $N = 50000$  firms who use the policy function  $P^*(\lambda, z)$ .

**SIMULATION INITIALIZATION.** The initial distribution  $\varphi_{i,0}(\lambda, z)$  for  $i = 1, 2$  is degenerate at  $(\lambda_0, z_0)$ . For each firm  $n \in \{1, 2, \dots, N\}$ , we assign it to be a firm of type  $\sigma_n = \sigma_1$  with probability  $\lambda_0$  and set time counter  $t$  to zero.

**V.a.** Given a firm's belief  $\lambda_{n,t}$ , let firm  $n$  set price  $P^*(\lambda_{n,t}, z_n)$ . We generate log sales by drawing log demand shocks  $\varepsilon_{n,t} \sim N(m, s^2)$  through:

$$q_{n,t} = -\sigma_n P^*(\lambda_{n,t}, z_n) + \mu_i + s^k + \varepsilon_{n,t}$$

where  $\mu_i = (\sigma_i - 1)\log(\bar{P}_i^k) + \log(\zeta_i)$ . Update firm  $n$ 's posterior to:

$$\lambda_{n,t+1} = B(\lambda_{n,t}, P^*(\lambda_{n,t}, z_n), q_{n,t}, S^k)$$

We apply exogenous death shocks  $\delta$  to each firm. If a firm exits, then we replace it with a new firm that we assign as type  $\sigma_1$  firm with probability  $\lambda_0$ . Its prior becomes  $\lambda_0$ .

**V.b.** We calculate  $\varphi_{i,t+1}(\lambda, z)$  for each  $i = 1, 2$  and stop the simulation when the distribution of beliefs settles in both measures of active firms or when the number of simulation periods exceed some upper bound  $T > 1$ , i.e.  $\sup_{\lambda \in (0,1), z \in Z} \|\varphi_{i,t+1}(\lambda, z) - \varphi_{i,t}(\lambda, z)\| < \Delta_\varphi$  for  $i \in \{1, 2\}$  and/or  $t = T$ . Otherwise, we set  $t := t + 1$  and repeat step **V.a.**

**VII.** Calculate  $\bar{P}_i^{temp}$  with the *simulated* density  $\tilde{\Phi}_i(\lambda)$ :

$$\bar{P}_i^{temp} = \left( \sum_{\lambda} P^*(\lambda, z)^{1-\sigma_i} \tilde{\Phi}_i(\lambda, z) \right)^{\frac{1}{1-\sigma_i}}$$

where  $\tilde{\Phi}_i(\lambda, z)$  is the empirical cross-sectional probability distribution function of beliefs and idiosyncratic productivity. Also, calculate the total amount of labor in the economy as:

$$H^{temp} = S^k \sum_{i=1}^2 \frac{\eta_i}{z} \left[ \frac{\sum_{\lambda, z} P^*(\lambda, z)^{-\sigma_i} \tilde{\Phi}_i(\lambda, z)}{\sum_{\lambda, z} P^*(\lambda, z)^{1-\sigma_i} \tilde{\Phi}_i(\lambda, z)} \right]$$

If  $\sup_i |\bar{P}_i^{temp} - \bar{P}_i^k| < \Delta_\varepsilon$  **and**  $|H - H^{temp}| < \Delta_\varepsilon$ , then stop; otherwise, we set  $\bar{P}_1^{k+1} = \bar{P}_1^{temp}$  and  $\bar{P}_2^{k+1} = \bar{P}_2^{temp}$ . Let  $\chi^{k+1} > \chi^k$  if and only if  $H - H^{temp} < 0$ . We update the counter to  $k := k + 1$  and repeat step **II.**

## B Robustness exercises and extensions to framework

### B.1 Bayesian learning with a continuum of types

Our baseline framework in section 3 features the simplest form of active learning with firms varying their price as a control. Even though a firm is only uncertain about its demand elasticity and its type can only be high or low, our menu cost model with active learning is already consistent with the life cycle patterns that we showed in section 2. Nevertheless, we show that the key patterns and incentives for active learning are preserved when we use a more elaborate form of learning.

Consider a monopolistically competitive producer with constant marginal costs  $c$  who is faced with a linear demand curve of the following form:

$$q = \alpha - \sigma p + \varepsilon$$

where the demand shock satisfies  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ . There are two key differences in this framework compared to the baseline. First, the firm faces uncertainty about the intercept  $\alpha$  and the slope  $\sigma$  of its demand curve. Second, the pair  $(\alpha, \sigma)$  is now part of a continuous parameter space. A firm's prior belief is thus specified by a probability density function on  $(\alpha, \sigma)$  over  $\mathbb{R}^2$ . This is denoted by  $f(\alpha, \sigma | \theta, \mathcal{Q})$  where  $\mathcal{Q}$  denotes its information set that consists of the history of previous realized sales.  $\theta$  parameterizes the distribution  $f$ . We specify the firm's initial prior over  $(\alpha, \sigma)$  to be a multivariate normal distribution that is parameterized by the mean vector  $(a, s)'$  and variance-covariance matrix  $\Sigma$ . The latter is symmetric and satisfies:

$$\Sigma = \begin{pmatrix} v_a & v_{as} \\ v_{as} & v_s \end{pmatrix}$$

Therefore, we have  $\theta = (a, s, \text{vec}(\Sigma))' = (a, s, v_a, v_{as}, v_s)'$ .

Given a prior distribution  $f(\alpha, \sigma | \theta_t, \mathcal{Q}_{t-1})$  at time  $t$  where  $\mathcal{Q}_{t-1} = \{q_1, q_2, \dots, q_{t-1}\}$  and after observing a realized sales value of  $q_t$ , a firm will update its prior to a posterior distribution according to Bayes' rule:

$$\begin{aligned} f(\alpha, \sigma | \theta_{t+1}, \mathcal{Q}_t) &= \frac{f(q_t | \alpha, \sigma, \theta_t, \mathcal{Q}_{t-1}) \cdot f(\alpha, \sigma | \theta_t, \mathcal{Q}_{t-1})}{f(q_t | \theta_t, \mathcal{Q}_{t-1})} \\ &\propto f(q_t | \alpha, \sigma, \theta_t, \mathcal{Q}_{t-1}) \cdot f(\alpha, \sigma | \theta_t, \mathcal{Q}_{t-1}) \end{aligned}$$

The family of Gaussian distributions is conjugate to itself with respect to a Gaussian likelihood function, so this means that the posterior function must be of the multivariate normal form as well. A standard application of the Kalman filter shows that:

$$\begin{pmatrix} a \\ s \end{pmatrix}_{t+1} = \begin{pmatrix} a \\ s \end{pmatrix}_t + \frac{\Sigma_t X_t}{X_t' \Sigma_t X_t + \sigma_\varepsilon^2} \left( q_t - X_t' \begin{pmatrix} a \\ s \end{pmatrix}_t \right) \quad (\text{K1})$$

$$\Sigma_{t+1} = \Sigma_t - \frac{\Sigma_t X_t X_t' \Sigma_t}{X_t' \Sigma_t X_t + \sigma_\varepsilon^2} \quad (\text{K2})$$

where  $X_t = (1, -p_t)'$ . A more direct derivation can be found in [Zellner \(1971\)](#). The function that changes the parameters from the prior distribution into their posterior counterparts, as a function of observed sales and a chosen price, is denoted by  $B : \Theta \times \mathcal{P} \times \mathbb{R}_+ \rightarrow \Theta$ . Thus, the above system of equations can be compactly written as  $\theta_{t+1} = B(\theta_t, p, q)$ . The ex ante expected profits are defined as:

$$\Pi(p; \theta) = \int_{\varepsilon \in \mathbb{R}} \int_{\alpha, \sigma \in \mathbb{R}^2} (p - c)(\alpha - \sigma p + \varepsilon) f(\alpha, \sigma | \theta, \mathcal{Q}_{-1}) q(\varepsilon; \sigma_\varepsilon^2) d(\alpha, \sigma) d\varepsilon$$

with  $q(\cdot; \sigma_\varepsilon^2)$  being a normal distribution with a mean of zero and a variance of  $\sigma_\varepsilon^2$ . Then, a firm's

Bellman equation can be written as:

$$V(\theta) = \max_{p \in \mathcal{P}} \left\{ \Pi(p; \theta) + \beta \int_{\varepsilon \in \mathbb{R}} \int_{\alpha, \sigma \in \mathbb{R}^2} V(B(\theta, p, \alpha - \sigma p + \varepsilon)) f(\alpha, \sigma | \theta, \mathcal{Q}_{-1}) q(\varepsilon; \sigma_\varepsilon^2) d(\alpha, \sigma) d\varepsilon \right\}$$

Under this setup, a firm chooses its optimal price by trading off two forces. To maximize its current profits, a firm chooses a price that maximizes myopic profits  $\Pi(p; \theta)$ . For a given prior  $\theta = (a, s, \text{vec}(\Sigma)')'$ , we can derive that the optimal myopic price equals:

$$\begin{aligned} p^{\text{my}}(\theta) &= \arg \max_{p \in \mathcal{P}} \Pi(p; \theta) \\ &= \frac{a + sc}{2s} \end{aligned}$$

However, a firm's price will affect its sales. The observed amount of sales in the future serves as an useful signal for the firm to update its prior beliefs. A firm internalizes this signal and thus needs to take into account how its price will affect its posterior beliefs. This consideration is also known as the trade-off between current control and estimation. More importantly, these incentives do not necessarily align with each other. The reasoning is as follows: For moderate beliefs, a firm prefers to choose a price that is not too extreme in order to maximize the myopic profits.<sup>48</sup> However, large deviations in a firm's price are more likely to result in large deviations in a firm's future sales which in turn means that its signals become more volatile and are thus more informative. In the end, a firm needs to strike a balance between maximizing strictly concave myopic profits and a convex continuation value.

To show this balance, we work out a numerical two-period version of the above framework. We denote  $\theta_t = (a_t, s_t, v_{a,t}, v_{as,t}, v_{s,t})$ . There are only two periods, thus in the second and last period, we must have:

$$\begin{aligned} V_2(\theta_2) &= \max_{p \in \mathcal{P}} \Pi(p; \theta_2) \\ &= \max_{p \in \mathcal{P}} (p - c)(a_2 - s_2 p) \\ &= (p^{\text{my}}(\theta_2) - c)(a_2 - s_2 p^{\text{my}}(\theta_2)) \\ &= \frac{(a_2 - s_2 c)(3a_2 - s_2 c)}{4s_2} \end{aligned}$$

---

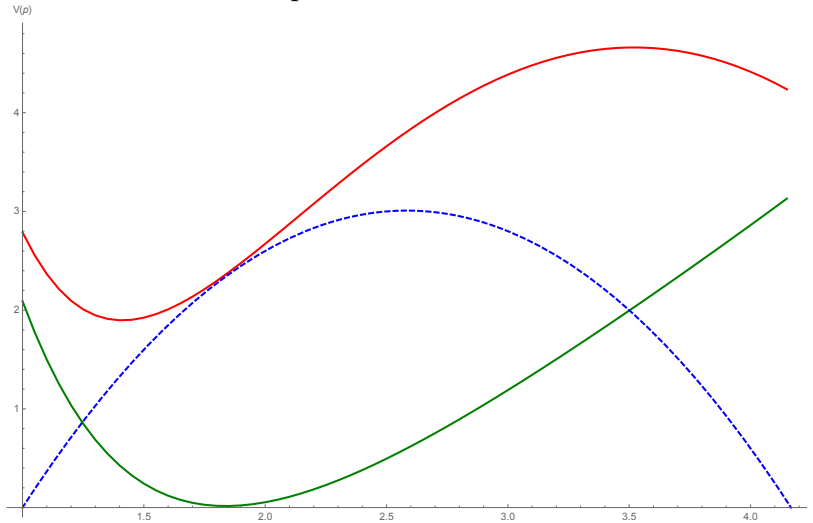
<sup>48</sup>Further, this price must exist and is unique since the profit function is strictly concave in  $p$  for every pair  $(\alpha, \sigma)$ .

By backward induction, we obtain:

$$\begin{aligned}
V_1(\theta_1) &= \max_{p \in \mathcal{P}} (p - c)(a_1 - s_1 p) + \beta \int_{\varepsilon \in \mathbb{R}} \int_{\alpha, \beta \in \mathbb{R}^2} \frac{(a_2 - s_2 c)(3a_2 - s_2 c)}{4s_2} f(\alpha, \sigma | \theta_1, q_1) q(\varepsilon; \sigma_\varepsilon^2) d(\alpha, \sigma) d\varepsilon \\
\text{s.t. } a_2 &= a_1 + \frac{(v_{a,1} + v_{as,1} p)(\alpha - \sigma p + \varepsilon - a - sp)}{v_{a,1} + 2v_{as,1} p + v_{s,1} p^2 + \sigma_\varepsilon^2} \\
s_2 &= s_1 + \frac{(v_{as,1} + v_{s,1} p)(\alpha - \sigma p + \varepsilon - a - sp)}{v_{a,1} + 2v_{as,1} p + v_{s,1} p^2 + \sigma_\varepsilon^2}
\end{aligned}$$

In the following example, we initialize the prior through  $\theta_1 = (5, -1.2, 0.5, -0.1, 2)'$  and normalize  $\sigma_\varepsilon^2 = 1$ . The graph below reflects the reasoning we just described.

Figure A.XVIII: Numerical Example of the Two-Period Model: Continuum of Types



Note: The figure shows the static profits (blue), the continuation value (green), and the total payoff (red) of the two-period model with a continuum of types. The y-axis represents the total payoff and the x-axis the price.

We only consider prices for which quantities are non-negative in expectation (with respect to the demand shock  $\varepsilon$  and the prior distribution), thus we define  $\mathcal{P} = [0, \frac{a_1}{s_1}]$  in this example. The blue, dashed myopic profits are concave as expected. If the firm ignores its incentives for estimation (i.e., changing its price to affect its posterior beliefs), then it is optimal to set  $p^{\text{my}}(\theta_1) = 2.58$ . However, setting a more extreme price delivers a more informative signal in the second period. This is reflected in the convex shape of the continuation value (depicted in green). In the end, a rational firm balances the trade-off between control and estimation. As a result, it maximizes the sum of myopic profits and its continuation value, which is depicted in the red line. The maximum of this function is obtained at  $p^*(\theta_1) = 3.5$ . By definition, the firm engages in active learning as  $p^{\text{my}}(\theta_1) \neq p^*(\theta_1)$ .

Note that the mechanics of active learning in this example with a continuum of types is identical to the two-period example we illustrated in section 3.2.1. In the baseline setup, a firm also faces

the trade-off between current control and estimation through concave myopic profits and a strictly convex continuation value. As a result, our results on the propagation of nominal shocks in section 4 should be robust to a more complicated version of active learning.

In fact, the incentives for active learning are stronger under a setup with a continuum of types. To understand this argument, we rely on the insights of [Kiefer and Nyarko \(1989\)](#). They show that under a setup with a linear demand curve all limiting beliefs and policy pairs  $(\bar{\theta}', \bar{p})$  must satisfy a set of three properties that we outline below:

$$\bar{\theta} = B(\bar{\theta}, \bar{p}, \alpha - \sigma\bar{p} + \varepsilon) \quad (\text{B1})$$

$$\Pi(\bar{p}, \bar{\theta}) = \max_{p \in \mathcal{P}} \Pi(p; \bar{\theta}) \quad (\text{B2})$$

$$\mathbb{E}(\alpha|\bar{\theta}) - \mathbb{E}(\sigma|\bar{\theta})\bar{p} = \alpha - \sigma\bar{p} \quad (\text{B3})$$

Equation B1 is also known as belief invariance and follows directly from the definition of a limiting belief. In the limit (if one exist), beliefs converge to a constant vector that is defined as the fixed point of the function  $B$  conditional on  $\bar{p}$ . If beliefs do not change in the limit, then there are no incentives to actively learn. As a result, the optimal policy must be the myopic one conditional on the limiting beliefs  $\bar{\theta}$  as described in equation B2. [Kiefer and Nyarko \(1989\)](#) refer to this policy as one-period optimization. Further, if prices are forever held at  $\bar{p}$ , then a firm will at least infer the true amount of sales associated at the price  $\bar{p}$ . Equation B3 is also known as the mean prediction property.

The solution  $(\bar{\theta}', \bar{p})$  that satisfies B1, B2 and B3 contains the correct limit belief but is in general not unique. [Wieland \(2000a\)](#) shows that any solution that contains incorrect limit beliefs must satisfy the following three properties:

**Perfect correlation.**  $\frac{\bar{v}_{as}^2}{\bar{v}_a \bar{v}_s} = 1$ .

**Uncertainty.**  $\bar{v}_a, \bar{v}_s > 0$ .

**Limit actions.**  $\bar{p} = -\frac{\bar{v}_{as}}{\bar{v}_s} = -\frac{\bar{v}_a}{\bar{v}_{as}}$ .

As a result, there is a set of incorrect, confounding beliefs under the continuum of types case. Recall from section 3 that a firm does not learn anything under the confounding belief and thus avoids setting prices that are equal to the confounding price. This is reflected by the discontinuity in policy function under the extreme active learning regime. Under a continuum of types, there are a multitude of such points. Thus, firms vary their prices more due to active learning in this case.

Another advantage of restricting our attention to active learning in which there are only two types,  $(\mu_1, \sigma_1)$  and  $(\mu_2, \sigma_2)$  with  $\sigma_2 > \sigma_1$ , is that equations B1, B2, and B3 can be used to show that there exists only one limit belief that does not converge to the truth (i.e.  $\lambda \notin \{0, 1\}$ ). This incorrect limit belief is equal to:

$$\bar{\lambda} = \frac{\sigma_2 \Delta \sigma c - \mu_2 (\sigma_1 + \sigma_2) + 2\mu_1 \sigma_2}{\Delta \sigma (\Delta \sigma c - \Delta \mu)}$$

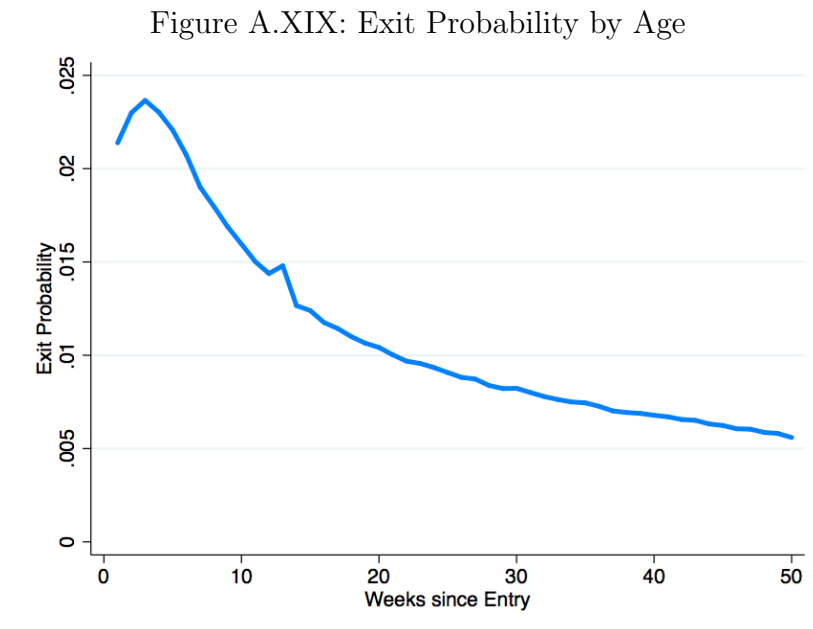
where  $c$  denotes a firm's marginal cost of production.

## B.2 Age-dependent exit rates

In our baseline framework, the exit rate is fixed at  $\delta > 0$ , which applies for each product. However, just like for firms (Caves, 1998), younger products are more likely to exit the market. Our assumption of a constant exit rate that is independent of the product's age could potentially bias our results on the propagation of nominal shocks. This is because the composition of products is biased towards younger products that experience a higher frequency and absolute size of price adjustments. In this section, we show that the assumption of a constant exit rate does not significantly bias the results generated by the baseline framework.

Therefore, we first show in the IRI Symphony data that product-level exit declines in age. Second, we propose an extension of our baseline framework to incorporate this observation. Third, we calibrate the extended framework and recalculate the real effects of a nominal shock.

In the first exercise, we compute the fraction of products that exit the market by age for each year and product category. Then, for each age bin, we average across years and product categories and plot these average exit rates by product age. The result can be found in the figure below that shows that exit rates indeed slope downward with age.



Note: The graph plots the average probability of exit of a UPC-store pair as a function of its age. We first compute the probability of exit for each category in the Symphony IRI data set. We then aggregate across categories using equal weights. The  $y$ -axis denotes the average probability exit in a given week, and the  $x$ -axis denotes the number of weeks the product has been observed in the data after it entered the market.

An alternative way of showing this fact is by estimating the product hazard function. There are



multiple ways of doing this estimation, but we adopt a parsimonious one. Under this method, we allow for a high degree of flexibility in the hazard rate by estimating it parametrically through the Weibull distribution. The hazard function is then given by:

$$h(a) = \lambda p a^{p-1}$$

where the scale parameter is denoted by  $\lambda$ . The shape parameter  $p$  indicates whether the hazard rate varies with age. A value  $p < 1$  means that the exit rates decline with a product's age whereas  $p = 1$  and  $\lambda = \delta$  corresponds to the exponential hazard function that we assume in our baseline framework. Thus, the Weibull distribution is flexible in that it allows for age-varying hazard rates. Another advantage of the Weibull specification is that it is straightforward to calibrate. Let  $T$  be a random variable that denotes the product's duration, then the following equalities hold whenever  $T \sim \mathcal{W}(\lambda, p)$ :

$$\begin{aligned}\mathbb{E}(T) &= \lambda^{-1/p} \Gamma(1 + p^{-1}) \\ V(T) &= \lambda^{-2/p} [\Gamma(1 + 2/p) - \Gamma(1 + p^{-1})^2]\end{aligned}$$

where  $\Gamma(\cdot)$  denotes the gamma function. The IRI data shows that the average amount of weeks that a product lasts in the market is 81.25 weeks. Furthermore, its variance is given by 8326.5. Then, we can form a system of two equations in the pair of unknowns  $(\lambda, p)$ . Its solution is given by  $\hat{\lambda} = 48.13$  and  $\hat{p} = 0.8922 < 1$ . Note that our calibrated value for  $p$  is not too far away from unity. This distance means that even though hazard rates decline, the product exit rates do not depend too strongly on age.

Note that the previous method relies on the structure of the Weibull distribution. Alternatively, we perform a non-parametric exercise. It is fairly difficult to obtain hazard functions non-parametrically, but we can still infer whether product-level exit rates depend on age in a non-parametric fashion. Recall that survival and hazard functions are related through the following identity:

$$-\log(S(a)) = \int_0^a h(\tau) d\tau$$

A concave, increasing cumulative hazard function then indicates that hazard rates decline with the product's age. This is useful as it is straightforward to obtain the survival function non-parametrically through the Kaplan-Meier estimator  $\hat{S}(a)$ .

To capture age-dependent exit rates in our framework, we extend the baseline model of section 3 by assuming that product-level exit rates depend on age as follows:

$$\delta(a) = \delta_0 \exp(-\delta_1 a)$$

The parameters  $\delta_0$  and  $\delta_1$  can then be chosen accordingly to match our observations from the above

graph. This estimation can be done by running a linear regression of average exit rates (in natural logs) on age. The estimated intercept and slope of this regression then correspond to  $\widehat{\delta}_0$  and  $\widehat{\delta}_1$ . We could also match the observed (cumulative) hazard rate function. In this case, we would add the two parameters  $\delta_0$  and  $\delta_1$  to our calibration.

### B.3 Age profile for product-level sales

In section 2, we showed that entering products play a substantial role in the aggregate economy. Approximately 45 percent of products in the US market entered in the last five years and they account for about 30 percent of total expenditures. However, entering products do not immediately reach these high levels of sales because, for example, the need to build customer bases. [Paciello, Pozzi and Trachter \(2014\)](#) argue that the pricing dynamics of firms are heavily influenced by customer retention concerns that are relatively more important for entering products. As a result, our quantitative results could be biased whenever we do not take into account that product sales require some time to be built up. In the following sections, we present two extensions to alleviate these concerns.

#### B.3.1 Demand shocks with age-dependent trend

The easiest, albeit mechanical, way of incorporating the fact that entering products' sales grow over time starting from a relatively low level is through an age trend in demand shocks. We add the following adjustment to the baseline framework. Assuming that the realization of taste shocks are independent across all groups and over time, we specify demand shocks to depend on a product's age through:

$$\alpha_t^i(k, a) = \omega^i(k, a)\alpha_t^i(k)$$

Then, the price index becomes:

$$P_{it} = \left( \int_{k \in J_i} \int_a \omega^i(k, a) p_t^i(k)^{1-\sigma_i} da dk \right)^{1/1-\sigma_i}$$

Recall that we are trying to capture the fact that entering products' sales start at a relatively low level and grow asymptotically towards some steady rate in a concave fashion. This steady rate is achieved fairly quickly in our dataset and occurs within the first three months after the product enters the market. As a result, we use the following functional form:

$$\omega^i(k, a) = \iota a^v$$

where we calibrate the initial level  $\iota$  and age-dependent slope  $v$  to match the age profile of sales.

Even though younger firms contribute less to output under this specification, their incentives to actively learn are higher given the prospects of higher sales in the future. In other words, the opportunity cost of poor sales at entry is low relative to future potential sales. These two forces contribute in different directions when measuring the response of real output to a nominal shock. As a result, under this specification, the effects on real output remain quantitatively similar to those obtained in our benchmark model.

### B.3.2 Customer base

In this section, we extend the canonical framework of [Golosov and Lucas \(2007\)](#) by adding a customer base. We show that such a model can rationalize the fact that product-level sales are dependent on age but this fact is not consistent with the documented life cycle patterns on product pricing.

Therefore, we model the customer base by incorporating external habits from the consumer side as is done in [Gilchrist, Schoenle, Sim and Zakrajšek \(2017\)](#). Under this setup, the aggregate consumption good  $C_t$  consists of a continuum of monopolistic competitive goods and is constructed as follows:

$$C_t = \left[ \int_0^1 \left( \frac{c_{it}}{b_{it-1}^\eta} \right)^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}$$

where  $b_{it}$  is the habit stock associated with good  $i$  at time  $t$ . The good-specific habit stock is assumed to be external: consumers take this level of stock as given. In addition to being more tractable, the assumption of external habits avoids the time-inconsistency problem of a firm setting its price associated with good-specific internal habits ([Nakamura and Steinsson, 2011](#)). Thus, we impose an exogenous law of motion for the external habit:

$$b_{it} = (1 - \delta^C) b_{it-1} + \delta^C c_{it}$$

where  $\delta^C$  denotes the depreciation rate of the customer base. Given the fact that consumers take the stocks of external habits  $\{b_{it}\}_i$  as given at time  $t$ , its good-specific demand can be derived as:

$$c_{it} = \left( \frac{p_{it}}{P_t} \right)^{-\sigma} (b_{it-1})^{\eta(1-\sigma)} C_t$$

The CES price index, adjusted for external habits, is denoted by:

$$P_t = \left( \int_0^1 (p_{it} b_{it-1}^\eta)^{1-\sigma} di \right)^{\frac{1}{1-\sigma}}$$

Identical to the baseline framework, we assume that each firm produces with a labor only production function that features constant returns to scale. Given a consumer's demand for good  $c_{it}$ , we can derive that a monopolistically competitive firm  $i$ 's profits are equal to:

$$\pi_{it} = \left( p_{it} - \frac{W_t}{z_{it}} \right) \left( \frac{p_{it}}{P_t} \right)^{-\sigma} (b_{it-1})^{\eta(1-\sigma)} C_t$$

Furthermore, we assume that firms are faced with a nominal rigidity in the form of a menu cost (denoted in units of labor). A firm's dynamic programming problem is summarized by the following Bellman equation:

$$V(b_{-1}, z, p_{-1}) = \max \{ V^A(b_{-1}, z), V^N(b_{-1}, z, p_{-1}) \}$$

where the value of adjusting is given by:

$$\begin{aligned} V^A(b_{-1}, z) &= \max_p \left( \frac{p}{P} - \frac{W}{zP} \right) \left( \frac{p}{P} \right)^{-\sigma} b_{-1}^{\eta(1-\sigma)} C - \psi \frac{W}{P} \\ &\quad + \beta \int_{z'} V \left( b, z', \frac{p}{1 + \tilde{\pi}} \right) dG(z', z) \\ \text{s.t. } b &= B(b_{-1}, p) = (1 - \delta^C) b_{-1} + \delta^C \left[ \left( \frac{p}{P} \right)^{-\sigma} b_{-1}^{\eta(1-\sigma)} C \right] \end{aligned}$$

which can be rewritten as:

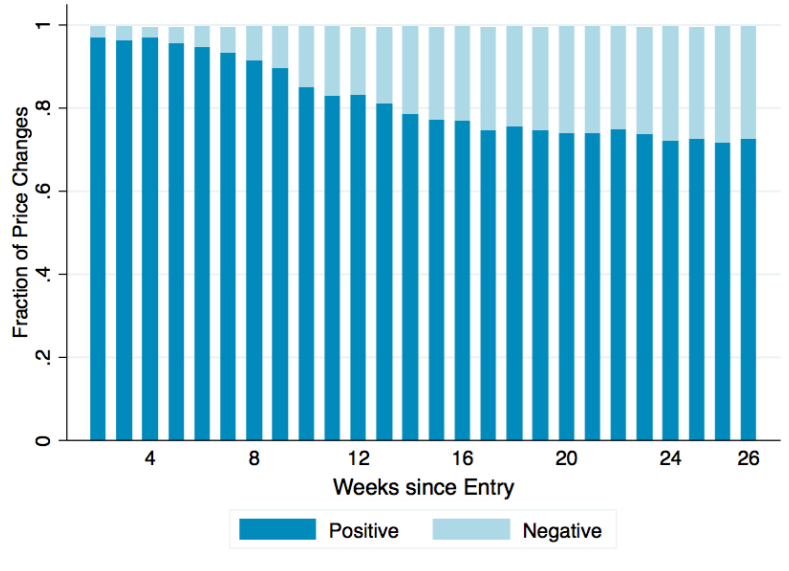
$$\begin{aligned} V^A(b_{-1}, z) &= \max_p \left( \frac{p}{S} - \frac{\omega}{z} \right) \left( \frac{p}{S} \right)^{-\sigma} b_{-1}^{\eta(1-\sigma)} \left( \frac{P}{S} \right)^{\sigma-2} - \frac{\psi \omega S}{P} \\ &\quad + \beta \int_{z'} V \left( B \left( b_{-1}, \frac{p}{1 + \tilde{\pi}} \right), z', \frac{p}{1 + \tilde{\pi}} \right) dG(z', z) \end{aligned}$$

The two crucial parameters that govern the customer base are  $\eta$  and  $\delta^C$ . To verify whether a standard price-setting model with customer base incentives are consistent with our stylized facts, we simulate the model by choosing the parameters  $\eta$  and  $\delta^C$  externally. [Foster, Haltiwanger and Syverson \(2016\)](#) structurally estimate these parameters and find values of 0.92 and 0.188 for  $\eta(1-\sigma)$  and  $\delta^C$  respectively. However, this depreciation rate is based on an annual basis. Our framework's unit of time is at the weekly level. Therefore, we set  $\delta^C$  to satisfy:

$$(1 - \delta^C)^{52} = 1 - 0.188$$

which gives  $\delta^C \simeq 0.003997$ . The following figure calculates the fraction of positive and negative price changes by product age in the simulated data.

Figure A.XX: Fraction of Positive and Negative Price Changes

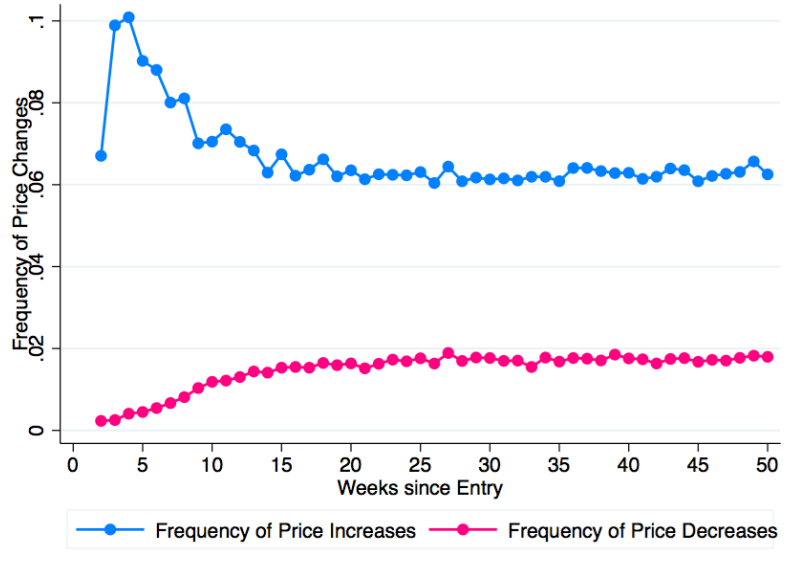


Note: The graph plots the share of regular price increases and regular price decreases conditional on adjustment. The data is generated by simulating the Golosov-Lucas model with a customer base under the calibration described in section B.3.2. It considers the first six months after the entry of a product in the model.

The figure shows that the majority of the price changes for young products is positive, which the data contradicts. In the IRI data, we observe that roughly 60 percent of all price changes are positive, and this fraction is largely stable over the product’s life cycle.

In a model with customer retention, firms have “invest” and “harvest” motives. By construction in a customer base model, current sales are dependent on the level of previous sales. This dependence means that recently entered products have relatively low prices that induce a high volume of sales and, hence, a large customer base. Once this customer base is built up to a sufficiently high level, a firm exercises its market power by increasing its price and generates profits. The simulated data indicates that customer bases are built up extremely fast. Firms set low prices to attract customers and then immediately exercise their market power afterward in a gradual fashion. This is also reflected in the frequency of price adjustments by product age. In the early stage of a product’s life cycle, firms’ incentives for harvesting are extremely large. Thus, they are willing to pay the menu cost to increase their prices. This incentive then slows down over time, which is consistent with the IRI data. However, we also observe in the data that the frequency of negative price adjustments weakly decreases with product age. This is contradicted by the customer base model because the frequency of negative price changes actually weakly increases.

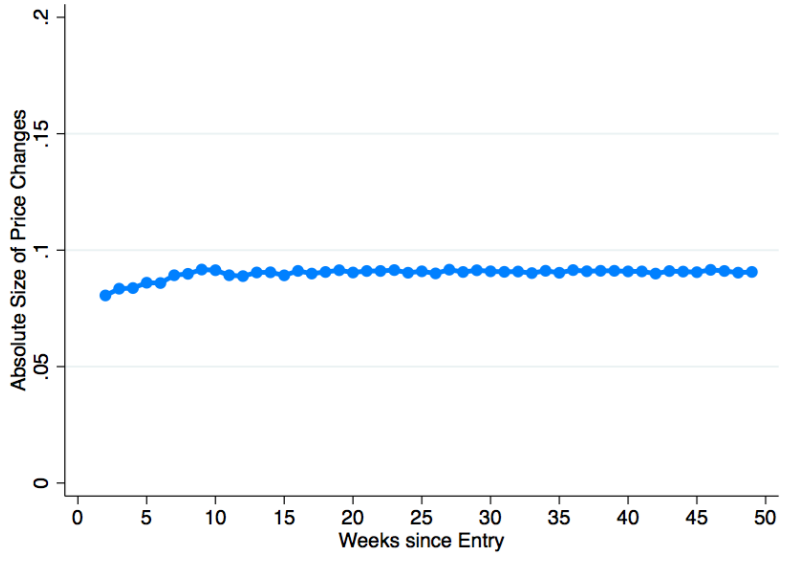
Figure A.XXI: Frequency of Price Increases and Decreases at Entry



Note: The graph plots the average weekly frequency of price adjustments of products entering the market. The data is generated by simulating the Golosov-Lucas model with a customer base under the calibration described in section B.3.2. The y-axis denotes the probability that the product adjusts prices in a given week and the x-axis denotes the number of weeks the product has been observed in the data after it entered the market. The blue line indicates the frequency of positive price adjustments and the red line the frequency of negative price adjustments.

We also show in the data that the absolute size of price changes declines with the product’s age. However, we see from the figure above that a customer base does not generate such a pattern. In fact, it seems to show that the absolute size of price changes is independent of a product’s age.

Figure A.XXII: Absolute Value of Price Changes at Entry



Note: The graph plots the average absolute size of price adjustments of entering products. The data is generated by simulating the Golosov-Lucas model with a customer base under the calibration described in section B.3.2. The y-axis is the absolute value of the log price changes in that week, and the x-axis denotes the number of weeks after the product entered.

## B.4 Endogenous entry over the cycle

Our baseline framework reflects a stationary environment in which the number of entrants is constant over time. Even though our baseline framework is successful in replicating the stylized facts that we show in section 2, it does not capture whether the magnitude of the previously mentioned amplification varies over the business cycle. In this section, we construct a fully dynamic version of our model to investigate whether cyclical changes in the extensive margin of products play an important role in the amplification of nominal shocks. Following work by Lee and Mukoyama (2015), aggregate productivity shocks are the source of aggregate fluctuations. In addition, the entry rate of products is endogenous, which allows it to vary with the aggregate state of the economy. In this section, we present the necessary ingredients to allow for a procyclical entry rate in our model.

Consumers and firms are identical as in the baseline framework. However, a firm's productivity now consists of two components: an idiosyncratic one as described in section 3 and an aggregate component  $Z_t$ . Aggregate productivity  $Z_t$  follows a symmetric, two state Markov chain. Thus, we have  $Z_t \in \{Z_L, Z_H\}$ . The transition matrix between high and low aggregate productivity is then characterized by:

$$\begin{bmatrix} \vartheta & 1 - \vartheta \\ 1 - \vartheta & \vartheta \end{bmatrix}$$

The average duration of a state is characterized by:

$$\sum_{\tau=1}^{\infty} \tau(1-\vartheta)\vartheta^{\tau-1} = \frac{1}{1-\vartheta}$$

As before, we assume that aggregate spending grows deterministically at the rate  $\tilde{\pi}$ , i.e.  $S_t = \exp(\tilde{\pi} \cdot t)$ . Recall that a price-adjusting incumbent firm has a two-dimensional idiosyncratic state. We denote this state by  $\mathbf{v}_t^i(k) = (\lambda_t^i(k), z_t^i(k))$ . Then, a firm's ex-ante expected profits can be written as a function of the idiosyncratic state  $\mathbf{v}_t^i(k)$  and the aggregate state  $\xi_t \equiv (S_t, Z_t)$ :

$$\Pi_t(p; \mathbf{v}_t^i(k), \xi_t)$$

A firm chooses a path of prices  $\{p_t^i(k)\}_{t \geq 0}$  to maximize the expected, discounted profits. The firm's problem in Bellman form is equal to:

$$V(\mathbf{v}, p_{-1}; \xi) = \max \{V^A(\mathbf{v}; \xi), V^{NA}(\mathbf{v}, p_{-1}; \xi)\}$$

where the value of adjusting and not adjusting are respectively given by:

$$\begin{aligned} V^A(\mathbf{v}; \xi) &= \max_{p \geq 0} \Pi(p; \mathbf{v}, \xi) - W \cdot \psi \\ &+ \mathbb{E}_{\xi'} \left[ \mathbf{q}(\xi, \xi') \lambda \int_{z'} \int_{\varepsilon} V \left( b_1(\lambda, \log(\frac{p}{1+\tilde{\pi}}), \varepsilon), z', \frac{p}{1+\tilde{\pi}}, \xi' \right) dF(\varepsilon) dG(z', z) \right. \\ &\left. + \mathbf{q}(\xi, \xi') (1-\lambda) \int_{z'} \int_{\varepsilon} V \left( b_2(\lambda, \log(\frac{p}{1+\tilde{\pi}}), \varepsilon), z', \frac{p}{1+\tilde{\pi}}, \xi' \right) dF(\varepsilon) dG(z', z) \right] \xi \\ V^{NA}(\mathbf{v}, p_{-1}; \xi) &= \Pi(p_{-1}; \mathbf{v}, \xi) \\ &+ \mathbb{E}_{\xi'} \left[ \mathbf{q}(\xi, \xi') \lambda \int_{z'} \int_{\varepsilon} V \left( b_1(\lambda, \log(\frac{p_{-1}}{1+\tilde{\pi}}), \varepsilon), z', \frac{p_{-1}}{1+\tilde{\pi}}, \xi' \right) dF(\varepsilon) dG(z', z) \right. \\ &\left. + \mathbf{q}(\xi, \xi') (1-\lambda) \int_{z'} \int_{\varepsilon} V \left( b_2(\lambda, \log(\frac{p_{-1}}{1+\tilde{\pi}}), \varepsilon), z', \frac{p_{-1}}{1+\tilde{\pi}}, \xi' \right) dF(\varepsilon) dG(z', z) \right] \xi \end{aligned}$$

where the stochastic discount factor is given by  $\mathbf{q}(\xi, \xi') = \beta \frac{u'(C')}{u'(C)}$ .

There is a pool of potential entrants. In the beginning of a period, everyone observes the aggregate state  $\xi_t = (S_t, Z_t)$ . Furthermore, every potential entrant is endowed with an idiosyncratic productivity  $z$  drawn from the exogenous distribution  $H$ . If a potential entrant wants to become a producer, she needs to pay a fixed entry cost  $c_E$ , which is denoted in units of labor. At entry only, we assume that the entrant is allowed to choose its price without incurring the menu cost. The value of becoming a producer then becomes:

$$V_t^E(z_t; \xi_t) = V_t^A(\lambda_0, z_t; \xi_t) + W_t \cdot \psi$$

This structure indicates that only those potential entrants with sufficiently high values for  $z_t$  can



actually enter the product market. In fact, there is a threshold value  $z_t^*$  that is defined by the free entry condition:

$$\begin{aligned} V_t^E(z_t^*; \xi_t) &= W_t \cdot c_E \\ &= \omega S_t \cdot c_E \end{aligned}$$

such that potential entrants become producers if and only if their drawn level of productivity satisfies  $z_t \geq z_t^*$ .

To analyze the model in general equilibrium, we need to consider an environment in which consumers and firms engage in optimal behavior while the markets for goods and labor clear. The optimization behavior is apparent from the representative consumer's first order conditions and firms' value functions. The market for goods clears by construction because we plug the optimal consumer demand into the firm's optimization problem. As a result, we only need to clear the labor market.

Let  $\varphi_t(\lambda, z, p_{-1})$  denote the labor demand of a firm with idiosyncratic state  $(\mathbf{v}, p_{-1}) = (\lambda, z, p_{-1})$ . Assuming the mass of potential entry in each period is one, its distribution at period  $t$  is denoted by  $\mu_t(\mathbf{v}, p_{-1})$ . Then, the quantity of labor demanded by incumbent producers is:

$$L_t^{d,p} = N_t \cdot \int_{\mathbf{v}, p} \varphi_t(\mathbf{v}, p_{-1}) d\mu_t(\mathbf{v}, p_{-1})$$

where  $N_t$  denotes the actual mass of potential entrants in period  $t$ . Furthermore, labor is used for the costs of entry. Thus, total labor demand  $L_t^d$  can be characterized as:

$$L_t^d = L_t^{d,p} + N_t \cdot (1 - H(z_t^*)) \cdot c_E$$

The above equation characterizes the optimal labor supply. Then, the market for labor clears when the labor supply equals labor demand.

Lastly, we describe how to calibrate the exogenous, aggregate productivity process. Following the RBC literature, we assume that the level of aggregate productivity can be well approximated by an autoregressive process. We use Fernald's (2014) quarterly utilization-adjusted time series for TFP and detrend it with the HP filter (using a smoothing parameter of 1,600). Then, we run a linear regression of this detrended series on its lagged counterpart and calculate the standard deviation of the residuals. These residuals are then interpreted as shocks to aggregate productivity. We find a value of 0.009 at the quarterly level. Converting this to the weekly level, we obtain  $\sigma_Z = \frac{0.009}{\sqrt{12}} \simeq 0.0026$ . We normalize the trend for aggregate productivity to unity and define a boom or bust as a one standard deviation increase or decrease from the trend respectively. As a result, we obtain  $Z_H = 1.0026$  and  $Z_L = 0.9974$ .<sup>49</sup> Furthermore, we assume that the average duration

---

<sup>49</sup>Vavra (2014) performs a similar exercise with real output per hours worked and finds a standard deviation for aggregate productivity shocks of 0.006 at the monthly level. This means a value of  $\sigma_Z = \frac{0.006}{\sqrt{4}} \simeq 0.003$  which is similar to what we find above.

of a boom or bust is 35 months, which is approximately 140 weeks. This indicates a value of the transition probability  $\vartheta = 1 - \frac{1}{140} \simeq 0.99286$ .