Price Setting with Customer Retention*

Luigi Paciello
Einaudi Institute for Economics and Finance and CEPR

Andrea Pozzi
Einaudi Institute for Economics and Finance and CEPR

Nicholas Trachter
Federal Reserve Bank of Richmond

November 4, 2013

Abstract

We study a model where customers face frictions when changing their supplier, generating sluggishness in the firm's customer base. Firms care about retaining customers and this affects their pricing strategy. We characterize optimal pricing in this model and estimate it using data on the evolution of the customer base of a large US retailer. The introduction of customer retention concerns reduces markups, more markedly for less productive firms. We show that our model delivers pro-cyclical markups, as well as heterogeneous pass-through of cost shocks. These results help explaining recent empirical evidence.

JEL classification:

Keywords:

---

*We thank Mark Aguiar, George Alessandria, Fernando Alvarez, Lukasz Drozd, Stefania Garetto, Mike Golosov, Bob Hall, Jonathan Heathcote, Hugo Hopenhayn, Eric Hurst, Pat Kehoe, Pete Klenow, Francesco Lippi, Ricardo Lagos, Erzo Luttmer, Kiminori Matsuyama, Guido Menzio, Dale Mortensen, Juan Pablo Nicolini, Ezra Oberfield, Guillermo Ordonez, Facundo Piguillem, Valerie Ramey, Leena Rudanko, Stephanie Schmitt-Grohe, Martin Uribe, Randall Wright, and participants at the 2nd Rome Junior Macroeconomics conference, 10th Philadelphia Search and Matching conference, ESSET 2013, the Minnesota Macro Workshop 2013, the Macro Faculty Lunch at Stanford, the MBF-Bicocca conference, the Ohio State University, and University of Tor Vergata for useful comments. Luigi Paciello thanks Stanford University for hospitality. The views expressed in this article are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1 Introduction

There is ample consensus that dynamics in the firms' customer base -i.e. the set of customers currently purchasing from a firm- are important determinants of performance, and that firms act to influence the evolution of their customer base (Foster et al. (2012)). However, how relevant customer markets are for the firms' optimal pricing policy is an open question (Blanchard (2009)). In this paper we study a model of price setting with customer retention concerns. We characterize the equilibrium of the model and estimate it using novel micro data. Using the estimated model, we study the implications of customer markets for firm's pricing policy. We show that introducing customer retention concerns into a standard price setting model has relevant consequences for equilibrium price dynamics: it substantially increases the pro-cyclicality of markup and decreases the pass-through of cost shocks.

We build on the seminal work on customer markets by Phelps and Winter (1970) and explicitly model the game between a customer and a firm. In particular, we model inertia in customer dynamics by introducing search frictions along the lines of Burdett and Coles (1997).[1] Customers demand an homogeneous good sold in different locations, each of them populated by a single firm. Every customer is matched to a particular firm at any point in time, and reacts to the pricing policy of a firm by deciding whether to exit the customer base of the firm and look for a new supplier. Customers incur in a cost if they change the firm they patronize, increasing inertia in the customer base.[2] In this framework, the firm considers its customers as an asset, and sets the price with no commitment and no discrimination taking into account the impact on consumer dynamics. Firms face a dynamic problem which is linked to two different margins of demand. The first one is the static *intensive margin* and concerns *how much* of the good the firm is able to sell to each of its customers, conditional on the customer buying from the firm. The second margin, which we label *extensive margin* of demand, concerns the decision of customers to exit the customer base. This dynamic margin relates to *how many* individuals decide to remain and buy from their old firm. Firms are characterized by different productivity levels, which evolve stochastically through time according to a Markovian process. The combination of search frictions on the consumer side and idiosyncratic firm productivity gives rise to equilibrium price dispersion, incentivizing customers to search.

We obtain several noteworthy results from the characterization of our model. First, we show that the optimal markup is a function of both the intensive and the extensive margin elasticity of demand. When customer retention concerns bind, the optimal markup is lower

[1] This approach is also used in the context of labor markets. See for instance Burdett and Mortensen (1998), Coles (2001) and Coles and Mortensen (2013).

[2] This cost can also be interpreted as switching cost in the spirit of Klemperer (1995).

than the markup that maximizes static profits. The higher the extensive margin elasticity, the lower the optimal markup. Moreover, the extensive margin is linked to a change in demand that last for several periods, as lost customers are hard to win back. Therefore, it weights more in the determination of prices than the intensive margin elasticity which only affects current demand. Even a small extensive margin elasticity can, therefore, have a large impact on optimal markups.

Looking at the cross-section of sellers within an industry, we show that shifts in productivity change the demand faced by the firm. In a Markovian equilibrium with persistence in productivity, current productivity carries information about future prices. If higher productivity reflects in lower prices, customers expect highly productive firms to offer better prices in the future. As a consequence, they are more reluctant to leave them and firms with higher productivity face lower extensive margin elasticity of demand and charge higher markup.

We complement our modeling effort with a two-pronged empirical analysis. From a descriptive point of view, we provide novel direct evidence of sluggishness in the customer base by looking at the evolution of the customer base of a large U.S. supermarket chain. Despite the growing attention towards the implications of customer retention, in fact, evidence in that regards is still mostly based on anecdotal or survey sources. Supermarket scanner data are particularly well suited to study customer base dynamics because they allow to track precisely customers' choice of the store where to buy. This allows us to document that there is persistence in a customer's choice of a retailer, and to show that a shopper decision to exit the customer base responds to the relevant price. On average, a one percent change in the price of the customer's typical basket of goods raises her likelihood of leaving the retailer by 0.014 percentage points. This estimate is important as it identifies the size of the friction determining customer dynamics in our model. Therefore, we target the extensive margin elasticity implied from our descriptive regression, as well as other standard macroeconomic targets, to quantify the model.

We use the estimated model to quantify the relevance of customer retention concerns for price dynamics in our model. As a benchmark, we compare to an alternative economy where we shut down the customer retention concerns, so that it approximates a model with standard constant elasticity of substitution (CES) preferences. At the parameter estimates the implied average markup is half of the equivalent statistic in the CES economy. Our model also generates more dispersion in markups and, therefore, less dispersion in prices. Finally, our model delivers strong comovement of markups with idiosyncratic productivity, consistent with evidence in Petrin and Warzynski (2012).

Our estimated model provides a laboratory to study the implications of customer markets for the response of markups to aggregate shocks. We quantify the cyclical behavior of markup simulating an aggregate technology and an aggregate demand shock. We show that customer retention concerns imply procyclical markup conditional on both aggregate technology and demand shocks, as conjectured by Rotemberg and Woodford (1991). These findings are in line with the empirical evidence provided in Nekarda and Ramey (2013) and would not generally arise in a standard CES model. The implications are quantitatively important: our model reduces the predicted business cycle fluctuations of output by 30% with respect to a setting where customer base concerns are not present.

In order to contrast the predictions of the model with abundant evidence from the international trade literature and in particular with recent empirical studies documenting that pass-through declines with productivity (Berman et al. (2012); Chatterjee et al. (2013)), we simulate the pass-through induced by a real exchange rate appreaciation. Our simulation delivers the same qualitative result.[3] However, we point out that this finding depends on restricting the analysis to a subsample of highly productive firms (the exporters). When we use the model to simulate the pass-through of a cost shock that hits a random subset of firms in the economy, and not only the most productive ones, we show that the implications are reversed: low productivity firms pass-through less than high productivity ones.

Customer markets were first analyzed in the context of macroeconomics quantifiable models by Phelps and Winter (1970), and by Rotemberg and Woodford (1991), who modeled the flow of customers as a function of the price posted by the firm. We microfound these approaches by having customer dynamics arising endogenously by solving the game between firms and customers. The literature on "deep habits" (Ravn et al. (2006), Nakamura and Steinsson (2011)) represents an alternative way to generate persistence in demand by introducing habits in consumption.

Our analysis of the implications of customer base preservation for pricing and markup ties into a growing body of literature using models where the market share of the firm is sluggish to study a number of issues such as pricing-to-market (Alessandria (2009), Drozd and Nosal (2012)) and firm investment (Gourio and Rudanko (2011)). We focus on the influence of customer base concerns on firm price setting as Burdett and Coles (1997), Menzio (2007) and Kleshchelski and Vincent (2009). Burdett and Coles (1997) study the relationship of equilibrium price dispersion with firm size in a model with endogenous entry of firms, while we allow for idiosyncratic shocks to productivity and study the pass-through of cost shocks. Menzio (2007) focuses on optimal price setting with commitment and asymmetric information between sellers and buyers, whereas we study a model with symmetric information and no commitment. Differently from Kleshchelski and Vincent (2009) we study an equilibrium

[3]Examples of other models consistent with the same predictions are Atkeson and Burstein (2008), Melitz and Ottaviano (2008), or Garetto (2012).

where firms are not symmetric allowing us to explore the relationship between markup, pass-through and productivity. None of these papers analyze the reaction to aggregate shocks.

By showing that consumers are influenced by prices in their decision of breaking long term relationships with suppliers, we add to the literature using scanner data to document empirical regularities in pricing and shopping behavior. A series of contributions (Aguiar and Hurst (2007), Coibion et al. (2012), Kaplan and Menzio (2013)) integrates store and customer scanner data to show that intensity of search for lower prices is depends on income and opportunity cost of time. We view our contribution as complementary to theirs as we instead focus on the elasticity of search to prices.

The rest of the paper is organized as follows. In Section 2 we lay out the model and characterize the equilibrium. Section 3 presents the data and descriptive evidence of the relationship between customer dynamics and prices. In Section 4 we discuss the estimation of the model. Results on markup dynamics and pass-through of cost shocks are presented in Section 5 and Section 6, respectively. Section 7 concludes.

2 The model

The economy is populated by a measure one of firms producing an homogeneous good, and a measure Γ of customers.

Firms. We use the superscript j, $j \in [0,1]$, to index a firm. The production technology of the good is linear in the unique production input, ℓ, and depends on the firm specific productivity z^j. That is, $y^j = z^j \ell$. We let w denote the marginal cost of the input ℓ, p denote the price of the good, and $\pi(p,z)$ the profit function. We assume that $\pi(p,z)$ is continuously differentiable in p and single-peaked. We assume that firm specific productivity z^j is i.i.d. across firms. Let $z \in [\underline{z}, \bar{z}]$ with $\bar{z} > \underline{z} > 0$, and $\mu(z)$ denote the unconditional density of productivity z. Also let $\mu(z'|\, z)$ denote the conditional density function of z' given z. The productivity shock is persistent in the sense that the cumulative distribution function, $\int_{\underline{z}}^{s} \mu(z'|\, z) dz'$, is weakly decreasing in z.

Customers. We use the superscript i, $i \in [0, \Gamma]$, to index a customer. Let I denote the income level of a customer each period. Let $d(p)$, $v(p)$ denote the static demand and customer surplus functions which, as in standard models, only depends on the current price p. We assume that: (i) $d'(p) < 0$, $d''(p) \geq 0$, with $\lim_{p\to\infty} d(p) = 0$; (ii) $\varepsilon_d(p) \equiv -\partial \ln d(p)/\partial \ln p \geq 1$; (iii) $\varepsilon_d'(p) \geq 0$, and (iv) $v'(p) < 0$; $v''(p) \leq 0$. Assumption (i) states that the demand function is decreasing and convex in prices and that the demand approaches zero as the price diverges, (ii) is required for positive markups, assumption (iii) implies that the demand elasticity increases with prices, while assumption (iv) simply states that the surplus of the customer is decreasing and concave in the price. Assumptions (i)-(iv) are typical properties arising in standard macro models with CRRA utility functions and CES demand across a large set of varieties. In Appendix A.1, we derive properties (i)-(iv) from first principles.

Notice that both demand and customer surplus functions do not depend on the identity of the customer i. Nevertheless, customers will face different realizations of both demand and surplus given they will face different prices p because customers are matched to different producers. Moreover, notice that we allow for the elasticity of demand $\varepsilon_d(p)$ to depend on the price p to nest recent models of variable markups and incomplete cost pass-through.[4] This allows to compare our mechanism to the existing literature.

Search and matching. Customers draw a random search cost $\psi \in [\underline{\psi}, \infty)$, drawn randomly each period from the same distribution $G(\psi)$, where $g(\psi) \equiv \partial G(\psi)/\partial\psi$; we restrict our attention to density functions that are continuous in $\left(\underline{\psi}, \infty\right)$. In the remaining of the paper we assume that $\underline{\psi} = 0$. If the customer incurs in the utility cost ψ, she exits the customer base of her current firm and is randomly matched to a new one, where the probability of matching with a particular type of firm is proportional to its customer base size. Matching occurs instantaneously and we allow customers to search at most once per period. Finally, we assume no recall of previously visited firms.

Timing of events. A period starts with all customers distributed across firms; we let m^j_{t-1} denote the mass of customers matched with firm j at the beginning of period t. The timing of events is the following: (i) productivity shocks z^j_t are realized and firm j posts price p^j_t to all its customers, (ii) each customer draws her search cost ψ^i_t and observes $\{p^j_t, z^j_t, m^j_{t-1}\}$, (iii) each customer decides between searching for a new firm -after incurring in search cost ψ^i_t- and being assigned randomly within period t, or remaining matched to firm j -and not incurring in cost ψ^i_t-, and (iv) customer surplus $v(p^j_t)$ and profits $\pi(p^j_t, z^j_t)$ are realized. We refer to (iii) as the extensive margin of demand, where each customer decides whether to remain in the customer base of firm j or exit, while we refer to (iv) as the intensive margin of demand.

2.1 The game between customers and firms

In this section we discuss the solution to the customers' search and firms' price-setting problems. Given the timing of the model, a firm and the customers matched to it play an anonymous sequential game, where the firm plays first by posting a price and customers best respond to it. We look for a Markov Perfect equilibrium where strategies are a function of the current state. Because neither the firm nor its customers can commit to future actions, they

[4] See Atkeson and Burstein (2008), Melitz and Ottaviano (2008), and Berman et al. (2012) for a review.

take as given the continuation value. This continuation value is encoded in the customer base size m and productivity level z, where we omitted the time index t and identity index j to ease on notation. Notice that there would time inconsistency if instead firms were setting a path for prices. Let $\tilde{\mathcal{P}}(m, z) : [0, \infty) \times [\underline{z}, \bar{z}] \to [0, \infty)$ be the pricing strategy of the firm mapping m and z to the space of prices. Also, let $\tilde{\chi}(p, z, m, \psi) : [0, \infty) \times [\underline{z}, \bar{z}] \times [0, \infty) \times [0, \infty) \to \{0, 1\}$ be the exit strategy of the customer mapping the price p, the productivity z, the customer base m, as well as the search cost ψ to $\{0, 1\}$, where $\tilde{\chi} = 1$ denotes the case of exiting the customer base of the firm.[5]

In our setup, as we will show later, the customer base enters multiplicatively the firm's problem, i.e. the value function of the firm is homogeneous in m. As a result, we will construct an equilibrium where strategies do not depend on the customer base. Therefore, $\mathcal{P}(z) : [\underline{z}, \bar{z}] \to [0, \infty)$ denotes the pricing strategy of the firm mapping z to the space of prices. Also, $\chi(p, z, \psi) : [0, \infty) \times [\underline{z}, \bar{z}] \times [0, \infty) \to \{0, 1\}$ denotes the exit strategy of the customer mapping z, p and ψ to $\{0, 1\}$.

2.2 The problem of a customer

Consider a customer buying goods from a given firm j. Let $\bar{V}(p, z)$ be the value function for a customer who decided to stay and buy from firm j, paying a price p for each unit of the good in the current period, and facing a pricing policy $\mathcal{P}(\cdot)$ in future states. We have

$$\bar{V}(p, z) = v(p) + \beta \int_{\underline{z}}^{\bar{z}} \left[\int_0^\infty \max\left\{ \bar{V}(\mathcal{P}(z'), z'), \hat{V} - \psi \right\} \; g(\psi) \; d\psi \right] \mu(z' | \, z) \; dz' \; , \tag{1}$$

where $\hat{V}$ is the outside option to the customer. Consider the search decision of a customer who draws a search cost ψ. The exit strategy of the customer is $\chi(z, p, \psi) = 1$ if $\bar{V}(p, z) < \hat{V} - \psi$, and $\chi(z, p, \psi) = 0$ otherwise. Because the value of exiting is decreasing in ψ and the value of not exiting does not depend on ψ, the exit strategy takes the form of a trigger, $\bar{\psi}$, such that the customer exits if $\psi < \bar{\psi}$. We next characterize the trigger policy.

Let $\bar{\mathcal{P}}(z)$ denote the function that solves $\bar{V}(\bar{\mathcal{P}}(z), z) \; = \hat{V}$, so that no customer decides to exit the customer base of firm j with productivity z if $p \leq \bar{\mathcal{P}}(z)$.[6] Using equation (1) we

[5] As noted in Coles (2001), supporting this type of equilibria requires imposing assumptions on off-equilibrium paths. To see this notice that in equilibrium firms are supposed to choose the same price whenever the productivity is the same. However, because future prices are taken as given in the current period, a firm has incentives to deviate from this price marginally and customers should not react by exiting the customer base of the firm as this small price change only lasts for one period. To avoid this type of behavior we assume that if customers observe a price which is not consistent with the equilibrium price they exist the customer base of the firm.

[6] The existence of $\bar{\mathcal{P}}(z)$ is a direct implication of $\bar{V}(p, z)$ being strictly decreasing in p. Because $v(p)$ is strictly decreasing in p, a direct application of the Contraction Mapping Theorem implies that $\bar{V}(p, z)$ is a

get that $\bar{\mathcal{P}}(z)$ solves

$$v\left(\bar{\mathcal{P}}(z)\right) = \hat{V} - \beta \int_{\underline{z}}^{\bar{z}} \left[\int_0^{\infty} \max\left\{\bar{V}(\mathcal{P}(z'), z'), \hat{V} - \psi\right\}\ g(\psi)\ d\psi\right] \mu(z'|\, z)\ dz'\ . \tag{2}$$

Notice that the variable $\bar{\mathcal{P}}(z)$ ranks the different firms according to the continuation value for the customer: when comparing two firms selling the good at the same current price p, the customer prefers the firm characterized by higher $\bar{\mathcal{P}}(z)$. Likewise, notice that $\bar{\mathcal{P}}(z)$ is the highest price the firm can charge so that no customer searches for anew firm.

Next we define the threshold for the search cost ψ below which customers decide to exit the customer base of a given firm. Let $\bar{\psi}(p, z)$ denote the threshold rule such that a customer decides to exit if $\psi \leq \bar{\psi}(p, z)$, so that we have

$$\bar{\psi}(p, z) = \begin{cases} 0 & \text{if } p \leq \bar{\mathcal{P}}(z)\ , \\ v(\bar{\mathcal{P}}(z)) - v(p) & \text{otherwise}\ , \end{cases} \tag{3}$$

and therefore the exit strategy is

$$\chi(p, z, \psi) = \begin{cases} 1 & \text{if } \psi \leq \bar{\psi}(p, z)\ , \\ 0 & \text{otherwise}\ . \end{cases}$$

The next lemma characterizes the threshold rule $\bar{\psi}(p, z)$.

Lemma 1 *For all $p > \bar{\mathcal{P}}(z)$, the threshold rule $\bar{\psi}(p, z)$ is strictly increasing in p. Moreover, (i) if $\bar{\mathcal{P}}(z) = \bar{\mathcal{P}}$ then $\bar{\psi}(p, z) = \bar{\psi}(p)$, (ii) if $\bar{\mathcal{P}}(z)$ is increasing in z then $\bar{\psi}(p, z)$ is increasing in z.*

The proof of Lemma 1 follows from $v(p)$ being strictly decreasing in p. The first part of the Lemma shows that the probability that a customer exits the customer base of a firm increases with its price p. Moreover, it implies that, when comparing two firms charging the same price p in the current period, a customer is more likely to exit the customer base of the firm with lower value of $\bar{\mathcal{P}}(z)$.

2.3 The problem of the firm

In Section 2.2 we concluded that a customer exits the customer base of the firm she currently patronizes when, for a given price p, her search cost ψ is low enough; we described this optimal behavior by the threshold function $\bar{\psi}(p, z)$. In this section we use that customers follow this policy to derive the optimal price set by firms.

strictly decreasing in p. The fact that no customer decides to exit if $p \leq \bar{\mathcal{P}}(z)$ follows immediately.

The customer base of a firm; m. Recall that m_{t-1} denotes the mass of customers at the beginning of a period currently buying from firm j. Also, let δ denote the arrival rate of new customers. The evolution of the customer base is

$$m_t \;=\; m_{t-1}\,\left[1+\delta-G\left(\bar{\psi}(p,z)\right)\right] \;\equiv\; m_{t-1}\,\Delta(p,z)\;, \tag{4}$$

where $z = z_t^j$ and $p = p_t^j$. The term $m_{t-1}\delta$ accounts for the inflow of new customers, while the term $m_{t-1}\left(1-G\left(\bar{\psi}(p,z)\right)\right)$ accounts for the mass of customers that decided to stay with the firm. In the end, $\Delta(p,z)$ measures the customer base growth of the firm. The linearity of the production function, coupled with the arrival rate of new customers being proportional to the customer base size of the firm, implies Gibrat's Law. That is, firm's growth is independent of its size (Luttmer (2010)).

Let

$$\varepsilon_m(p,z) \equiv -\frac{p}{\Delta(p,z)}\frac{\partial\Delta(p,z)}{\partial p} = -v'(p)p\,\frac{G'(\bar{\psi}(p,z))}{1-G(\bar{\psi}(p,z))+\delta} \geq 0\;, \tag{5}$$

denote the elasticity of the growth rate of the customer base to price, where the last equality follows from equation (3) and the definition of $\Delta(p,z)$. We refer to this object as the *extensive* margin demand elasticity as it governs the amount of customers within the firm's customer base. Likewise, we refer to $\varepsilon_d(p)$ as the *intensive* margin demand elasticity as it controls the expenditure of those customers that decided to stay. Notice that aggregate quantity sold by firm j is $q_t(p,z) \equiv m_{t-1}\;\Delta(p,z)\;d(p)$ so that overall demand elasticity is

$$\varepsilon_q(p,z) = \varepsilon_d(p) + \varepsilon_m(p,z)\;.$$

Profits. Recall that $\pi(p,z)$ denotes nominal period profits per customer of a firm charging price p, with productivity z,

$$\pi(p,z) \;=\; d(p)\;\left(p-\frac{w}{z}\right)\;, \tag{6}$$

from where it is immediate to see that $\pi(p,z)$ is strictly increasing and strictly concave in z. Given the assumptions we made regarding the demand function $d(p)$, the profit function is continuously differentiable in p. We also made assumptions regarding the single-peakness of the profit function. With this in mind, let $\hat{\mathcal{P}}(z)$ denote the price that maximizes static profits for a firm with current productivity z, i.e., $\partial\pi(\hat{\mathcal{P}}(z),z)/\partial p = 0$. We have that

$$\hat{\mathcal{P}}(z) = \frac{\varepsilon_d(p)}{\varepsilon_d(p)-1}\;\frac{w}{z}\;. \tag{7}$$

In Appendix A.1 we show that the existence of a unique maximizer $\hat{\mathcal{P}}(z)$ follows when the utility function of customers is CRRA with CES demand across a large set of varieties.

The value function of a firm. Let $\tilde{F}(z_t, m^j_{t-1})$ be the value function for a firm with current productivity z_t and customer base m^j_{t-1}, so that $\tilde{F}(z_t, m^j_{t-1})$ solves

$$\tilde{F}(z_t, m^j_{t-1}) \;=\; \max_p \; m^j_t \; \pi(p, z_t) \;+\; \beta \int_{\underline{z}}^{\bar{z}} \tilde{F}(z_{t+1}, m^j_t) \; \mu(\, z_{t+1}|\, z_t) dz_{t+1} \; ,$$

An application of the Contraction Mapping Theorem implies that the value function for a firm is homogeneous of degree one in m, i.e., $\tilde{F}(z, m) = m \; \tilde{F}(z, 1) \equiv m \; F(z)$, where $F(z)$ solves

$$F(z) \;=\; \max_p \; \Delta(p, z) \left(\pi(p, z) + \beta \int_{\underline{z}}^{\bar{z}} F(z')\mu(\, z'|\, z) dz' \right) \; , \tag{8}$$

where the profit per customer function $\pi(p, z)$ satisfies equation (6) and we used equation (4). Also, to ease on notation, we let $z = z_t$ and $z' = z_{t+1}$. Notice that, because the size of customer base does not appear in the problem presented in equation (8), two firms with different customer base size face the same problem and therefore charge the same price; this happens because firms care about the growth rate of the customer base, not about its size.

Two important issues arise from inspection of the firm's problem presented in equation (8). The first one is that the effective discount rate, $\Delta(p, z)\beta$, is such that the operator for $\hat{F}(z)$ is not necessarily a contraction; this happens because, for some z, the effective discount rate can be above one. To address this we make sure that the discount factor is low enough such that the effective discount rate is always below one; this restriction guarantees that the problem is a contraction. The second issue is that the problem of the firm is not necessarily continuously differentiable nor globally concave in p. This implies that the first order condition is not sufficient for the optimal price schedule; this stems from $\bar{\psi}(p, z)$, and its effect on the effective discount rate and period payoff.

Let $\Pi(p, z) \equiv \pi(p, z) + \beta \int_{\underline{z}}^{\bar{z}} F(z')\mu(\, z'|\, z) dz'$ denote the value of a customer for the firm. It is composed of two terms. The first term, $\pi(p, z)$ accounts for the profits the customer brings to the firm in the current period. The second term, $\beta \int_{\underline{z}}^{\bar{z}} F(z')\mu(\, z'|\, z) dz'$, accounts for the value that the customer brings to the firm in the future. Rewriting equation (8) we get that

$$F(z) = \max_p \; \Delta(p, z) \; \Pi(p, z) \; ,$$

which shows that the firm's optimal price trades-off size of the customer base $\Delta(p, z)$, which is decreasing in p, with the value per customer $\Pi(p, z)$, which attains its maximum at $\hat{\mathcal{P}}(z)$. Moreover, $\Delta(p, z)$ is regulated by the extensive margin demand elasticity $\varepsilon_m(p, z)$ while

$\Pi(p, z)$ is regulated by the intensive margin demand elasticity $\varepsilon_d(p)$.

The next proposition characterizes the solution to the firm's maximization problem, and discusses cases in which the first order condition applies.

Proposition 1 *Let $\mathcal{P}(z)$ denote the set of prices maximizing the value of a firm with current productivity z and let $\tilde{\mathcal{P}}(z)$ denote the set of prices solving the following expression,*

$$\frac{\partial \pi(p, z)}{\partial p} \; \Delta(p, z) = -\frac{\partial \Delta(p, z)}{\partial p} \; \Pi(p, z) \; . \tag{9}$$

Then $\mathcal{P}(z) = \tilde{\mathcal{P}}(z) = \{\hat{\mathcal{P}}(z)\}$ if and only if $\hat{\mathcal{P}}(z) \leq \bar{\mathcal{P}}(z)$. If, instead, $\hat{\mathcal{P}}(z) > \bar{\mathcal{P}}(z)$, then $\mathcal{P}(z) = \tilde{\mathcal{P}}(z) \subset [\bar{\mathcal{P}}(z), \hat{\mathcal{P}}(z))$. Moreover, if $\partial\Delta(p, z)/\partial p$ is continuous and $\bar{\mathcal{P}}(z) < \hat{\mathcal{P}}(z)$, then $\mathcal{P}(z) = \tilde{\mathcal{P}}(z)$ and $\tilde{\mathcal{P}}(z) \subset (\bar{\mathcal{P}}(z), \hat{\mathcal{P}}(z))$.

A proof of the proposition can be found in Appendix A.2. The next lemma gives conditions under which the first order condition in equation (9) is necessary and sufficient for an optimum.

Lemma 2 *If $g(\underline{\psi}) = 0$, $\Delta(p, z)$ is continuously differentiable and the first order condition in equation (9) is necessary for an optimum. If furthermore, the hazard rate of ψ is increasing, there exists a value $\bar{\delta} > 0$ such that if $\delta < \bar{\delta}$, then $\varepsilon_m(p, z)$ is increasing in p, and $\tilde{\mathcal{P}}(z)$ is a singleton.*

See Appendix A.3 for a proof. From now on we consider the case where the assumptions of Lemma 2 are satisfied. Notice that productivity z has two distinct effects on prices in our model. An important result of Proposition 1 is that the optimal price of a firm is never above the price that maximizes per-customer static profits, $\hat{\mathcal{P}}(z)$. First, productivity has the standard *direct effect* on prices: higher productivity induces lower prices as the marginal benefit on per-customer profits of a price increase is decreasing in both p and z. Second, productivity has also an *indirect effect* on prices through $\bar{\mathcal{P}}(z)$: in equilibrium $\bar{\mathcal{P}}(z)$ depends on current productivity of a given firm, as it determines the probability distribution of future prices in that firm. Thus productivity in our model affect both costs and demand structure of a firm. As a result, if $\bar{\mathcal{P}}(z)$ increases with productivity z, then $\mathcal{P}(z)$ is not necessarily a monotonic function of current productivity level as more productive firms might face a less elastic demand function than less productive firms. By this argument, everything else being equal, more productive firms would like to charge a higher price. Thus, depending on which effects prevails, the optimal pricing policy might not be monotonic in productivity. This is possible in those cases where customers value more firms with higher productivity, i.e. $\bar{\mathcal{P}}(z)$

increasing in z, and the sensitivity in the growth in the customer base to prices decreases fast when $\bar{\mathcal{P}}(z)$ increases.

The optimal price $\mathcal{P}(z)$ solves equation (9). Using the demand elasticities definitions we can rewrite this expression as,

$$\frac{p}{w/z} = \frac{\varepsilon_d(p) + \alpha_\pi(p,z)\varepsilon_m(p,z)}{\varepsilon_d(p) - 1 + \alpha_\pi(p,z)\varepsilon_m(p,z)} \ , \tag{10}$$

where $\alpha_\pi(p,z) \equiv \Pi(p,z)/\pi(p,z) \in (1,\infty)$. The equation shows that the optimal markup, $p/(w/z)$, is affected by (i) the intensive margin demand elasticity $\varepsilon_d(p)$, and (ii) the extensive margin elasticity $\varepsilon_m(p,z)$ weighted by the relative importance of the value of retaining the marginal customer with respect to her static value, $\alpha_\pi(p,z)$. As a result, this expression shows that the optimal price trades-off short term profits vs. long term value accrued from the marginal customer. The next proposition presents a set of comparative statics.

Proposition 2 *The optimal price of a firm is (i) decreasing in the intensive margin elasticity $\varepsilon_d(p)$, (ii) decreasing in the extensive margin elasticity $\varepsilon_m(p,z)$, (iii) decreasing in $\alpha_\pi(p,z)$, and (iv) $p = \frac{\varepsilon_d(p)}{\varepsilon_d(p)-1}\frac{w}{z} = \hat{\mathcal{P}}(z)$ if and only if $\varepsilon_m(p,z) = 0$ and $p < \frac{\varepsilon_d(p)}{\varepsilon_d(p)-1}\frac{w}{z} = \hat{\mathcal{P}}(z)$ if $\varepsilon_m(p,z) > 0$. Finally, (v) for a given total demand elasticity $\varepsilon_q(p,z) \equiv \varepsilon_m(p,z) + \varepsilon_d(p)$, the higher $\varepsilon_m(p,z)$ the lower the optimal price.*

The proof of the proposition is straightforward and therefore omitted. Part (i) and (ii) of the proposition simply states that the higher either the intensive or extensive margin elasticities, the lower the price charged by a firm. Part (iii) of the proposition states that the optimal price decreases with the relative importance of the long term value of a customer relative to the short term value, which is mostly governed by the discount factor β. Part (iv) states that, when customer base concerns are not binding, the optimal price depends only on the intra-temporal elasticity of demand and the markup coincides with that one chosen by a monopolist facing elasticity $\varepsilon_d(p)$. When customer base concerns are present, i.e., when $\varepsilon_m(p,z) > 0$, the optimal markup is below that one following from static profit maximization; this happens because firms value retaining customers, and customers are hard to retain. Part (v) of the proposition states that the larger the extensive margin elasticity share of the aggregate demand elasticity, the lower the optimal price.

An immediate result is that if a firm with productivity z choosing the price that maximizes static profits $\hat{\mathcal{P}}(z)$ faces no customer base concerns, then the optimal price for this firm is $\hat{\mathcal{P}}(z)$.

As discussed, the extensive demand margin elasticity $\varepsilon_m(p,z)$ plays a crucial role in the optimal price determination. Therefore, understanding what shapes up this object is useful.

We already established that $\varepsilon_m(p,z) \geq 0$, and discussed conditions under which $\varepsilon_m(p,z)$ is increasing in p. The next lemma discusses conditions under which $\varepsilon_m(p,z)$ is increasing in z.

Lemma 3 *If the assumptions of Lemma 2 are satisfied, and $\partial\bar{\psi}(p,z)/\partial z \leq 0$, then $\varepsilon_m(p,z)$ is decreasing in z.*

The proof of the lemma is straightforward and therefore omitted. The lemma states that if the threshold rule $\bar{\psi}(p,z)$ decreases with z, then the extensive margin elasticity is decreasing in the productivity level z. Notice that $\bar{\psi}(p,z)$ decreases with z when customers expect firms with higher productivity today to charge lower prices in the future and, as a result, under this assumption the extensive margin elasticity falls as productivity increases. An example is when optimal prices are monotonic in productivity, and productivity has some persistence. Finally, notice that both derivatives of $\varepsilon_m(p,z)$ with respect to p and z crucially depend on the hazard rate of ψ being increasing. This assumption implies that, in relative terms, the density of customers exiting to the fraction remaining in the customer base of the firm increases with $\bar{\psi}$.

2.4 Equilibrium

In this section we define a stationary equilibrium and we discuss some of its general properties. Let $\mathcal{P}^*(z) \subseteq \mathcal{P}(z)$ denote the candidate equilibrium price at productivity level z. Let $M(z)$ measure the amount of customers currently buying good 1 from firms with productivity z. The endogenous inflow of customers per period δ is such that the measure $M(z)$ is stationary and satisfies the following equations:

$$M(z') = \int_{\underline{z}}^{\bar{z}} M(z)\ \Delta(\mathcal{P}^*(z), z)\ \mu(z'|z)\ dz\ , \tag{11}$$

$$\Gamma = \int_{\underline{z}}^{\bar{z}} M(z)\ dz\ . \tag{12}$$

Definition 1 *A stationary equilibrium is (i) a pricing policy $\mathcal{P}^*(z)$, (ii) an invariant distribution $M(z)$ and arrival rate δ, (iv) a function for the threshold rule $\bar{\psi}(\cdot,\cdot)$, (v) the customer outside option satisfying $\hat{V} = \int_{\underline{z}}^{\bar{z}} \bar{V}(\mathcal{P}^*(z), z)\ \mu(z)\ dz$, customers search optimally (equation (3))), firm's optimal pricing solves equation (8), and dynamics of customer base satisfy equations (11).*

We now start discussing equilibrium characterization, which we will do in more detail in Section 2.5. To this end, let $\mathcal{Z}_1$ denote the set of productivity levels at which the firm's optimal price maximizes static profits and let $\mathcal{Z}_2$ denote the set of productivity levels at which the firm's optimal price is below the price that maximizes static profits. That is, $\mathcal{Z}_1 = \{z \text{ such that } \mathcal{P}^*(z) = \hat{\mathcal{P}}(z)\}$ and $\mathcal{Z}_2 = \{z \text{ such that } \mathcal{P}^*(z) < \hat{\mathcal{P}}(z)\}$.

Proposition 3 *In equilibrium we have (i) $\mathcal{P}^*(z) \leq \hat{\mathcal{P}}(z)$ for all z, and (ii) $\mathcal{Z}_1 \neq \emptyset$ and $\mathcal{Z}_2 \neq \emptyset$. Then, for any two productivity levels $\{z, y\}$, $z \in \mathcal{Z}_1$ and $y \in \mathcal{Z}_2$, we have that $\Delta(\mathcal{P}^*(z), z) = 1 + \delta > \Delta(\mathcal{P}^*(y), y) > \delta$, and for some y, $\Delta(\mathcal{P}^*(y), y) < 1$.*

A proof of the proposition can be found in Appendix A.4. Part (i) of the proposition states that no firm prices above the price that maximizes static profits; this is an immediate implication of Proposition 1. Part (ii) of the proposition shows that there is a set of productivity levels where the optimal price maximizes static profits; this happens because for productivity levels lying in the set $\mathcal{Z}$, customer retention concerns are not present and as a result $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z) < \bar{\mathcal{P}}(z)$. It also shows that there is a set of productivity levels where the optimal price is strictly below the price that maximizes static profits. The latter follows from the assumption that $\underline{\psi} = 0$, so that price dispersion necessarily implies that some customers will exit the customer base of firms with highest prices. As a result, relative to the standard model of price setting under monopolistic competition where the optimal price is $\hat{\mathcal{P}}(z)$, the presence of customer retention concerns generate lower average prices and markups.

Proposition 4 *If the assumptions of Lemma 2 are satisfied, an equilibrium exists.*

A proof of the proposition can be found in Appendix A.5. The assumptions of Lemma 2 ensures that the firm's maximization problem has a unique solution, and that the optimal price is a continuous function of productivity. The proof follows by applying Brouwer fixed point theorem.

In the next two Remarks we characterize two useful limiting cases. In the first Remark we analyze the equilibrium where customer base concerns are no present, which happens when $\underline{\psi} \to \infty$. Because of no customer base concerns, the model in this case reduces to the standard price setting problem under monopolistic competition. In the second Remark we discuss the equilibrium where there is no productivity dispersion.

Remark 1 *Suppose that the search cost is divergent, i.e. $\underline{\psi} \to \infty$. Then, in equilibrium: (i) the optimal price maximizes static profits, $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z)$ for all z, and (ii) customers do not find it optimal to exit the customer base of their current firm.*

The proof of the Remark follows immediately since, because the search cost diverges, firms do not face customer retention concerns and therefore find it optimal to charge the price that maximizes static profits. Formally, when $\underline{\psi} \to \infty$, we have that $\bar{\mathcal{P}}(z) \to \infty$. It is immediate that $\hat{\mathcal{P}}(z) \leq \bar{\mathcal{P}}(z)$ so that $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z)$.

Remark 2 *Let* $\underline{z} = \bar{z} = z_0$*, and* $\mu(z_0) = 1$*. Then,* $\mathcal{P}^*(z) = \bar{\mathcal{P}}(z) = \hat{\mathcal{P}}(z)$*. As a result,* $\Delta(\mathcal{P}^*(\bar{z}), z) = 1$ *so that firms enjoy a constant customer base. Moreover, if* $G(0) = 1$*, the equilibrium price is* $\mathcal{P}^*(\bar{z}) \in \left[W/\bar{z}, \hat{\mathcal{P}}(\bar{z})\right]$*.*

A proof of the Remark can be found in Appendix A.6. The Remark shows how our model relates to Diamond (1971). It shows that when firms are homogeneous in their productivity level they all charge the price that maximizes static profits. Moreover, because every firm charges the same price, customers have no incentives to search so that the customer base is constant.

The next proposition discusses the equilibrium properties when optimal prices are monotonic in productivity.

Proposition 5 *If* $\mathcal{P}^*(z)$ *decreases monotonically with* z *the equilibrium is such that* $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z)$ *for all* $z \geq \hat{z}$*, and* $\mathcal{P}^*(z) < \hat{\mathcal{P}}(z)$ *for all* $z < \hat{z}$*, where* $\hat{z} \in (\underline{z}, \bar{z})$ *is unique.*

A proof can be found in Appendix A.7. A special case where the assumptions of the Proposition are satisfied is when productivity shocks are i.id..

In the next section we provide examples where the assumptions of Proposition 4 are satisfied, and explore the implications of our mechanism for markups.

2.5 Equilibrium prices when productivity follows a simple process

In this section we explore some properties of the model using a simple productivity process. The process we consider is the following

$$z' = \begin{cases} z & \text{with probability } \rho \\ \varepsilon & \text{with probability } 1-\rho \end{cases} \tag{13}$$

where ε is a random draw from $\mu(\varepsilon)$. This process encompasses the case of i.id. productivity shocks, when $\rho = 0$, and entails some persistence when $\rho > 0$. The next proposition characterizes the properties of optimal prices under equation (13).

Proposition 6 *Let* z *follow the process given by equation (13), then the equilibrium prices are such that: (i)* $\mathcal{P}^*(z)$ *decreases monotonically with productivity* z*; (ii)* $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z)$ *for all* $z \geq \hat{z}$*, and* $\mathcal{P}^*(z) < \hat{\mathcal{P}}(z)$ *for all* $z < \hat{z}$*; (iii) markups, i.e.* $\mathcal{P}^*(z)/(w/z)$*, increase with* z*.*

Point (i) stems from the "promised" continuation value to customers being necessarily non-decreasing in the current productivity z, given the process for z. Point (ii) follows by applying Proposition 4. Point (iii) follows because monotonicity of equilibrium prices guarantees that $\varepsilon_m(\mathcal{P}^*(z), z)$ and $\varepsilon_d(\mathcal{P}^*(z))$ decrease with z, and the process in equation (13) makes $\alpha_\pi(\mathcal{P}^*(z), z)$ inversely proportional to equilibrium profits $\pi(\mathcal{P}^*(z), z)$, and thus decreasing with z. Then by using equation (10) (iii) follows. For a detailed proof see Appendix A.8.

Finally, it is interesting to study the role of ρ for optimal markups. When productivity shocks are i.i.d. $\varepsilon_m(p, z)$ does not depend on z directly but only through the contemporaneous price p. The direct effect of z on $\varepsilon_m(p, z)$ amplifies the impact of variation in productivity onto variation in equilibrium markups. The larger the persistence ρ, the more important this channel.

3 Empirical evidence

We complement the theoretical analysis with an empirical investigation that relies on cashier register data from a large US supermarket chain on grocery purchases of households holding the chain loyalty card. The empirical analysis has two purposes: in this section we exploit the data to document that, as the model would predict, fluctuations in the price of the good sold by a firm influence a customer decision to remain in its customer base. Next, we will use the data to estimate the model and quantify the importance of customer base concerns in shaping firm price setting.

3.1 Data and variable construction

For every trip made at the supermarket chain by a panel of households between June 2004 and June 2006, we have information on date of the trip, store visited and list of goods (identified by their Universal Product Code, UPC) purchased, as well as quantity and price paid. The data are ideal to study customer base dynamics since they loyalty card allows to track precisely when a particular household is shopping at the supermarket chain. Moreover, the fact that all the households in our sample own a loyalty card of the chain implies that they can be thought of as regular customers. This makes it more plausible that they would face some cost of switching to a different retailer even in a setting, like the supermarket industry, where there are no formal contracts binding the customer-retailer relationship.

The micro data on household purchases are complemented by data on store sales for the two years in the sample. We use information on weekly revenues and quantities sold for a sample of stores representative of the different prices areas to compute prices for each UPC in the sample. The construction of the price variable is therefore analogous to that in Eichenbaum et al. (2011) and is subject to the same caveats. The price variable varies weekly matching the timing of price adjustment for the retailer. For the analysis we retain only UPC's for which we have complete time series of prices for the 104 weeks in the sample span.

For the purpose of our exercises we have to construct two key variables: (i) an indicator signaling when the household is leaving the chain's customer base (i.e. the household stops shopping at the chain), and (ii) the price of the household basket. Here we briefly describe the procedure followed to obtain them, further details are included in Appendix B.

We consider every customer shopping at the retailer in a given week as belonging to the chain's customer base in that week. When a particular household has not shopped at the chain for eight or more consecutive weeks, we deduce that she has left the customer base. The 8-weeks window is a conservative choice. Every household in the data holds a loyalty card which suggests that they are not casual shoppers at the chain. We made sure of that by selecting only households shopping there at least 48 times over the two years in the sample. Regular customers are unlikely to experience that long of a spell without shopping for reasons other than having switched to another chain (e.g. consuming their inventory). Indeed, customers in our sample make an average of 157 shopping trips at the chain over the two years; if those trips were uniformly distributed that would imply visiting a store of the chain six times per month. The average number of days elapsed between consecutive trips is close to four and the 99th percentile is 24 days, roughly half the length of the absence we require before inferring that a household is buying its grocery at a competing chain.

We construct the price of the basket of grocery goods usually purchased by the households in a fashion similar to Dubois and Jodar Rosell (2010). We identify the goods belonging to a household's basket using scanner data on items the household purchased over the two years in the sample. The price of its basket in a particular week is then the average of the weekly prices of the goods included in the basket, weighted by their expenditure share in the household budget. Since households differ in their choice of grocery products and in the weight such goods have in their budget, the price of the basket is household specific.

3.2 Price and customer base dynamics

We estimate a linear probability model where the dependent variable is an indicator for whether the household has left the customer base of the retail chain in a particular week. Our regressor of interest is the logarithm of the price of the basket of grocery goods usually purchased by the households (p).

In Table 1, we report results of regressions of the following form,

$$Exit_{it} = b_0 + b_1 p_{it} + X_i' b_2 + \varepsilon_{it} \tag{14}$$

In the regression we include year-week fixed effects to account for time-varying drivers of the decision of exiting the customer base common across households and control for observable characteristics through inclusion of household's demographics matched from Census 2000. Finally, we include as regressors the logarithm of the price of the basket in the first week in the sample and the standard deviation of price changes for each household over the sample period. These are meant to control for differences in the composition of the basket across shoppers. For example, some customers may purchase product categories more prone to promotions than others and experience more intense price fluctuations as a result.

The regressions aim at testing the existence of a correlation between the decision of exiting the customer base of the chain and the price of the basket at the chain, relative to what that price would be at the competition. However, since we are using price data from a single chain, we cannot exclude that some of the variation in p_{it} comes from aggregate cost shocks affecting all the retailers (e.g. a spike in the price of aluminium that makes soda cans more expensive for every retailer). Aggregate cost shocks do not change the relative price and, therefore, should not trigger exit from the customer base. Failing to remove aggregate shocks should not affect the sign of the correlation but add noise to it, biasing against finding a significant result. Moreover retailer specific shocks (i.e. a shift in the cost of all the goods sold by the chain providing the data), despite being a valid source of identification, do not contribute to it in our setting. In fact, since we only use data from one chain, their effect is absorbed by the time dummies. The correlation of interest is then identified by good-retailer specific shocks as those triggered, for example, by the expiration of a contract between the chain and a manufacturer. The availability of data from several stores helps this identification strategy by introducing variation in the cost of supplying the store due to logistics (e.g. distance from the warehouse) which will hit differently goods with different intensity in delivery cost (e.g. refrigerated vs. non refrigerated goods).

Unlike in standard demand analysis we are not modeling the household's choice of the retailer where to shop but rather the decision of whether to leave a given retailer or to stick with it. Endogeneity driven by unobserved store characteristics correlated with both price and the propensity to leave is not an issue in this context. Price could still be endogenous if: i) the retailer knows about household idiosyncratic shocks -unobservable to us- that affect the decision of changing retailer; ii) the retailer can use this knowledge to adjust accordingly the price faced by the household. Whereas it is conceivable that the supermarket observes variables predicting the exit from the customer base of groups of households, it is unlikely that it can react with targeted prices. In fact the basket of different households will at least partially overlap making it impossible to fine tune the basket price faced by some households without affecting the price of others.

A final important difference with standard demand equation is in the interpretation of the price coefficient. According to our model, the price plays a dual role: it represents the cost of purchasing the basket at the retailer today but, as stressed by its dependence on productivity. It also signals the convenience of purchasing from the retailer in the future. Therefore, the coefficient cannot be used to predict the customer reaction to a random, unrelated to productivity shocks, change in prices. Furthermore, the decision to exit the customer base is also influenced by the productivity level z, which is customer-time specific. Because we do not include a measure of productivity in the specification, the price coefficient should be regarded as averages across productivity levels, which follows from a stationary interpretation of the regression.

Table 1: Effect of price on the probability of exiting the customer base

Exiting: Missing at least 8 consecutive weeks	**(1)**	**(2)**	**(3)**
$\log(P_i)$	0.012*	0.014*	0.013*
	(0.007)	(0.007)	(0.007)
Num. competitors		-0.000	
		(0.001)	
Walmart entry			0.041
			(0.044)
Observations	77,815	71,049	66,182

Notes: An observation is a household-week pair. The sample only includes households who prominently shop at stores for which we have complete price data for all the UPCs they purchase. We trim from the sample the top and bottom 1 % in the distribution of the number of trips over the two years. Demographic controls rely on a subsample of households for which information on the block-group of residence was provided and include as regressors ethnicity, family status, age, income, education, and time spent commuting (all matched from Census 2000) as well as distance from the closest outlet of the supermarket chain and distance from the closest competing supermarket (provided by the retailer). The logarithm of the price of the household basket in the first week in the sample and the standard deviation of changes in the log-price of the household basket over the sample period are included as a controls in all specifications. Week-year fixed effects are also always included. Standard errors are in parenthesis. ***: Significant at 1% **: Significant at 5% *: Significant at 10%.

Column (1) of Table 1 documents that a 1% increase in the growth rate of individual specific weekly prices is associated with 0.012 percentage points increase in the probability that the household leaves the chain to patronize a rival grocer. In the analysis we have so far abstracted from the behavior of the competitors of the chain. In part this is justified by the fact that in the theoretical framework we introduce in section Section 2 we assume that search is indirect. Therefore, pricing and promotion policies of competitors should not affect the customer's decision of leaving the chain. Indeed, under the lens of the model only price shocks idiosyncratic to the chain should affect the decision of leaving implying that aggregate shock can only bias us against finding a significant price effect. Furthermore, in order to believe that this omission is driving our result we should assume that the retailer systematically raises its prices when competition increases rather than lowering them. Nevertheless, we address the issue of competition in the last two columns of Table 1. The specification in column 2 includes the number of competing supermarket stores, calculated using information from Reference US, in the zipcode of residence of a household. In column 3 we use data from Holmes (2011) to construct a dummy signaling the opening of a Walmart super-center in the zip-code of the store where the customer is shopping. In both cases the main result does not change. We also assess whether our finding is robust to defining exit from the customer base in a more restrictive fashion, only considering spells of eight or more consecutive weeks without shopping at the chain. This alternative definition reduces the number of exits observed and leads to a lower but still significant effect of prices on the decision to change supplier.

4 Calibration

In this section we discuss how we parameterize the model. Some parameters of the model we borrow them directly from external sources, while we estimate other by a method of moments.

We first choose preferences that microfound the demand and surplus functions, $d(p)$ and $v(p)$. Let customers derive utility from a composite of two goods, d and n. We label the composite good by c, where $c = \left(\omega\, d^{\frac{\theta-1}{\theta}} + (1-\omega)\, n^{\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\theta-1}}$, $\theta > 1$. We let the utility function of customers being $u(c) = c^{1-\gamma}/(1-\gamma)$. We interpret good d as the good produced by the firms described above while good n is produced by a competitive representative firm with linear production and unitary labor productivity. We let n act as a numeraire. The budget constraint of a customer is $pd + n = I$. Details on the derivation of $d(p)$ and $v(p)$ can be found in Appendix A.1.

Income comes from labor proceeds and profits. Profits are divided equally across the different customers, and we assume that each customer supplies an equal amount of hours ℓ. We determine ℓ in Appendix A.9 by solving the maximization problem of customers facing disutility from working. Given perfect competition in the production of good n, and given the linear production technology, the equilibrium wage, w, has to equal marginal product of labor in such sector, and thus $w = 1$.

Imputed parameters. We set the time period to be one month and we set the discount rate to be $\beta = 0.997$ (i.e. an yearly interest rate of 4%). We set the relative risk aversion parameter to $\gamma = 2$ which is a standard value. We set $\omega = 0.5$ so that goods d and n have the same weight in preferences.[7]

As it is well known prices do not adjust every period. To address this fact we use the process presented in equation (13), with $\varepsilon \sim N(0, \sigma)$. This process is mean reverting, where $1 - \rho$ is the probability of mean reversion. Given that in our model the equilibrium price depends solely on the productivity z, the probability that a firm changes her price is also $1 - \rho$. Goldberg and Hellerstein (2009) provides, for the PPI (i.e. Producer Price Index) a measure of the average monthly frequency of price changes (0.32). Interpreting these prices as the cost of production for consumption goods allows us to calibrate the parameter ρ, obtaining $\rho = 0.68$.

Parameters requiring calibration. We assume that the search cost is drawn from a Gamma distribution with shape parameter ζ, and scale parameter λ. We have to obtain estimates for these two parameters, as well as for the parameter governing the elasticity of substitution across different types of goods, θ, and the volatility of innovations in productivity, σ.

To identify the parameters of the search cost distribution we exploit the estimates of the relationship between price and probability of exiting the customer base discussed in Section 3. The marginal effect of price on exit probability (the parameter b_1 in equation (14)) is informative about the level of the search costs and thus we use it to identify the scale parameter λ. The shape parameter ζ is proportional to the coefficient of variation of the search cost distribution. In the model, higher dispersion of search costs implies higher cross-firms variation in the extensive margin elasticity. In the data, we measure this variation by fitting a spline to equation (14), allowing the price marginal effect on the probability of exit to vary for different terciles of price levels. Then ζ is identified by matching the difference between the average marginal effect between the top and the bottom terciles measured in the data and in the model.

We select volatility of innovations in productivity, σ, so that the implied price dispersion in the model matches that reported by Kaplan and Menzio (2013) for bundles of homogeneous packaged goods. Finally, we choose the parameter θ so that the average intensive margin elasticity is 5, a value in the range of those used in the macro literature.[8]

We define $\Omega \equiv [\zeta\ \lambda\ \theta\ \sigma]'$ as the vector of parameters to be estimated, and denote by $v(\Omega)$ the vector of the theoretical moments evaluated at Ω, and by v_d their empirical counterparts. We search for Ω that minimizes the quadratic form $(v_d - v(\Omega))'\ (v_d - v(\Omega))$. Table 2 reports the results.

[7] In the economic application that we will perform in the next section, we will interpret the numeraire good n as a non-tradable, so that $\omega = 0.5$ will roughly correspond to an average share of non-tradables in the consumption basket of 50%.

[8] See Nakamura and Steinsson (2010) for a discussion.

Table 2: Parameter estimates

	Value	Target
Volatility of productivity innovations, σ	0.195	Price dispersion: 8%
Elasticity of substitution, θ	6.72	Avg. intensive margin elasticity: 5
Distribution of cost, $g(\psi) \sim \; Gamma(\zeta, \lambda)$		
Shape parameter, ζ	4	Inter-tercile difference in marginal effect: 1.8%
Scale parameter, λ	0.55	Average marginal effect: 1.4%

5 Markup: quantification and cyclicality

A large literature has investigated theoretically and empirically the business cycle properties of markups.[9] Studying the impact of customer markets for firms' price setting strategy has long been suggested as a potential explanation of markup movements (e.g. Phelps and Winter (1970), Blanchard (2009)). In this section we study the predictions of our model of customer markets for equilibrium markups.

We are well positioned to assess the cyclicality of markups in customer markets models for two reasons. First, we solve for equilibrium markups in a model where the dynamics of customers is endogenously determined by customers' search decision, microfounding the setups of Phelps and Winter (1970) and Rotemberg and Woodford (1991, 1999). Second, as discussed in Section 4, we exploit micro data that allow to discipline our exercise and to quantify the impact of customer retention concerns on markups dynamics.

As a benchmark, we compare the estimates of our model with customer retentions concerns (henceforth, "Baseline economy") to the estimates obtained from a model where we shut down the customer retention incentive. We do so by letting search costs diverge to infinity (i.e., $\lambda \to \infty$) so that customers would never want to search for a new firm. We choose θ so that the resulting average elasticity of total demand (i.e., $\int_{\underline{z}}^{\bar{z}} \varepsilon_q(\mathcal{P}^*(z), z)\mu(z)dz$) is the same as in our Baseline economy. Since this alternative model is analogous to the standard CES preferences widely used in the macro literature, we will refer to it as "CES economy".

The remaining of the section is organized as follows. We first study the stationary distribution of markups in our economy, and explore how equilibrium markups vary in response to idiosyncratic shocks to cost. We later introduce aggregate shocks in the model, and study

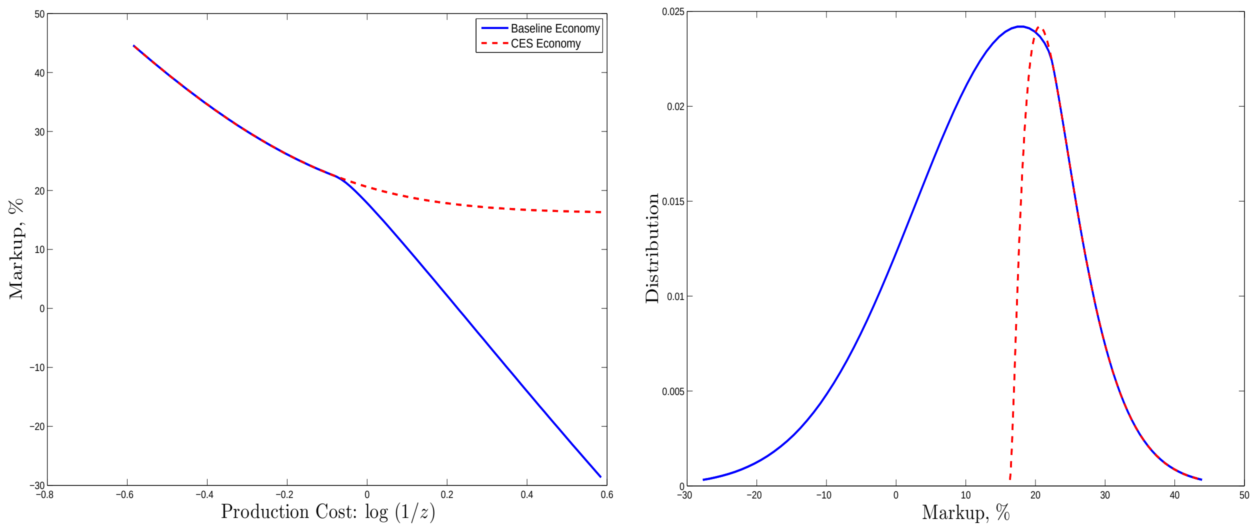
[9] See Nekarda and Ramey (2013) and Hall (2012) for a review of the empirical evidence and Blanchard (2009) for a review of theoretical contributions.

the response of markups to aggregate productivity and demand shocks.

5.1 Equilibrium markup dynamics to idiosyncratic cost variation

In this section we describe the distribution of equilibrium markups with respect to the idiosyncratic productivity shocks of our model, and study their cyclical properties.

Figure 1: Equilibirum Markups



Note: The blue solid line plots the optimal equilibrium markups in the model at our baseline parameters. The red dashed line plots the pass-through in the CES economy with $\lambda \to \infty$.

The left panel of Figure 1 compares the equilibrium markups in the Baseline and in the CES economy. Our theoretical results identify two sets of firms: "low cost" firms that do not face customer retention concerns and "high cost" firms for which the threat of losing customers is significant. In our estimates the threshold for production cost above which firms do face customer base concerns roughly coincides with the average production cost.

Two main features emerge. First, for given production cost, "low cost" firms charge the same markup in the Baseline and in the CES economy, while "high cost" firms choose strictly lower markup in the Baseline economy. This follows from the customer retention margin binding for "high cost" firms in the Baseline economy, incentivazing them to charge lower markups. The right panel of Figure 1 documents the implications for the distribution of markups. The distribution of markups in our model displays substantial negative skewness, with 11% of firms charging a negative markup. The average markup in the Baseline economy is 17%, lower than the average markup predicted by the CES economy (i.e. 25%). Finally, the behaviour of "high cost" firms increases the dispersion in markups from 6% in the CES economy to 14% in the Baseline economy.

Second, the optimal markup is strictly decreasing in production cost in both economies but, for "high cost" firms, it decreases faster in the Baseline economy than in the CES economy. There is, in fact, a common feature to both models that gives rise to a negative comovement between markups and production cost. This depends on the intensive elasticity of demand, $\varepsilon_d(p)$, being increasing in p.[10] Given that equilibrium prices are monotonically increasing in production cost, it follows that firms with higher production cost face higher elasticity of demand, so that optimal markups are decreasing in production cost.

However, there is an additional reason why markups decrease with production cost in the Baseline economy. The incentives to increase prices following an increase in production cost are lower because doing so implies persistent losses in customers and, therefore, in demand. Moreover, a persistent increase in firm specific cost of production is associated to higher expected future prices, and thus a worsened ability to retain customers. As a result, the average elasticity of markups to idiosyncratic productivity shocks is 69% in the Baseline model, against 28% in the CES economy.

Given that firms' idiosyncratic productivity and output comove in equilibrium, the pro-cyclicality of markups is magnified by customer retention concerns: a 1% increase in output is associated, on average, to a 0.5% increase in markup in the Baseline model, whereas the increase is only 0.09% in the CES economy. The pro-cyclicality of markups with respect to idiosyncratic productivity variation is consistent with recent empirical evidence from the industrial organization literature (Petrin and Warzynski (2012)).

5.2 Markup dynamics to aggregate shocks

So far we have focused on the cyclical properties of markups with respect to idiosyncratic variation in production costs. In this section we investigate the business cycle properties of markups in our model with customer retention concerns. There is no consensus in the literature on the business cycle properties of markups. Using direct estimates of marginal cost, Bils (1987) finds that markups are countercyclical, whereas Nekarda and Ramey (2013) find a positive correlation between markups and output.[11] In order to do so, we need to

[10] With CES preferences the demand of good i depends on the relative price p_i/P. With a finite number of goods in the basket of the consumer, an increase in p_i, also increases the price of the basket, P, thus reducing the overall increase in p_i/P and effect on demand. The effect on P is larger, the higher the weight of good i in the basket, that is the lower the price p_i and the higher its demand. Therefore, the elasticity of demand $\varepsilon_d(p)$ increases in p.

[11] A different approach to recover marginal cost is followed by Bils et al. (2012) and Gopinath and Itskhoki (2011).

extend our model so that we can study the general equilibrium effects of aggregate shocks. We proceed as follows. We augment our model with a representative household choosing labor supply (as we did in in Section 4 to calibrate the income of consumers), and savings so to determine the equilibrium interest rate. Firms are owned by the household, and use the household discount factor to evaluate future payoffs. We also introduce government spending as a fraction g of household's total income, financed through lump-sum taxes on the representative household. We assume for simplicity that government spending is of no value to the consumers in the economy, and that the government faces a problem similar to the one of the private sector in terms of shopping decisions.[12] We assume that government spending in steady state is zero and that there are no foreseen aggregate shocks, so that in steady state the augmented model is equivalent to the model estimated in Section 4.[13]

We consider the following experiment. We start the economy in steady state at $t = 0$. A one time unforeseen aggregate shock A_0 hits the economy at $t = 0$ before economic agents make their decisions. The shock dies out with an AR(1) process in logs, i.e. $\log(A_t) = \rho_a \log(A_{t-1})$ for $t > 0$. After the realization, agents have perfect foresight of the aggregate shock and its dynamics. We solve for the equilibrium dynamics of the distribution of markups and output following the aggregate shock until the economy converges back to the initial steady state. We separately study two types of aggregate shocks: an aggregate supply shock shifting productivity z for all the firms in the economy by the same amount, and a demand shock in the form of an increase in government spending g.

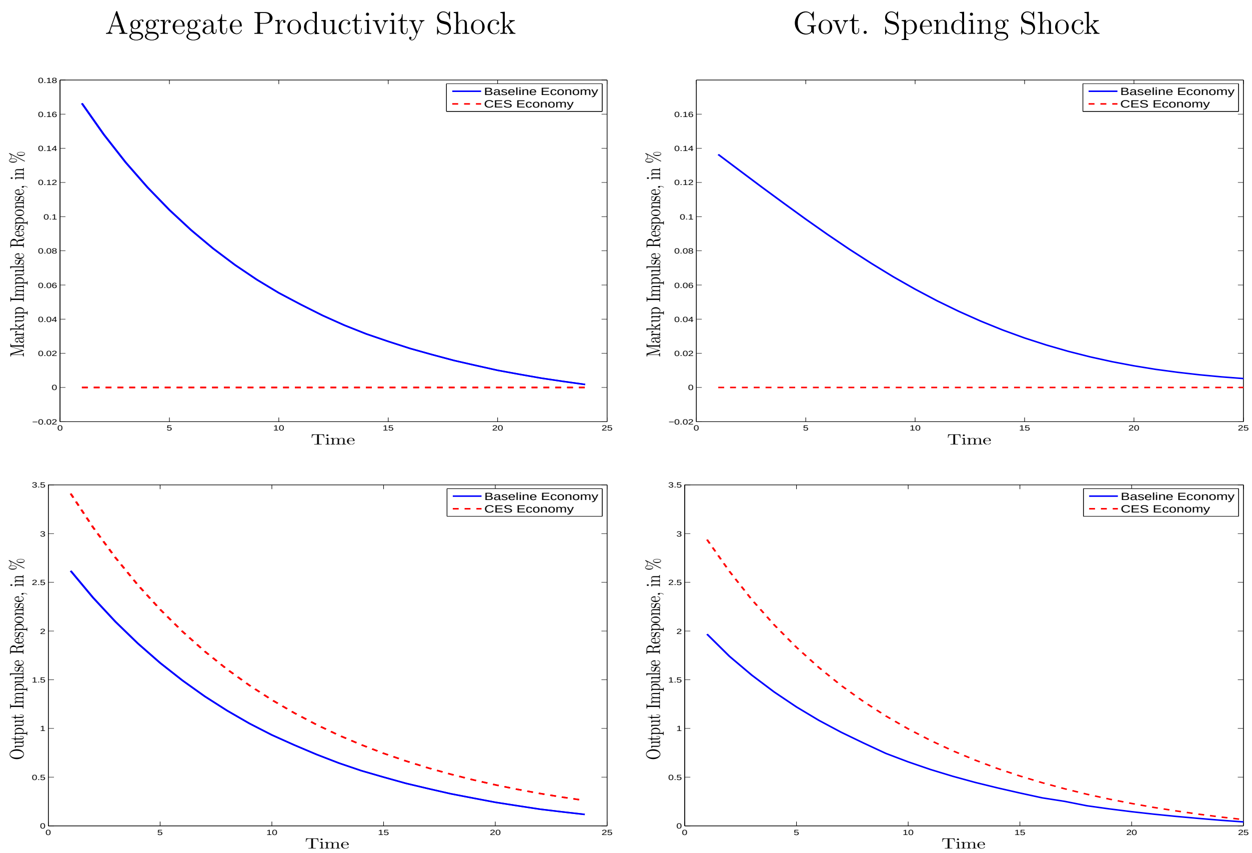
Figure 2 plots the impulse responses of average equilibrium markup (first row) and output (second row) in response to a 5% increase in aggregate productivity (left panel), and an increase in government spending (right panel) equivalent to 5% of steady state household's income. The shocks are persistent, and follow an AR(1) process with an autocorrelation coefficient equal to $\rho_a = 0.9$. In particular, each panel reports, for each period $t \geq 0$, the cross-sectional average markup and cross-sectional average output expressed in log-deviation from the steady state. We report predicted impulse responses both in our Baseline economy and in the CES economy.

Markups are pro-cyclical in the Baseline economy, both in response to productivity and government spending shocks, while they do not respond to the aggregate shocks in the CES economy. This model-based prediction is consistent with recent empirical evidence by Nekarda and Ramey (2013) who indeed show that markups and output positively comove

[12] The government is composed of a mass Γ of shoppers, with the same disposable income, i.e. $g\ I$, who decide when to search for a new supplier of a good d subject to the same cost of the private sector, and how much to spend on good d and good n. For the firms it makes no difference whether the customer is a government or a private consumer.

[13] See Appendix A.10 for details.

Figure 2: The response of average markups and output to aggregate shocks



Note: The blue solid line plots the cross-firms average of equilibrium output of good d in log-deviations from the steady state. The red dashed line plots the cross-firms average of equilibrium markup on good d in log-deviations from the steady state.

in response to these types of shocks. Such result also validates Rotemberg and Woodford's (1991) conjectures about the pro-cyclicality of markups in customer markets models. As a consequence of the different markup dynamics, the response of output to productivity and government spending shocks are substantially dampened in the Baseline model relatively to the CES economy: the cumulated output response to the productivity and government spending shocks are, respectively, 28% and 31% smaller than the corresponding values in the CES economy.[14] The mechanism through which aggregate shocks affect markups is different from the case of idiosyncratic shocks. Aggregate shocks do not affect the relative position of firms in the productivity distribution, which is the main driver of firm responses to idiosyncratic shocks. However, firms still want to react to aggregate productivity and demand

[14]The cumulated output response is given by $\int_0^\infty \hat{y}_t dt$ where $\hat{y}_t$ is output at t in log-deviations from the steady state.

shocks because such shocks change the expected returns on customers. In fact, a mean-reverting increase in productivity or government spending creates expectations of decreasing demand and profits relatively to the current period. This lessens the opportunity-cost of losing customers in the current period.

Markups to do respond to aggregate shocks in the CES economy because such shocks have no impact on marginal cost of production. In fact, due to perfect competition in the numeraire sector and labor market, the equilibrium wage is equal to the productivity of labor in the numeraire sector. Therefore, the equilibrium wage does not to the government spending shock, while it respond one for one to the aggregate productivity shock as the latter is also hitting the productivity of the numeraire sector. Notice that a similar result would be obtained in the CES economy even if the marginal cost were responding to the aggregate shock if the number of goods in the consumer basket goes to infinity as in the standard Dixit-Stiglitz preferences.

Finally, while we showed that, in absence of other types of frictions, customer retention concerns give rise to pro-cyclical markups, we see this result as complementing existing studies on the cyclicality of markups. Gopinath and Itskhoki (2011) discuss alternative frictions in price setting that may give rise to cyclical markups, as for instance nominal rigidities. Studying the interaction of customer retention concerns with other types of frictions is an interesting avenue for future research on price dynamics.

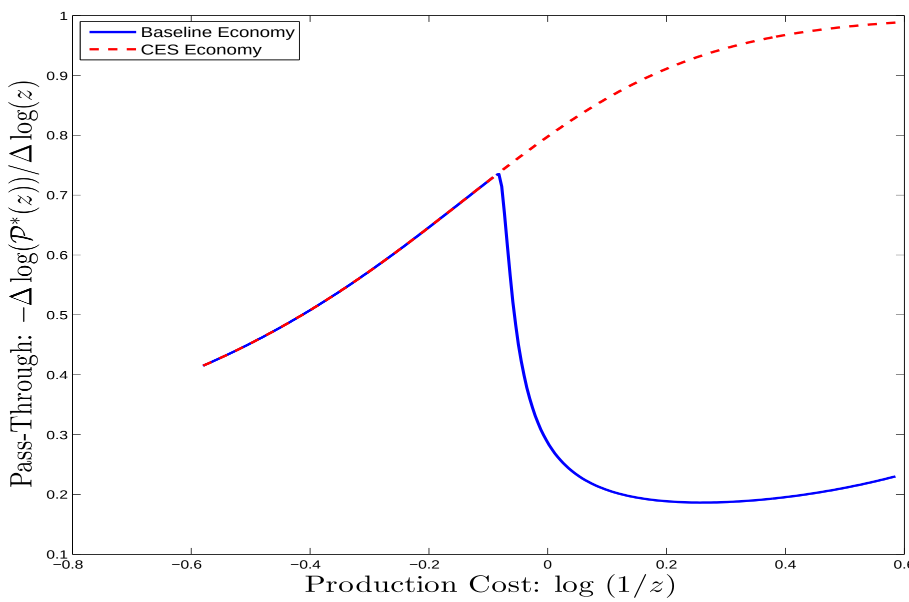
6 Pass-through of cost shocks

The presence of customer retention concerns makes it harder for firms to raise prices. It follows that our model provides an interesting setting to analyze the decision to pass-through a cost shock. We begin by explaining how firms react to an idiosyncratic shock to their productivity to illustrate the forces at work. Such shocks, however, are rarely observed in data. Therefore, we move to analyze the response to a different class of shocks: those hitting a subset of firms in the economy. The latter scenario can be interpreted as a movement in the real exchange rate between two countries and allows to compare the predictions of our model with the evidence on exchange rate pass-through available in the literature.

6.1 The pass-through of idiosyncratic cost shocks

One distinctive property of our model is that it predicts not only that pass-through is heterogeneous in productivity but also that the relationship between pass-through and productivity is non monotonic. Two forces determine pass-through of cost shocks. The first one has to do with the strength of the customer retention motive. Firms at higher risk of losing customers (i.e. those with higher extensive margin elasticity) have lower incentives to pass-through cost shocks. Given the relationship between extensive margin elasticity and productivity, this margin implies that pass-through is increasing in productivity. The second force is a feature of CES demand and goes in the opposite direction.[15] In the standard CES economy, firms with high cost pass-through more than firms with low cost of production. The intensive margin elasticity increases with the price charged by the firm. Higher productivity firms charging lower prices also face higher intensive margin elasticity, and pass-through less.[16]

Figure 3: Equilibirum Pass-through of an idiosyncratic cost shock



Note: The blue solid line plots the equilibrium pass-through to a marginal increase in production cost in the model at our baseline parameters. The red dashed line plots the pass-through in the CES economy with $\lambda \to \infty$.

In Figure 3 we display the equilibrium pass-through caused by a marginal increase in the production cost $(-\Delta\log(\mathcal{P}^*(z))/\Delta\log(z))$ as a function of the initial level of production cost. The continuous line portrays the relationship in the model with customer retention concerns; the dotted line represents the response in a CES economy.

[15]Examples of other models consistent with the same predictions are Atkeson and Burstein (2008), Melitz and Ottaviano (2008), or Garetto (2012).

[16]This is due to the concavity of $\varepsilon_d(p)$ which implies that the marginal sale reduction triggered by a price increase is smaller the higher the price. Since in the CES economy prices are monotonic in cost, it follows that high cost-high price firms choose to pass-through more. Similar consideration explain why, if we look within the two regions below and above the average cost, even in the model with customer retention concerns pass-through rises with cost.

The shape of the relationship in the baseline model can be understood comparing different sets of firms. First we contrast the behavior of firms whose production cost is low enough that customer retention concerns do not bind ("low cost" firms) with that of the remaining firms ("high cost" firms). Across these two groups of firms the first force discussed above dominates, so that "low cost" firms that pass-through more.[17] Next, we can compare pass-through of firms with different productivities but both located in the group for which customer retention concerns do not bind. In this case the CES effect prevails and pass-through decreases with productivity. The difference in average pass-through rate predicted by the baseline model (40%) and the CES one (78%) is therefore entirely driven by the behavior of high cost firms.

6.2 The pass-through of real exchange rate shocks

In this section, we consider the pass-through of an unexpected permanent change in production cost affecting a *subset* of the firms in the economy on the ground that this scenario is analogous to a shock to the real exchange rate. In fact, an appreciation of the real exchange rate -defined as the price the domestic numeraire good in units of the foreign numeraire- implies *de facto* an increase in labor cost for a subset of firms in the market (the *foreign* firms) relatively to another group (the *domestic* firms).[18]

The pass-through of idiosyncratic shocks we documented so far was less readily comparable with outside estimates for external validation. In fact, these shocks are rarely observed and, therefore, their pass-through is hardly documented. Simulating the effect of a real exchange rate shock allows to stack the predictions of our model against the findings of a large literature that has analyzed empirically the associated pass-through and the effects of tariffs on competition between foreign and domestic producers.[19]

We consider our economy in steady state at period t_0, as calibrated in Section 4. We assume that a fraction $\tilde{J} \subset [0,1]$ of the producers of good d are foreign firms, whereas the rest are domestic firms. We hit foreign firms with an unexpected and unforeseen permanent shock to production cost in the form of a scaling factor $(1+\tau)$, so that the marginal cost of production of a generic firm $j \in \tilde{J}$ goes from $w/z^j_{t_0}$ to $(1+\tau)w/z^j_{t_0}$. This shock is realized after the firm has learned about idiosyncratic productivity z but before pricing and customer's exit

[17] For instance if we compare the bottom and the top decile of the productivity distribution, we find that the passthrough in on average 22% for the former group and 55% for the latter.

[18] Given that we assume perfect competition in the market for the numeraire, the equilibrium wage coincides with the price of the numeraire.

[19] A shock to the effective cost of a subset of the players in the industry is the salient characteristic of a number of other real world scenarios. For instance, variations in state sales tax would affect local online sellers but not those located out-of-state, as the latter cannot be compelled to collect it (Einav et al. (forthcoming)). A similar effect is generated by the introduction of size-contingent employment protection legislation (Acemoglu and Angrist (2001), Schivardi and Torrini (2008)).

decisions are taken. Since we interpret our exercise as a perturbation of the equilibrium in a particular industry, we disregard general equilibrium effects on labor markets and income.

We set the fraction of firms in the economy hit by the shock to 12%, which matches the share of imports into U.S. personal consumption expenditure (Hale and Hobijn (2011)). It is well established that firms participating in export are more productive (Eaton et al. (2011)). To mimic this selection mechanism we assume that the distribution of idiosyncratic marginal cost of production, $\mu(z)$, is different for foreign and domestic producers. In particular, the mean of the cost distribution of foreign firms is 15% lower than that of domestic firms.[20]

Table 3: The average long-run response to a τ change in real exchange rate

	$\tau = 1\%$	$\tau = 5\%$
All firms	71%	72%
Top tercile	58%	60%
Bottom tercile	87%	87%

Notes: The pass-through is computed as the log-change in the long-run price or demand of each producer relatively to the absolute value of τ. Then we take an average across the different producers in a given set of firms (all firms, the most and least productive third), using the weight each producer has into aggregate consumption.

In Table 3 we report the long run pass-through and the elasticity of markup in response to a permanent 1% and 5% appreciation of the real exchange rate. The average pass-through of a 1% shock is 71%, which only goes up slightly when we consider a larger shock. This estimate is in the range reported by Hellerstein (2008) for wholesaler in the beer industry but below her estimated pass-through for retailers. At the top tercile of the productivity distribution of the foreign firms, the pass-through is below 60%; whereas at the bottom third of the same distribution the pass-through is almost 90%. These results are consistent with the evidence presented in Berman et al. (2012). Using data on French firms, they show that the pass-through of a real exchange rate shock on export prices is decreasing in productivity.[21]

In order to gauge to what extent these empirical results about exchange rate pass-through extend to generic cost shocks we repeated the experiment hitting with the cost shock τ a

[20] The gap between the average productivity of foreign and domestic firms is created in such a way that the average productivity of the economy as a whole is the same as in the baseline case. This is important to mantain the ratio between the scale of the search cost -which we are not reestimating for the sake of this exercise- and productivity constant.

[21] Berger and Vavra (2013) also find supporting evidence for heterogeneity in pass-through in BLS data, showing that individual items with high price change variance have greater exchange rate pass-through.

random subset of firms, representative of the overall distribution of productivity. In this case we find that the pass-through among "high cost" firms is higher than among "low cost" firms. This is analogous to results we obtained when we analysed the idiosyncratic cost shock. Therefore, it appears that selection of firms affected by the shock matters for the resulting heterogeneity in pass-through.

A final interesting result from this exercise is that not only foreign but also domestic firms adjust their prices, even though they have not been directly hit by the shock. Foreign firms are passing-through an actual cost shocks and their adjustment is limited because they are constrained by retention concerns. In fact, they are becoming less productive, which leads them to face a higher elasticity of the extensive margin of demand. It follows that foreign firms enjoy lower markup after the exchange rate shock, they raise their price less than proportionally and reduce their markup. "High cost" domestic firms increase prices without having experienced any cost shift, out of a pure feedback effect. The appreciation of the real exchange rate improves their position with respect to foreign firms and reduces their extensive margin elasticity. Since these firms choose a price in the increasing portion of the profit function, their reaction is to raise their price leading to higher markup. In contrast, "low cost" domestic firms do not respond to the shock as neither their demand nor their cost is affected. Garetto (2012) uses data on the European car industry to investigate the effects of an exchange rate shocks on the prices of domestic producers. Her findings are consistent with the predictions of our model. She documents that domestic firms adjust their prices following exchange rate fluctuations and that "high cost" producers react more than "low cost" ones.

7 Conclusions

The customer base is an important determinant of firm performance. Introducing customer retention consideration into standard models can improve our understanding of firm pricing behavior. We setup and estimate a model where firms face sticky demand and use it to explore the implications of this feature for the cyclicality of markups and cost-passthrough.

In our setting, there are two margins of demand adjustment. Customers can adjust both the quantity they purchase in response to price fluctuations and decide to leave and shop elsewhere. Because of search frictions, lost customers are hard to gain back and firms have an incentive to care about retaining their customers. We characterize the equilibrium of the model and show that customer retention concerns introduce a dynamic element into optimal markup so that markups depend on firm productivity.

We use scanner data on households' purchases at a U.S. supermarket chain to provide direct evidence that customers do respond to variation in the price of their consumption basket. We also exploit the data to estimate the key parameters of the model and provide a quantification of the effect of customer retention concerns on firm pricing.

We show that our model implies pro-cyclicality of markups not only to idiosyncratic shocks but also to aggregate ones (both technology and government spending). This is a feature documented in the data and we show that it cannot be reproduced by a standard CES model. Our model also matches the empirical evidence on heterogeneous pass-through of real exchange rate shocks but at the same time invites caution in extrapolating this finding to the pass-through of other types of cost shocks.

We notice that product market frictions is the only price setting friction in our economy, and that we study such friction in isolation from other possible relevant frictions for price dynamics. We showed that customer retention concerns give rise to pro-cyclical markups. This obviously does not prevent other types of frictions, such as nominal rigidities for instance, to introduce an element of counter-cyclicality in markups. Moreover, complementing our model with nominal rigidities may also allow to study the impact of monetary shocks on the real economy.

Finally, in this paper we have focused on the impact of the retention margin on price setting. However, the possibility to acquire customers through prices would induce incentives on price setting similar to the ones from customer retention. Extending our model to allow both for acquisition and retention of customers to depend on firm pricing, and measuring the role of both for markup dynamics, is an interesting topic for future research.

References

Acemoglu, D. and Angrist, J. (2001). Consequences of employment protection? the case of the americans with disabilities act, *Journal of Political Economy* **109**: 915–957.

Aguiar, M. and Hurst, E. (2007). Life-cycle prices and production, *American Economic Review* **97**(5): 1533–1559.

Alessandria, G. (2009). Consumer search, price dispersion, and international relative price fluctuations, *International Economic Review* **50**(3): 803–829.

Atkeson, A. and Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices, *American Economic Review* **98**(5): 1998–2031.

Berger, D. and Vavra, J. (2013). Volatility and pass-through, *Working paper*.

Berman, N., Martin, P. and Mayer, T. (2012). How do different exporters react to exchange rate changes?, *The Quarterly Journal of Economics* **127**(1): 437–492.

Bils, M. (1987). The cyclical behavior of marginal cost and price, *American Economic Review* **77**(5): 838–55.

Bils, M., Klenow, P. J. and Malin, B. A. (2012). Reset price inflation and the impact of monetary policy shocks, *American Economic Review* **102**(6): 2798–2825.

Blanchard, O. (2009). The state of macro, *Annual Review of Economics* **1**(1): 209–228.

Burdett, K. and Coles, M. G. (1997). Steady state price distributions in a noisy search equilibrium, *Journal of Economic Theory* **72**(1): 1 – 32.

Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment, *International Economic Review* **39**(2): 257–73.

Chatterjee, A., Dix-Carneiro, R. and Vichyanond, J. (2013). Multi-product firms and exchange rate fluctuations, *American Economic Journal: Economic Policy* **5**(2): pp. 77–110.

Coibion, O., Gorodnichenko, Y. and Hong, G. H. (2012). The cyclicality of sales, regular and effective prices: Business cycle and policy implications, *IMF Working Papers 12/207*, International Monetary Fund.

Coles, M. G. (2001). Equilibrium wage dispersion, firm size and growth, *Review of Economic Dynamics* **4**(1): 159–187.

Coles, M. G. and Mortensen, D. T. (2013). Equilibrium labor turnover, firm growth and unemployment, *Working paper*.

Diamond, P. A. (1971). A model of price adjustment, *Journal of Economic Theory* **3**(2): 156–168.

Drozd, L. A. and Nosal, J. B. (2012). Understanding international prices: Customers as capital, *American Economic Review* **102**(1): 364–95.

Dubois, P. and Jodar Rosell, S. (2010). Price and brand competition between differentiated retailers: A structural econometric model, *CEPR Discussion Papers 7847*, C.E.P.R. Discussion Papers.

Eaton, J., Kortum, S. and Kramarz, F. (2011). An anatomy of international trade: Evidence from french firms, *Econometrica* **79**(5): 1453–1498.

Eichenbaum, M., Jaimovich, N. and Rebelo, S. (2011). Reference prices and nominal rigidities, *American Economic Review* **101**(1): 234–262.

Einav, L., Knoepfle, D., Levin, J. and Sundaresan, N. (forthcoming). Sales taxes and internet commerce, *Amer* .

Einav, L., Leibtag, E. and Nevo, A. (2010). Recording discrepancies in nielsen homescan data: Are they present and do they matter?, *Quantitative Marketing and Economics* **8**(2): 207–239.

Foster, L., Haltiwanger, J. and Syverson, C. (2012). The slow growth of new plants: Learning about demand?, *Working Papers 12-06*, Center for Economic Studies, U.S. Census Bureau.

Garetto, S. (2012). Firm's heterogeneity and incomplete pass-through, *mimeo*, Boston University.

Goldberg, P. and Hellerstein, R. (2009). How rigid are producer prices?, *Working Papers 1184*, Princeton University, Department of Economics, Center for Economic Policy Studies.

Gopinath, G. and Itskhoki, O. (2011). In search of real rigidities, *NBER Macroeconomics Annual 2010, Volume 25*, NBER Chapters, National Bureau of Economic Research, Inc, pp. 261–309.

Gourio, F. and Rudanko, L. (2011). Customer capital, *NBER Working Papers 17191*, National Bureau of Economic Research, Inc.

Hale, G. and Hobijn, B. (2011). The u.s. content of made in china, *FRBSF Economic Letter* (Aug 8).

Hall, R. E. (2012). The cyclical response of advertising refutes counter-cyclical profit margins in favor of product-market frictions, *Working Paper 18370*, National Bureau of Economic Research.

Hellerstein, R. (2008). Who bears the cost of a change in the exchange rate? pass-through accounting for the case of beer, *Journal of International Economics* **76**: 14–32.

Holmes, T. (2011). The diffusion of wal-mart and economies of density, *Econometrica* **79**: 253–302.

Kaplan, G. and Menzio, G. (2013). Five facts about price dispersion, *Working paper*, University of Pennsylvania.

Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade, *Review of Economic Studies* **62**(4): 515–39.

Kleshchelski, I. and Vincent, N. (2009). Market share and price rigidity, *Journal of Monetary Economics* **56**(3): 344–352.

Luttmer, E. G. (2010). Models of growth and firm heterogeneity, *Annual Review of Economics* **2**(1): 547–576.

Melitz, M. J. and Ottaviano, G. I. P. (2008). Market size, trade, and productivity, *Review of Economic Studies* **75**(1): 295–316.

Menzio, G. (2007). A search theory of rigid prices, *PIER Working Paper Archive 07-031*, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Nakamura, E. and Steinsson, J. (2010). Monetary non-neutrality in a multisector menu cost model, *The Quarterly Journal of Economics* **125**(3): 961–1013.

Nakamura, E. and Steinsson, J. (2011). Price setting in forward-looking customer markets, *Journal of Monetary Economics* **58**(3): 220–233.

Nekarda, C. J. and Ramey, V. (2013). The cyclical behavior of the price-cost markup, *mimeo*, UC San Diego.

Petrin, A. and Warzynski, F. (2012). The impact of research and development on quality, productivity and welfare, *mimeo*.

Phelps, E. S. and Winter, S. G. (1970). Optimal price policy under atomistic competition, *in* G. Archibald, A. A. A. and E. S. Phelps (eds), *Microeconomic Foundations of Employment and Inflation Theory*, New York: Norton.

Pistaferri, L. (2003). Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply, *Journal of Labor Economics* **21**(3): 729–754.

Ravn, M., Schmitt-Grohe, S. and Uribe, M. (2006). Deep habits, *Review of Economic Studies* **73**(1): 195–218.

Rotemberg, J. J. and Woodford, M. (1991). Markups and the business cycle, *NBER Macroeconomics Annual 1991, Volume 6*, NBER Chapters, National Bureau of Economic Research, Inc, pp. 63–140.

Rotemberg, J. J. and Woodford, M. (1999). The cyclical behavior of prices and costs, *NBER Working Papers 6909*, National Bureau of Economic Research, Inc.

Schivardi, F. and Torrini, R. (2008). Identifying the effects of firing restrictions through size-contingent differences in regulation, *Labour Economics* **15**: 482–511.

A Technical Appendix

A.1 Derivation of assumptions from first principles

In this section we propose a micro founded model that can give rise to the assumptions we made on the paper regarding customer's demand $d(p)$ and surplus $v(p)$.

Let customers derive utility from a large number of varieties $N > 1$ according to $u(C) = C^{1-\gamma}/(1-\gamma)$, where $C = \left(\sum_{n=1}^{N} c_n^{\frac{\theta-1}{\theta}}\right)^{\frac{\theta}{\theta-1}}$ is a CES consumption aggregator, where c_n denotes the consumption level of each variety, and $\theta > 1$. The way to connect this setup with the one used in the paper is to consider $d(p) = c_n(p_n)$ and $v(p) = u(C(p_n))$, where p_n denotes the price of variety n. That is, our model can be interpreted as evaluating the competition in the industry producing variety n, where firms differ in their productivity.

Each period the customer maximizes,

$$\max_{\{c_n\}_{n=1}^{N}} \; u(C) \text{ subject to } \sum_{n=1}^{N} p_n c_n = I \; ,$$

from where we obtain the standard first order condition for variety n,

$$u'(C)\frac{\partial C}{\partial c_n}\frac{\theta-1}{\theta}c_n^{\frac{\theta-1}{\theta}-1} = \lambda p_n \text{ for all } \; n,$$

where λ denotes the Lagrange multiplier of the budget constraint. Operating with the first order conditions we obtain that

$$c_n = \frac{I}{P}\left(\frac{p_n}{P}\right)^{-\theta} \quad , \quad P \equiv \left(\sum_{n=1}^{N} p_n^{1-\theta}\right)^{\frac{1}{1-\theta}} \quad ,$$

where we have written the demand for variety n, c_n, as a function of its own price, p_n, and the price level, P. Notice that the price level is such that $P\,C = I$.

We start by discussing the properties of the demand function $d(p)$, which in this setup maps to evaluating the properties of $c_n(p_n)$. It is immediate to see that the demand for variety n converges to zero as its price diverges to infinity. That is, $\lim_{p_n\to\infty} c_n = 0$, which follows directly from the expression for c_n. We now show that when the number of varieties is large, the demand for variety n is decreasing and convex in its price. To this end, it proves useful to compute the following derivatives,

$$\begin{aligned}
\frac{\partial P}{\partial p_n} &= \left(\frac{P}{p_n}\right)^{\theta} \equiv A(p_n), \\
\frac{\partial c_n}{\partial p_n} &= \frac{c_n}{p_n}\left[-\theta + (\theta-1)\ A(p_n)^{\frac{\theta-1}{\theta}}\right] ,
\end{aligned}$$

where $A(p_n) > 0$ and $A(p_n) = N^{\theta}$ in a symmetric equilibrium. Notice that in the symmetric equilibrium, if N is large, we have that $\frac{\partial c_n}{\partial p_n} < 0$ and $\frac{\partial^2 c_n}{\partial p_n^2} > 0$, consistent with the demand function $d(p)$ being decreasing and convex in p. Moreover, because c_n and the price index P are twice continuously differentiable in prices and number of varieties N, the result also applies more generally away from the symmetric equilibrium.

Direct computations using the definition of c_n and price index P provide that the elasticity of demand of variety n is given by

$$\varepsilon_d(p_n) = -\frac{\partial \ln c_n}{\partial \ln p_n} = \theta - (\theta-1)\ \frac{c_n p_n}{I},$$

where $c_n p_n = \left(\sum_{i=1}^{N}\left(\frac{p_i}{p_n}\right)^{1-\theta}\right)^{-1}$ which, in a symmetric equilibrium, is positive when $N\theta > \theta - 1$. For example, this condition is guaranteed to apply for large N. In particular, as N diverges to infinity we get that $\varepsilon_d(p_n) = \theta$, so that when there are infinite many varieties the demand elasticity is constant. Moreover, notice that

$$\frac{\partial \varepsilon_d(p_n)}{\partial p_n} = (\theta-1)^2 \frac{1}{p_n}\left(\sum_{i=1}^{N}\left(\frac{p_i}{p_n}\right)^{1-\theta}\right)^{-1}\left[1 - \left(\sum_{i=1}^{N}\left(\frac{p_i}{p_n}\right)^{1-\theta}\right)^{-1}\right] ,$$

which in a symmetric equilibrium is equal to $(1/p_n)(\theta-1)^2(1-1/N)/N > 0$.

Now we evaluate $v(p)$. In this micro funded setup this maps into exploring the effect of p_n on $u(C)$. Recall that, from the construction of the price index P, $P\ C = I$, so that

$$\begin{aligned}
\frac{\partial C}{\partial p_n} &= -\frac{c_n}{P} , \\
\frac{\partial^2 C}{\partial p_n^2} &= -\left[\frac{\partial c_n}{\partial p_n}\frac{1}{P} - \frac{c_n}{P^2}A(p_n)\right] ,
\end{aligned}$$

Then,

$$
\begin{aligned}
\frac{\partial u(C)}{\partial p_n} &= C^{-\gamma}\frac{\partial C}{\partial p_n} < 0 \ , \\
\frac{\partial^2 u(C)}{\partial p_n^2} &= -C^{-\gamma-1}\left(\frac{\partial C}{\partial p_n}\right)^2\left[\gamma - \frac{C}{\frac{\partial C}{\partial p_n}}\frac{\frac{\partial^2 C}{\partial p_n^2}}{\frac{\partial C}{\partial p_n}}\right] \\
&= -C^{-\gamma-1}\left(\frac{\partial C}{\partial p_n}\right)^2\left[\gamma + \theta\left(1 - A(p_n)^{\frac{1-\theta}{\theta}}\right) - 2\right]
\end{aligned}
$$

so that $\frac{\partial^2 u(C)}{\partial p_n^2} \leq 0$ if $\gamma + \theta\left(1 - A(p_n)^{\frac{1-\theta}{\theta}}\right) - 2 \geq 0$. For example, in the symmetric equilibrium, where $A(p_n) = N^\theta$, the required condition can be rewritten as $\gamma + \theta\left(1 - N^{1-\theta}\right) \geq 2$, which shows that a sufficient condition (in the symmetric equilibrium) is that $\gamma \geq 2$.

We now explore the existence of a unique solution that maximizes the profit function of the firm. This involves proving two different things. First, that there exists a unique solution to $\partial\pi(p,z)/\partial p = 0$. Second, that this solution is a maximum (i.e., that the profit function is strictly concave).

The first derivative of the profit function with respect to the price reads,

$$
\frac{\partial \pi(p_n,z)}{\partial p_n} = c_n\left[1 - \varepsilon_d(p_n)\left(1 - \frac{w/z}{p_n}\right)\right] \ ,
$$

where a solution to $\frac{\partial\pi(p_n,z)}{\partial p_n} = 0$ exists and it is unique if $\frac{p_n}{w/z} = \frac{\varepsilon_d(p_n)}{\varepsilon_d(p_n)-1}$ has a unique solution. Let $h_1(p_n) \equiv \frac{p_n}{w/z}$ and $h_2(p_n) \equiv \frac{\varepsilon_d(p_n)}{\varepsilon_d(p_n)-1}$. Notice that $h_1(p_n)$ is continuous, strictly increasing, with $h_1(0) = 0$ and $\lim_{p_n\to\infty} h_1(p_n) = \infty$. Also, because $\varepsilon_d(p_n)$ is continuous and increasing, $h_2(p_n)$ is continuous, decreasing, with $\lim_{p_n\to\infty} h_2(p_n) = \theta$. It the follows that, for any number of varieties N, there exists a unique price solving $\frac{\partial\pi(p_n,z)}{\partial p_n} = 0$.

We now show that this unique price maximizes the firm's profits. To this end we show that in a symmetric equilibrium, for large N, the profit function evaluated at this price is concave. The second derivative of the profit function with respect to p_n reads,

$$
\frac{\partial^2\pi(p_n,z)}{\partial p_n^2} = -\frac{c_n}{p_n}\left[\varepsilon_d(p_n)\left(1 - \varepsilon_d(p_n)\left(1 - \frac{w/z}{p_n}\right)\right) + p_n\varepsilon_d'(p_n)(1 - \frac{w/z}{p_n}) + \varepsilon_d(p_n)\frac{w/z}{p_n}\right] \ .
$$

Notice that, in a symmetric equilibrium, c_n, p_n, $\varepsilon_d(p_n)$, and $\varepsilon_d'(p_n)$ are continuous in N. We will use this fact to prove that for large N the profit function is concave at the price maximizing static profits. Notice that when N diverges to infinity the second derivative reduces to

$$
\left.\frac{\partial^2\pi(p_n,z)}{\partial p_n^2}\right|_{N\to\infty} = -\frac{c_n}{p_n}\left[\theta\left(1 - \theta\left(1 - \frac{w/z}{p_n}\right)\right) + \theta\frac{w/z}{p_n}\right] \ ,
$$

because $\lim_{N\to\infty} \varepsilon_d(p_n) = \theta$ and $\lim_{N\to\infty} \varepsilon'_d(p_n) = 0$. Moreover, the markup $p_n/(w/z)$ can be obtained from equalizing the first derivative to zero. The markup in this case is $\theta/(\theta-1)$ and, as previously discussed, it is unique. Therefore,

$$\left.\frac{\partial^2 \pi(p_n, z)}{\partial p_n^2}\right|_{N\to\infty} = -\frac{c_n}{p_n}(\theta - 1) < 0 \ ,$$

so that when there are infinite many varieties, under the symmetric equilibrium the profit function has a unique maximizer, and it equalized the first derivative of the profit function to zero. Moreover, because c_n, p_n, $\varepsilon_d(p_n)$, and $\varepsilon'_d(p_n)$ are continuous in N, it is also the case that, in a symmetric equilibrium, $\frac{\partial^2 \pi(p_n,z)}{\partial p_n^2} < 0$ for large N. In the end, we concluded that if there is a large number of varieties, the profit function is concave, and $\partial\pi(p,z)/\partial p = 0$ characterizes its maximizer.

A.2 Proof of Proposition 1

First we prove that if $\hat{\mathcal{P}}(z) \le \bar{\mathcal{P}}(z)$ then $\mathcal{P}(z) = \tilde{\mathcal{P}}(z) = \{\hat{\mathcal{P}}(z)\}$ follows. By definition of $\hat{\mathcal{P}}(z)$ and $\bar{\mathcal{P}}(z)$, and given that firm's value is increasing in both profits per customer and customer base, $\hat{\mathcal{P}}(z) \le \bar{\mathcal{P}}(z)$ implies $\mathcal{P}(z) = \hat{\mathcal{P}}(z)$. Given that for all $p \le \bar{\mathcal{P}}(z)$, $\frac{\partial\Delta(p,z)}{\partial p} = 0$, the first order condition in equation (9) reduces to $\frac{\partial\pi(p,z)}{\partial p} = 0$, which is uniquely satisfied at $p = \hat{\mathcal{P}}(z)$. Therefore, $\hat{\mathcal{P}}(z) \in \tilde{\mathcal{P}}(z)$. Finally, given that $\pi(p,z)$ is strictly concave in p, attaining its maximum at $\hat{\mathcal{P}}(z)$, and $\Delta(p,z)$ non-increasing in p, we have that $\tilde{\mathcal{P}}(z) = \{\hat{\mathcal{P}}(z)\}$. Next, we show that the converse is true. Assume by contradiction that $\hat{\mathcal{P}}(z) > \bar{\mathcal{P}}(z)$ and $\mathcal{P}(z) = \tilde{\mathcal{P}}(z) = \{\hat{\mathcal{P}}(z)\}$. Because $\frac{\partial\Delta(p,z)}{\partial p} < 0$ for any $p > \bar{\mathcal{P}}(z)$, and given the definition of $\hat{\mathcal{P}}(z)$, we have that $\hat{\mathcal{P}}(z)$ cannot satisfy equation (9) , so $\hat{\mathcal{P}}(z) \notin \tilde{\mathcal{P}}(z)$, which constitutes a contradiction of $\tilde{\mathcal{P}}(z) = \{\hat{\mathcal{P}}(z)\}$. Given that $\Delta(p,z)$ is strictly decreasing in p for $p > \bar{\mathcal{P}}(z)$, and given that $\frac{\partial\pi(p,z)}{\partial p} = 0$ at $p = \hat{\mathcal{P}}(z)$ and continuous, the value function of the firm is strictly decreasing at $p = \hat{\mathcal{P}}(z)$. As a result, there exists another value of $p < \hat{\mathcal{P}}(z)$ that gives higher value to the firm, contradicting $\hat{\mathcal{P}}(z) \in \mathcal{P}(z)$.

We next show that if $\hat{\mathcal{P}}(z) > \bar{\mathcal{P}}(z)$, then $\mathcal{P}(z) \subset [\bar{\mathcal{P}}(z), \hat{\mathcal{P}}(z))$. First we show $\mathcal{P}(z) < \hat{\mathcal{P}}(z)$. Suppose in the contradiction that $\mathcal{P}(z) > \hat{\mathcal{P}}(z)$, then given that $\Delta(p,z)$ is strictly decreasing in p for $p > \bar{\mathcal{P}}(z)$ and given that $\frac{\partial\pi(p,z)}{\partial p} < 0$ at any $p > \hat{\mathcal{P}}(z)$ and continuous, implies that there exists another value of $p < \mathcal{P}(z)$ that gives higher value to the firm, contradicting optimality of $\mathcal{P}(z) > \hat{\mathcal{P}}(z)$. Moreover, given results above, we have that $\mathcal{P}(z) < \hat{\mathcal{P}}(z)$ (that is, $\hat{\mathcal{P}}(z) \notin \mathcal{P}(z)$). Finally, notice that for $p < \bar{\mathcal{P}}(z)$, $\Delta(p,z)$ is constant and profits per customer are strictly increasing in p, so that $\mathcal{P}(z) \ge \bar{\mathcal{P}}(z)$.

We next show that if the elasticity of the customer base is continuous, then $\mathcal{P}(z) \subset (\bar{\mathcal{P}}(z), \hat{\mathcal{P}}(z))$ and $\mathcal{P}(z) \subset \tilde{\mathcal{P}}(z)$. Start by noting that that the requirement for the elasticity being continuous is that $g(0) = 0$. First, for any $p < \bar{\mathcal{P}}(z)$, the firm's value function is strictly increasing in p because (i) the profit function is strictly increasing in p given that $\bar{\mathcal{P}}(z) < \hat{\mathcal{P}}(z)$, (ii) the customer base is constant. Moreover, because $\Delta(p,z)$ is continuously differentiable with $\lim_{p \uparrow \bar{\mathcal{P}}(z)} \frac{\partial \Delta(p,z)}{\partial p} = 0$, the value function is also strictly increasing in p at $p = \bar{\mathcal{P}}(z)$. As a result, $\mathcal{P}(z) \subset (\bar{\mathcal{P}}(z), \hat{\mathcal{P}}(z))$. Finally, given the value function is strictly increasing in p at $p = \bar{\mathcal{P}}(z)$, and strictly decreasing at $p = \hat{\mathcal{P}}(z)$, and because the value function is continuously differentiable in p, then there is at least a maximum in $(\bar{\mathcal{P}}(z), \hat{\mathcal{P}}(z))$, and such maximum has to satisfy the first order condition, so that $\mathcal{P}(z) \subset \tilde{\mathcal{P}}(z)$.

A.3 Proof of Lemma 2

We first show that the solution to the firm's problem is interior. We then show that when $g(\underline{\psi}) = 0$, the first order condition has at least one solution. After this we show that if the hazard rate of $G(\cdot)$ is increasing in ψ and δ is small, the solution to the first order condition is unique.

To see that the solution to the firm's problem is interior notice first that $\lim_{p\to\infty} \Delta(p,z)\Pi(p,z) < \Delta(\hat{\mathcal{P}}(z), z)\Pi(\hat{\mathcal{P}}(z), z)$. This implies that the optimal price would never diverge to ∞ as there exists a profitable deviation; for example, pricing so as to maximize static profits. Likewise, $\lim_{p\to 0} \Delta(p,z)\Pi(p,z) < \Delta(\hat{\mathcal{P}}(z), z)\Pi(\hat{\mathcal{P}}(z), z)$ if $\hat{\mathcal{P}}(z) < \bar{\mathcal{P}}(z)$ and $\lim_{p\to 0} \Delta(p,z)\Pi(p,z) < \Delta(\bar{\mathcal{P}}(z), z)\Pi(\bar{\mathcal{P}}(z), z)$ if $\hat{\mathcal{P}}(z) > \bar{\mathcal{P}}(z)$. This shows that the optimal price is interior.

We not turn to prove that, under the assumptions of the Lemma, the optimal price is characterized by the first order condition. It proves useful to define the following objects,

$$
\begin{aligned}
H_1(p,z) &\equiv \frac{\partial \Pi(p,z)}{\partial p}\frac{1}{\Pi(p,z)} = \frac{\partial \pi(p,z)}{\partial p}\frac{1}{\Pi(p,z)} ,\\
H_2(p,z) &\equiv -\frac{\partial \Delta(p,z)}{\partial p}\frac{1}{\Delta(p,z)} = \frac{\partial \bar{\psi}(p,z)}{\partial p}\frac{g(\bar{\psi}(p,z))}{1-G(\bar{\psi}(p,z))}\frac{1-G(\bar{\psi}(p,z))}{1-G(\bar{\psi}(p,z))+\delta} ,
\end{aligned}
$$

so that the first order condition in equation (9) can be rewritten as $H_1(p,z) = H_2(p,z)$.

We first show that there exists at least one price p that solves $H_1(p,z) = H_2(p,z)$. Notice that $H_1(p,z)$ is strictly decreasing in p with $H_1(\hat{\mathcal{P}}(z), z) = 0$. By Lemma 1, the function $H_2(p,z)$ is continuous in p for $p > \bar{\mathcal{P}}(z)$. It is immediate to see, because $\bar{\psi}(p,z) = 0$ for every $p \leq \bar{\mathcal{P}}(z)$, that $H_2(p,z)$ is continuous in p for $p < \bar{\mathcal{P}}(z)$. Continuity of $H_2(p,z)$ for every p requires continuity at $p = \bar{\mathcal{P}}(z)$. That is, it requires that $H_2(\bar{\mathcal{P}}(z)) = \lim_{p\uparrow \bar{\mathcal{P}}(z)} H_2(p,z) = \lim_{p\downarrow \bar{\mathcal{P}}(z)} H_2(p,z)$, which is satisfied if $g(\underline{\psi}) = 0$. Under the assumption that $g(\underline{\psi}) = 0$, because $H_2(p,z)$ is continuous, with $H_2(p,z) \geq 0$ for all p, it follows that there exists at least one price p which solves $H_1(p,z) = H_2(p,z)$.

We now show that the solution to the first order condition is unique when the hazard rate of $G(\cdot)$ is increasing and δ is small. Because $H_1(p,z)$ is strictly decreasing in p, it suffices to prove that $H_2(p,z)$ is strictly increasing in p for $p > \bar{\mathcal{P}}(z)$, so that $H_1(p,z)$ and $H_2(p,z)$ intersect only once. To prove that $H_2(p,z)$ is strictly increasing in p for $p > \bar{\mathcal{P}}(z)$, it is useful to define a new function, $\tilde{H}_2(p,z) \equiv \frac{\partial \bar{\psi}(p,z)}{\partial p} \frac{g(\bar{\psi}(p,z))}{1-G(\bar{\psi}(p,z))}$, where $\tilde{H}_2(p,z) = \lim_{\delta \to 0} H_2(p,z)$. Notice that, by continuity of $\frac{\partial H_2(p,z)}{\partial p}$ with respect to δ, for any $\varepsilon > 0$ there exists a $\bar{\delta}$ such that for any $\delta < \bar{\delta}$ we have $\left| \frac{\partial H_2(p,z)}{\partial p} - \frac{\partial \tilde{H}_2(p,z)}{\partial p} \right| < \varepsilon$. Notice that Lemma 1 provides that $\frac{\partial \bar{\psi}(p,z)}{\partial p} > 0$ for $p > \bar{\mathcal{P}}(z)$. This, together with the assumption that $\frac{g(\bar{\psi}(p,z))}{1-G(\bar{\psi}(p,z))}$ (i.e., increasing hazard rate) implies that $\tilde{H}_2(p,z)$ is strictly increasing in p for $p > \bar{\mathcal{P}}(z)$. In the end, because $\tilde{H}_2(p,z)$ in this range, there exists a value $\bar{\delta}$ such that if $\delta < \bar{\delta}$ also $H_2(p,z)$ is strictly increasing in this range. Then, the first order condition has a unique solution.

A.4 Proof of Proposition 3

Part (i) of the proposition follows immediately from Proposition 1.

We now prove part (ii) of the proposition. It is useful to recall that $\hat{V} = \int_{\underline{z}}^{\bar{z}} \bar{V}(\mathcal{P}^*(z), z)\mu(z)dz$ and that a customer exits the customer base of the firm if $\hat{V} - \psi > \bar{V}(\mathcal{P}^*(z), z)$. We first show that in equilibrium there is at least one productivity level where firms experience customers leaving, which implies that $\mathcal{Z}_2 \neq \emptyset$. We prove this by contradiction, by assuming that no firm experience customers exiting its customer base. If no customer exits from the customer base of any firm we need that $\bar{V}(\mathcal{P}^*(z), z) = \hat{V}$ for all z. An immediate implication is that there cannot be price dispersion, with $\mathcal{P}^*(z) = \mathcal{P}^*$ for all z. Then, because prices are independent of productivity, $\bar{\mathcal{P}}(z) = \bar{\mathcal{P}} = \mathcal{P}^*$ for all z. Notice that $\bar{\mathcal{P}} = \hat{\mathcal{P}}(\bar{z})$. If $\bar{\mathcal{P}} < \hat{\mathcal{P}}(\bar{z})$, by Proposition 1, firms with productivity $\bar{z}$ would deviate and charge a price above $\bar{\mathcal{P}}$. If $\bar{\mathcal{P}} < \hat{\mathcal{P}}(\bar{z})$, also by Proposition 1, firms with productivity $\bar{z}$ would deviate and charge a price below $\bar{\mathcal{P}}$. In both cases, $\bar{V}(\mathcal{P}^*(z), z) \neq \hat{V}$ for all z. Because $\bar{\mathcal{P}} = \hat{\mathcal{P}}(\bar{z})$ and $\hat{\mathcal{P}}(z)$ decreasing in productivity, it is the case that $\hat{\mathcal{P}}(\underline{z}) > \bar{\mathcal{P}}$. Then, by Proposition 1, firms with productivity $\underline{z}$ deviate to charge a price strictly above $\bar{\mathcal{P}}$, so that there is price dispersion and at least there is one productivity level at which firms experience a customer outflow and $\mathcal{P}^*(z) < \hat{\mathcal{P}}(z)$. Then, $\mathcal{Z}_2 \neq \emptyset$. This also implies that $\mathcal{Z}_1$ is non-empty. This follows because, if some firms' optimal price is such that $\bar{V}(\mathcal{P}^*(z), z) < \hat{V}$, then for another set of firms it has to be the case that $\bar{V}(\mathcal{P}^*(z), z) > \hat{V}$. Because no one exits the customer base of these firms, this implies that $\hat{\mathcal{P}}(z) < \bar{\mathcal{P}}(z)$ so that, by applying Proposition 1, $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z)$.

Notice that, for any $z \in \mathcal{Z}_1$, it is the case that $\mathcal{P}^*(z) \leq \bar{\mathcal{P}}(z)$. This implies that, for any $z \in \mathcal{Z}_1$, $\Delta(\mathcal{P}^*(z), z) = 1 + \delta$. For any $y \in \mathcal{Z}_2$, because $\mathcal{P}^*(y) > \bar{\mathcal{P}}(y)$, we have that $\Delta(\mathcal{P}^*(y), y) = 1 + \delta - G(\bar{\psi}(\mathcal{P}^*(y), y)) \in (\delta, 1+\delta)$. Finally, given that $\int_{\underline{z}}^{\bar{z}} \Delta(\mathcal{P}^*(z))\mu(z)dz = 1$, because for any $z \in \mathcal{Z}_1$ we already concluded that $\Delta(\mathcal{P}^*(z), z) > 1$, it has to be the case that for some $y \in \mathcal{Z}_2$, $\Delta(\mathcal{P}^*(y), y) < 1$.

A.5 Proof of Proposition 4

The outline of the proof is the following. We start by showing that the firm pricing decision $\mathcal{P}(z)$ is continuous in z and $\bar{\mathcal{P}}(\cdot)$. Then, we show that, under the assumptions of the proposition, $\bar{\mathcal{P}}(\cdot)$ is continuous in z and firm's optimal pricing $\mathcal{P}(z)$. As a consequence, we combine all these results to obtain a continuous mapping from prices $\mathcal{P}(z)$ into itself, and apply Brouwer's fixed-point theorem to show that there exists a solution $\mathcal{P}^*(z)$.

Let $\mathcal{T}_1(z, \bar{\mathcal{P}}(z))$ denote the mapping from productivity z and the threshold price $\bar{\mathcal{P}}(z)$ to optimal pricing $\mathcal{P}(z)$.

Claim 1 *If the assumptions of Lemma 2 are satisfied the mapping $\mathcal{T}_1(\cdot,\cdot)$ is unique and continuous in both arguments.*

The proof follows immediately from Proposition 1, equation (3) and Lemma 2. Because the first order condition is continuous in p and $\bar{\mathcal{P}}(z)$, $\mathcal{T}_1(\cdot,\cdot)$ is continuous in both arguments.

Let $\mathcal{T}_2(z, \mathcal{P})$ denote the mapping from productivity z and the optimal price function $\mathcal{P}(\cdot)$ to threshold pricing $\bar{\mathcal{P}}(z)$.

Claim 2 *The mapping $\mathcal{T}_2(\cdot,\cdot)$ is unique and continuous in both arguments.*

The proof follows immediately from equations (1)-(2).

Let $\mathcal{T}(z, \mathcal{P}) \equiv \mathcal{T}_1(z, \mathcal{T}_2(z, \mathcal{P}))$ denote the function that maps productivity z and a pricing function $\mathcal{P}(\cdot)$ into the space where $\mathcal{P}(\cdot)$ belongs.

Claim 3 *$\mathcal{T}(\cdot,\cdot)$ is continuous in both arguments.*

The proof follows because the composition of continuous functions is continuous.

Notice, given that the set of z is compact and given Proposition 1, we have that $\mathcal{P}^*(z)$ lies in a compact set. Moreover, we know that $\mathcal{P}(z) \in [0, \hat{\mathcal{P}}(\underline{z})]$, so that the operator $\mathcal{T}(z, \mathcal{P})$ maps prices $\mathcal{P}$ from the set $[0, \hat{\mathcal{P}}(\underline{z})]$ to a subset of it. By applying Brouwer's fixed-point theorem, we obtain existence of a solution $\mathcal{P}^*(z) = \mathcal{T}(z, \mathcal{P}^*)$ for all z.

A.6 Proof of Remark 2

We first provide a proof of the first part of the Remark. Under the assumptions of the remark every firm has the same productivity z at every point in time. As a result, as previously discussed, every firm chooses the same price $\mathcal{P}^*(z)$. We prove the statement in two steps. In the first step we conjecture a solution and then verify that it is an equilibrium. In the second step we show that there are no other solutions.

Step 1: Conjecture that firms choose $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z)$. It is immediate to see that $\bar{\mathcal{P}}(z) = \hat{\mathcal{P}}(z)$. Because $\mathcal{P}^*(z) = \bar{\mathcal{P}}(z)$, a direct application of Proposition 1 validates the conjecture.

Step 2: We now show by contradiction that $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z)$ is the unique price schedule in equilibrium. There are two cases: one were firms price above $\hat{\mathcal{P}}(z)$ and one were they price below. For the first case, conjecture that $\mathcal{P}^*(z) > \hat{\mathcal{P}}(z)$. Again, it is immediate to see that $\bar{\mathcal{P}}(z) > \hat{\mathcal{P}}(z)$. However, by Proposition 1, given that $\hat{\mathcal{P}}(z) < \bar{\mathcal{P}}(z)$ firms should deviate and price at $\hat{\mathcal{P}}(z)$. Therefore, $\mathcal{P}^*(z) > \hat{\mathcal{P}}(z)$ cannot happen in equilibrium. For the second case, conjecture that $\mathcal{P}^*(z) < \hat{\mathcal{P}}(z)$. Now we have that $\bar{\mathcal{P}}(z) < \hat{\mathcal{P}}(z)$. However, by Proposition 1, given that $\hat{\mathcal{P}}(z) > \bar{\mathcal{P}}(z)$ a firm's optimal price should satisfy $\bar{\mathcal{P}}(z) < \mathcal{P}^*(z) < \hat{\mathcal{P}}(z)$. Therefore, $\mathcal{P}^*(z) < \hat{\mathcal{P}}(z)$ cannot happen in equilibrium.

The proof of the second part of the corollary is straightforward and follows from Bertrand's competition.

A.7 Proof of Proposition 5

It proves useful to define the following object,

$$\tilde{V}(z) = \int_{\underline{z}}^{\bar{z}} \left[\int_0^\infty \max\left\{ \bar{V}(\mathcal{P}^*(z'), z'), \hat{V} - \psi \right\} \ g(\psi)\ d\psi \right] \mu(z'|\, z)\ dz' \ .$$

So that the value for a remaining customer is $\bar{V}(p,z) = v(p) + \beta\tilde{V}(z)$. Notice that, because $\mathcal{P}^*(z)$ is decreasing in z, $v(p)$ decreasing in p, and the process for the productivity shock displaying persistence, we have that $\bar{V}(\mathcal{P}^*(z), z)$ is strictly increasing in z. It is then immediate to get that $\bar{\mathcal{P}}(z)$ is strictly decreasing in z. Furthermore, all of these together with the price that maximizes static profits $\hat{\mathcal{P}}(z)$ being strictly decreasing in z, imply the existence of the unique threshold $\hat{z}$ where firms with productivity above $\hat{z}$ charge $\hat{\mathcal{P}}(z)$ and those below the threshold charge a price below the price that maximizes their static profits. Finally, Proposition 3 provides that $\hat{z}$ is interior.

A.8 Proof of Proposition 6

It proves useful to define the following object,

$$\tilde{V}(z) = \int_{\underline{z}}^{\bar{z}} \left[\int_0^\infty \max\left\{ \bar{V}(\mathcal{P}(z'), z'), \hat{V} - \psi \right\} \ g(\psi)\ d\psi \right] \mu(z'|\, z)\ dz' \ .$$

So that the value for a remaining customer is $\bar{V}(p,z) = v(p) + \beta\tilde{V}(z)$.

1) Productivity shocks are i.i.d., $\rho = 0$. Points (i)-(ii). We start by showing that $\bar{\mathcal{P}}(z) = \bar{\mathcal{P}}$. Under the assumptions of the proposition we have that $\mu(z'|\, z) = \mu(z'|\, y)$ for all z, y. As a result, we get that $\tilde{V}(z) = \tilde{V}$. Because the continuation value of being matched with any firm equals $\tilde{V}$, using equation (2) we get that $\bar{\mathcal{P}}(z) = \bar{\mathcal{P}}$. Notice that this result immediately implies that $\varepsilon_m(p, z) = \varepsilon_m(p)$.

To prove that there exists a unique value $\hat{z}$ recall that $\hat{\mathcal{P}}(z)$ is decreasing in z. This, together with Proposition 3 and $\bar{\mathcal{P}}(z)$ constant in z, immediately implies the existence of the unique threshold $\hat{z}$ where firms with productivity above $\hat{z}$ charge $\hat{\mathcal{P}}(z)$.

We now prove that the pricing schedule $\mathcal{P}^*(z)$ is monotonic. We do this in three steps:

Step 1: firms charging $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z)$. For every firm with productivity $z \in \mathcal{Z}_1$ the optimal price $\mathcal{P}^*(z)$ decreases monotonically when $z \in \mathcal{Z}_1$ because $\hat{\mathcal{P}}(z)$ decreases monotonically.

Step 2: firms charging $\mathcal{P}^*(z) < \hat{\mathcal{P}}(z)$. Consider two productivity levels z, z' with $z > z'$. Conjecture, in the contradiction, that $\mathcal{P}^*(z) > \mathcal{P}^*(z')$. There are two cases. Case 1: suppose that $\mathcal{P}^*(z') > \hat{\mathcal{P}}(z)$. Here, because $\partial\pi(p, z)/\partial p < 0$ for all $p > \hat{\mathcal{P}}(z)$ and $\bar{\mathcal{P}}(z) = \bar{\mathcal{P}}$ for all z , a firm with productivity z has a deviation that increases her value by charging price $\mathcal{P}^*(z')$. Case 2: suppose that $\mathcal{P}^*(z') < \hat{\mathcal{P}}(z)$. Here, because $\partial\pi(p, z')/\partial p > 0$ for all $p < \hat{\mathcal{P}}(z')$ and $\bar{\mathcal{P}}(z) = \bar{\mathcal{P}}$ for all z, a firm with productivity z' has a deviation that increases her value by charging $\mathcal{P}^*(z)$. Hence, in both cases we get a contradiction. Therefore, $\mathcal{P}^*(z)$ decreases monotonically when $z \in \mathcal{Z}_2$.

Step 3: the firms with lowest productivity in set $\mathcal{Z}_1$ and firm with highest productivity in set $\mathcal{Z}_2$. Consider the firm with the lowest productivity in $\mathcal{Z}_1$, which we label by z, and the firm with highest productivity belonging to set $\mathcal{Z}_2$, which we label by z'. We now show that $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z) < \mathcal{P}^*(z')$. Consider, in the contradiction, that $\mathcal{P}^*(z') < \mathcal{P}^*(z)$. Notice that (i) the firms with productivity z do not face customer retention concerns and (ii) $\partial\pi(p, z')/\partial p > 0$ for any $p < \hat{\mathcal{P}}(z')$. Because of (i) and (ii), it is immediate that a firm with productivity z' has a deviation that increases her value by setting a price equal to $\mathcal{P}^*(z) = \hat{\mathcal{P}}(z)$.

2) Persistent productivity shocks, $\rho \in (0, 1)$. Points (i)-(ii). We will prove the statement by contradiction. Pick two productivity levels $z_l < z_h$. Conjecture that $\mathcal{P}^*(z_l) \leq \mathcal{P}^*(z_h)$. We rewrite $\tilde{V}(z)$ as follows,

$$\tilde{V}(z) = (1 - \rho)\ \hat{V} + \rho \int_0^\infty \max\left\{\bar{V}(\mathcal{P}^*(z), z), \hat{V} - \psi\right\}\ g(\psi)\ d\psi$$

from where it is clear that if $\mathcal{P}^*(z_l) \leq \mathcal{P}^*(z_h)$ then, using equation (2), it is necessary that $\bar{\mathcal{P}}(z_l) \geq \bar{\mathcal{P}}(z_h)$. We now show that when $\bar{\mathcal{P}}(z_l) \geq \bar{\mathcal{P}}(z_h)$, it is the case that $\mathcal{P}^*(z_l) > \mathcal{P}^*(z_h)$, which constitutes a contradiction.

Notice that we can rewrite equation (8) as $F(z) = \max_p H_1(p,z) + H_2(p,z)$ where

$$
\begin{aligned}
H_1(p,z) &\equiv \frac{\Delta(p,z)}{1-\beta\rho\Delta(p,z)}\pi(p,z) \ , \\
H_2(p,z) &\equiv \frac{\Delta(p,z)}{1-\beta\rho\Delta(p,z)}\beta(1-\rho)\mathcal{F} \ ,
\end{aligned}
$$

where $\mathcal{F}$ is a constant. Notice that, using equation (3), $\arg\max_p H_1(p,z_H) < \arg\max_p H_1(p,z_L)$ if $\bar{\mathcal{P}}(z_l) \geq \bar{\mathcal{P}}(z_h)$. Also, notice that $H_2(p,z)$ is constant for any $p \leq \bar{\mathcal{P}}(z)$ and then strictly decreasing. As a result, it follows immediately that, as long as $\bar{\mathcal{P}}(z_l) \geq \bar{\mathcal{P}}(z_h)$, the $\arg\max H_1(p,z_H) + H_2(p,z_H) < \arg\max H_1(p,z_L) + H_2(p,z_L)$ so that $\mathcal{P}^*(z_H) < \mathcal{P}^*(z_L)$, which constitutes a contradiction. Then, $\mathcal{P}^*(z)$ decreases monotonically with z.

Because $\mathcal{P}^*(z)$ is decreasing in z, $\bar{\mathcal{P}}(z)$ is strictly increasing in z: this follows from the definition of $\tilde{V}$ and $\bar{\mathcal{P}}(z)$. To prove that there exists a unique value $\hat{z}$ recall that $\hat{\mathcal{P}}(z)$ is decreasing in z. This, together with Proposition 3 and $\bar{\mathcal{P}}(z)$ strictly increasing in z, immediately implies the existence of the unique threshold $\hat{z}$ where firms with productivity above $\hat{z}$ charge $\hat{\mathcal{P}}(z)$.

3) Point (iii) (for both cases). Monotonicity of $\mathcal{P}^*(z)$ and $\bar{\mathcal{P}}(z)$, together with Lemma 2 and Lemma 3 imply that $\partial\bar{\psi}(p,z)/\partial z \leq 0$ so that $\varepsilon_m(\mathcal{P}^*(z),z)$ decreases with z. The process in equation (13) implies $\alpha(p,z) = 1 + \beta\rho + \beta(1-\rho)\mathcal{F}/\pi(p,z)$. Finally, $\pi(\mathcal{P}^*(z),z)$ increasing in z guarantees that $\alpha(p,z)$ decreases with z. Then by using equation (10) Point (iii) follows.

A.9 Closing the model: a simple model of labor choice

We describe the workings of the labor market determining the equilibrium level of income I. We assume that each period a representative household chooses labor supply to solve the following problem

$$
\max_{\ell} \int_{\underline{z}}^{\bar{z}} \left(v(\mathcal{P}^*(z)) - \int_0^{v(\bar{\mathcal{P}}(z)) - v(\mathcal{P}^*(z))} \psi \ g(\psi) d\psi \right) \frac{M(z)}{\Gamma} dz - \frac{\ell^{1+\phi}}{1+\phi} \ , \tag{15}
$$

subject to

$$
\begin{aligned}
v(\mathcal{P}^*(z)) &= \frac{\left(I \ \left(\omega^{1-\theta}\mathcal{P}^*(z)^{1-\theta} + (1-\omega)^{1-\theta}\right)^{-\frac{1}{1-\theta}} \right)^{1-\gamma}}{1-\gamma} \ , \\
I &= w \ \ell + \int_{\underline{z}}^{\bar{z}} \pi(\mathcal{P}^*(z), z) \ \frac{M(z)}{\Gamma} \ dz \ ,
\end{aligned}
$$

and equation (3). The representative household takes prices and aggregate profits as given, and has the same measure of the customers, i.e Γ. The representative household chooses the labor supply ℓ and distributes labor proceeds equally across its customers. The parameter representing the disutility of labor (ϕ) is set to 1.43 so that the Frisch elasticity of labor supply is equal to 0.7 (see Pistaferri (2003)).

A.10 Augmenting the model with government spending

For simplicity, we assume that the government behaves exactly as the consumers in our economy, i.e. it is allocated a fraction g of household's income, i.e. $gI\Gamma$, and distributes proceeds equally across different shoppers. In the steady state we assume $g = 0$. Government purchases are of no value to the household. Thus the only difference between the representative household and the government in our model is that the representative household also makes a labor decision, taking the lump-sum transfer by the government given, while the government' shoppers take similar decisions as the shoppers in the household. The impact of higher government spending on the household is that it decreases disposable income for given labor proceeds and profits, while it leaves firms' demand unchanged.

The lump-sum tax on the household is such that the government budget balances every period, i.e. $\tau = gI\Gamma$. As a consequence, if $d(\mathcal{P}^*(z))$ and $n(\mathcal{P}^*(z))$ are the demand of the good d and the numeraire good by household' shoppers when matched to a producer of good d with productivity z, the corresponding demand by the government's shoppers are $d(\mathcal{P}^*(z))\ g/(1-g)$ and $n(\mathcal{P}^*(z))\ g/(1-g)$ respectively.

The equilibrium in the labor markets requires that labor supply equals labor demand, i.e.

$$\ell = (1 + \frac{g}{1-g}) \left(\int_{\underline{z}}^{\bar{z}} \frac{d(\mathcal{P}^*(z))}{z} \ \frac{M(z)}{\Gamma} \ dz + n(\mathcal{P}^*(z)) \right) \ . \tag{16}$$

Finally, given that the representative household owns the firms, the relevant T-periods ahead stochastic discount factor in the firm maximization problem is given by

$$Q_{t+T|t} = \beta^T \ \frac{\int_{\underline{z}}^{\bar{z}} (C(\mathcal{P}^*_{t+T}(z)))^{-\gamma} M(z) dz}{\int_{\underline{z}}^{\bar{z}} (C(\mathcal{P}^*_{t}(z)))^{-\gamma} M(z) dz} \ , \tag{17}$$

where $U(C) = C^{1-\gamma}/(1-\gamma)$, and $C(\mathcal{P}^*_t(z)) = I \ \left(\omega^{1-\theta}\mathcal{P}^*_t(z)^{1-\theta} + (1-\omega)^{1-\theta}\right)^{-\frac{1}{1-\theta}}$ is the consumption basket of consumers buying good d at price $\mathcal{P}^*_t(z)$.

B Data sources and variables construction

B.1 Data sources

The empirical evidence presented in Section 3 is based on two data sources provided by a large supermarket chain that operates over 1500 stores across the US. We exploit information on weekly store sales between January 2004 and December 2006 for a panel of over 200 stores located in 10 different states. For each good (identified by its UPC) carried by the stores in those weeks, the data report total amount grossed and quantity sold.

In addition to store level data, we have information on grocery purchases at the chain between June 2004 and June 2006 for a panel of over 11,000 households. For each grocery trip made by a household, we observe date and store where the trip occurred, the collection of all the UPC's purchased with quantity and price paid. The data include information on the presence and size of price discounts but do not generally report redemption of manufacturer coupons. The geographical dispersion of the households mirrors that of the store data: our customers live in some 1,500 different zipcodes across 10 states. Data are recorded through usage of the loyalty card; the retailer is able to link loyalty cards belonging to different memebr of the family to a single household identifier. Purchases made without using the card are not recorded. However, the chain ensures that the loyalty card has a high penetration by keeping to a minimum the effort needed to register for one. Furthermore, nearly all promotional discount are tied to ownership of a loyalty card, which provides a strong incentive to use it. Another potential drawback of the data is that we only follow households when purchasing at stores of a single, albeit large, supermarket chain. Other data sources on the same industry, like the Nielsen Homescan database, rely on households themselves scanning the barcodes of the items purchased once they return home after a trip and can therefore track them shopping at a plurality of competing firms. On the other hand, cash register data contain significantly less measurement error than databases relying on home scanning (Einav et al. (2010)).

B.2 Variables construction

Exit from customer base. The dependent variable in the regression presented in equation (14) is an indicator for whether a customer is exiting the customer base of the chain. With data on grocery purchases at a single retail chain it is hard to definitively assess whether a household has abandoned the retailer to shop elsewhere or it is simply not purchasing grocery in a particular week, for instance because it is leaving off its inventory. In fact, we observe households when they buy grocery at the chain but do not have any information on their shopping at competing grocers. To circumvent this problem, we focus on a subsample of households who shop frequently at the chain. For them we can plausibly assume that sudden long spells without trips represent instances in which the household has left the chain and is fulfilling grocery needs shopping at one of its competitors. Operationally, we select households who made at least 48 trips at the chain over the two years spanned in the sample, implying that they would shop on average twice per month at the chain. When such households do not visit any supermarket store of the chain over at least eight consecutive weeks, we assume that the customer is shopping elsewhere. The *Exit* dummy is constructed so that it takes value of one in correspondence to the last visit at the chain before a spell of eight or more weeks without shopping there. Table 4 summarizes shopping behavior for households in our sample. It is immediate to notice that a 8-weeks spell without purchase is unusual, as customers tends to show up frequently at the stores. This strengthens our confidence that customer missing for such a long period have indeed switched to a different retailer.

Table 4: Descriptive statistics on customer shopping behavior

	Mean	*Std.dev.*	*25th pctile*	*75th pctile*
Number of trips	157	141	65	208
Days elapsed between consecutive trips	4.1	7.4	1	5
Frequency of exits	0.004	0.065		
Items in the basket	289.5	172.4		

Price of the basket. The household level scanner data report information on the price paid conditional on a certain item having been bought by the customer. Therefore, if we do not observe at least one household in our sample buying a given item in a store in a week, we would not be able to infer the price of the item in that store-week from the household panel data. However, the store level data allow us to calculate unit value prices every week for every item in sale in a given store, whether or not that particular UPC was bought by one of the households in our data. Unit value prices are computed using data on revenues and quantities sold as

$$UVP_{stu} = \frac{TR_{stu}}{Q_{stu}} \ ,$$

where TR represent total revenues and Q the total number of units sold of good u in week t in store s.

As explained in Eichenbaum et al. (2011) this only allows to recover an average price for goods that were on promotion. In fact the same good will be sold to loyalty card carrying customers at the promotional price and at full price to customers who not have or use a loyalty card. Without information on the fraction of these two types of customers it is not possible to recover the two prices separately. Furthermore, since prices are constructed based on information on sales, missing values can originate even in this case if no unit of a specific item is sold in a given store in a week. This is, however, an unfrequent circumstance and involves only rarely purchased UPC's, which are unlikely to represent important shares of the basket for any of the households in the sample. For the analysis, we only retain UPC's with at most two non consecutive missing price observations and impute price for the missing observation interpolating the prices of the contiguous weeks.

The retail chain applies different prices in different geographic areas and supplied weekly data on revenues and quantities sold by UPC for 270 stores that are representative of the different price areas. Households shop in one (or a subset) of some 1,500 stores and we have to devise a way to match the store a household visits to the price areas to which it belongs. However, we have no information on how the chain divides its markets into price areas. A possible solution is to infer in which price areas the store(s) visited by a household are located by comparing the prices contained in the household panel with those in the store data. In principle the household data should give information on enough UPC prices in a given week to identify the price area representative store whose pricing they are matching. However, even though two stores belonging in the same price area should have the same prices, they may not have the same unit value prices if the share of shoppers using the loyalty card differs in the two stores. Therefore, we choose to restrict attention to the 1,336 households whose most frequently visited store is one of the representative stores. Since the 270 representative stores are not selected following any particular criterion, the resulting subsample of households is not subject to any type of selection.

We are interested in observing whether households change their grocery supplier in response to fluctuations in the price of the basket of goods they purchase. To this end, we construct a price index summarizing for each customer in every week the price of the collection of goods she regularly buys. We include in a customer basket all the UPC's she purchased over the two years of data and construct the price of the basket for household i in week t by taking the average of the weekly prices of all the UPC's the customer purchased over the two years weighted by the share of her expenditure they represent. Namely:

$$p_{it} = \sum_{u \in U^i} w_{iu} p_{ut}, \quad w_{iu} = \frac{\sum_t E_{iut}}{\sum_{u \in U^i} \sum_t E_{iut}}$$

where U^i is the set of all the UPC's (u) purchased by household i during the sample period.

We choose to calculate the weights using the expenditure share of the UPC over the two years in the sample. This can lead to some inaccuracy in identifying the goods the customer cares for at a given point in time. For example, if a customer bough only Coke during the first year and only Pepsi during the second year of data, our procedure would have us give equal weight to the price of Coke and Pepsi throughout the sample period. If we used a shorter time interval, for example using the expenditure share in the month, we would correctly recognize that she only cares about Coke in the first twelve months and only about Pepsi in the final twelve months. However, weights computed on short time intervals are more prone to bias induced by pricing. For example, a two-weeks promotion of a particular UPC may induce the customer to buy it just because of the temporary convenience; this would give the UPC a high weight in the month. The effect of promotion is instead smoothed when we compute weights using expenditure over the entire sample period.

Table Table 5 reports descriptive statistics on the change in price of the basket.

Table 5: Descriptive statistics on basket price changes

	Mean	*Std.dev.*
Δp	-0.0001	0.043
$\|\Delta p\|$	0.029	0.031
$\%\|\Delta p\| > 1\%$	73.5	
$\%\|\Delta p\| > 5\%$	16.8	
$\%\|\Delta p\| > 10\%$	4.1	