Firm Dynamics and Pricing under Customer Capital Accumulation*

Sonia Gilbukh† NYU Stern

Pau Roldan‡ New York University

January 4, 2018

Job Market Paper

CLICK HERE FOR THE LATEST VERSION

Abstract

This paper analyzes the macroeconomic implications of customer capital accumulation at the firm level. We build an analytically tractable search model of firm dynamics in which firms of different sizes and productivities compete for customers by posting pricing contracts in the product market. Cross-sectional price dispersion emerges in equilibrium because firms of different sizes and productivities use different pricing strategies to strike a balance between attracting new customers and exploiting incumbent ones. Using micro-pricing data from the U.S retail sector, we show that our mechanism can rationalize empirical correlations between store sales and relative prices, and the growth dynamics of stores across sizes. We then calibrate our theory to match long-run moments from the cross-sectional distribution of sales and prices, and use our estimated model to explain sluggish aggregate dynamics and cross-sectional heterogeneity in the markup response to aggregate shocks. Finally, we show that our estimated model offers an explanation for the secular decline in business dynamism and the rise in the average markup experienced in the U.S. since the early 1980s.

JEL codes: D21, D83; E2; L11

Keywords: Customer Capital; Product Market Frictions; Directed Search; Firm Dynamics; Dynamic Contracts; Price Dispersion.

*We are indebted to Ricardo Lagos for his invaluable guidance and support throughout this project. Our most sincere gratitude to Jess Benhabib, Boyan Jovanovic, and Edouard Schaal, for continued discussions and suggestions which helped improve this paper greatly. We have also benefited from comments by Isaac Baley, Jaroslav Borovička, Luís Cabral, Laurent Cavenaile, Diego Daruich, Mark Gertler, Simon Gilchrist, Julian Kozlowski, Virgiliu Midrigan, Simon Mongey, Joshua O'Donnell, Jesse Perla, Michael Peters, Thomas Philippon, Mariona Segú, Tom Schmitz, Venky Venkateswaran, Ewout Verriest, and Basil Williams, and from seminar and conference participants at NYU, FRB of St. Louis, SED 2017, the 2017 European Search and Matching Network conference, the 2017 North American Summer Meetings of the Econometric Society, and the 1st Catalan Economic Society Conference. Pau Roldan gratefully acknowledges the hospitality of the FRB of St. Louis where part of this research was completed. Errors are only ours. First draft: December 2016.

†New York University Stern School of Business, 44 W 4th Street, New York, NY 10012.

‡**Job Market Paper.** New York University, 19 W 4th St., New York, NY 10012. Corresponding author: pau.roldan@nyu.edu.

1 Introduction

Firm heterogeneity is key for explaining the relationship between firm performance and macroeconomic flows. Firms of different sizes and ages experience persistently different growth paths along their life cycle. In particular, newly established businesses typically start out small relative to their more mature competitors, and this gap takes time to close (e.g. Dunne *et al.* (1988), Caves (1998), Cabral and Mata (2003)).

A large theoretical literature, inspired by the seminal work of Jovanovic (1982) and Hopenhayn (1992), has traditionally attributed this evidence to a process of selection on the basis of productivity differences among firms, and has analyzed how these may in turn shape firm and industry dynamics in various meaningful ways. However, this interpretation has been recently challenged by a number of studies showing that, because empirical patterns of firm growth are usually based on revenue data (which cannot easily disentangle output prices from quantities), the productivity-based view of firm heterogeneity may confound selection on technological productivity with selection on profitability. As more disaggregated data have become available over subsequent years, new empirical evidence has shown that large cross-sectional differences in revenue across firms remain after controlling for heterogeneity in productivity, suggesting that differences in firm performance are stemming, to a great extent, from differences in firms' idiosyncratic demand. For instance, Hottman *et al.* (2016) have recently shown that most variation in the firm size distribution is attributable to variation in demand components (e.g. firms' "appeal" such as quality and taste, and product scope), while the contribution of marginal costs and technological differences plays only a minor role. In manufacturing, Foster *et al.* (2008, 2016) have documented that, even though new and well-established firms exhibit very different behavior, the productivity advantage of entrants is only small and it dissipates within the first few years of operation.[1]

Hence, the evidence suggests an important demand-side channel of variation: firm investment in demand accumulation could account for the differential performance of businesses of similar productivities but different sizes. In this paper, we formalize these ideas by developing an equilibrium theory of firm dynamics in frictional product markets with aggregate and idiosyncratic shocks in which there is a meaningful role for a demand accumulation process at the firm level. We interpret this process as the formation of a customer base. The model is that of a frictional product market in which a fixed mass of ex-ante identical buyers must search for sellers of a certain homogenous product, and the latter post price contracts intended to attract new potential customers. Sellers of equal productivity are ex-post heterogeneous in the number of buyers that they sell the product to, since their choice of the contract endogenously determines the inflow of new potential buyers, and thus the rate at which sellers are able to accumulate demand over time. Outside of the market, inactive firms must pay fixed entry (or market penetration) costs to reach their first customer.

Even though the model's dynamics are rich, the environment admits a recursive representation whereby sellers post complete, long-term *recursive contracts* for their current buyers. Recursive contracts specify

[1] These observations are not unique to the U.S. economy. Other studies exploring the demand component of firm dynamics for different countries include Carlsson *et al.* (2014) (Sweden), Pozzi and Schivardi (2016) (Italy), Hong (2017) (France), Kaas and Kimasa (2016) (Germany), and Kugler and Verhoogen (2012) (Colombia). Eaton *et al.* (2014) show that similar demand considerations are also prevalent in the dynamics of exporting firms.

a price level to be paid contemporaneously by all incumbent customers of the firm, and a set of continuation promises that state the life-time utility that buyers can expect to obtain under each and every possible future size of the firm if they remain matched. New buyers of the firm immediately become captive because, due to a reputational concern, the seller commits to delivering the promised price schedule moving forward. Unmatched buyers, on the other hand, trade off the ex-post gains from matching to the ex-ante probability of joining the customer base, as they internalize the endogenous probability with which each supplier's size changes through the posted contract. Though we assume no commitment on the buyer's side, the customer nevertheless remains loyal to the firm because the promised continuation payoffs compensate her for the opportunity cost of searching for other suppliers. Hence, valuable customer relationships emerge endogenously, as forward-looking buyers must internalize the future path of prices and thereby the future evolution of the firm that is implied by those pricing decisions.

In equilibrium, sellers strike an optimal balance between instantaneous revenues (via high prices) and future market shares (via continuation values). The way this trade-off is resolved depends on the size of the seller's customer base. In equilibrium, the sign of the correlation between prices and firm size is not built in, and it depends on the degree of frictions in the market.[2] When costs to market penetration are relatively high, small sellers optimally decide to promise high continuation utilities in order to generate a high probability of quickly expanding their base and raise enough resources to afford the entry cost. Because of product market congestion effects, the customer capital accumulation process takes time. As firms mature and approach their stationary size, they lower their future promises and raise the price as they increasingly prefer to exploit their customer base at the expense of lowering the speed at which their market share accumulates. As a result, their markups tend to increase as they grow in size, and the firm's rate of growth slows down. When entry costs are relatively low, however, the firm might instead be willing to lower its prices as it grows, because it has a weaker preference for rapid growth at the early stages of its life cycle. In either scenario, these endogenous forces of customer acquisition are counteracted by per-customer separation and exit shocks, meaning that firms converge to a stationary size even if there are no decreasing returns in technology. This gives rise to both price and firm size dispersion, as well as well-defined and right-skewed firm and customer distributions.

To solve for the optimal pricing contract, we show that the policy that maximizes the seller's expected value is equivalent to the optimal contract from a *joint surplus* perspective. In the latter formulation, the pricing contract maximizes the sum of valuations across incumbent buyers and seller, and the price level can be thought of as establishing a surplus-splitting rule between the agents involved. The equivalence between the seller's and the joint surplus problems is important because it reduces the dimensionality of the state space considerably, and renders a partial analytical characterization of the equilibrium dynamics. Furthermore, we formally show that a Markov perfect equilibrium is constrained-efficient. This allows us to interpret the model as a theory of *efficient markups*, in which sellers' use of prices leads to a socially optimal allocation of customers across different product markets.

The key behind the analytical tractability is that we can describe the equilibrium allocation indepen-

[2] In this sense, we abstain from taking a stance ex-ante on the active empirical debate regarding the dynamics of firm-level prices, where the literature has found mixed evidence. Foster *et al.* (2008, 2016) and Piveteau (2017) claim that prices are increasing in the firm's tenure in the market, while Berman *et al.* (2017) find that they are slightly decreasing. Fitzgerald *et al.* (2017) find no dynamics of prices, and attribute growth in quantities to advertising and marketing expenditures.

dently of the distribution of agents across states and time. We accomplish this because the search equilibrium is *block-recursive*, a common property of models of directed search (e.g. Shi (2009), Menzio and Shi (2010, 2011)) implying that, in order to evaluate payoffs, buyers and sellers need not keep track of the distribution of agents across states over time. Thus, the firm distribution can be derived independently of the optimal contracting problem, and the dynamics of firms and prices along the stationary solution, as well as out of steady state, can be characterized without the need for approximation methods. Applying these insights to a frictional product market with aggregate shocks is a theoretical contribution of this paper.

After presenting our model, we turn to the data to discipline the behavior of prices across firm size. An important empirical challenge is that, because sellers in the model belong to the same narrowly defined product market, testing its predictions requires the use of highly granular data that contains separate information on revenues and quantities. For this reason, we use highly disaggregated product-level pricing data for the U.S. retail sector for the period 2001-2007 and exploit variation across store size. We document the growth and pricing profiles of sellers of different sizes within closely defined product markets. Specifically, we show that stores with larger volumes of sales tend to set higher prices for their products, relative to the average price across all stores within the same product market segment. Through the lens of our model, this pattern occurs because these small stores are trying to accumulate customers. Moreover, these smaller stores tend to experience higher rates of growth on average, in accordance with the predictions of the model. One potential concern is that these observations may be driven by pure store age effects, as store age and size are likely to be positively correlated. For instance, relative prices could be increasing with size because young sellers offer price discounts in order to learn about unknown idiosyncratic demand components. While this type of interpretation is plausible, we show that the size effect still survives after controlling for the store's age.

We then proceed to quantify our model in order to study the aggregate implications of customer capital accumulation through firm-level pricing strategies. Using simulated method of moments, we calibrate the model to stationary moments of the distribution of relative prices and sales, which we take from our sample of micro-pricing data from the U.S. retail sector. The model provides a good match to measures of price dispersion and the correlation between prices and sales that we document in the data. Using the estimated model, we then analyze the aggregate response of the economy to both aggregate demand and aggregate supply shocks. In this exercise, we find both level and distributional effects. First, we propose a new channel of transmission explaining the incomplete pass-through of shocks to prices: in the wake of negative shocks to their profits, firms are able to front-load their contracts by charging higher prices today and lowering the utility promised to their customers in the future. At the heart of this result is the observation that, when hit by a shock, firms must trade-off immediate losses to future market shares, which they can do by appropriately rebalancing prices and promised utilities in the pricing contract. We also describe an important role for demand shocks on firm pricing. Shocks that raise the marginal propensity to consume by buyers generate a boom in demand and additional entry of firms. Since new firms enter small and charge relatively low prices, increased competition lowers the average price level of the economy, and as a result the average size and growth rate of firms increases. Moreover, since firms' size dynamics are slow-moving, our model can offer a rationale for the sluggish response of macroeconomic variables to aggregate shocks, and a channel of amplification that could potentially be important as a mechanism of transmission of nominal shocks.

Finally, we use the model to study the co-movement of two secular trends in the U.S. at frequencies lower than the business cycle: (i) the steady decline in business dynamism, and (ii) the secular increase in market power. In particular, it is a well-documented fact that the entry rate of firms across different industries has dramatically declined since the early 1980s, a process that has been coupled with an increase in the average size of firms (e.g. Pugsley and Şahin (2015)). Over the same time period, most industries have become more concentrated, and the average markup in the U.S. has dramatically increased by a factor of more than three, according to DeLoecker and Eeckhout (2017). Further, this increase in market power has been more pronounced in the upper tail of the distribution of markups. In the last part of the paper, we investigate the relationship between these two phenomena using our estimated model. In the estimated model, a decline in the entry rate implies an increase in the average size of firms and a subsequent increase of the average markup in the economy, since larger firms set relatively higher prices. Moreover, since the decline in firm entry implies a shift of the firm size distribution toward higher-markup firms, the increase in concentration can give rise to an increase in dispersion at the top of the markup distribution, in line with the empirical evidence presented by DeLoecker and Eeckhout (2017).

Outline The remainder of the paper proceeds as follows. Section 2 summarizes related literature and details our main contributions. In Section 3 we present our model of customer acquisition, pricing, and firm dynamics, including the derivation of the firm size distribution, and the equilibrium efficiency result. Section 4 discusses the main mechanism, shows that search frictions can deliver different profiles for prices, and explains the role of each central assumption. Section 5 describes our application to the U.S. retail sector, and proceeds to the calibration of the model and its quantitative results, including the response of the economy to aggregate shocks. Section 6 discusses the long-run rise in average markups and the secular decline in business dynamism in the U.S. through the lens of our calibrated model. Section 7 presents extensions to the baseline model, and Section 8 concludes. The Appendix includes supplementary tables and figures, all the proofs, and some additional theoretical results.

2 Related Literature

There is a large amount of survey evidence that suggests that the customer base of a firm and its pricing decisions are tightly linked. Blinder *et al.* (1998) show that the vast majority of firms report having implicit contracts with their customers, and that these contracts are a major source of price stickiness. For a variety of different countries, other studies such as Hall *et al.* (1997), Cason and Friedman (2003), Renner and Tyran (2004), and Apel *et al.* (2005), present similar survey evidence showing that customer loyalty is a sensitive concern for price-setting firms.[3] More recently, using the same pricing dataset that we use in Section 5.1, Paciello *et al.* (2016) are able to identify customer-retailer transactions and demonstrate that customer attrition rates are on average low over long spells (i.e. a retailer's customer base is typically

[3] See also Fabiani *et al.* (2004) for exhaustive cross-country evidence in the Euro Area. There is also a large literature in Marketing showing that there exists a large degree of persistence in consumer inertia and brand preferences (for a review of this literature, see Bronnenberg and Dubé (2017)).

sticky).

Our theory is primarily related to a long tradition of building customer capital into macroeconomic models of firm pricing. Early attempts by Phelps and Winter (1970), Bils (1989), and Rotemberg and Woodford (1991, 1999), analyzed pricing behavior under customer retention concerns. In these papers, firms face an exogenously-given law of motion for the customer base. A number of papers have further developed a variety of reasons why customers may be locked into a repeated-purchase relationship in the first place. For instance, Klemperer (1987, 1995) and Kleshchelski and Vincent (2009) propose that customers face switching costs, which can be broadly understood as the transaction costs associated with switching to a competitor, or the costs in terms of utility when the consumer has developed a loyalty toward a certain brand. In a similar vein, Ravn *et al.* (2006) and Nakamura and Steinsson (2011) consider that customers form good-specific habits for consumption, and for this reason have a preference for repeating purchases with the same sellers.[4] While the literature has traditionally resorted to reduced-form formulations for customer capital formation, we contribute by offering a micro-foundation whereby customers become captive. In our model, it is the seller's commitment to the pricing contract (because of, for example, reputational concerns) which naturally gives rise to these long-lasting relationships.

Regardless of the reason, the common insight in the literature is that when customers are locked into their suppliers, demand becomes forward-looking. In this situation, prices not only fulfill the usual *distributive role* of splitting gains from trade between buyers and sellers, as in a standard Walrasian economy, but may also play an *allocative role* and determine the duration of customer-seller relationships or the likelihood that new ones form. Consequently, the optimal price of the static profit maximization problem may differ from the dynamic one because firms must solve a dynamic trade-off between exploiting their current customers (by setting high prices) and attracting new customers in the future (by setting low prices). In short, low prices today serve as a tool to guarantee larger market shares in the future.[5]

While our model shares these features with the literature, an important focus of our work are the implications that customer-seller relationships have on firm dynamics, including firm growth, entry, and exit, in an equilibrium model with aggregate shocks. In this dimension, our paper is also related to the literature that has introduced a role for various types of firm intangibles into models of firm and industry dynamics.[6] The effects of intangibles on different aspects of the aggregate economy are well-understood, including labor wedges (Gourio and Rudanko (2014a)), aggregate productivity (McGrattan and Prescott (2014) and McGrattan (2015)), and household behavior (Hall (2008)). A number of papers have further analyzed how expenditures on intangibles may shape the evolution of firms and industries.

[4] As an application of this approach, Gilchrist *et al.* (2016) show that the inter-temporal pricing behavior of firms in customer markets interacts with their degree of financial constraints, and can rationalize the mild disinflation episode experienced in the United States during the Great Recession.

[5] If firms are not committed to the price path, however, a well-known time-inconsistency problem arises: firms promise low prices to attract customers and, once these customers become captive, sellers renege on their earlier promises and charge a higher price. For instance, Nakamura and Steinsson (2011) show that, in this case, repeated interaction can lead to the development of implicit contracts which, through a set of properly defined trigger strategies, can prevent prices from increasing beyond a certain upper bound. We show that, when there is commitment on the seller's side (e.g. the firm faces reputational concerns), a similar type of contractual environment can emerge.

[6] Firm intangibles are a substantial share of firms' expenditures, and in the U.S. as much as 7.7% of GDP is devoted to marketing, with advertising expenditures alone averaging about 2.2% since the early 1980s (see e.g. Arkolakis (2010)). More recently, Bhandari and McGrattan (2017) have estimated the value of aggregate private-business "sweat equity" (e.g. firm investment into building customer bases, client lists, and related intangibles) to be 0.65 times GDP.

Atkeson and Kehoe (2005) show that organizational capital (i.e. investment in new technologies, new markets, and new and higher-quality products) can drive the life-cycle of plants, and Hsieh and Klenow (2014) argue that these processes may account for differences in plant-specific TFP between different countries. Another class of papers, including Alessandria (2009), Drozd and Nosal (2012), Eaton *et al.* (2014), Arkolakis (2016), and Piveteau (2017) study how consumer search and costs to market penetration can rationalize certain patterns of trade and firm growth among exporting firms, while Dinlersoz and Yorukoglu (2012) study the effects of information dissemination to customers for industry dynamics.

The paper that we most relate to is Gourio and Rudanko (2014b), who analyze the timing of firm responses to investment shocks by augmenting a neoclassical firm investment model with a search model of the product market in which firm use price discrimination by offering a discount on new customers. In more recent work, Rudanko (2017) uses a related setting to study the role of both discriminatory and non-discriminatory pricing for firm growth, with a focus on time-inconsistent seller behavior under different commitment protocols in monopolist markets. Like both of these papers, we interpret customer acquisition as a search-and-matching process in a frictional product market. Different from Gourio and Rudanko (2014b), where firm growth is limited by convex adjustment costs to customer acquisition, we limit firm expansion through the interaction between the search frictions and our structure with dynamic long-term contracts with commitment. Indeed, we find that there is a limit to firm growth even when the firm's technology features constant returns to scale. As discussed in Section 4, this allows for a flexible dependence between firm size and price, which can be either positive or negative. Relative to Rudanko (2017), we focus on the case of commitment, which gives rise to efficient firm dynamics (Proposition 3). Moreover, unlike either of these studies, we analyze firm pricing and customer dynamics in the presence of aggregate shocks. An important emphasis of our work is on the cross-sectional heterogeneous response and incomplete pass-through of prices and markups in response to these shocks (Section 5.3).[7] Luttmer (2006) and Fishman and Rob (2003) also study the implications of customer acquisition for the firm size distribution, but those papers do not allow for a meaningful role for prices. In contrast, like us, Paciello *et al.* (2016) study the implications of customer markets for the cross-sectional price distribution and the pass-through of shocks, but while they study the pricing problem of firms with *retention* concerns, we offer a complementary view whereby firms use prices to *attract* customers.[8]

We contribute to the aforementioned literature by providing a link between market shares and firm dynamics in customer markets. In particular, a prevailing feature in the data is that the growth rate of firm size is size-, and age-, dependent (e.g. Sutton (1997), Caves (1998), and Rossi-Hansberg and Wright (2007)). Further, the size distribution is right-skewed in the data (e.g. Luttmer (2007)). These

[7] Another difference with Gourio and Rudanko (2014b) is that we assume no price discrimination between customers. However, this assumption is not key to generate firm dynamics or price and firm distributions. For a full discussion on this issue, see Section 7.2. To cite more examples in the literature of intangible and industry dynamics: Kaas and Kimasa (2016) embed the Gourio and Rudanko (2014b) framework into a frictional labor market to study the joint dynamics of prices and wages; Perla (2016) studies the implications of product sorting by uninformed consumers on the industry life cycle and the degree of market competition; Bai *et al.* (2012) incorporate a frictional goods market into a representative-agent neoclassical economy to study the role of demand shocks; and Petrosky-Nadeau and Wasmer (2015) combine the goods market friction with frictions in the credit market to analyze distortions in the labor market.

[8] Methodologically, another innovation of our framework relative to Paciello *et al.* (2016) is that, to obtain analytical tractability, we do not need to assume that the growth rate of firms is independent of the size of the customer base. Indeed, the fact that firm growth is inherently a function of the firm's current size is a key aspect of our theory.

stylized patterns of growth, which we will document for our sample of retail firms in Section 5.1, can be rationalized by our model. Under a certain parametrization of the model, small firms promise relatively low prices, thereby attracting more customers and generating a higher likelihood of growing. In this case, larger firms instead prefer to exploit their customer base, typically by charging higher prices, thereby growing slower or even shrinking on average. This generates a right-skewed firm distribution: since larger firms are visited less frequently and lose proportionally more customers than smaller firms, the probability that a firm grows to be large is relatively low, and this generates a fat right tail. As we will show, the size distribution in the micro data similarly exhibits a fat right tail.

The link between firm dynamics and prices is also supported by a number of studies relating empirically demand-side fundamentals to the determination of prices at the firm level. Peters (2016) and Kugler and Verhoogen (2012) find a positive correlation between output prices and size at the plant level for Indonesian and Colombian firms, respectively, while Carlsson *et al.* (2014) find, using Swedish micro data, that a substantial component of output price variation remains unexplained after accounting for productivity differences. DeLoecker and Eeckhout (2017) have found that smaller firms charge lower markups relative to competitors within their industry, and DeLoecker (2011), DeLoecker and Warzynski (2012), and DeLoecker *et al.* (2016) perform similar analyses in the context of exporting firms for different countries, concluding that markups contribute to differences in revenue productivity. While we do not take a stand ex-ante on the relation between prices (or markups) and size, we rely on these observations to justify our demand-driven theory of firm dynamics.

Because we use search frictions to obtain a non-degenerate cross-sectional price distribution, this paper also contributes to the search literature on equilibrium price dispersion. Empirically, newly available micro-level evidence has shown that there is substantial price dispersion for identical goods sold at a given time and market (e.g. Kaplan and Menzio (2015)), an observation that we also document in Section 5.1. Theoretically, search models of price dispersion have proliferated since the work by Butters (1977), Varian (1980), and Burdett and Judd (1983). More recently, Menzio and Trachter (2015) and Kaplan *et al.* (2016) have shown that price dispersion can emerge from buyer heterogeneity in situations in which sellers can price-discriminate. In our model, in contrast, buyers are identical and there is no price discrimination. Instead, it is ex-post differences between firms which give rise to different price levels. While a similar argument is made in Burdett and Coles (1997) and Menzio (2007), these papers do not discuss the implications of customer capital for the evolution of the firm size distribution, nor do they analyze the implications of customer accumulation at the aggregate level.

Finally, our paper is methodologically related to search-and-matching models with large firms, where most advances have been made in the context of labor markets. We embed *directed* search into a model of firm dynamics in the spirit of Elsby and Michaels (2013), Kaas and Kircher (2015), and Schaal (2017).[9] Particularly, we combine two technical insights from this literature. First, we exploit the property of block recursivity, which allows for a tractable characterization of the firm size distribution and its dynamics. This property implies that agents do not have to carry distributions as state variables in their optimization problems even though the model incorporates aggregate dynamics, thereby allowing us to study out-of-steady-state transitions in response to aggregate shocks. Secondly, we make use of dynamic long-term

[9] For a recent survey of directed search theory, see Wright *et al.* (2017).

contracts (e.g. Moscarini and Postel-Vinay (2013), Schaal (2017)), which greatly reduce the dimensionality of the state space as they allow us to condense the full forward-looking pricing problem into an amenable recursive form.

3 Model

This section develops a directed search model of customer and firm dynamics with aggregate and idiosyncratic shocks in which sellers must post pricing contracts in order to attract consumers. The key assumption in the model is that contracts are long-term in nature, as sellers perfectly commit to the terms of trade. As this commitment is internalized by agents, a dynamic trade-off emerges for both sellers and buyers between the added value of new customers and the loss of profits on incumbent ones. As we shall see, this mechanism is at the core of equilibrium firm and pricing dynamics.

3.1 Environment

Time is continuous and goes on forever, with a time instant indexed by $t \in \mathbb{R}_+$. The aggregate state of the economy is indexed by a time-varying random variable φ taking values in a discrete and finite support $\Phi := \{\underline{\varphi} < \cdots < \overline{\varphi}\}$, with cardinality $|\Phi| = k_\varphi \geq 2$. The aggregate state is the source of exogenous aggregate demand and/or supply fluctuations in the economy. We assume φ follows a homogenous continuous-time Markov chain with generator matrix $\mathbf{\Lambda}_\varphi := \left[\lambda_\varphi(\varphi'|\varphi)\right]$, where $\lambda_\varphi(\varphi'|\varphi)$ denotes the intensity rate of a φ-to-φ' transition.[10]

Demographics

The economy is populated by a mass-one continuum of risk-neutral, infinitely-lived, ex-ante identical *buyers*, and a continuum of risk-neutral *firms* (sometimes referred to as *sellers*). While the total mass of buyers is exogenous and normalized to unity, the composition of buyers across aggregate states and between types (described below) is endogenous. The total measure of firms, in contrast, is not fixed exogenously but determined in equilibrium. Buyers and sellers both discount future payoffs with a common and exogenous rate, $r > 0$. All payoffs and payoff-relevant states are public information among all agents.[11]

There is a single homogenous, indivisible, and perishable good in the economy. Buyers and sellers must participate in a search-and-matching market in order to engage in trade because the product market is frictional: searchers cannot coordinate into finding a match with certainty at any given instant. The product market frictions are meant to capture congestion effects in product markets with customer

[10] For all $\varphi \in \Phi$, the following properties hold: $\lambda_\varphi(\varphi|\varphi) \leq 0$, $\lambda_\varphi(\varphi'|\varphi) \geq 0$ for any $\varphi' \neq \varphi$, and $\sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi) = 0$. These properties are definitional of continuous-time Markov processes (e.g. Norris (1997), Chapters 2 and 3). The rates additionally satisfy the condition $\sum_{\varphi' \neq \varphi} \lambda_\varphi(\varphi'|\varphi) < +\infty$, $\forall \varphi$ (i.e. when any given state φ is visited, the economy always spends a non-zero measure of time in it).

[11] Faig and Jerez (2005) and Shi (2016) introduce search models in which private information about buyers' payoffs generates customer relationships. Menzio (2007) analyzes the dynamics of prices when there is, instead, private information about the cost structure of sellers.

anonymity in reality. One interpretation is that there exist informational asymmetries regarding product characteristics, or some aspects of supply that are unknown to the potential customer (e.g. the exact location of seller-price pairs). Another interpretation is that sellers may face inventory and/or capacity constraints, and are unable to simultaneously serve a large amount of buyers (as in Burdett *et al.* (2001)). In any case, these considerations lead businesses to invest in reputation-building in order to overcome those frictions.[12]

Buyers value the consumption of the good by the same fixed utility flow, $v > 0$. At any instant in time, a buyer is said to be *active* if she is matched with a firm and is consuming the good, and *inactive* if she is unmatched and searching for a seller at a cost, c. These parameters possibly depend on the aggregate state of nature, φ. Since v and c relate directly to buyers' preferences, this state-dependence incorporates the possibility of aggregate demand shocks into the model.[13] We also assume no buyer is ever allowed to borrow against its future income.

Sellers belong to one of two groups: incumbent (or *active*) sellers, and potential entrant (or *inactive*) sellers. At any given time t, a typical incumbent seller has a customer base of $n_t \in \mathbb{N} := \{1, 2, 3, \dots\}$ customers, which we subsequently call the *size* of the seller. Each seller is also characterized by the realization of an idiosyncratic productivity level z, taking values on a discrete and finite support $\mathcal{Z} := \{\underline{z}, \dots, \overline{z}\}$ of cardinality $|\mathcal{Z}| = k_z \geq 2$. Like the aggregate state, the idiosyncratic state follows a continuous-time Markov chain with generator matrix $\mathbf{\Lambda}_z := \left[\lambda_z(z'|z)\right]$, where $\lambda_z(z'|z)$ denotes the transition rate from z to z'.[14] The realization of the idiosyncratic state is observable and public information.

An incumbent seller's output is constrained by the size of its customer base. Since the good is indivisible, and because there is no benefit in leaving customers unserved, the number of units sold by the seller equals the number of customers in the base, with each customer consuming one unit. The seller also faces operating variable flow costs of $\mathcal{C}(n; z, \varphi)$, which depend on the idiosyncratic state (n, z), as well as possibly the aggregate state φ. Further, we make the following assumptions:

Assumption 1 *The following properties hold for all* $(z, \varphi) \in \mathcal{Z} \times \Phi$*:*

(i) $\mathcal{C}$ *is a continuous, increasing, and time-invariant function of* n*, with* $\mathcal{C}(n; z, \varphi) \geq 0$ *and* $\mathcal{C}(0; z, \varphi) = 0$.

(ii) $\mathcal{C}(n; z, \varphi)$ *is weakly convex in all* $n \in \mathbb{N}$.

Assumption 1 imposes mild regularity conditions on the firms' technology. In particular, it states that firm profits are continuous in firm size. The curvature of $\mathcal{C}$ with respect to n determines the degree of returns to scale in the firm's technology. For now, we need not make an explicit assumption in this respect besides a *weak* form of convexity. Indeed, as we shall see, equilibrium firm-level prices are size- (and productivity-) dependent even when marginal costs are constant in n. In the estimation section, we will re-introduce the notion of convexity in $\mathcal{C}$ for quantitative purposes only.

Besides serving their customers, incumbent sellers post prices in the product market. Posting a price bears no explicit cost for an incumbent. Incumbent sellers exit the market (and enter the pool of potential

[12] Informational frictions in the product market are the preferred interpretation of Faig and Jerez (2005), Gourio and Rudanko (2014b), and Foster *et al.* (2016), among others. Perla (2016) provides a micro-foundation for this view based on the evolution of buyers' consideration sets.

[13] The source of variation in shopping disutility can be thought of as reflecting the cyclical nature of household shopping behavior, which has been documented by Petrosky-Nadeau *et al.* (2016) for the United States.

[14] Once again, for all $z \in \mathcal{Z}$ we have: $\lambda_z(z|z) \leq 0$; $\lambda_z(z'|z) \geq 0$ for any $z' \neq z$; $\sum_{z' \in \mathcal{Z}} \lambda_\varphi(z'|z) = 0$; and $\sum_{z' \neq z} \lambda_\varphi(z'|z) < +\infty$.

entrants) in either one of two ways: because they go bankrupt, at a constant exogenous rate $\delta_f > 0$, or if they separate from their last remaining customer (because the buyer abandons the firm), at an exogenous rate $\delta_c > 0$.[15] These events are assumed to be mutually independent, and orthogonal to the idiosyncratic and aggregate shocks.

Like incumbent firms, inactive firms are posting prices in order to attract customers and start operating in the product market. Unlike them, however, they must incur an entry cost $\kappa > 0$ for doing so, which possibly depends on the aggregate state of nature, φ. A firm must pay this cost every time it has lost all its customers and intends to re-enter the market, so κ can be thought of as a proxy for the fixed costs of an advertising campaign that has the objective to reach the first customer of the firm. More broadly, κ can be understood as a cost to market penetration, in the sense of Arkolakis (2010). Sellers who successfully attract their first customer (and thus start operating with $n = 1$) draw an initial productivity level $z_0 \in \mathcal{Z}$ from some distribution π_z, where $\pi_z(z) \geq 0$, $\forall z \in \mathcal{Z}$, and $\sum_{z\in\mathcal{Z}} \pi_z(z) = 1$. We assume that there is free entry of firms into the product market.

Pricing Contracts

All agents are able to direct their search in the following sense. Sellers announce price contracts in order to attract buyers. Buyers, on the other hand, can perfectly observe the posted contract and are able to discern the identity (i.e. the size n and productivity z) of the firm who is posting it.

When firms post prices to attract customers, a potential contractual relationship is thus formed. For a customer-seller match formed at time t, a *price contract* is defined as a sequence $(p_{t+j} : j \geq 0)$, which specifies the price level at each tenure length $j \geq 0$ of the match, conditional on no separation. Contracts are complete and fully state-contingent. Thus, every element p_{t+j} of the contract is contingent on the history of aggregate and the firm's idiosyncratic states up to date $t + j$. Since all the relevant states are public, then $p_{t+j} = p(n^{t+j}; z^{t+j}, \varphi^{t+j})$, $\forall j, t$.

The contractual environment is as follows. On the demand side, we assume no commitment to the contract, in that matched buyers can costlessly transition to inactivity if they so desire (though in equilibrium this will not occur because of the subsequent additional cost c of re-sampling firms).[16] On the sellers' side, we make two key assumptions. First, unlike the buyer, the seller fully commits to the contract that is posted. This means that contracts with captive customers cannot be revised by the firm for the duration of the match, and contracts have to comply with the firm's prior promises.[17] Second, we assume anonymity among buyers, in that the firm is unable to price-discriminate between new and old customers, and thus cannot index the contract to the identity of each buyer.[18] This implies that, when setting a price path optimally, the firm must internalize that any additional revenue from expanding the number of customers comes at the expense of potentially lowering the average revenue from the incumbent base.

[15] In Section 7.1 we show how to endogenize the customer separation rate δ_c.

[16] More specifically, there are endogenous switching costs for buyers: customer loyalty emerges endogenously because of the opportunity cost (i.e. forgone contracted-upon expected value) of leaving the seller.

[17] A possible interpretation of this assumption is that firms have a reputational concern, so that reneging on previous promises entails unaffordable costs for them. We shall discuss the role of this assumption in Section 4.

[18] In Section 7.2 we discuss the implications of relaxing the no discrimination assumption.

Product Markets

As is customary in the directed search literature, a sufficient statistic for each long-term pricing contract is the promised life-time value that the contract delivers in expectation to the buyer at the point in time when the match is formed and the contract is initiated. We denote this value by x, let $\mathcal{X} = [\underline{x}, \overline{x}] \subseteq \mathbb{R}_+$ be the set of feasible values, and assume that all sellers advertising the same value x compete in all such contracts. Moreover, buyers cannot coordinate their decisions among themselves. Up to the observable idiosyncratic state (n, z), sellers offering the same value x are virtually indistinguishable to the buyer. Thus, x effectively indexes a *product market segment* (or *sub-market*).

Each seller can simultaneously post, and each buyer can simultaneously search, in at most one sub-market. Within each $x \in \mathcal{X}$, and given a realization $\varphi \in \Phi$ of the aggregate state of nature, a certain mass $B(x; \varphi) \in [0, 1]$ of buyers apply to the contract, which is posted by a mass $S(x; \varphi) \geq 0$ of sellers. Because buyers cannot screen sellers within a market x, *within*-market search is random. A market is then said to be *active* (or *open*) if:

$$\theta(x; \varphi) := \frac{B(x; \varphi)}{S(x; \varphi)} > 0$$

where $\theta(x; \varphi)$ is the buyer-to-seller ratio in market segment x, also referred to as the *market's tightness*. Importantly, agents take the mapping $\theta : \mathcal{X} \times \Phi \to [0, +\infty)$ as given when directing their search toward specific offers. This is relevant because expected payoffs within a market can be fully evaluated using the tightness measure: in a typical $x \in \mathcal{X}$, a single applicant obtains offer x at the endogenous Poisson arrival rate $\mu\big(\theta(x; \varphi)\big) \geq 0$, while a seller successfully finds an applicant for offer x at the Poisson arrival rate $\eta\big(\theta(x; \varphi)\big) \geq 0$, where $\eta(\theta) = \theta\mu(\theta)$.

Further, we impose the following regularity conditions:

Assumption 2 *The meeting rates satisfy:*

(i) $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ *and* $\mu : \mathbb{R}_+ \to \mathbb{R}_+$ *are twice continuously differentiable and time-invariant functions;*

(ii) η *is increasing and concave;* μ *is decreasing and convex;*

(iii) For some decreasing $h : \mathbb{R}_+ \to \mathbb{R}_+$*, define the composition* $f = \eta \circ \mu^{-1} \circ h$*. Then, the function* $f(x)(\widehat{x} - x)$ *is concave for all* $x \in [0, \widehat{x}]$ *and* $\widehat{x} > 0$*;*

(iv) $\eta(0) = \lim_{\theta \nearrow +\infty} \mu(\theta) = 0$*, and* $\lim_{\theta \nearrow +\infty} \eta(\theta) = \lim_{\theta \searrow 0} \mu(\theta) = +\infty$*.*

The first two restrictions guarantee that the problems of the buyer and the seller are well-defined; assumption *(iii)* is a restriction on the composition $\eta \circ \mu^{-1}$ guaranteeing that the price-posting problem of the seller has a unique interior solution; finally, part *(iv)* imposes a transversality condition on the meeting rates.

A common micro-foundation of these assumptions is to suppose that each market $x \in \mathcal{X}$ is endowed with a constant-returns-to-scale matching function $M(B, S)$ that is equipped with the appropriate Inada conditions. Pairwise matching then requires that $\eta(\theta) = M(\theta, 1)$ and $\mu(\theta) = \eta(\theta)/\theta$. Intuitively, the seller's meeting rate is found as the measure of meetings per seller, and because of congestion effects in the product market, longer queues of applicants for a contract yield lower (respectively, higher) rates of matching for the buyer (respectively, the seller).

Recursive Formulation

We seek to solve for the *symmetric Markov perfect equilibrium* of this economy. We narrow attention to this class of equilibria in the following sense. *Markov-perfection* means that the equilibrium policies depend solely on the firm's vector of payoff-relevant states $(n, x; \mathbf{s})$, where henceforth we use $\mathbf{s} = (z, \varphi)$ to denote the vector of exogenous (idiosyncratic and aggregate, respectively) states. We look for a *symmetric* equilibrium in the sense that all firms within the same product market x choose to post the same contract. This is a consequence of the assumption that there is competition within each sub-market, and the fact that the firm's states are fully observable. Finally, we restrict our attention to a *stationary* environment, in which policies are time-varying only insofar as they are state-dependent. Thus, subsequently we drop time subscripts unless otherwise needed.[19]

Because a dynamic pricing contract is a time path and thus a large and potentially complex object, we exploit the property of stationarity to propose the following recursive formulation. We define a *recursive dynamic contract* for a firm in state $(n, x; \mathbf{s})$ as the object:[20]

$$\boldsymbol{\omega} := \left\{p, \mathbf{x}'(n'; \mathbf{s}')\right\}$$

The elements of a recursive contract $\boldsymbol{\omega}$ are the following. First, the contract specifies the price p that is to be charged to each one of the n incumbent customers of the firm. Second, the contract specifies the vector $\mathbf{x}'(n'; \mathbf{s}') \subseteq \mathcal{X}$ of continuation payoffs that are promised by the firm to each buyer on the next stage, i.e. under every possible size $n' \in \{n-1, n, n+1\}$ and exogenous state $\mathbf{s}' \in \mathcal{Z} \times \Phi$. Hence, by stationarity, conditional on a fixed exogenous state $\mathbf{s}$ (respectively, a size n), contracts are rewritten every time the seller changes sizes (respectively, productivity), and they remain in place for as long as the firm's state does not change (i.e. $x'(n'; \mathbf{s}') = x$ when $n' = n$ and $\mathbf{s}' = \mathbf{s}$).[21] Notice, finally, that the contract is not indexed to the aggregate distribution of agents across states. This is an implication of the property of *block recursivity*, which we take as given and we discuss in some detail in Section 3.5.

3.2 Buyer's Problem

Inactive Buyers

Let us now describe the value functions of each type of agent in the economy. If a buyer is presently inactive, let its expected value be $U^B(\varphi)$ in state $\varphi \in \Phi$. The buyer enters the sub-market that offers the

[19] While the Markov structure provides a lot of tractability to the setting and will allow for some sharp results, we should stress that this is by no means the only equilibrium of the economy. Similarly, though we do not rule out the possibility of non-stationary equilibria, we focus our analysis on the stationary solution for simplicity. If agents were risk-averse, for instance, non-stationary solutions would emerge naturally. One such case is the model of the labor market with job-to-job transitions presented in Shi (2009), where workers' utilities have some curvature and firms optimally choose a front-loaded wage profile in time to entice workers to remain matched to them.

[20] We subsequently use bold characters for vectors, light characters for scalars, and capitalized characters for matrices.

[21] As we shall see, there is an equilibrium one-to-one mapping between continuation utilities and price levels, so p also remains fixed for as long as the firm does not experience a change in its state vector.

highest valuation, and therefore:

$$U^B(\varphi) = \max_{\widehat{x}(\varphi)\in\mathcal{X}} u^B\big(\widehat{x}(\varphi);\varphi\big) \tag{1}$$

where $u^B\big(x;\varphi\big)$ is the value of searching in market x, satisfying the Hamilton-Jacobi-Bellman (HJB) equation:

$$ru^B\big(x;\varphi\big) = -c(\varphi) + \mu\big(\theta(x;\varphi)\big)\Big(x - u^B\big(x;\varphi\big)\Big) + \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)\Big(u^B\big(x;\varphi'\big) - u^B\big(x;\varphi\big)\Big) \tag{2}$$

for any $x \in \mathcal{X}$.[22] Equation (1) states that the inactive buyer searches in the product market that promises the highest expected value, $\widehat{x}(\varphi)$. The value of entering market x incorporates the search cost $c(\varphi) > 0$, and the option value of matching with any one firm within said market. The meeting rate depends on how "crowded" the marketplace is, as measured by the prevailing tightness schedule $\theta(x;\varphi)$. This tightness is taken by agents as a given function mapping $\mathcal{X}$ to $\mathbb{R}_+$. In case of a successful match, and because sellers can only meet at most one customer every instant, the buyer will instantly join the seller's customer base. The last additive term in equation (2) incorporates the expected change in value due to a change in the aggregate state, from φ to some φ', occurring at rate $\lambda_\varphi(\varphi'|\varphi)$.[23]

Since inactive buyers choose to apply to the highest-valuation offers, active markets must be solutions to the buyer's search problem. Therefore:

$$\forall(x,\varphi)\in\mathcal{X}\times\Phi: \;\; u^B\big(x;\varphi\big) \leq U^B(\varphi), \;\; \text{with equality if, and only if, } \theta(x;\varphi) > 0$$

This says that a market either maximizes ex-ante payoffs for the inactive buyer, or it remains unvisited. In equilibrium, a non-zero measure of markets is open, and we let $\mathcal{X}^*(\varphi) := \big\{x\in\mathcal{X} : \theta(x;\varphi) > 0\big\} \subseteq \mathcal{X}$ be the *equilibrium* set of markets in state $\varphi\in\Phi$. Hence, for any given aggregate state $\varphi\in\Phi$, we have:

$$\mu\big(\theta(x;\varphi)\big)\Big(x - U^B(\varphi)\Big) = \Gamma^B(\varphi) \tag{3}$$

for all $x\in\mathcal{X}^*(\varphi)$, where we have defined:

$$\Gamma^B(\varphi) := c(\varphi) + rU^B(\varphi) - \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)\Big(U^B(\varphi') - U^B(\varphi)\Big) \tag{4}$$

as the opportunity cost of matching for the buyer in equilibrium market x. Intuitively, equation (3) describes how inactive buyers sort in equilibrium, stating that all active market segments equalize the expected option value of searching to the opportunity cost of matching. Thus, all equilibrium markets make inactive buyers ex-ante indifferent.

For a given value of inactivity $U^B(\varphi)$, this ex-ante revenue-equalization condition uniquely pins down

[22] For a derivation of all the HJB equations in the main text, see Appendix D.1.

[23] Note that we assume that the inactive buyer returns to market x if unsuccessful in his search. As we argue shortly, this entails no loss in generality. We should also point out that notation has been economized in two ways here. First, since the value of inactivity is itself an equilibrium object, we write $\theta(x;\varphi)$ when in fact we mean $\theta(x;\varphi,U^B(\varphi))$. Second, since market tightness is taken as given by the agent, $u^B\big(x;\varphi\big)$ is actually short-hand notation for $u^B\big(x;\varphi,\theta\big)$, where θ here is a function mapping from $\mathcal{X}\times\Phi$ to $\mathbb{R}_+$. Similar concise notation will be used throughout the paper.

the market tightness of any market in equilibrium. Importantly, equation (3) defines the $\theta(\cdot;\varphi)$ schedule over the entire support $\mathcal{X}$, and thus it is used by agents to form beliefs about market tightness on *both* equilibrium and off-equilibrium markets. This restriction, which is implicit in the bulk of the competitive search literature, imposes a form of trembling-hand stability in beliefs, and ensures the existence of a stable rational-expectation equilibrium.[24] In particular, no firm (or coalition of firms) can possibly make a profitable off-equilibrium deviation, for in this case beliefs dictate that buyers would remain indifferent and thus the equilibrium allocation would be unaffected. Although we recognize the possibility that other type of equilibria may exist under alternative specifications of agents' beliefs, in what follows we will focus on this type of perfect-foresight equilibrium for the sake of tractability.

With these remarks in place, we note that an implication of equation (3) is that, for each $\varphi \in \Phi$, $\theta(x;\varphi)$ is an increasing function of $x \in \mathcal{X}$. This result is intuitive: more ex-post profitable offers attract a larger mass of buyers per seller, while sellers offering less favorable contracts to the buyer can expect to find a match sooner. In equilibrium, firms design contracts for which a low buyer meeting rate μ can be compensated with a high enough promised expected continuation value x. Further, the buyer-to-seller ratio is increasing in $U^B(\varphi)$: when the inactive buyers' outside option is higher, contracts must offer more attractive deals in order to compensate for the opportunity cost of matching.

Active Buyers

Consider now a customer who is currently consuming the homogeneous good from a firm of size $n \in \mathbb{N}$ and idiosyncratic productivity $z \in \mathcal{Z}$, under contract $\boldsymbol{\omega} = \{p, \mathbf{x}'(n';\mathbf{s}')\}$. The contract delivers the promised value x to the customer, and it specifies the current price p and the new continuation promises $\mathbf{x}'(n';\mathbf{s}')$, to be delivered by the seller after a n-to-n' and/or s-to-s$'$ transition.

The value for the buyer is given by the following HJB equation:

$$\begin{aligned} rV^B(n,\boldsymbol{\omega};\mathbf{s}) = v(\varphi) - p + (\delta_f+\delta_c)\Big(U^B(\varphi) - V^B(n,\boldsymbol{\omega};\mathbf{s})\Big) + (n-1)\delta_c\Big(x'(n-1;\mathbf{s}) - V^B(n,\boldsymbol{\omega};\mathbf{s})\Big) \\ + \eta\Big(\theta\big(x'(n+1;\mathbf{s});\varphi\big)\Big)\Big(x'(n+1;\mathbf{s}) - V^B(n,\boldsymbol{\omega};\mathbf{s})\Big) \\ + \sum_{z'\in\mathcal{Z}} \lambda_z(z'|z)\Big(x'(n;z',\varphi) - V^B(n,\boldsymbol{\omega};\mathbf{s})\Big) + \sum_{\varphi'\in\Phi} \lambda_\varphi(\varphi'|\varphi)\Big(x'(n;z,\varphi') - V^B(n,\boldsymbol{\omega};\mathbf{s})\Big) \end{aligned} \tag{5}$$

where $U^B(\varphi)$ solves equation (1). The right side of equation (5) has different additive terms. In the first line: the first term, $v - p$, shows flow surplus for the agreed-upon price $p \in \boldsymbol{\omega}$; the second term states the possibility of separation, due to either the destruction of the firm or the destruction of the match, in which case the customer ceases to consume and becomes inactive; the third term includes the event in which any customer of the firm (except for the buyer in question) separates, in which case the firm becomes size $n-1$ and changes the promised value to $x'(n-1;\mathbf{s}) \in \boldsymbol{\omega}$ for all those customers that remain captive. The second line is the expected change in value due to the firm successfully attracting a customer with its currently posted offer, in which event the seller becomes size $n+1$ and implements

[24] For an in-depth discussion of the game-theoretical foundations of this assumption, see Galenianos and Kircher (2009, 2012).

value $x'(n+1;\mathbf{s}) \in \boldsymbol{\omega}$. Because the seller cannot differentiate between the n incumbent customers and the newcomer, this event affects the match value for all captive buyers in the same way. The likelihood of the event depends upon how tight market $x'(n+1;\mathbf{s})$ is. Finally, the last line of equation (5) includes the change in value due to an exogenous shock, whether of idiosyncratic (first term) or aggregate (second term) nature.

Importantly, equation (5) shows the sense in which the customer must anticipate the future path of prices. When the buyer is captive and the seller is subject to size or productivity changes, the customer must internalize how the seller will optimally redesign the contract under the new state. This meaningful forward-looking aspect of demand thus arises endogenously because the seller is committing to its customers. Let us now describe how the seller optimally chooses to do so.

3.3 Seller's Problem

Incumbent Sellers

Consider a seller with idiosyncratic productivity $z \in \mathcal{Z}$ who is endowed with $n \in \mathbb{N}$ captive customers. This seller currently follows the price strategy set up by its past contracts, under which its customers agreed to trade in exchange for a promised value of x. The problem of such a seller, whose expected value is denoted by $V^S(n,x;\mathbf{s})$, is to select a new contract $\boldsymbol{\omega} = \{p, \mathbf{x}'(n';\mathbf{s}')\}$ for all of its n customers so as to maximize the life-time value:

$$\begin{aligned} rV^S(n,x;\mathbf{s}) = \max_{\boldsymbol{\omega}\in\Omega} \Bigg\{ & pn - \mathcal{C}(n;\mathbf{s}) + \delta_f\Big(V_0^S(\varphi) - V^S(n,x;\mathbf{s})\Big) + n\delta_c\Big(V^S\big(n-1,x'(n-1;\mathbf{s});\mathbf{s}\big) - V^S(n,x;\mathbf{s})\Big) \\ & + \eta\Big(\theta\big(x'(n+1;\mathbf{s});\varphi\big)\Big)\Big(V^S\big(n+1,x'(n+1;\mathbf{s});\mathbf{s}\big) - V^S(n,x;\mathbf{s})\Big) \\ + \sum_{z'\in\mathcal{Z}} \lambda_z(z'|z)\Big(V^S\big(n,x'(n;z',\varphi);z',\varphi\big) - V^S(n,x;\mathbf{s})\Big) & + \sum_{\varphi'\in\Phi} \lambda_\varphi(\varphi'|\varphi)\Big(V^S\big(n,x'(n;z,\varphi');z,\varphi'\big) - V^S(n,x;\mathbf{s})\Big)\Bigg\} \end{aligned} \tag{6}$$

where $V_0^S(\varphi)$ denotes the value of having no customers (which we derive below).[25] The right side of equation (6) has the following components. The term $\big[pn - \mathcal{C}(n;\mathbf{s})\big]$ is the seller's flow profits, composed of revenue from selling n units, net of operating costs. The next term on the first line is the expected change in value if the seller goes bankrupt, in which case she instantly loses all customers and enters the pool of potential entrants. The third additive term states that the seller faces the probability that any one of its n customers separates from the match, in which case the seller shrinks down to size $(n-1)$ and delivers the promised value $x'(n-1;\mathbf{s}) \in \boldsymbol{\omega}$. The second line shows that, by posting a new offer $x'(n+1;\mathbf{s}) \in \boldsymbol{\omega}$, the seller attracts a certain mass of buyers and faces a probability of increasing its size to $n+1$. When making a new offer, the seller understands the sorting behavior of buyers across states for different promised values through the equilibrium θ schedule. In the event of a successful match, the seller would implement the new continuation value, and its state vector would transition from $(n,x;\mathbf{s})$ into $\big(n+1, x'(n+1;\mathbf{s});\mathbf{s}\big)$. Finally, the value of the firm could change exogenously because of a state

[25] The object $\Omega := \mathbb{R} \times [\underline{x}, \overline{x}]^{\overline{k}}$ denotes the set of admissible contracts, and $\overline{k} \equiv 3k_z k_\varphi - 1$. For $n = 1$, we note that $x'(n-1;\mathbf{s}) = \emptyset$, $\forall \mathbf{s} \in \mathcal{Z} \times \Phi$, and denote $V^S(n-1, x'(n-1;\mathbf{s});\mathbf{s})$ by $V_0^S(\varphi)$.

transition from $\mathbf{s} = (z, \varphi)$ to either (z', φ) or (z, φ'), as captured by the last two terms in equation (6).

When choosing a contract $\boldsymbol{\omega}$, a seller in state $(n, x; \mathbf{s})$ is constrained by the following condition:

$$V^B(n, \boldsymbol{\omega}; \mathbf{s}) \geq x \tag{7}$$

Equation (7) is a *promise-keeping (PK) constraint* guaranteeing that, in its choice of the contract, the seller honors the promises that were made in the past: the value that each buyer of the firm obtains under the contract must be weakly greater than the value x that was promised to her. This constraint is in place due to our commitment assumption on the seller's side.

Potential Entrants

To conclude with the description of the model's environment, let us describe the problem of an outside firm. These firms have no customers (i.e. $n = 0$) and, unlike incumbents, they must incur a flow set-up cost $\kappa > 0$ in order to post an initial contract. Prior to start selling the good, they must also draw an initial productivity level z_0 from the π_z distribution. For each possible realization $z_0 \in \mathcal{Z}$, the contract is the object $\{x'(1; z_0, \varphi)\}$, specifying the utility promised to the first customer of the firm under state (z_0, φ). Thus, the potential entrant chooses amongst a menu of contracts, $\boldsymbol{\omega}_0(\varphi) := \{\mathbf{x}'(1; z_0, \varphi)\}_{z_0 \in \mathcal{Z}}$, contingent on each realization of productivity at entry. Note, in particular, that the potential entrant's contract does not specify a price level for the first customer, for this choice is made ex-post, i.e. once the customer has been acquired (and the seller faces problem (6) for $n = 1$).

The ex-ante value of the potential entrant in aggregate state φ is, therefore:

$$rV_0^S(\varphi) = -\kappa(\varphi) + \sum_{z_0 \in \mathcal{Z}} \pi_z(z_0) v_0^S(z_0, \varphi) + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi) \Big(V_0^S(\varphi') - V_0^S(\varphi) \Big) \tag{8}$$

This value is composed of the set-up flow cost κ (first additive term), the expected value of posting a contract under productivity draw z_0 (second term), and the expected change in the ex-ante value of entry for a change in the aggregate state (third term). We have defined the expected value of entry for a firm under a z_0 draw by:

$$v_0^S(z_0, \varphi) := \max_{x' \in \mathcal{X}} \ \eta\big(\theta(x'; \varphi)\big) \Big(V^S\big(1, x'; z_0, \varphi\big) - V_0^S(\varphi) \Big) \tag{9}$$

Once again, the firm understands how inactive buyers sort across markets, as the $\theta(\cdot; \varphi)$ schedule is taken as given. Note that, because this firm does not yet have any customers at the time of choosing the contract, the entrant's problem is not subject to a PK constraint.

We assume free entry into the product market for the first customer. Since the total mass of sellers adjusts freely, this assumption implies that, in equilibrium, more firms will enter the economy as long as the expected value of posting a contract exceeds the set-up cost $\kappa(\varphi) > 0$. As more potential entrants flood into the market, this expected value is pushed down to the entry cost. Therefore, in an equilibrium with positive entry in all aggregate states, it must be the case that:

$$\forall \varphi \in \Phi : \qquad V_0^S(\varphi) = 0$$

Since, by construction, firms enter with one customer, the free-entry condition pins down the average market tightness among firms of size one in the cross-section of initial productivity levels. In particular, from equation (8), for all states $z_0 \in \mathcal{Z}$ such that $\theta\big(x_1'(z_0,\varphi);\varphi\big) > 0$, it must be the case that:

$$\kappa(\varphi) = \sum_{z_0 \in \mathcal{Z}} \pi_z(z_0)\eta\Big(\theta\big(x_1'(z_0,\varphi);\varphi\big)\Big)V^S\big(1, x_1'(z_0,\varphi); z_0, \varphi\big) \tag{10}$$

for any given $\varphi \in \Phi$, where $x_1'(z_0,\varphi)$ solves problem (9) in state (z_0,φ).

3.4 Optimal Contract

In this section, we derive and describe the properties of the optimal contract for a typical firm. Our main result is that, since contracts are complete, and sellers and buyers can engage in revenue-neutral transfers schemes, the profit-maximizing contract leads to an allocation of utilities in which the *joint surplus* (i.e. the sum of the expected values of a seller and all of its customers) is maximized. Moreover, for any allocation that maximizes the joint surplus, there always exists a price that maximizes the seller's profit subject to the PK constraint. Thus, the seller's and the joint surplus problems are equivalent. As we shall see, this opens up a great simplification of the state space, and it renders the equilibrium computationally tractable.

Joint Surplus Problem

To start, consider a typical firm whose state vector is $(n, x; \mathbf{s})$, where recall that $\mathbf{s} := (z,\varphi)$ collects the exogenous states. As seen in the last section, the optimal contract $\boldsymbol{\omega} = \big\{p, \mathbf{x}'(n';\mathbf{s}')\big\}$ can be obtained as the solution to the problem of the seller, described in (6). A standard monotonicity argument reveals that sellers will offer the lowest values to their buyers such that the seller's promises are still honored, and so the PK constraint (7) will hold with equality. Thus, to economize on notation, for the remainder of the paper we write x (a predetermined state variable) in place of $V^B(n,\boldsymbol{\omega};\mathbf{s})$.

Next, define the *joint surplus* in a typical state $(n, x; \mathbf{s})$ as the sum of the seller's expected value from the match, $V^S(n,x;\mathbf{s})$, and the aggregate expected value for all the n customers of the firm:

$$W(n,x;\mathbf{s}) := V^S(n,x;\mathbf{s}) + nx$$

In Appendix B.1 we show that the joint surplus can be written in the following HJB representation:

$$\begin{aligned}(r+\delta_f)W(n,x;\mathbf{s}) = \max_{\mathbf{x}'(n';\mathbf{s}')} \Bigg\{ & n\bigg(v(\varphi) + (\delta_f+\delta_c)U^B(\varphi)\bigg) - \bigg(\mathcal{C}(n;\mathbf{s}) + \eta\Big(\theta\big(x'(n+1;\mathbf{s});\varphi\big)\Big)x'(n+1;\mathbf{s})\bigg) \\ & + \eta\Big(\theta\big(x'(n+1;\mathbf{s});\varphi\big)\Big)\Big(W\big(n+1,x'(n+1;\mathbf{s});\mathbf{s}\big) - W(n,x;\mathbf{s})\Big) + n\delta_c\Big(W\big(n-1,x'(n-1;\mathbf{s});\mathbf{s}\big) - W(n,x;\mathbf{s})\Big) \\ & + \sum_{z'\in\mathcal{Z}} \lambda_z(z'|z)\Big(W\big(n,x(z',\varphi);z',\varphi\big) - W(n,x;\mathbf{s})\Big) + \sum_{\varphi'\in\Phi} \lambda_\varphi(\varphi'|\varphi)\Big(W\big(n,x(z,\varphi');z,\varphi'\big) - W(n,x;\mathbf{s})\Big)\Bigg\}\end{aligned} \tag{11}$$

Intuitively, equation (11) represents the joint surplus as the present discounted value of the buyers' total surplus, net of the seller's total costs. On the first line, the term $n\big(v(\varphi) + (\delta_f+\delta_c)U^B(\varphi)\big)$ represents the aggregate flow surplus for all the n customers of the firm, which is composed of the sum of the per-customer utility from consumption, v, and the expected per-customer gains from separation, $(\delta_f + \delta_c)U^B(\varphi)$. The second component in parentheses is the total costs of the match for the seller, which include total operating costs, $\mathcal{C}(n;\mathbf{s})$, and the expected costs of offering a life-time continuation value of $x'(n+1;\mathbf{s})$ to the new incoming customer, adjusted by the endogenous rate at which a new customer joins the match. The second and third lines include the change in the expected joint surplus when the match shrinks (because any one of the n customers leaves, or the firm is destroyed), or grows (because a new customer joins). Finally, the last two terms incorporate expected changes in the joint surplus that are due to exogenous shocks to z and φ.

With this specification at hand, we can now state the main equivalence result:

Proposition 1 (Joint Surplus Problem) *The following properties hold:*

i. The firm's and the joint surplus problems are equivalent*:*

(a) Given a contract $\boldsymbol{\omega}^* = \{p, \mathbf{x}'(n';\mathbf{s}')\}$ that maximizes (6), $\mathbf{x}'(n';\mathbf{s}')$ is a solution to (11).

(b) Conversely, for every vector $\mathbf{x}'(n';\mathbf{s}')$ that solves (11), there exists a unique p for which $\{p, \mathbf{x}'(n';\mathbf{s}')\}$ is a solution to (6).

ii. The joint surplus is invariant to x, i.e. $W(n,x;\mathbf{s}) = W(n,\tilde{x};\mathbf{s})$, for all $x,\tilde{x} \in \mathcal{X}$, $n \in \mathbb{N}$, $\mathbf{s} \in \mathcal{Z} \times \Phi$.

The proof is in Appendix B.1. Part *i.* of Proposition 1 establishes that the contract that maximizes the seller's profits can be found by solving an alternative problem, given by (11). In this problem, the contract is designed so as to maximize the profits of all the parties involved in a utilitarian manner, provided that the seller extracts rents from each buyer up to the limit established by promise-keeping. Since the contract space is complete (that is, it specifies continuation promises for each and every possible future state), and both agents have linear preferences, there always exists a menu composed of a price and promised utility pair that, for any configuration of future states, redistributes rents among the seller and its customers in a payoff-maximizing manner. Moreover, because the seller commits to the terms of the contract, the allocation is unique in that no deviation from joint surplus maximization can deliver higher profits for the seller.

Part *ii.* of the proposition thus follows immediately from the first one, and clarifies why problem (11) is much simpler to solve than the firm's problem (in (6)). Since price and continuation promises map one-for-one, the maximized surplus is invariant to the rent-sharing components of the contract. Conveniently, this means that the problem can be split in two stages. In the first stage, the firm sets the vector of continuation promises $\mathbf{x}'(n';\mathbf{s}') \subseteq \mathcal{X}$ that maximizes the size of the surplus under every possible combination of future states. In the second stage, the price level is set so as to implement such an allocation, thereby splitting and distributing rents among the $(n+1)$ agents involved so that the seller keeps its promise (i.e. ensuring that PK binds in every state). Further, the surplus is also constant in the firm's previous promise, since x is a predetermined state that was chosen optimally in the prior stage of the firm. By Markov perfection and completeness, the size n and current exogenous state $\mathbf{s} = (z,\varphi)$ serve as sufficient statistics to determine the current surplus-maximizing policies. Thus, given $\mathbf{s}$, there exists a sequence $\{W_n(\mathbf{s})\}_{n=1}^{+\infty}$ of positive real numbers such that the joint surplus can be expressed as:

$$W_n(\mathbf{s}) = W(n, x; \mathbf{s}), \quad \forall n \in \mathbb{N}$$

As a result, the policy that solves problem (11) is not a function of x, and neither is the optimal price level. While the equivalence between the joint-surplus problem and the decentralized problem is a familiar result in the literature on complete contracts with commitment and transferrable utilities, here we show that it can also result from, and provide great analytical tractability to, a dynamic model with ex-post heterogeneity and meaningful firm dynamics.[26]

Characterization

Let us characterize the equilibrium policies that result from problem (11). Recall that, by ex-ante indifference, the option value of matching for the buyer is constant across markets and given by $\Gamma^B(\varphi)$ (equation (4)). Then, by equation (3) we know that:

$$\theta(x; \varphi) = \mu^{-1}\left(\frac{\Gamma^B(\varphi)}{x - U^B(\varphi)}\right) \tag{12}$$

for all $x \in \mathcal{X}$. By Assumption 2.*i* and continuity of θ on x, equation (11) describes the maximization of a continuous function over a compact support, so there exist promises $\{x_n^+(\mathbf{s}), x_n^-(\mathbf{s}), \mathbf{x}_n(\mathbf{s}')\}$ and a price level $p_n(\mathbf{s})$ solving the joint surplus problem, where the "+" (respectively, "−") superscript denotes the upsize (respectively, downsize) decision. Once again, note that we index these policies by n, but not x.

Stage 1. Continuation promises Let us begin with the choice of x^+. First, using equation (12) and differentiability of η, the following first-order condition is sufficient for optimality:[27]

$$\left.\frac{\partial \eta\big(\theta(x;\varphi)\big)}{\partial x}\right|_{x=x_n^+(\mathbf{s})}\Big(W_{n+1}(\mathbf{s}) - W_n(\mathbf{s})\Big) = \left.\frac{\partial \eta\big(\theta(x;\varphi)\big)}{\partial x}\right|_{x=x_n^+(\mathbf{s})} x_n^+(\mathbf{s}) + \eta\Big(\theta\big(x_n^+(\mathbf{s});\varphi\big)\Big) \tag{13}$$

Intuitively, the optimal continuation value $x_n^+(\mathbf{s})$ equates the expected marginal benefit of upgrading the size of the firm by one customer (left-hand side), to the expected marginal costs of such a transition (right-hand side). On the one hand, an increase by one dollar in the promised value x^+ increases the joint surplus by the amount $(W_{n+1} - W_n) > 0$ in case the seller makes a size transition. These gains must then be weighted by the marginal effect of x_n^+ on the likelihood that the firm meets a new customer. On the other hand, for every dollar spent on the new continuation value x_n^+, the seller incurs in two associated costs: first, the direct cost of delivering the new value to the additional customer, weighted by the change in the meeting rate; and second, the decrease in the price level, by $\eta\big(\theta(x_n^+;\varphi)\big)$ dollars, which is required by promise-keeping.

As for the choices of $\mathbf{x}(\mathbf{s}')$ and $x^-(\mathbf{s})$, note that these do not feature anywhere in equation (11) once we impose that the joint surplus is invariant to promised utilities (Proposition 1, part *ii.*). Therefore, $\{\mathbf{x}_n(\mathbf{s}'), x_n^-(\mathbf{s})\}$ cannot be determined by a surplus-maximizing condition similar to (13). Instead, these

[26] For an application of this idea to a rich firm-dynamics search model of the labor market, see Schaal (2017).

[27] Sufficiency obtains because the second-order condition follows from Assumption 2.*iii* specialized to $h(x) = \frac{\Gamma^B}{x - U^B}$ and $\widehat{x} = W_{n+1} - W_n$.

values are purely redistributive: the only dimension in which they matter is the price level, and thus they affect only the way in which the total surplus is split (i.e. the terms of trade). In particular, since beliefs are pinned down by equation (12), the firm's choice must be consistent with the sorting behavior of inactive buyers. By symmetry, the optimal downsizing choice for a size-n must be consistent with the optimal upsizing choice for a firm of size $(n-2)$, or $x_n^-(\mathbf{s}) = x_{n-2}^+(\mathbf{s})$. Similarly, when transitioning to another state, equation (12) and symmetry require that $x_n(\mathbf{s}') = x_{n-1}^+(\mathbf{s}')$. Therefore, a firm's optimal continuation utility is independent of the firm's state history. Moreover, by commitment, we know that firms must deliver on their outstanding promises. In sum:

$$x_n(\mathbf{s}) = x_{n-1}^+(\mathbf{s}) = x_{n+1}^-(\mathbf{s}), \quad \forall (n; \mathbf{s}) \in \mathbb{N} \times \mathcal{Z} \times \Phi$$

Thus, the set of active market segments in equilibrium is comprised of a collection of promised utility levels, $\mathcal{X}^* := \{x_n(z,\varphi) : (n,z,\varphi) \in \mathbb{N} \times \mathcal{Z} \times \Phi\}$, where each element $x_n(z,\varphi)$ is given by the solution to (13).

Clearly, in equilibrium we must have $x_n(z,\varphi) > U^B(\varphi)$, a direct implication of the inactive buyer's search problem: active markets must provide a positive option value to idle consumers, or else these markets would remain unvisited. More interestingly, as we will show in Section 4 by means of numerical examples, there exist parametrizations under which x_n can be *either an increasing or a decreasing sequence* in size n. Later on we shall argue, however, that the empirically relevant case is for x_n to be a *decreasing* sequence in n, namely $x_n(z,\varphi) \geq x_{n+1}(z,\varphi) \geq \cdots > U^B(\varphi)$. Under the latter parametrization, sellers with less customers write more attractive contracts from the point of view of (matched) customers. Hence, though indifferent ex-ante, buyers ex-post prefer to be matched to smaller firms. As sellers mature and their customer base expands, the promised utility declines as they increasingly prefer to extract more rents from each captive customer. Importantly, as we shall discuss in the next section, this observation is a result of the product market frictions, and does not hinge on the behavior of marginal costs across sizes. Indeed, the result holds even in parametrizations in which $\mathcal{C}(n)$ is linear in n. Figure A.1 in Appendix A depicts the different markets in equilibrium, for a given state $\mathbf{s} = (z,\varphi)$. All equilibrium markets are distributed on the θ schedule defined by buyer's ex-ante revenue equalization, and the sequence of markets is constructed inductively as described above. To grow, the seller makes a state-contingent promise that is strictly below the current valuation of buyers, depicted on the horizontal axis. In equilibrium, the resulting collection of markets make buyers indifferent ex-ante.

Once the equilibrium markets are pinned down, the remaining equilibrium objects readily follow. First, equilibrium market tightness levels are given by $\theta_n(z,\varphi) := \theta\big(x_n(z,\varphi);\varphi\big)$ via equation (12). Since $\theta(x;\varphi)$ is an increasing and continuous function of x (equation (3)), then θ_n inherits the size-dependence in x_n. When x_n is decreasing, smaller firms attract more buyers per unit of time by offering higher ex-post values, so the buyer-to-seller ratio is higher in those markets, and these firms grow relatively faster compared to other firms. Indeed, we can write the law of motion of the seller's customer base as:

$$n_{t+\Delta} - n_t = \begin{cases} 1 & \text{w/prob. } \eta\big(\theta_{n_t+1}(z,\varphi)\big)\Delta + o(\Delta) \\ -1 & \text{w/prob. } n_t\delta_c\Delta + o(\Delta) \\ -n_t & \text{w/prob. } \delta_f\Delta + o(\Delta) \\ 0 & \text{else} \end{cases} \tag{14}$$

where $\Delta > 0$ is small, and $o(\Delta)$ satisfies $\lim_{\Delta \searrow 0} \frac{o(\Delta)}{\Delta} = 0$. For small firms, the probability of attracting a new customer (first line) is relatively higher. As the firm grows, the attrition probability (second line) increases proportionally to the firm's size as the attraction probability decays due to a decrease in x_n, leading to slower firm growth. Eventually, conditional on survival, these differences in growth rates ensure that firms converge to a stationary size.

Stage 2. Prices Finally, the equilibrium price is given implicitly by the PK constraint, which binds with equality. First, we replace $V^B(n, \boldsymbol{\omega}_n; z, \varphi) = x_n(z,\varphi)$ and $\boldsymbol{\omega}_n = \big\{p_n(z,\varphi); \mathbf{x}_{n'}(z',\varphi') : (n',z',\varphi') \in \{n-1, n, n+1\} \times \mathcal{Z} \times \Phi\big\}$ in equation (5). Then, solving for $p_n(z,\varphi)$ we obtain:

$$\begin{aligned} p_n(z,\varphi) = & \underbrace{v(\varphi) - r x_n(z,\varphi)}_{\substack{\geq 0 \\ \text{Baseline component}}} + \underbrace{\delta_f\big(U^B(\varphi) - x_n(z,\varphi)\big)}_{\substack{\leq 0 \\ \text{Exit component}}} + \underbrace{\eta\big(\theta_{n+1}(z,\varphi)\big)\big(x_{n+1}(z,\varphi) - x_n(z,\varphi)\big)}_{\substack{\leq 0 \\ \text{Growth component}}} \\ & + \underbrace{n\delta_c\left(\frac{U^B(\varphi) + (n-1)x_{n-1}(z,\varphi)}{n} - x_n(z,\varphi)\right)}_{\substack{\leq 0 \\ \text{Separation component}}} + \underbrace{\sum_{z' \in \mathcal{Z}} \lambda_z(z'|z)\Big(x_n(z',\varphi) - x_n(z,\varphi)\Big)}_{\text{Idiosyncratic-shock component}} + \underbrace{\sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)\Big(x_n(z,\varphi') - x_n(z,\varphi)\Big)}_{\text{Aggregate-shock component}} \end{aligned} \tag{15}$$

The optimal price level for a firm of type (n, z) can be decomposed into the following parts. The first one, $v - rx_n$, is the price level that would prevail if, in the absence of any exogenous shock, each customer were to stay matched forever with its seller and the firm did not change size going forward. We call this term the *baseline price level*.[28] The remaining terms in (15) introduce the necessary adjustments for possible changes in firm states. These adjustments persuade the customers to accept the terms of trade at the margin imposed by the firm's promise-keeping.

To provide intuition, consider the parametrization under which $x_n \geq x_{n+1}$. First, the firm offers a price reduction of $\delta_f\big(U^B - x_n\big) \leq 0$ to compensate the customer for the expected loss in value in the event that the firm exits the market. We label this the *exit component*. Second, the term $\eta(\theta_{n+1})(x_{n+1} - x_n) \leq 0$ is a compensation for the possibility that the firm grows. The pre-existing customer requires a compensation because in that case the total surplus is split among more buyers and each individual customer obtains a lower share. This compensation is thus labeled as the *growth component*. Third, the firm adjusts the price for the possibility of customer separation: a reduction in size lowers the seller's value and has a pecuniary externality on all the customers that remain matched, so the price must again be adjusted to remain compatible with the seller's commitment. We call this term the *separation component*. If a sep-

[28] Indeed, in that case we would have $p_n = v - rx_n$, so $x_n = \int_0^{+\infty} e^{-rt}(v - p_n)\mathrm{d}t$, the PDV of perpetually obtaining the fixed surplus $(v - p_n)$.

aration occurs, then the separating customer obtains U^B, and the remaining non-separating customers each obtain x_{n-1}. This amounts to an average value of $\frac{U^B+(n-1)x_{n-1}}{n}$ per customer, which is the expected change in the per-customer value due to a separation. Finally, the last two terms in equation (15) adjust the price level for expected changes in the exogenous states.

In sum, as in the original intuition behind Rotemberg and Woodford (1991), firms use prices as a way to invest into larger market shares in the future. In Section 4 we will discuss the different price effects that may be present in equilibrium, provide intuition for the dependence on size, and describe the qualitative properties of the solution.

3.5 Distribution Dynamics

To close the equilibrium, we need to describe the dynamics of the distribution of agents. The equilibrium of the economy described above features heterogeneous agents making forward-looking decisions and sorting into distinct product markets in the presence of both idiosyncratic and aggregate shocks. The distribution of agents across markets in turn depends on the aggregation of such decisions. Yet, the characterization of individuals' decisions has been silent on the exact composition of buyers and sellers across market segments, or the evolution of this distribution. This property is known as *block-recursivity*.

In our model, block recursivity arises from two key ingredients. On the one hand, we assume that search is directed, and thus sellers' offers are not contingent on the identity of the applicant (in particular, they are not contingent on the applicant's outside option). As a result, market tightness, which embodies agents' distributions, serves as a sufficient statistic for both sellers and buyers when making decisions, and allows them to not have to forecast the evolution of aggregates over future states of the economy. On the other hand, the ex-ante revenue-equalization condition across all markets among inactive buyers (equation (3)), and free entry of firms in the different entry markets $\{(n,z) : n = 1, z \in \mathcal{Z}\}$ (equation (10)), together imply that the equilibrium tightness on each market adjusts to be consistent with agents' beliefs.[29] This has allowed us to inductively construct the entire sequence of buyer-to-seller ratios without ever having to specify the exact composition of agent types within each market segment. Thus, the equilibrium policy functions depend on the aggregate state $\varphi \in \Phi$, but not on the distribution of agents across individual states (n, z). Because market tightness is a sufficient statistic to evaluate payoffs in this economy, the model allows for the description of distribution dynamics (on and off equilibrium) by means of flow equations (below), and its numerical solution does not require approximation techniques such as those of Krusell and Smith (1998), which are typically needed in models with aggregate shocks. This makes our environment particularly apt to study aggregate product market dynamics.

Let us now present the aggregate dynamics of the model. Let $S_{n,t}(z) \geq 0$ be the total measure of firms of size n with idiosyncratic productivity $z \in \mathcal{Z}$ at time $t \geq 0$. Recall that all such firms are seeking new customers in market $x_{n+1}(z, \varphi)$. Therefore, letting $B^I_{n+1}(z, \varphi)$ be the measure of (inactive and searching)

[29] Kaas and Kircher (2015) exploit similar insights to obtain tractability. An alternative approach would have been to dispense of the indifference condition among inactive buyers, and assume instead free entry of firms across all contracts (not only in the market of single-unit firms). This is the approach usually followed by the literature of directed search with aggregate shocks and on-the-job search (e.g. Menzio and Shi (2010, 2011) and Schaal (2017)). Moscarini and Postel-Vinay (2013) develop similar tools for firm-dynamics models of random search.

buyers within market $x_{n+1}(z, \varphi)$, market tightness must guarantee that:

$$B^I_{n+1,t}(z, \varphi) = \theta_{n+1}(z, \varphi) \cdot S_{n,t}(z) \tag{16}$$

at every $t \geq 0$ for all $n \in \mathbb{N}$. Using that $\eta(\theta) = \theta\mu(\theta)$, equation (16) can be written $\mu\big(\theta_n(z, \varphi)\big) B^I_{n,t}(z, \varphi) = \eta\big(\theta_n(z, \varphi)\big) S_{n-1,t}(z)$, stating that the measure of inactive buyers who become customers of a (n, z)-type firm is equal to the measure of sellers of productivity z and size $n-1$ who acquire an additional customer.

Similarly, let $B^A_{n,t}(z)$ be the measure of customers that are matched with firms of type (n, z) at time t. By construction, we have:

$$B^A_{n,t}(z) = nS_{n,t}(z) \tag{17}$$

at any $t \geq 0$. The measures of inactive and active buyers must add up to the total mass of buyers in the economy at all times, and thus:

$$\forall \varphi \in \Phi,\ \forall t \geq 0 : \quad \underbrace{\sum_{n=1}^{+\infty} \sum_{z \in \mathcal{Z}} B^A_{n,t}(z)}_{=B^A_t} + \underbrace{\sum_{n=1}^{+\infty} \sum_{z \in \mathcal{Z}} B^I_{n,t}(z, \varphi)}_{=B^I_t} = 1 \tag{18}$$

This equation establishes an aggregate feasibility constraint, stating that the unit mass of buyers must be either matched with a firm and consuming, or looking for one.

We have made $B^I_n(z, \varphi)$ depend explicitly on both φ and time t because this is a jump variable that also evolves smoothly over time for each given state. Indeed, the measure of customers looking to be matched with a specific type of seller responds instantaneously to the aggregate state to guarantee that the indifference condition among unmatched buyers (equation (3)) is met in all states of nature. In contrast, $S_{n,t}(z)$ is a stock variable, as it varies with t but does not respond instantaneously to changes in φ. Through equation (17), $B^A_{n,t}(z)$ changes smoothly over time, while through equation (16) market tightness $\theta_n(z, \varphi)$ must jump instantaneously in response to aggregate shocks. Yet, by the block-recursivity property, we do not explicitly index it by t, for it remains constant along each aggregate state. The mass of potential entrants, denoted $S_{0,t}(\varphi)$, jumps following a φ-shock, and otherwise evolves smoothly due to sellers flowing in and out of inactivity in the transition. Finally, because the evolution of the measure of customers is always continuous, by equation (18) the distribution of inactive buyers searching on each market must jump with each regime switch in such a way for the *aggregate* measure of inactive buyers to adjust only smoothly over time. Other economic aggregates, such as the average price level, average markup, and aggregate output, inherit this property as well.

In sum, while our policy functions are jump variables, the distributions of agents respond slowly to aggregate shocks. Because of this slow adjustment, the model features sluggish aggregate dynamics. Figure A.2 in the Appendix provides a graphical and comprehensive depiction of all possible transitions. Mathematically, the dynamics of sellers over idiosyncratic states can be summarized by a set of Kolmogorov Forward (KF) equations taking values on a discrete support $\mathbb{N} \times \mathcal{Z}$.[30]

First, for $n = 1$ we have:

[30] The derivation of the KF equations to follow can be found in Appendix D.2.

$$\partial_t S_{1,t}(z) = \pi_z(z)\eta\big(\theta_1(z,\varphi)\big)S_{0,t}(\varphi) + 2\delta_c S_{2,t}(z) + \sum_{\tilde{z}\neq z}\lambda_z(z|\tilde{z})S_{1,t}(\tilde{z}) - \Big(\delta_f + \delta_c + \eta\big(\theta_2(z,\varphi)\big) + \sum_{\tilde{z}\neq z}\lambda_z(\tilde{z}|z)\Big)S_{1,t}(z) \quad (19)$$

for any $z \in \mathcal{Z}$, where ∂_t is the partial derivative operator with respect to time. This equation represents flows into and out of state $(1, z)$. Inflows (first line) are given by those successful entrants that draw productivity z upon entry, and by the share of incumbents that are either of type z and size $n = 2$ and lose one customer, or that have one customer and transition into the productivity state z from some $\tilde{z} \neq z$. Outflows (second line) are given by firms in state $(1, z)$ that either die, lose their only customer, gain a second customer, or transition to a distinct productivity state, $\tilde{z} \neq z$. The aggregate state enters the law of motion only implicitly through its influence on the jump dynamics of S_0 and θ_1. Therefore, the dynamics of S_1 are smooth (i.e. not indexed by φ).

Similarly, for any $n \geq 2$:

$$\partial_t S_{n,t}(z) = \eta\big(\theta_n(z,\varphi)\big)S_{n-1,t}(z) + (n+1)\delta_c S_{n+1,t}(z) + \sum_{\tilde{z}\neq z}\lambda_z(z|\tilde{z})S_{n,t}(\tilde{z}) - \Big(\delta_f + n\delta_c + \eta\big(\theta_{n+1}(z,\varphi)\big) + \sum_{\tilde{z}\neq z}\lambda_z(\tilde{z}|z)\Big)S_{n,t}(z) \quad (20)$$

The interpretation is similar to the previous equation: flows into state (n, z), $n \geq 2$, are given by the share of firms of size $(n-1)$ that obtain their n^{th} customer, the share of firms of size $(n+1)$ that lose one customer, and the share of size-n firms that transition into productivity level z from some state $\tilde{z} \neq z$; outflows are given by firms that either die, lose or gain a customer, or experience a productivity shock.

Finally, the measure of potential entrants, $S_{0,t}(\varphi)$, evolves according to the following ODE:

$$\partial_t S_{0,t}(\varphi) = \delta_f \mathcal{S}_t + \delta_c \sum_{z\in\mathcal{Z}} S_{1,t}(z) - \sum_{z_0\in\mathcal{Z}} \pi_z(z_0)\eta\big(\theta_{1,t}(z_0,\varphi)\big)S_{0,t}(\varphi) \quad (21)$$

where $\mathcal{S}_t := \sum_{n=1}^{+\infty}\sum_{z\in\mathcal{Z}} S_{n,t}(z)$ is the total measure of *incumbent* firms (i.e. firms with one or more customers). The usual intuition applies, with the particularity that entering firms must now draw an initial productivity level at random, $z_0 \sim \pi_z$.

Equations (19)-(21) offer a full characterization of out-of-steady-state dynamics, and thanks to the property of block recursivity we can analyze the transition paths of the economy in response to aggregate shocks (Section 5). We can further specialize these equations to obtain the *time-invariant* distribution of firms by equating flows in and out of every possible state: $\partial_t S_{n,t}(z) = 0, \forall(n, z) \in \mathbb{N}\times\mathcal{Z}$.[31] The following result ensures that the dynamics are convergent:

[31] In general, an analytical solution for the stationary distribution does not exist. One exception is the economy without shocks and $\delta_f = 0$. In this case, the flow equations amount to a continuous-time Markov chain with reflection bound at zero and exponentially distributed transition times (sometimes called a *birth-death process*), where transition rates are endogenous and state-dependent. Appendix D.2 derives the analytical solution for this particular case.

Proposition 2 (Stability) *Given an equilibrium allocation, the dynamical system represented by the flow equations (19), (20), and (21) is stable, and converges to an invariant distribution for each aggregate state $\varphi \in \Phi$.*

For the proof, see Appendix B.2. In Appendix D.2 we then show how to derive the aggregate equilibrium measures of agents in the stationary solution explicitly.

3.6 Equilibrium Definition and Efficiency

We are now ready to define an equilibrium:

Definition 1 *A Recursive Equilibrium is, for each aggregate state $\varphi \in \Phi$, a set of value functions $V^S : \mathbb{N} \times \mathcal{X} \times \mathcal{Z} \to \mathbb{R}_+$ and $V^B : \mathbb{N} \times \Omega \times \mathcal{Z} \to \mathbb{R}_+$; a value of inactivity $U^B(\varphi) \in \mathbb{R}$; a joint surplus $W : \mathbb{N} \times \mathcal{Z} \to \mathbb{R}_+$; a contract $\boldsymbol{\omega}_n(\mathbf{s}) = \{p_n(\mathbf{s}), \mathbf{x}_n(\mathbf{s}'), x_n^+(\mathbf{s}), x_n^-(\mathbf{s})\}$ for incumbent firms, and $\boldsymbol{\omega}_0(\varphi) = \{x_1(z_0, \varphi)\}_{z_0 \in \mathcal{Z}}$ for potential entrant firms; a decision rule $\widehat{x}(\varphi)$ for inactive buyers and a promised utility $x_n(\mathbf{s})$ for active buyers of firms of type $(n, z) \in \mathbb{N} \times \mathcal{Z}$; a market tightness function $\theta(\cdot, \varphi) : \mathcal{X} \to \mathbb{R}_+$; aggregate measures of agents: $\{S_0(\varphi), \mathcal{S}(\varphi), B^A(\varphi), B^I(\varphi)\}$; and a distribution of sellers and buyers: $\{S_n(z), B_n^A(z), B_n^I(z, \varphi) : (n, z) \in \mathbb{N} \times \mathcal{Z}\}$; such that: (i) the value functions solve (5) and (6), $U^B(\varphi)$ is the fixed point of the free-entry condition (9)-(10), and the joint surplus $W_n(z, \varphi)$ solves (11); (ii) the entrant's contract $\boldsymbol{\omega}_0(\varphi)$ solves problem (8), and the incumbent's contract $\boldsymbol{\omega}_n(\mathbf{s})$ is such that $x_n^+(\mathbf{s})$ satisfies (13), $x_n^-(\mathbf{s}) = x_{n-1}(\mathbf{s})$, and $p_n(\mathbf{s})$ is given by (15); (iii) $\widehat{x}(\varphi)$ solves the inactive buyer's problem, (1)-(2), and sellers promise utility $x_n(\mathbf{s}) = x_{n-1}^+(\mathbf{s})$; (iv) market tightness $\theta(x; \varphi)$ is consistent with the sorting behavior of inactive buyers, (3); and (v) aggregates and the distribution of agents satisfy the flow equations described in Section 3.5.*

To compute the decentralized recursive equilibrium, one can show that the joint surplus problem defines a contraction in a space of vector-valued function, and obtain the value of inactivity $U^B(\varphi)$ as the fixed point of the free entry problem.[32] This insight is instructive for the numerical implementation (Appendix C.1), which exploits the nested fixed-point nature of the problem to solve for equation (11) via value function iteration on W and a bisection step on U^B. Existence of the recursive equilibrium is, however, notoriously harder to show given the rich structure of the model. Particularly, a non-trivial requirement for block recursivity is that there be non-negative entry of firms in all aggregate states. This condition effectively sets bounds on the exogenous state processes so that the free entry condition, and therefore ex-ante revenue equalization across markets and by extension the block recursivity property, can be met in all states of nature.

Proposition 3 below states that the recursive equilibrium is constrained-efficient. In particular, it establishes that the decentralized allocation maximizes aggregate welfare subject to the cross-sectional and dynamic properties of the distribution of agents described in Section 3.5. In our environment, the planner chooses distributions of buyers and sellers, as well as market tightness levels, across all states and time, in order to maximize:

$$\mathbb{E}_0 \int_0^{+\infty} e^{-rt} \Bigg\{ -\kappa(\varphi_t) S_{0,t} + \sum_{n_t=1}^{+\infty} \sum_{z_t \in \mathcal{Z}} \Big(v(\varphi_t) B_{n_t,t}^A(z_t) - \mathcal{C}(n_t; z_t, \varphi_t) S_{n_t,t}(z_t) - c(\varphi_t) B_{n_t,t}^I(z_t) \Big) \Bigg\} \mathrm{d}t \quad (22)$$

[32] A preliminary proof of this result is available upon request.

Aggregate welfare thus equals the present discounted sum of consumption gains by active buyers, net of search costs by inactive buyers, and production and entry costs by firms. Using this definition of welfare, we then establish:

Proposition 3 (Efficiency) *The allocation of the decentralized equilibrium coincides with the planner's solution.*

For the proof, see Appendix B.3. This result implies that our model features efficient firm dynamics and efficient pricing behavior. In particular, markup dispersion is necessary to optimally split the gains from trade among buyers and sellers, as prices in our environment serve to efficiently direct buyer search toward specific product markets. The result is in contrast to models explaining dispersion in firm-level revenue through resource misallocation (e.g. Hsieh and Klenow (2009)). While we do not rule our other interpretations, our setting demonstrates that this type of dispersion may also be generated through efficient pricing.

4 Understanding the Mechanism

Before turning to the empirical and quantitative parts of the paper, this section presents a discussion of the qualitative properties of the equilibrium, with an emphasis on how product market frictions lead firms of different sizes to set different combinations of price and promised utilities, and to experience different subsequent growth paths along their life cycle. We finish the section with a discussion of the main modeling assumptions and describe the role that each of them plays in equilibrium.

4.1 Qualitative Features

To describe the qualitative features of the economy, we begin with a useful result: under a standard parametrization of the meeting rates, we can obtain an analytical characterization of the joint surplus. In particular, for the remainder of the paper we will use the Cobb-Douglas matching function:

$$\mu(\theta) = \theta^{\gamma-1}$$

with $\eta(\theta) = \theta\mu(\theta)$, where $\gamma \in (0,1)$ is the matching elasticity. The following proposition summarizes the solution to the joint surplus in this case.

Proposition 4 (Analytical Solution of the Joint Surplus) *For each $(z,\varphi) \in \mathcal{Z} \times \Phi$:*

(a) The joint surplus $W_n(z,\varphi)$ solves the following second-order difference equation:

$$W_{n+1}(z,\varphi) = W_n(z,\varphi) + U^B(\varphi) + \left(\frac{\Gamma^B(\varphi)}{\gamma}\right)^{\gamma} \left(\frac{\Gamma_n^S(z,\varphi)}{1-\gamma}\right)^{1-\gamma} \tag{23}$$

where $\Gamma^B(\varphi)$ is given by (4), and $\{\Gamma_n^S(z,\varphi)\}_{n=1}^{+\infty}$ is given in equation (B.4.4) of Appendix B.4.

(b) The buyers' promised utility is given by $x_{n+1}(z,\varphi) = \gamma\Big(W_{n+1}(z,\varphi) - W_n(z,\varphi)\Big) + (1-\gamma)U^B(\varphi)$.

The proof is given in Appendix B.4. Proposition 4 shows that, in spite of the rich dynamics of the model, the solution to the joint surplus problem can be expressed analytically for each realization of the shock.[33] The result has an intuitive interpretation. First, recall from equation (3) that $\Gamma^B(\varphi)$ is the expected value from matching for a buyer, which is equalized across markets (and, therefore, independent of (n, z)). Likewise, in the Appendix we argue that $\Gamma^S_n(z, \varphi)$ can be interpreted as the expected match value for a seller in market (n, z). Therefore, result (23) says that, with a Cobb-Douglas matching function, the equilibrium marginal net gain in joint surplus for each additional customer is a convex combination of the *expected* match surplus that accrues to the seller, Γ^S, and that which accrues to the new customer of the firm, Γ^B. The matching elasticity parameter governs how the surplus is shared between the seller and the new customer. In particular, after a $n \to (n+1)$ transition, the new customer of the firm absorbs a fraction $0 < \gamma < 1$ of the gains in the total net surplus from the new match:[34]

$$\underbrace{x_{n+1} - U^B}_{\substack{\text{Incoming customer's ex-post} \\ \text{net gains from matching}}} = \gamma \underbrace{\left(W_{n+1} - W_n - U^B\right)}_{\substack{\text{Total net gains in joint surplus} \\ \text{from the size expansion}}}$$

In turn, the added value for the seller is given by:

$$\underbrace{V^S_{n+1} - V^S_n}_{\substack{\text{Seller's ex-post net gain} \\ \text{from new customer}}} = \underbrace{(1-\gamma)\left(W_{n+1} - W_n - U^B\right)}_{\substack{\text{[A] Surplus extracted} \\ \text{directly out of new match}}} + \underbrace{n\left(x_n - x_{n+1}\right)}_{\substack{\text{[B] Surplus transferred from all} \\ \text{pre-existing customers}}}$$

In words, after a $n \to (n+1)$ transition, the seller absorbs the remainder share $(1-\gamma)$ of the net gain in the joint value from the new customer (term [A]), and some additional surplus that the seller extracts from each of the n pre-existing customers (term [B]), whose individual value has now decreased from x_n to $x_{n+1} \leq x_n$.

Let us now explain intuitively why the seller's size affects its incentives to build a customer base in the first place. For this, we proceed in two steps: first, we show how, give a size n, the seller uses prices and promised utilities as complementary instruments to extract customer rents; second, we provide intuition for the dependence between promised utilities and size, and the direction of this correlation.

Price versus promised utility Let us first show how the price is affected by size changes through adjustments in the promised utility. For this, consider taking partial derivatives to equation (15) around the (x_n, x_{n+1}) equilibrium promises. Respectively, this yields:

$$\left.\frac{\partial p_n}{\partial x}\right|_{x=x_n} = -\left(r + \delta_f + \eta(\theta_{n+1}) + n\delta_c\right) < 0 \tag{24a}$$

[33] We must note that this is not a *closed-form* solution, in the sense that the equilibrium object U^B features in equation (23). It is also worth noting that the existence of such a solution is not specific to the Cobb-Douglas case. For instance, a second-order difference equation for W also emerges under the CES function $\mu(\theta) = (1+\theta^\gamma)^{-1/\gamma}$, where $\gamma > 1$ (proof available upon request).

[34] We suppress state dependence here to alleviate notation. This does not affect the intuition in any significant way.

$$\left.\frac{\partial p_n}{\partial x}\right|_{x=x_{n+1}} = \underbrace{\eta(\theta_{n+1})}_{[\text{C}]>0} + \underbrace{\left.\frac{\partial \eta(\theta_{n+1})}{\partial x}\right|_{x=x_{n+1}}}_{[\text{D}]>0} \underbrace{(x_{n+1} - x_n)}_{[\text{E}]<0} \lesseqgtr 0 \tag{24b}$$

Equation (24a) shows that there is always a negative relation between the utility paid to each incumbent buyer and the price each one is charged. This is because the PK constraint binds in equilibrium: higher prices are detrimental to the customer's valuation. This dependence is strongest for the smallest and the largest firms, where firm size is more likely to change per instant of time, and weakest (though still negative) for sellers close to their stationary size.

More interesting is the effect of a seller's promise to a potentially *incoming* customer on the price that the seller charges to its currently *incumbent* customers (equation (24b)). This marginal effect has two additive terms of opposite sign. The first part (term [C]) says that each additional dollar that the seller offers to the new potential customer must be financed through revenue obtained from an increase in the price that is currently being paid by the pre-existing customers. In expectation, this additional dollar is worth $\eta(\theta_{n+1})$, as this is the effective probability with which a new customer actually joins the match. At the same time, offering an additional dollar to the new potential customer raises the probability with which a new customer is successfully attracted (as captured by term [D], of positive sign by equation (3)). In such an event, we have argued above (term [B]) that the seller captures the difference $(x_n - x_{n+1}) > 0$ from each and every one of the n captive buyers. Therefore, part [E] says that each individual incumbent customer can expect its own valuation to *decrease* by exactly this amount in the event of an increase in size. By promise-keeping, the seller must therefore provide a compensation on the price, equal to the product of [D] and [E], for each additional dollar raised in the contract, in order to entice the customer to remain matched even in the prospect of a size increase.

Intuition for the size dependence It remains to argue why x_n could be a decreasing function of n in the first place. To make the argument transparent, consider that marginal costs are constant in n (i.e. $\mathcal{C}(n) \propto n$), so that all size dependence emerges only from the search frictions.

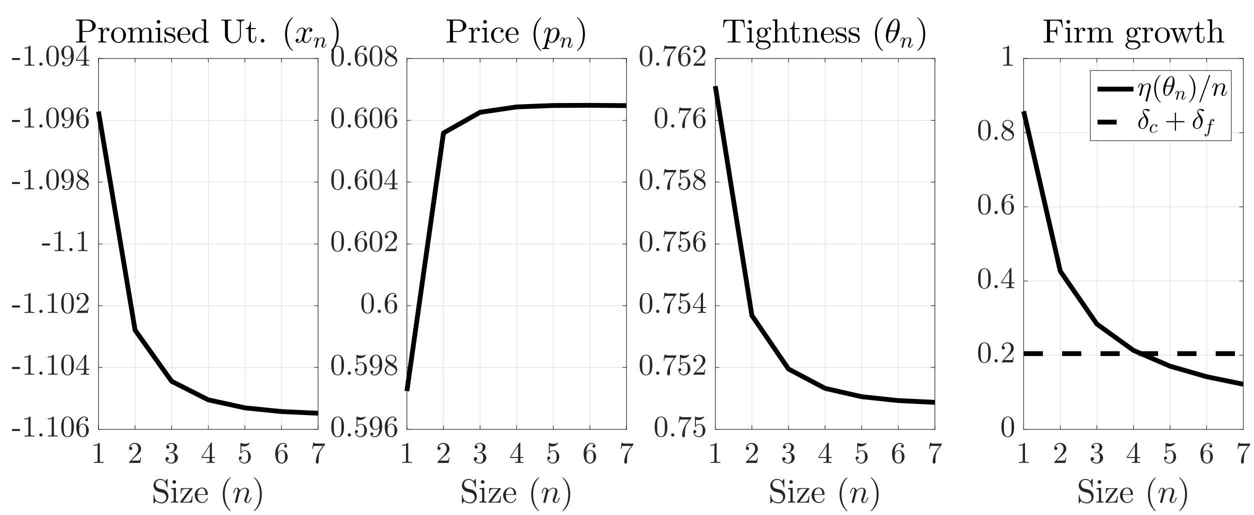

Figure 1: Numerical Example: Promised utility, price, market tightness, and firm growth, as a function of size, for the simple model with no exogenous (z, φ) shocks, and a constant marginal cost (i.e. $\mathcal{C}(n) \propto n$). Firm growth has been decomposed between the rate of customer attraction (solid line) and that of customer attrition (dashed line).

Figure 1 shows, using a numerical example, that prices can be strictly increasing in size even in an environment with linear costs. The intuition for this result is a combination of the entry cost for sellers and the search cost for buyers. When a seller has no customers, it has to pay the $\kappa > 0$ cost to enter. It does so by promising a certain utility x'_1 to its first customer. This offer must satisfy $x'_1 > U^B$, or else the customer would prefer to remain unmatched, and must ensure that, if a match occurs, the expected value at entry pays for the entry cost, or $\kappa = \eta(\theta(x'_1))V_1^S$. After the seller has acquired its first customer, the κ cost becomes sunk, and the seller must post an offer x'_2 to attract a second customer. Again, $x'_2 > U^B$ is needed to prevent separation. Moreover, in the event of a successful match, the surplus changes from $W_1 = V_1^S + x'_1$ to $W_2 = V_2^S + 2x'_2$. Since the number of customers has grown proportionally, but customers must still absorb part of the added value to agree to remain matched, the seller now reduces the offer to each customer slightly, $x'_2 < x'_1$, to be able to absorb part of the leftover gains. This intuition carries over for larger sizes: customers are enticed to remain matched, as they still receive compensation against the costly search state in the form of expected utilities, but the seller lowers the promised as it grows because the growth in the base is linear and the seller needs to raise resources quickly to overcome the high costs to market penetration.

How does this argument depend on the degree of frictions? An important parameter for the size dependence is the size of entry costs κ.[35] Figure 2 uses the same parameter values as Figure 1, but with a much lower value for κ. The path of prices is now different. First, the seller must still enter with a high promise to the first customer in order to generate a high enough probability of entering. But because entry into the market is now cheaper, the seller is not as concerned about raising resources quickly to make up for the costs of market penetration, and is now willing to back-load its promises and offer lower prices for larger sizes.

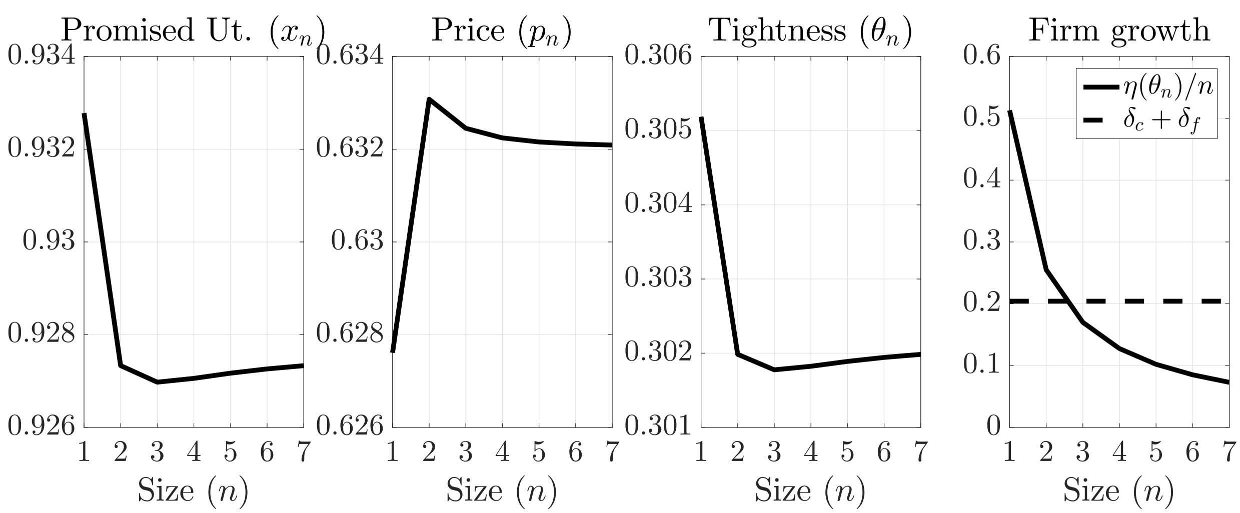

Figure 2: Same as Figure 1, but with a lower value for κ.

Therefore, importantly, the positive correlation between prices and sizes is not built into the model. Indeed, there is an active empirical debate in the literature about the direction of this correlation. Using plant-level data from the U.S. manufacturing sector, Foster *et al.* (2008, 2016) claim that prices are increasing in tenure in the market, while Berman *et al.* (2017) find, using customs data, that prices are slightly

[35] For comparative statics analysis with other parameters, see Section D.5 in the Appendix.

decreasing for export markets. In contrast, Fitzgerald *et al.* (2017) find no dynamics of prices after adjusting for selection. In light of this evidence, our model thus abstains from taking a stance ex-ante, and allows for both increasing and decreasing price paths. In the next section, however, we will document that prices increase in size in our sample, and that this effect is still significant after controlling for tenure in the product market.

4.2 Discussion of the Model's Main Assumptions

To complete our qualitative analysis, let us now discuss our main modeling assumptions. Our model is somewhat stylized and uses some restrictions, particularly on the contractual environment. Arguably, the three most relevant assumptions are: (i) active customers cannot search for other sellers without having to pay the search cost; (ii) sellers cannot discriminate across customers; and (iii) the seller credibly commits to the pricing plan. Let us discuss the role of each one of these in turn, and argue that the qualitative properties of the model do not hinge on these modeling choices.

1. Endogenous Separations An important assumption that has been made for tractability is that customers cannot bypass the costly inactive state when they separate (either voluntarily or due to a shock) from their seller. Allowing for endogenous seller-to-seller transitions would incorporate an additional dimension into the firm's pricing decisions. Besides the rent-extraction trade-off between incoming customers and the current base, the firm would now have solve an attraction-attrition trade-off: a more ex-post profitable contract for inactive buyers may enhance the chance of a customer match, but also increase the likelihood of a voluntary separation. We propose how to endogenize this margin in Section 7.1, and discuss the technical challenges it presents.

2. Price Discrimination Secondly, we have assumed that sellers cannot price discriminate across different customers. While this assumption is realistic for most major sectors of the economy, especially those in which sellers face a large number of potential buyers (such as retail, our application in Section 5.1), it may not be apt for certain others, for instance industries in which personalized buyer-seller relationships may explicitly develop (newspapers, cell phone and internet services, commercial banks, etc.). Gourio and Rudanko (2014b) propose a model for these type of relationships, show that sellers attract buyers by offering a price discount on their first-ever transaction, and study the implications of this pricing behavior for firm investment. Though the focus of our paper is different, it is still worth emphasizing that our environment does *not* nest the Gourio and Rudanko (2014b) result. In fact, allowing for price discrimination not only preserves the block-recursivity property, which is key for providing tractability, but it also preserves firm dynamics and price dispersion. Importantly, however, assuming discriminatory contracting results into a new feature (always within the Markov-perfect equilibrium class): equilibrium multiplicity in the form of price indeterminacy. We explain these results in detail in Section 7.2.

3. No Commitment The third main contractual assumption is that of perfect commitment on the seller's side. Intuitively, long-term contracts are a stand-in for a reputational concern on the side of the firm. By promising to deliver a utility level, the seller can balance the price with the continuation value to lure customers into remaining matched. In turn, customers understand that the firm does not price gouge, and they remain loyal to their seller in order to avoid going through the costly inactive state. As discussed briefly in Section 3.2, the market tightness schedule $\theta : \mathcal{X} \times \Phi \to \mathbb{R}$ is taken as given by all agents, which sets rationality in beliefs in the sense that if a seller were to deviate from its pricing plan, all inactive buyers would remain indifferent between the new off-equilibrium offer and the remaining ones being offered on-path. Importantly, if we were to dispense of the commitment assumption on the seller side, we would lose block recursivity and, thereby, all the attractive analytical features of the equilibrium. The reason for this is that, due to a time-inconsistency problem, firms would engage in potentially multiple forms of pricing strategies, all of which could be sustained in equilibrium under appropriately designed "implicit contracts", paired with trigger strategies on the buyer side (see Nakamura and Steinsson (2011) for a discussion). The implicit contracts literature (going back to Baily (1974) and Azariadis (1975)) shows that these type of contracts exhibit history-dependence, which in our framework would break our recursive formulation. Moreover, sellers would need to keep track of the distribution of buyers in order to understand how to best lock-in buyers across markets, as ex-ante revenue equalization would fail to hold. For these reasons, seller's commitment is a *sine qua non* for our set-up.

5 Quantitative Analysis

Let us now turn to the quantitative part of the paper. First, we provide empirical evidence from the U.S. retail sector and analyze correlations between seller sizes and prices. Then, we proceed to the computational implementation of the equilibrium, present the calibration exercise, and study the aggregate implications of the model in order to illustrate the effects of customer capital accumulation on the micro- and macro-economic effects of aggregate shocks.

5.1 Data

Estimating our model requires the use of disaggregated data that allows us to observe prices and quantities separately. Fortunately, these type of data have become increasingly available over the last few years as macroeconomists have drawn renewed attention to the analysis of the micro structure of markets. In this section, we use micro-pricing data on the U.S. retail sector. In this interpretation of the theory, sellers are stores, and buyers are private consumers. We then argue that small stores in our sample (i) experience higher growth rates, and (ii) set lower prices relative to other firms within the same product market. We also show that this size effect is not driven by a store age effect, meaning that our customer accumulation interpretation of seller dynamics is compatible with competing theories based on store age that would deliver similar patterns. Additionally, this section will serve to document some salient features of the distributions of relative prices and sales, which we will use in the next section to calibrate our model.

Although the model is general and can be applied to different sectors of the economy, we view the retail sector as fitting well with its basic features. We will use data of unique high granularity which will allow us to focus on narrowly-defined homogenous products that are sold by sellers of different sizes within the same market segment, in accord with the environment of the model. Moreover, the types of goods in the data are non-durable consumption products that, as in the model, are likely to engage customers and sellers into repeated purchases and thereby lock them into lasting relationships. The fact that retail stores face a potentially large number of customers implies that the customer anonymity assumption likely provides a good approximation for the bulk of observed store transactions. Finally, under this framing of the model, we interpret the buyer's search cost as a proxy for the transport and/or information costs associated to finding and/or switching stores for customers.

We use weekly micro-level data from the IRI Symphony scanner data set.[36] The whole data set is large, spanning a period of 12 years (from the first week of January 2001 to the last week of December 2012), and containing revenue and quantity information for over 5,000 retail (drug and grocery) stores over 50 Metropolitan Statistical Areas (MSA) in the U.S. The data are automatically generated by retailers themselves through their point-of-sale systems, so a caveat of the dataset is that we do not observe overall consumer expenditures. Products are grouped into 31 broad categories, with each product defined at the Universal Product Code (UPC) level.[37] Because of the large amount of information, we focus our attention only on two large geographical markets (New York and Los Angeles) in the period 2001-2007, and consider 15 of the 31 product categories.[38]

Although the IRI does not explicitly report prices for individual transacted products, the weekly average price can be backed out by taking the ratio of the value of sales to the number of units sold for each item of the store. That is, we define:

$$P_{usm,t} = \frac{TR_{usm,t}}{Q_{usm,t}}$$

as the average retail price of UPC u within week t, in store s and (geographic) market m, where TR denotes the total dollar value of revenues from sales, and Q denotes the units of the product that are sold. Throughout, we consider only transactions at stores with unique identifiers at each UPC $\times$ market $\times$ week cluster. We also restrict our sample to only those products that are commonly available across stores and not only sold in specific establishments. Specifically, given the overall number of stores in our sample, we choose to drop those goods that are sold in less than 10 stores in every given week and market.[39] Finally, in the absence of a theory of price discounts, we focus only on *regular* prices by filtering out of the sample those products that are on sale. A convenient feature of the data is that products are flagged whenever they go on promotion, which means that we need not employ a filtering algorithm as in Nakamura and Steinsson (2008) but we can rather exclude flagged products directly.[40] Table A.1 in

[36] The data are available for request at `https://www.iriworldwide.com/en-US/solutions/Academic-Data-Set`. For documentation, see Bronnenberg *et al.* (2008). Recent studies in economics studying related issues that use the IRI Symphony include Alvarez *et al.* (2014), Gagnon and López-Salido (2014), and Coibion *et al.* (2015).

[37] The UPC is an array of numerical digits that is uniquely assigned to a given item, and it constitutes the highest level of disaggregation available for a product. The description of products is very detailed, including information about the brand, flavor, and several packaging attributes.

[38] The 15 categories of consideration are: Beer, Blades, Carbonated Beverages, Cigarettes, Coffee, Cold Cereal, Deodorant, Diapers, Frozen Pizza, Frozen Dinners, Household Cleaners, Hotdogs, Laundry Detergent, Margarine and Butter, and Mayonnaise.

[39] To further eliminate outliers, we also drop stores with non-positive sales, transactions with prices above \$100 (which approximately account for the top .02% of the price distribution in the full sample), and cases with multiple observations at the store$\times$market$\times$week$\times$UPC level, which we deem as mis-reported transactions.

[40] A "promotion" is defined by the IRI as a temporary price reduction of 5% or greater. Sales are quite unresponsive to the

Appendix A shows some descriptive statistics of our data before and after applying these restrictions.

To study the degree to which the same good is sold at different prices by stores of different size profiles, we follow the literature (e.g. Kaplan *et al.* (2016)) by focusing on a *relative* measure of prices. That is, we define:

$$\widehat{p}_{usm,t} = \log P_{usm,t} - \frac{1}{N^S_{um,t}} \sum_{s=1}^{N^S_{um,t}} \log P_{usm,t} \tag{25}$$

where $N^S_{um,t}$ is the number of stores selling good u in market m and week t. In words, $\widehat{p}_{usm,t}$ indicates the log-deviation in the price of good u in store s relative to the average price across all stores selling that good in the week and market of interest. Price dispersion is measured as the average standard deviation of $\widehat{p}_{usm,t}$ across stores, markets, and time. In our full sample, dispersion at the barcode level is high (15.73%), in line with previous studies using similar micro pricing data from different sources (e.g. Kaplan and Menzio (2015)). The restricted sample has a lower dispersion (10.55%), as a result of having eliminated price outliers and uncommon goods. Figure A.3 shows the distribution of relative prices in our sample, alongside that of normalized sales (the ratio of store-level sales to its mean) and store sales growth rates. Table A.2 presents summary statistics for these distributions. We observe that the store size distribution has a fat right tail, which accounts for the high dispersion in normalized sales.

For a measure of the store's average relative expensiveness, we average relative prices across products sold within the store:

$$\widehat{e}_{sm,t} = \frac{1}{N^U_{sm,t}} \sum_{u=1}^{N^U_{sm,t}} \widehat{p}_{usm,t} \tag{26}$$

where $N^U_{sm,t}$ is the number of goods that are sold by store s in the corresponding week and market of interest. The variable $\widehat{e}_{sm,t}$ describes how prices in the store compare, on average, relative to those of the seller's competitors.

We can now examine how stores of different sizes price their products differently. Our baseline measure of size is store-level sales (in logs), which we compute as the sum of dollar revenues across all the products sold by a given store, in a given week and market. We are then interested in describing the effects of store size on relative prices and sales growth rates. Formally, we adopt the following empirical specification:

$$y_{usm,t} = \alpha_0 + \mathbf{X}^{\top}_{sm,t}\boldsymbol{\beta} + \gamma_a + \gamma_s + \gamma_t + \gamma_m + \varepsilon_{usm,t} \tag{27}$$

for store s in week t and (depending on the specification) product u, where $y_{usm,t}$ stands for the dependent variable of interest. To isolate all sources of aggregate time variation that may be common across stores, we control for time fixed effects γ_t. Further, we control for store characteristics by including store fixed effects, γ_s, and market-specific common across time and stores, γ_m. We will run regressions with and without store-age fixed effects γ_a, in order to verify that the size effects are not driven by age.[41]

business cycle, as documented by Coibion *et al.* (2015) for the IRI data, and therefore excluding do not change our life-cycle results significantly.

[41] The IRI reports (masked) retailer identifiers, which we exploit to track the tenure spell of different stores within the

Covariates in $\mathbf{X}_{sm,t}$ include our measure of size at both store and product levels.

We run three main specifications for our dependent variable: the growth rate of product sales within the store, the relative price at the UPC level, and the store's relative expensiveness. Table 1 presents results for the corresponding coefficients on store size.

	(1a) $\Delta\log(Sales_{us})$	(1b) $\Delta\log(Sales_{us})$	(2a) $\widehat{p}_{us}$	(2b) $\widehat{p}_{us}$	(3a) $\widehat{e}_s$	(3b) $\widehat{e}_s$
$\log(Sales_s)$	–.05674***	–.05545 ***	.01805***	.01588***	.01667***	.01449***
	(.00075)	(.00077)	(.00023)	(.00023)	(.00008)	(.00008)
Age FE	✗	✓	✗	✓	✗	✓
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
R^2	.1371	.1373	.1541	.1550	.8862	.8917
Obs. (millions)	43.4	43.4	59.81	59.81	59.81	59.81

Table 1: Coefficient values for size regressions, with size proxied by sales. Source: IRI Symphony weekly data, 2001–2007. Notes: Subscript s denotes store, us denotes UPC within store. All regressions are controlled by log-sales at the store-product level. *Fixed Effects* include store, market, and time fixed effects. Standard errors in parentheses, clustered at the Store×UPC level. Significance levels: *=10%; **=5%; ***=1%.

The results show that there exist significant differences in growth rates and prices across comparable stores of different sizes. Columns (1a), (2a), and (3a) show the pure size effect, without controlling for store age. Columns (1b), (2b), and (3b) show results for the same specifications in which age fixed effects have been included, to demonstrate that the size effect remains significant. The latter implies that our customer accumulation interpretation of store dynamics does not conflict with other views on pricing behavior along the seller's life cycle (e.g. learning about demand through price experimentation).

Columns (1a)-(1b) are our baseline store growth regressions, showing that larger firms (in terms of total weekly revenue) experience lower rates of per-product sales growth. In particular, a 1 percentage-point increase in store sales slows down sales growth by 0.055 percentage points on average. In the model, this effect comes from the pricing behavior of smaller sellers, who set relatively lower prices and attract customers at a faster pace. Empirically, we also find that relative prices are positively correlated with store size. In this case, an increase in store sales by 1% is associated with a per-product price that is, on average, about 0.016 percentage points higher than the average price for that product across all stores selling in the week and market of consideration (column (2b)). Finally, columns (3a)-(3b) show the size effect on relative prices on average across all products sold within the store. We see that an increase in 1 percentage point in store-level sales is associated with an increase in the store's expensiveness index of about 0.015 points.[42]

Though we do not claim that these results unveil a causal relationship between prices, sales, and incentives to accumulate customers, we use these correlations to discipline the type of relationship that the estimated version of the model should feature between firm size and prices.

market×week observation of interest. For a given market m and week t, the age of store s is the number of weeks $a_{sm,t} = 1, 2, \ldots, t$ elapsed since the store's first entry into the market.

[42] A potential concern in these regressions is that large and small chains behave differently in terms of pricing, a dimension that our model does not capture. To avoid comparing very large chains to medium and small ones, we run our regressions for only those chains whose number of stores lies below the median number of stores per chain in the sample. Table A.3 shows that the results are qualitatively similar.

5.2 Estimation

Parametrization

Let us proceed with the estimation of the model. The first step is to parametrize the cost function of firms and establish the structure of the exogenous shocks. For the former, we choose a convex function that is separable in firm size:

$$\mathcal{C}(n; z, \varphi) = w(z, \varphi) \cdot n^{\psi} \tag{28}$$

where $w(z, \varphi) > 0$ is a scale parameter, and $\psi \geq 1$ is a curvature parameter controlling for the degree of returns to scale in technology. When marginal costs are increasing in size ($\psi > 1$), there is a natural upper bound on firm size for each state, given by $n^*(z, \varphi) \equiv \left(\psi w(z, \varphi)/v\right)^{\frac{1}{1-\psi}}$, beyond which the static flow surplus $\pi_n(z, \varphi) = nv - \mathcal{C}(n; z, \varphi)$ is strictly decreasing and the seller does not want to grow further.[43] The scale parameter changes across firms and aggregate states, and for simplicity we assume a linear form in both arguments so that $w(z_i, \varphi_j) = w z_i \varphi_j$, for $i = 1, \ldots, k_z$ and $j = 1, \ldots, k_{\varphi}$, where $w > 0$ controls the optimal scale. This specification of shocks is isomorphic to multiplicatively-related idiosyncratic and aggregate TFP shocks in the production function, a standard approach in the search-and-matching firm-dynamics literature (e.g. Kaas and Kircher (2015)).

On the other hand, we must specify the structure of the exogenous shocks, z and φ. As these variables evolve over a discrete grid in the model, in principle we should estimate the value of all the transition rates in the underlying generator matrices. For each shock $s \in \{z, \varphi\}$, this would require the estimation of $k_s(k_s - 1)$ additional parameters, a potentially large number. To reduce the parameter space, in practice we assume that the exogenous shocks follow continuous-time analogues of AR(1) (so-called Ornstein-Uhlenbeck) processes, which we in turn approximate on finite grids using the Tauchen (1986) method.[44] This reduces the estimation of the shocks to only two parameters: a persistence parameter ρ, and a volatility parameter σ.

Calibration Strategy

Our calibration strategy is to match aggregate moments related to store dynamics in the U.S. retail sector as well as average long-run moments across all years of our sample of micro-pricing data presented above.

The model is quite parsimonious, with 11 free parameters that need to be identified. Of these, 9 are deep parameters: $\left(v, r, \delta_f, \delta_c, w, \psi, \kappa, \gamma, c\right)$, corresponding to the value of consumption, the time discount rate, the separation rates of firms and consumers, the scale and curvature parameters of the operating cost function, the entry cost for new sellers, the matching elasticity, and the search cost for inactive buyers, respectively. On top of this, we must set values for the persistence and dispersion parameters of the exogenous productivity state process: (ρ_z, σ_z). We assume that z can take up to $k_z = 25$ different values. We do not estimate the aggregate shocks φ because the spirit of our calibration exercise is to estimate an

[43] Even though, as we saw in Section 4, the existence of a stationary size does not hinge on the curvature in the cost function, the parameter ψ will help us pin down the size dependence in prices more easily.
[44] Full details can be found in Section C.2 of the numerical appendix.

economy in its long-run equilibrium. These shocks will be re-introduced in Section 6, where we discuss our application to markup trends within the context of the calibrated economy.

External identification The parameters (v, r, δ_c) are calibrated outside the model. The value of consumption is normalized to $v = 1$, so that the consumption good serves as the numeraire of the economy. The discount rate is set to $r = 0.05$, corresponding to a discount factor of approximately 95% annually. Finally, the exogenous separation probability is set to match a 0.044% weekly customer turnover rate (corresponding to $\delta_c \approx 0.2041$ at our yearly frequency), which implies that customer relationships last a bit more than 4 and a half years on average. We take this value from Paciello *et al.* (2016), who estimated it using the same database that we have used in our empirical analysis. The number falls within the range of values reported by Gourio and Rudanko (2014b), who survey the marketing literature on customer relationships for both contractual and non-contractual settings and find that turnover rates range between 10% and 25% annually, depending on the sector.

Internal identification We are left with the following free parameters: (i) the firm exit rate δ_f; (ii) the cost scale w; (iii) the firm entry cost κ; (iv) the buyer search cost c; (v) the matching elasticity parameter γ; (vi) the cost curvature parameter ψ; and (vii) the autocorrelation and dispersion parameters (ρ_z, σ_z). Because of the high non-linearity of the model, identifying each parameter separately is hard in our environment, though we can provide some intuition for how each parameter is informative of each specific moment. Methodologically, we estimate the parameters jointly by matching a combination of aggregate and firm-level long-run moments via Simulated Method of Moments (SMM). To implement this procedure, we use an algorithm that randomly searches in the parameter space, and then employ an unweighted minimum-distance criterion function that compares empirical moments to model-implied moments from both the stationary solution and from simulated data.

For the stationary solution, we solve a nested fixed-point algorithm that uses a bisective step to solve for the value of inactivity, U^B. Appendix C.1 outlines the details of this method. To obtain moments from simulated data, we compute paths for many distinct firms over $T = 100$ years of data which we discretize with time steps of equidistant length $\Delta = 0.01$ each.[45] All firms are drawn from the stationary distribution $\{g_n(z) : (n, z) \in \mathbb{N} \times \mathcal{Z}\}$ at time $t = 0$ and evolve endogenously through simulated Markov chains that replicate the state dynamics described in Section 3.5. To allow for convergence of the distribution, we drop the first half of the time sample when computing the average simulated moments. For the productivity distribution π_z, from which entrants draw their initial productivity level, we use the ergodic distribution implied by the calibrated Markov chain for z.

The set of targeted moments can be grouped into two broad categories: (i) aggregate moments, and (ii) firm-level moments related to the long-run distribution of sales and prices. At the aggregate level, we target an average annual entry rate of 8.9%, which we compute for our IRI sample as the average across years 2001-2007 of the ratio of stores aged 52 weeks or less to the total number of existing stores within that year (see Table A.4 in the Appendix). We define the entry rate in the model as the ratio of *actual* entrants to the total mass of incumbents. The exit rate is the measure of firms who either die or lose their

[45] Our baseline results report simulations for 1,000 firms. However, we have tried with up to 10,000 firms and obtained quantitatively similar results.

last remaining customer. By equation (21), this means:

$$EntryRate = \frac{S_0}{\sum_{n,z} S_n(z)} \sum_{z_0 \in \mathcal{Z}} \pi_z(z_0)\eta\big(\theta_1(z_0)\big) \quad \text{and} \quad ExitRate = \delta_f + \delta_c \frac{\sum_z S_1(z)}{\sum_{n,z} S_n(z)} \tag{29}$$

These formulas hold in and out of steady state, but are equal to each other in the absence of φ shocks, so the entry rate in the data helps us identify the exogenous exit rate δ_f in the model. At the aggregate level, we also target the cross-sectional average markup. Because measuring markups in the data usually requires a stand on market structure and the demand curve faced by firms, estimates vary substantially in the literature depending on the empirical methodology, the industry of consideration, and the overall sample. Using firm-level data, typical estimates range from about 10% to as much as 50% or more.[46] Since we are calibrating our model to long-run average moments of the U.S. retail sector, we choose to target a markup of 39%, a number that we impute from the average ratio of gross margin to sales in the retail sector for our sample period (2001-2007). We obtain this number from the latest Annual Retail Trade Report of the U.S. Census Bureau.[47] To be consistent with the empirical target, in the model we compute *measured* markups as the sales-weighted average of the ratio of price to marginal cost across firms:

$$\overline{m} = \sum_{n \in \mathbb{N}} \sum_{z \in \mathcal{Z}} m_n(z); \qquad \text{with } m_n(z) := s_n(z) \frac{p_n(z)}{mc_n(z)} \tag{30}$$

where $s_n(z) = \frac{np_n(z)}{\sum_{n,z} np_n(z)}$ is the sales share of type (n, z) firms, and $mc_n(z) := \mathcal{C}(n; z) - \mathcal{C}(n-1; z)$ is the marginal cost of this type of firms. Though many parameters affect the average markup, γ is the most relevant one, as it governs how the gains from trade are shared between the customers and their seller (recall our discussion in Section 4).

At the store level, we target several moments of the long-run distribution of prices and sales that we compute from our IRI sample. The cost parameters (ψ, w) determine firm profitability across sizes, so they play an important role in determining the degree by which firms of similar productivity choose to set different prices for the same product. We choose to target two moments that relate to this dimension of heterogeneity. First, we target the time-series average standard deviation of relative prices (equation (25)), our baseline measure of price dispersion, equal to 10.55% in the data. Second, we target the inter-decile range in the distribution of relative prices between the tenth percentile and the median relative price, equal to 1.1215 (see Table A.2). The reason we target this measure of left-tail dispersion is because we estimate the model so the bulk of the population of firms charges low prices relative to the average price (resulting from right-skewness in the stationary size distribution). Matching the lower part of the price distribution is therefore important because our ultimate goal is to understand the macroeconomic

[46] Using a model-free approach and manufacturing data at the plant level from Slovenia, DeLoecker and Warzynski (2012) find that median markups range from 10% to 28%, depending on the specification, while DeLoecker *et al.* (2016) find even more variation, from 15% up to 43%, using similar methods for India. In a study spanning 1981-2004, Christopoulou and Vermeulen (2008) report an average markup of 37% for the Euro area, and of 32% for the U.S., and find that these estimates vary substantially at the sectoral level.

[47] The data are freely available at https://www.census.gov/retail/index.html. The average gross margin is about 28%, implying an average markup of $.28/(1-.28) \approx .39$. For comparison, Hottman (2017) estimates average markups in the U.S. retail sector and finds slightly lower numbers, in the range 29-33%.

implications of pricing when firms accumulate demand.

Next, we need to discipline the parameters of the exogenous productivity process, z. Having matched price dispersion measures, we are now interested in variation across productivity levels for fixed seller size. Thus, we target the yearly autocorrelation in normalized store-level sales (pinning down the persistence ρ_z), and the dispersion in the distribution of normalized sales (pinning down the volatility σ_z). Finally, we need to calibrate the search cost for buyers, c, and the market penetration cost for sellers, κ. As we discussed in Section 4, these parameters are important to pin down the dependence between firm size and firm price, which determines two key aspects of firm dynamics: (i) the growth rate of firms across sizes; and (ii) firms' stationary size. For the former, we target the correlation between store product-level sales growth rates and the relative price of those products. The correlation is negative, consistent with the idea that smaller sellers grow faster and set lower prices relative to their competitors within the same product market. Regarding (ii), we target the stationary size of sellers in the data. In the model, we measure the average size of firms as the mean number of units sold per firm. Since each customer consumes one unit, the average size is (see e.g. Luttmer (2006)):

$$\overline{L} = \left(\sum_{n \in \mathbb{N}} \sum_{z \in \mathcal{Z}} \frac{1}{n} L_n(z) \right)^{-1}; \qquad \text{with } L_n(z) := \frac{n S_n(z)}{\sum_{n,z} n S_n(z)} \tag{31}$$

where $L_n(z)$ is the fraction of active buyers that are customers of sellers of type (n, z). In our sample, the average number of units sold of each product within a store is 12.4 in volume-equivalent terms,[48] so we target this number in the estimation to make average firm size comparable between data and model.

Estimation Results

The full set of calibrated parameter values is presented in Table 2, and the result of the calibration exercise in terms of moment-matching is presented in Table 3. The model's fit is reasonably good, being able to explain both aggregate entry rates and average markups accurately, as well as both average and left-tail dispersion in relative prices. Note that the model slightly under-predicts dispersion in normalized sales, probably as the result of outliers in the data. On the other hand, the correlation between sales growth and relative prices in the model is a little too strong relative to its empirical counterpart. This likely reflects the presence of factors attenuating the relationship between prices and sales in the data that cannot be captured by the model.[49]

Figure A.5 plots the joint surplus, the pricing policy function, the measured markups (computed using the sales-weighted measure of equation (30)) and the promised utility, in the space of seller sizes (n) and productivities (z), for the calibrated set of parameters. Recalling that higher values of z correspond to higher average costs (recall equation (28)), we find that matches with more customers and higher productivity levels (i.e. lower values for z) earn a larger surplus. Moreover, we find that the pricing policy is increasing in the size of the customer base, and decreasing in productivity. Even though marginal costs

[48] The IRI sample provides a conversion system whereby units of different product categories can be made comparable. We use this standardization for this calculation.

[49] To get a visual idea of identification, Figure A.4 in the Appendix plots each calibrated moment against the distribution across different model simulation runs that results from our parameter search algorithm. We see that, with a few exceptions, the calibrated moment is close to the median of this distribution.

Parameter	Value	Description	Source / Target
		Calibrated externally	
v	1	Value of consumption	Normalization
r	0.05	Discount rate	5% annual risk-free rate
δ_c	0.2041	Separation rate	Paciello *et al.* (2016)
		Estimated internally	
δ_f	0.0738	Firm exit rate	Annual store entry rate
γ	0.5339	Matching elasticity	Average markup
ψ	1.4044	Cost curvature	Standard deviation of relative prices
w	0.1510	Cost scale	p50-p10 inter-decile range in relative prices
c	0.5457	Buyer search cost	Average store size
κ	1.6214	Firm entry cost	Correlation between sales growth and relative price
ρ_z	0.0751	Persistence of z	Autocorrelation in normalized store sales
σ_z	0.1034	Volatility of z	Dispersion in normalized store sales
Time frequency	Annual		

Table 2: Full set of calibrated parameters in the baseline estimation. Notes: The parameters (ρ_z, σ_z) correspond to the Euler-Maruyama equation (C.2.1) of the Ornstein-Uhlenbeck process for z. See Appendix C.2 for details.

are higher for larger firms (as $\psi > 1$), measured markups are still increasing in size, i.e. larger sellers introduce relatively higher margins over their unitary costs into their prices. In Figure A.6 we plot the distribution of normalized sales and that of the seller's customer base that result from the simulation of the economy under the calibrated set of parameters. The figure demonstrates that our model can generate an invariant firm size distribution with a fat right-tail in both firm revenues and firm output that resembles its empirical counterpart (see Figure A.3 in the Appendix). We also show (panel (c)) the age distribution, to demonstrate that most small sellers in the economy are young, as argued in the size-age regressions of the empirical analysis (Table 1).

Moment	Model	Data	Data Source
A. Aggregate moments			
Annual entry rate	0.087	0.089	IRI (Table A.4)
Average markup (2001-07)	1.388	1.383	U.S. Census
B. Store-level moments			
sd(Relative prices)	0.1072	0.1055	IRI (Table A.1)
p50-p10 IDR relative prices	1.1224	1.1215	IRI (Table A.1)
Average store size	10.73	12.44	IRI
corr(Sales growth, Relative price)	–0.023	–0.007	IRI
ac(Normalizes sales)	0.854	0.828	IRI
sd(Normalized sales)	0.6	0.474	IRI

Table 3: Targeted moments: model versus data. Notes: Average markup is weighted by sales shares. *IDR* means inter-decile range. *sd*, *corr*, and *ac* mean "standard deviation", "correlation", and "autocorrelation", respectively.

Validation

To validate the results of our calibration, we assess the model's performance on non-targeted moments. We look at two sets of moments. First, we check the model's performance on other measures of relative price dispersion, namely inter-decile ranges between the first and ninth deciles, and fifth and ninth deciles. The results are in Panel A of Table A.5. The model's predictions regarding price dispersion above and below the median are in line with the micro-pricing data.

We also look at the model's ability to generate quantitatively correct predictions for the behavior of price *changes*. Although our model does not provide a theory of price stickiness (prices are sticky because they are indexed to slow-moving states, not because of costly price adjustment), it is still worth asking how the model performs in terms of the frequency and the size of the price adjustments that we see in our IRI sample. For this exercise, in the model we compute micro-price statistics along the stationary solution using the theoretical results derived in Appendix D.3. In the data, we define the weekly frequency of price changes within a store and market of interest as the share of goods sold by that store in that week that experience a price change.[50] For the moments of the distribution of price changes, we look at the absolute value of log differences. Finally, we annualize frequencies and rates in the model and the data for the sake of comparison.

Panel B in Table A.5 reports the simulated moments and their empirical counterparts. The calibrated model does a good job in predicting the empirical frequency of price changes, even though these moments were not targeted. Therefore, the model also predicts relatively well the median price durations, though the average duration is not accurately predicted as the distribution of price durations in the model is not sufficiently skewed.[51] Finally, the model predicts several moments of the distribution of expected price changes, especially the average price change and the median. Moreover, the model can explain about one third of the dispersion in the size of price changes, even though it was not calibrated for this purpose.

5.3 The Response to Aggregate Shocks

In this section, we analyze the role of aggregate supply and demand shocks at both the macroeconomic level as well as in the cross-section of firms. The purpose of this exercise is to understand the role that customer pricing heterogeneity can have on macroeconomic transmission. In particular, we seek to identify how firms' incentives to accumulate customers can generate substantial amplification on macroeconomic aggregates through both level and distributional effects, as well as imperfect pass-through of shocks to prices.

[50] We focus only on *regular* price changes, which we define (following Coibion *et al.* (2015)) as changes in prices that are larger than 1% or \$0.01 in absolute value for products that are neither entering nor coming out of promotion, and whose initial price is less than, or equal to, \$5. For non-promotional goods with initial prices higher than \$5, this threshold is set at 0.5%. These criteria eliminate small price changes that may possibly be due to rounding or reporting errors. Moreover, in order to filter out temporary price reductions that may not have been captured by the sales flag provided by the IRI, we exclude price changes that return to their initial level within 3 weeks after the initial change.

[51] To transform frequency f to duration d, we use the formula $d = -\frac{1}{\log(1-f)}$. See details in Appendix D.3. For medians, we apply the formula directly on the median frequency to obtain the median duration. For means, we first use the formula to compute the implied duration for each store and price, and then take the mean.

Supply Shocks

Starting from the stationary equilibrium of the calibrated economy, we first study the response to a 1% transitory shock to the marginal cost, i.e. a positive φ-shock to $\mathcal{C}(n; z, \varphi)$ parametrized as in equation (28). As with idiosyncratic shocks z, we assume that the aggregate state φ follows a mean-reverting process in logs (details of this implementation can be found in Appendix C.2). The shock hits at time t_0, and the process mean-reverts without any further shocks for all $t > t_0$.[52] The shock is fully-anticipated in the sense that agents incorporate the expectation about changes in their value functions due to the aggregate shock (recall the HJB equations presented in Section 3).

Figure 3 presents the results. The response of the economy to the aggregate shock combines both level and compositional effects. First, due to an exogenous increase in the cost of serving each customer (panel (a)), the flow payoff in joint surplus falls on impact (panel (b)). To mitigate the effects on their own profits, sellers lower the continuation utility that they promise to deliver to each customer going forward. Panel (c) shows, in particular, that active buyers are hit harder than sellers by the productivity shock. As a result of a lower promised utility, firms attract less inactive buyers, as their ex-ante value from matching is now lower. Consequently, the average tightness in the market falls (panel (d)), and with it the expected probability of firm growth. Interestingly, prices increase in response to the shock, but the pass-through is incomplete (panel (e)). The increase in prices is due to the fact that, when faced with an adverse shock to their costs, sellers choose to re-balance their contracts by *front-loading payments* from their buyers. They implement this by choosing to exploit their customers more today (through a high price) and forego some market shares in the future (through lower promised utility x). This strategy also explains the incomplete pass-through: as the shock is smoothed out inter-temporally via these two contracting instruments, the price response is muted. Note, moreover, that this dynamic re-balancing of payments momentarily increases flow sales in spite of the decrease in the extensive margin of demand, though this increase is only temporary (panel (f), solid line). Flow profits decrease, in contrast, as the rise in sales is overwhelmed by the increase in costs (panel (f), dashed line).

To explain the behavior of measured markups (panel (l)), we must first understand the compositional effects of the shock. First, in response to the decrease in demand, the rate of inactive buyers gradually increases (panel (g)), so firms start to shrink on average (panel (h)).[53] Thus, the firm size distribution shifts to the left and, since the risk of exiting is higher for smaller firms, the aggregate exit rate goes up (panel (i)). Interestingly, and despite the fall in the average size, the entry rate goes up as well (panel (j)), even more so than the exit rate does, which in turn explains that the number of inactive sellers decreases in response to the shock (panel (k)). The reason why the entry rate spikes up is that tightness in the *entry market* (where $n = 1$) has increased, making the ex-post rate of acquiring the first customer higher. The entry market thus behaves differently to any market for incumbents (where $n \geq 2$), as the latter type of market sees its tightness decrease in response to the shock (recall our intuition in the above paragraph). The reason for this behavior is that, in order to enter into the economy, potential sellers must raise enough resources to pay for the fixed market penetration cost, κ. While these costs have remained

[52] Throughout this section, the aggregate state process is implemented with $k_\varphi = 25$ grid points. To obtain smooth responses in the value and policy functions, we use interpolation with cubic splines.

[53] Note that these effects are purely compositional: they unfold with time, and there is no change on impact.

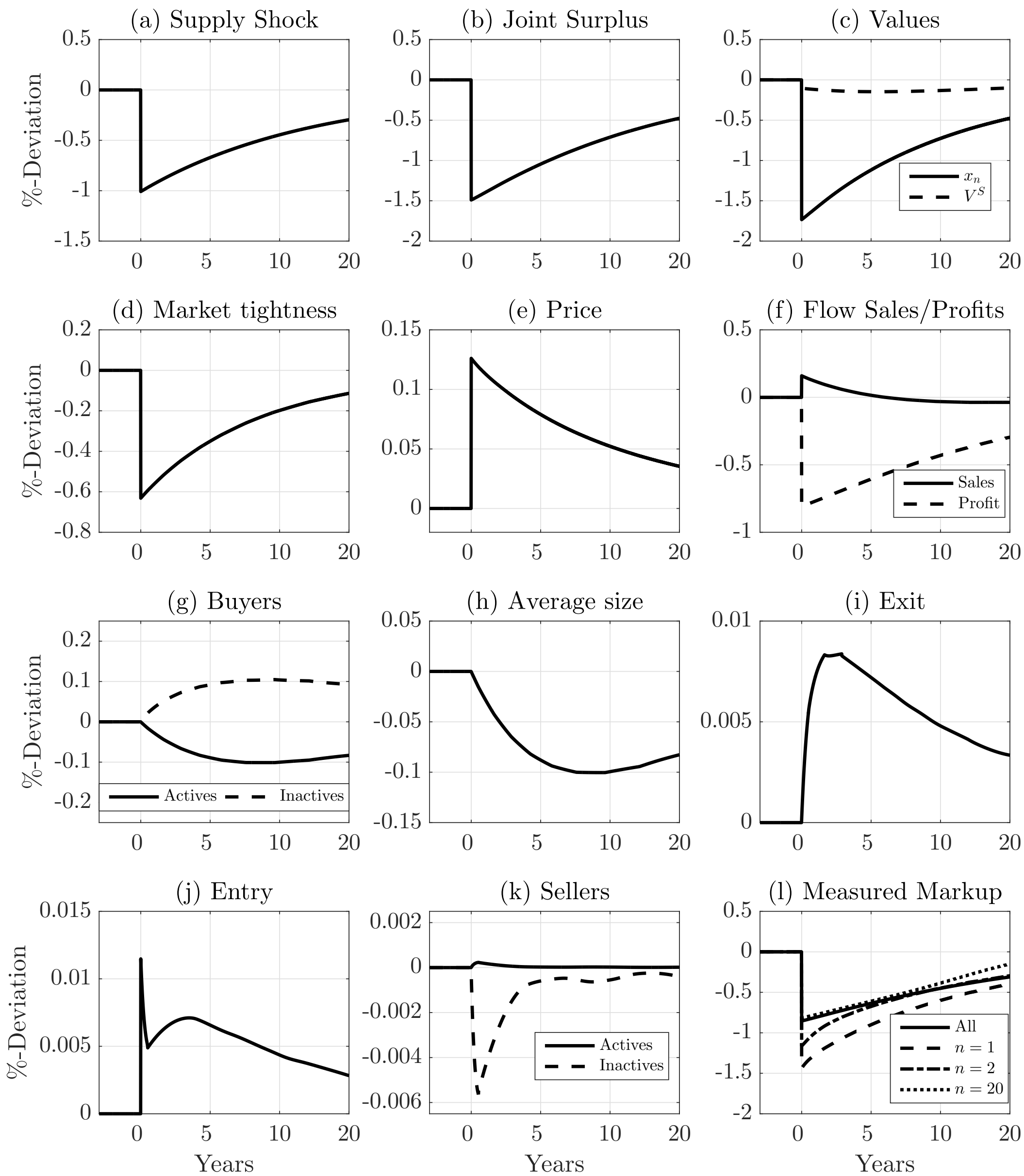

Figure 3: Impulse responses of selected variables to a negative and temporary 1% shock to aggregate productivity (i.e. an increase in the marginal cost). Notes: All responses expressed in %-deviations from steady-state. The shock hits at date $t_0 = 0$. Responses are smoothed out with the use of cubic splines. Panel (a) depicts the path of the exogenous state. Panels (b) to (f) depict cross-sectional averages using the simulated distribution of firms over idiosyncratic states. That is, for any policy or value function $f(n, z)$, we plot the %-deviation of $N_t^{-1} \sum_{n,z} f(n, z) m_t(n, z)$, where $m_t(n, z)$ is the count of firms of type (n, z) at time t, and $N_t := \sum_{n,z} m_t(n, z)$ is the total count of incumbent firms. The average number of customers per firm in panel (h) is computed using equation (31). Panels (i) and (j) are computed using equation (29). Panel (l) is computed using equation (30).

constant, every inactive buyer's ex-ante match value has worsen, so the seller must now raise the initial promised compensation (x_1) sufficiently in order to guarantee that the same entry costs are still being recouped in expectation.[54]

Overall, the response in the measured markup (computed as a sales-weighted average, from equation (30)) is explained mostly through shifts in the distribution of firms. On the one hand, the price level increases for all firms through the rebalancing mechanism explained above, which puts *upward* pressure on markups. On the other hand, because measured markups and size are positively correlated in the calibrated economy (recall Figure A.5(c)), the increase in the relative measure of small firms means that the contribution of low-markup firms to the aggregate markup is now relatively more important. This puts *downward* pressure on markups. Therefore, the cyclicality of measured markups is, in principle, ambiguous. In our calibrated economy, however, the latter compositional effect dominates the former level effect, and markups on average decrease. Interestingly, this has also implications for the cross-sectional response. Panel (l) shows the markup response for three different firm sizes, and shows that there exists substantial heterogeneity between the very small firms and larger firms. In particular, the smallest firms ($n = 1$) respond to the shock by almost twice as much as larger firms ($n = 20$) do. Similar features have been documented in the data. For instance, Hong (2017) has found differential responses of markups across firm sizes, with smaller firms displaying more elastic responses to output shocks, as in our model.

To illustrate the idea that sellers smooth out the effects of the adverse shock inter-temporally by depressing future demand, Figure 4 shows how the response varies with the average duration of customer relationships, as measured by δ_c. In particular, we compare the baseline economy (solid), with an economy in which the duration of customer relationships is one-third shorter (dashed). In line with our intuition, the figure shows that the response is dampened when customer relationships are shorter (that is, when the customer separation rate δ_c is higher). This is because, when sellers expect their customers to remain captive for a shorter time, sellers care more about their present profits, so promised utility is less depressed in the future (panel (a)) and prices react more today (panel (c)). As a result, the shock has a smaller impact on sellers' value (panel (b)). Moreover, the effects of the shock on prices and continuation utilities become less persistent. Finally, the fact that the price passthrough becomes less incomplete as δ_c increases means that the absolute response of the average markup (panel (d)) is weaker. In the limit as δ_c gets very large, markups would be acyclical to marginal cost shocks, as prices would respond one-for-one and promised utilities would remain unaffected by the shock.

Demand Shocks

Recent research has emphasized the relevance that consumer shopping behavior may have on macroeconomic dynamics. For instance, Bai *et al.* (2012) incorporate a frictional goods market into a representative-agent neoclassical economy to study the role of demand shocks, and observe that the latter are akin to productivity shocks because an increase in demand induces search and hence an output boom. Petrosky-Nadeau and Wasmer (2015) further show that goods market frictions can provide additional persistence

[54] Note that this feature emerges because firms are forced to enter with one buyer in our setting. In a model with free entry into *every* product market, the entry rate would likely move in the direction of the shock.

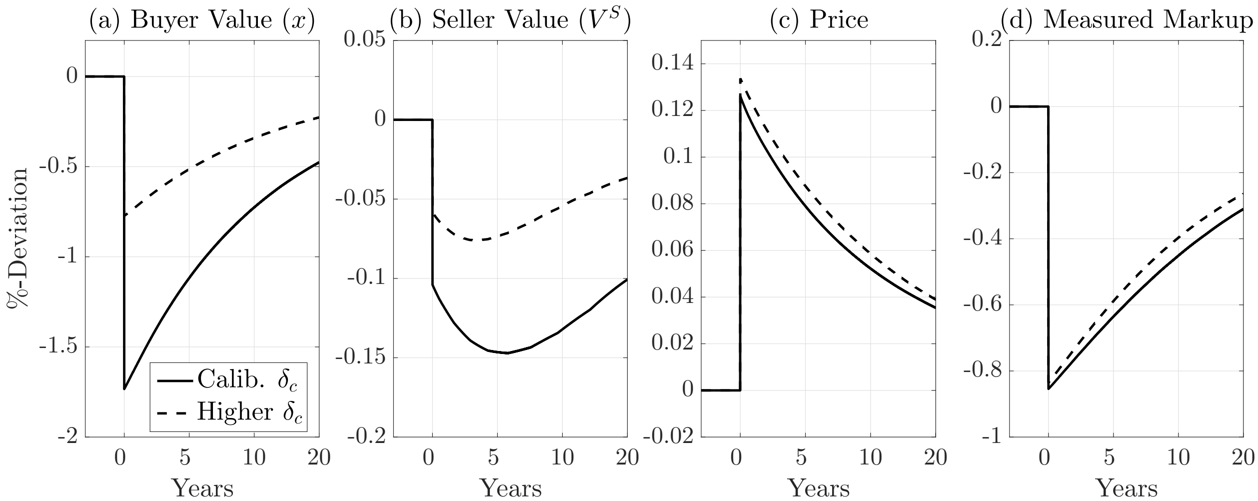

Figure 4: Impulse responses of selected variables to a negative and temporary 1% shock to aggregate productivity (i.e. an increase in the marginal cost), for different values of δ_c. Notes: See Figure 3.

to aggregate shocks, and Paciello *et al.* (2016) show that they can provide additional amplification. These are because an increase in search effort raises demand elasticity and leads firms to charge lower prices. In this section, we argue that the underlying size dynamics and the forces of firm entry can provide additional insights into the aggregate response of the economy to aggregate demand shocks.

We consider a shock to the instantaneous utility from shopping, v. In particular, we implement a 1% negative shock to v at time t_0, and let the process mean-revert without any further shocks for all $t > t_0$. We choose an autocorrelation for the φ process implying a half life of about three years, following Paciello *et al.* (2016) and in line with estimates by Bai *et al.* (2012).[55]

Figure 5 presents the results. A negative shock to the utility from consumption leads to a decrease in the number of buyers looking for a seller, since consumption is worth less. Because the buyers' outside option has relatively improved, firms lower the promised utility in an attempt to smooth the effects of the shock. Once again, the burden of the shock is passed almost entirely to the customer: the seller's value decreases only slightly (panel (c), dashed line), and it is the decrease in the value of the buyer (panel (c), solid line) which accounts for the bulk of the drop in joint surplus (panel (b)). As a result of the decrease in demand, market tightness drop on impact (panel (d)), and the decrease in the matching rate leads firms to progressively shrink in size.

In this respect, the demand shock is akin to a productivity shock (Figure 3), in line with the intuition in Bai *et al.* (2012). In particular, the compositional effects are similar, with a left-ward shift in the firm size distribution accounting for the increase in the exit rate and in the relative contribution of high-markup firms to the aggregate markup. However, note that the behavior of prices in response to the demand shock is remarkably different. First, the incomplete pass-through that we observed in the case of a supply shock, and which was due to firms optimally tilting their pricing contract toward more immediate payoffs through higher prices today, is no longer present here. A shock to the marginal propensity to consume has a one-to-one impact on the extensive margin of demand because of linearity in consumers'

[55] In particular, Paciello *et al.* (2016) use a quarterly autocorrelation of 0.98 for their demand shock. Using the Euler-Maruyama method described in equation (C.2.1) (Appendix C.2), in our calibration at a yearly frequency where each year is discretized by $1/\Delta = 100$ sub-periods, a 0.95 quarterly autocorrelation means $\rho_\varphi = \frac{1-0.98^{4\Delta}}{\Delta} \approx 0.0808$.

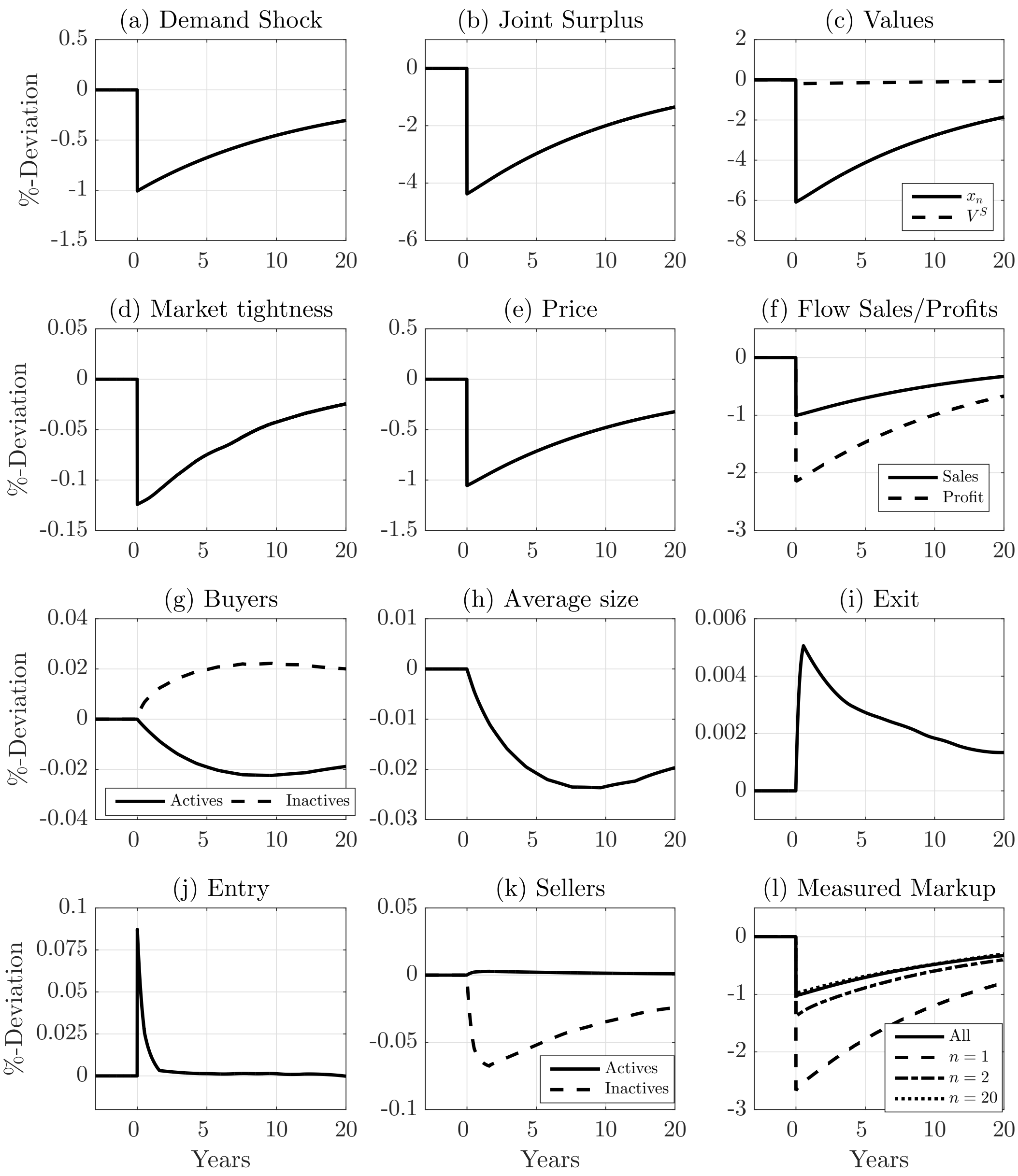

Figure 5: Impulse responses of selected variables to a negative and temporary 1% shock to aggregate demand (i.e. a decrease in the utility of consumption v), expressed in %-deviations from steady-state. Notes: See Figure 3.

preferences (note p_n is linear in v in equation (15)). This means that all the adjustment has to be made along promised utilities, which respond a lot more to the shock compared to the case of supply shocks.

The level effects are, once again, accompanied by interesting compositional effects. As a result of the demand shock, the relative attractiveness of small firms improves, as markups decrease relatively more for these firms (see panel (l)). This induces a short-lived spike in the entry of small firms, which puts further downward pressure on prices through an increase in competition. At the same time, the entry of (small) sellers causes a surge in the contribution of these firms to the aggregate price level. On the other end of the distribution, in contrast, the shock decreases the relative mass of large firms, and the contribution by larger, higher-markup firms does not respond as much to the shock. Thus, the cyclicality in the response of measured markups is partly explained by compositional shifts in the firm size distribution, whereby the entry of new firms with low prices amplifies the response of the economy to aggregate shocks. Indeed, note that small-firm markups are more responsive (by a factor of about two) to the demand shock compared to the supply shock. This, in turn, highlights the role that consumer search may have in explaining macroeconomic transmission.

The Margins of Price Adjustment

We have described above how firms of different sizes respond to aggregate shocks to demand and productivity. To further understand which are the main margins of price adjustment, we now decompose the price response by the different additive terms that we identified in equation (15). Recall that, according to this decomposition, sellers incorporate the equilibrium transitions into the price level in the form of compensations that ensure that the seller's utility promises are delivered in equilibrium.

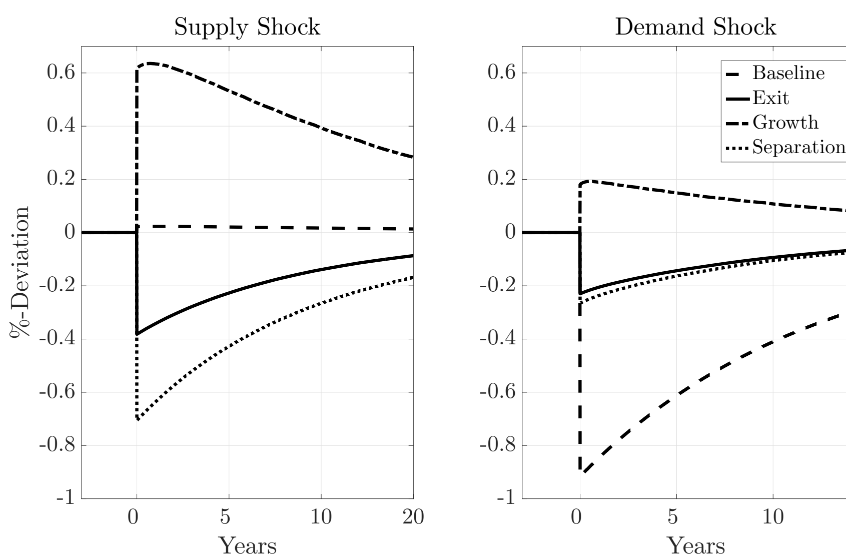

Figure 6: Impulse responses to negative and temporary 1% supply and demand shocks (same as Figures 3-5). Decomposition of the average price response across the different components identified in equation (15), where each component is averaged using the theoretical distribution of sellers across states. The exogenous shock adjustments (last two terms of equation (15)) are not being plotted.

Figure 6 shows the response of each fundamental component, for the same supply and demand shocks introduced above. We note, first, that the response along the "baseline" component of price is a lot more elastic in the case of a demand shock, overwhelming the remaining components and ultimately explaining why the price response is pro-cyclical after a demand shock and countercyclical after a supply shock. Again, this is due to the fact that a demand shock to consumer preferences has an instantaneous effect on the price through a fall in the consumers' contemporaneous willingness to pay. In the case of the supply shock, this effect is muted by the contractual mechanism of incomplete pass-through that we have discussed at length above. It is thus this difference which explains the different cyclical nature of the average price level in each case.

We also observe that, in both cases, the "growth" component is countercyclical, while the "exit" and "separation" components always co-move with the direction of the shock. For the former, the reason is that, even though sellers respond to both types of shocks by cutting promised utilities and, therefore, decreasing their probability of growing, smaller sellers make relatively bigger cuts as demand is more elastic for these type of firms. This means that the relative value of an additional customer (the object $x_{n+1} - x_n$) goes up after a negative shock, and thus sellers overcompensate customers in their price level for the eventuality of growing. In contrast, sellers cut down their "exit" and "separation" compensations, as both of these become less likely after a negative aggregate shock, no matter its nature.

6 The Secular Increase in Markups

In a recent study, DeLoecker and Eeckhout (2017) have documented a secular increase in the average markup in the U.S. since the early 1980s. Using panel data for publicly traded firms across all sectors, they find that average markups have been steadily rising from about 20% in 1980 to nearly 70% today (see Figure 7, left). These changes have occurred mostly within, rather than between, industries. Namely, the revenue share of top firms producing goods of comparable quality has steadily increased over time, and the distribution of markups has experienced a gradual shift toward (and particularly amongst) high-markup firms.[56] As in our model, firm size and markups in their data are positively correlated within narrowly defined industries.

Their analysis further unveils that the steep increase in the average markup has been accompanied by a rise in market concentration across all major sectors. This observation is complementary to a well-documented secular decline in business dynamism: since the early 1980s, the U.S. economy has experienced a persistent decline in firm entry and a compositional shift toward larger and older firms (see e.g. Pugsley and Şahin (2015)). In particular, the start-up rate (or the fraction of entering firms to the total number of firms) has declined from about 12% in the early 1980s to about 8% by 2012. In the retail sector specifically, entry rates have declined from 12.3% in 1980, to 8.6% in 2012 (see Figure 7, right).

We interpret this evidence through the lens of our model. The spirit of the exercise is not to provide an explanation to the decline in the entry rate per se, but rather to connect it to: (i) a substantial increase in the average markup; and (ii) an increase in dispersion on the upper tail of the markup distribution. Using the calibrated economy, we present these results in two ways: first, we study the effects of a rise in

[56] These observations have recently drawn a lot of attention among scholars. Autor *et al.* (2017a,b) argue that the increase in sales and employment concentration across all major sectors has been driven by the top (so-called "superstar") firms, experiencing the most dramatic increases in sales shares within their industries. Gutiérrez and Philippon (2017) argue that the decline in competition explains why firms tend to under-invest relative to Tobin's Q.

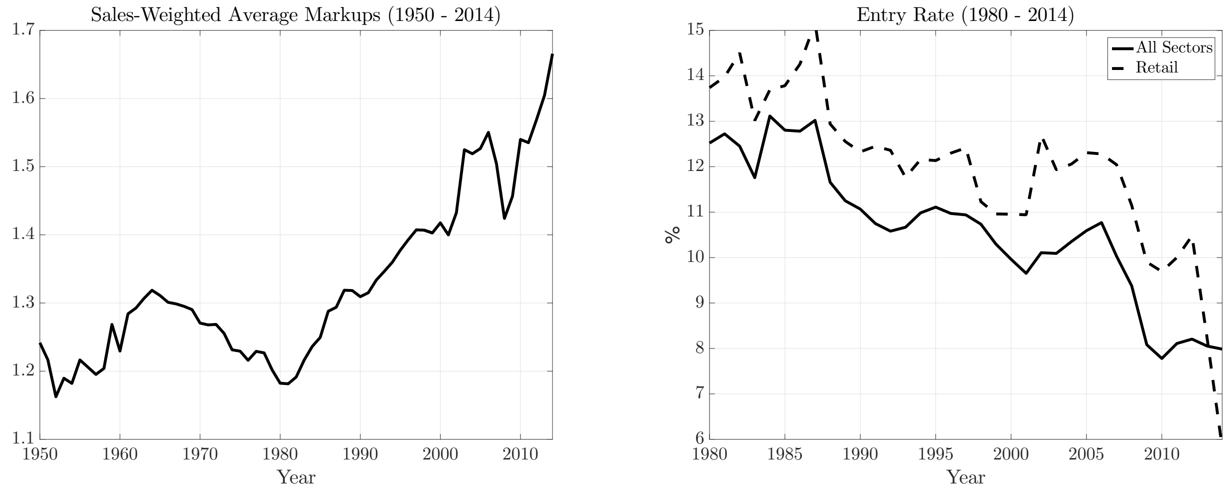

Figure 7: *Left:* Average markup in the firm cross-section across all industries for the U.S. (1950 – 2014). Markups are weighted by firm-level market shares in sales. Source: DeLoecker and Eeckhout (2017), using Compustat data. We thank Jan Eeckhout for kindly sharing this time series with us.
Right: Entry rates (in %) across all sectors (solid line) and for the Retail Trade sector (dashed line), for the U.S. (1980 – 2014). Entry rates are defined as the share of all firms that are aged 1 year or less. Source: Business Dynamics and Statistics (BDS) of the U.S. Census Bureau. Link: http://www.census.gov/ces/dataproducts/bds/data.html.

the entry cost κ across stationary solutions of the model; then, we study the transition between the two extreme steady states to understand how the markup distribution adjusts slowly over time.

Figure 8 shows the first set of results, i.e. the effects of changes in the entry cost κ on the stationary solution of the economy, at the calibrated set of parameters. Using this steady state comparison, shifts in the entry cost of firms can rationalize both the secular trend in the average markup of the economy and the rise in the within-industry markup coupled with an increase in concentration in the firm size distribution. In particular, the lower-right panel of Figure 8 shows that the increase in the average markup is driven by firms at the top of the markup distribution, as the contribution of large firms to the average markup becomes more and more important relative to that of smaller firms as the entry cost increases. Varying the entry cost parameter, we observe that the model can generate an increase in markups from 20% to more than 60%, i.e. of similar magnitude as in the data. As for the entry rate, the model generates a decline from about 8.75% to about 8.1%. Finally, we see that these effects are detrimental to welfare, as measured by equation (22) in steady state. This is in part because the increase entry cost is directly reducing resources, but also because of the distributional effects on buyers and sellers that come from the increase in market power.

[Transitional dynamics are currently work in progress]

7 Extensions

In this section, we relax some central assumptions of the baseline model of Section 3. In Section 7.1 we outline how to endogenize customer separations. In Section 7.2, we relax the assumption of no price discrimination across customers, and show that the model preserves its main structure though the predictions may change substantially.

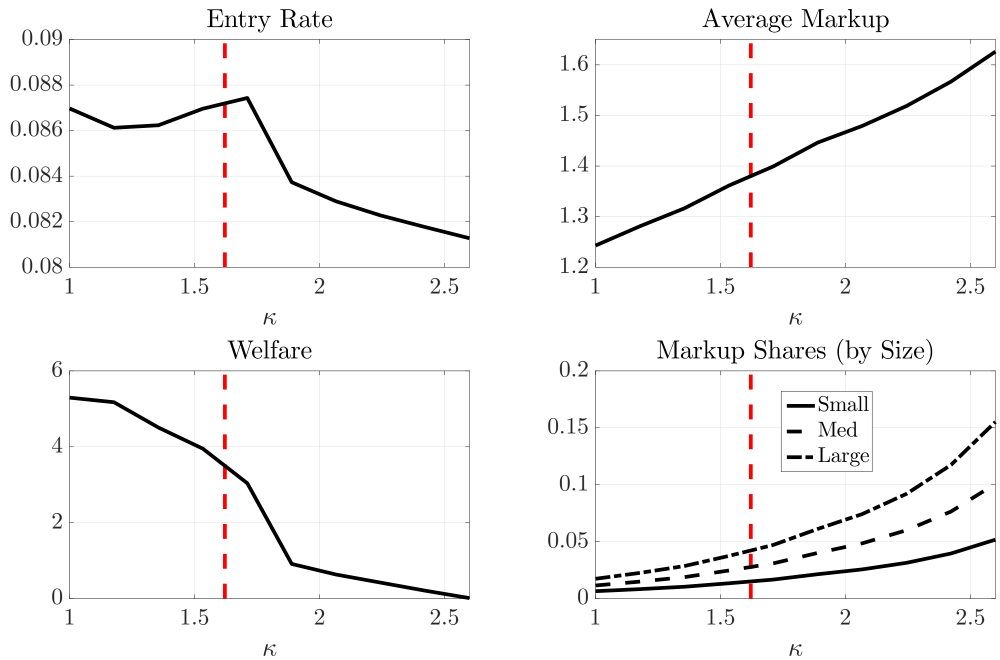

Figure 8: Entry rate, sales-weighted average markup, welfare, and markup shares by firm size, in the stationary solution of the calibrated economy, for different values of the entry cost (κ). The calibrated value of κ is marked with a dashed vertical line in all plots. *Welfare* is computed using equation (22) in steady state. *Average markup* is computed using equation (30). *Markup shares (by size)* plots some of the additive terms of equation (30), for different values of n (small, medium, and large), and z fixed at the median value.

7.1 Endogenous Customer Separations

To introduce customer firm-to-firm transitions, we can model customer search explicitly.[57] While we assume that there is still an exogenous risk $\delta_c > 0$ of separation for each customer (in which case the buyer must go through the inactivity stage), additionally we now add the possibility that customers search, and potentially endogenously separate, while on the match. We assume that active buyers do not face a cost of search, as they do not discontinue their consumption when transitioning from one seller to the next.

Introducing this additional dimension into our model is not at all straightforward. Endogenous buyer transitions across sellers could break the ex-ante indifference condition among inactive buyers, which in our baseline setting is key to pin down equilibrium market tightness. In order to preserve the block-recursive structure, one remedy would be to assume free entry across all markets on the *seller* side (e.g. Schaal (2017)). This would change the environment substantially, so we leave it for future work.

To outline the basic setting of endogenous customer search within the baseline model, we thus must turn off the aggregate shocks. The search decision of a customer currently obtaining value V^B is:

$$\max_{x \in [V^B, \overline{x}]} \mu(\theta(x))\big(x - V^B\big)$$

Note that the matched buyer only considers offers that deliver an expected value that weakly dominates the current perceived utility, V^B. Denoting the solution by $\widehat{x}(n, \boldsymbol{\omega}; z)$ for a customer matched with

[57] Appendix D.4 shows an alternative way of endogenizing the separation rate.

firm of type (n, z) under contract $\boldsymbol{\omega}$, the first-order condition reads:

$$\left(\widehat{x}(n, \boldsymbol{\omega}; z) - V^B(n, \boldsymbol{\omega}; z)\right) \frac{\partial \mu(\theta(x))}{\partial x}\bigg|_{x=\widehat{x}(n,\boldsymbol{\omega};z)} = -\mu\Big(\theta\big(\widehat{x}(n, \boldsymbol{\omega}; z)\big)\Big) \tag{32}$$

Intuitively, the inactive buyer trades off the expected option value of transitioning (left-hand side) to the rate at which this offer can be obtained (right-hand side). Since we focus on equilibria in which market tightness is an increasing function of promised utilities, it is not difficult to show (e.g. Shi (2009)) that $\widehat{x}(n, \boldsymbol{\omega}; z)$ is increasing in $V^B(n, \boldsymbol{\omega}; z)$. In words, the more profitable a match is ex-post, the higher the offer for which the customer will apply next. Therefore, customers separate according to their initial state, and climb up on the utility ladder. This effect tends to shift the mass of customers (and therefore firms) toward higher promised utilities, and thus acts as a countervailing force to the equilibrium dynamics of the baseline model: when the sellers offering the worst terms of trade lose customers, they need to start setting up more favorable contracts.

The risk of endogenously losing customers must now be incorporated into the pricing decisions of firms. The buyers' and firm's HJB equations are identical to (5) and (6), respectively, except we now replace δ_c by the "effective" customer separation rate:

$$\widehat{\delta}_c(n, \boldsymbol{\omega}; z) := \delta_c + \mu\Big(\theta\big(\widehat{x}(n, \boldsymbol{\omega}; z)\big)\Big)$$

Therefore, the value functions now incorporate that all customers alike can endogenously transition to other sellers if the terms of their current contract are not attractive enough for them. The market tightness must now incorporate that the pool of searching buyers is composed of both inactive as well as active buyers. Since active buyers only search for strictly better markets, and there is a single such market they may possibly transition to, we have:

$$\theta_n(z) = \frac{1}{S_{n-1}(z)}\left(B_n^I(z) + B_{\iota(n)}^A(z)\right)$$

for any $n \geq 1$, where $\iota(n) \in \mathbb{N}$ is the size of the firm that a customer seeking to transition to a size-n firm is currently matched with, i.e. the solution to $x_n(z) = \widehat{x}_{\iota(n)}(z)$.

7.2 Price Discrimination

The assumption of no price discrimination across different customers is not key to generate efficient firm dynamics. We argue that, so long as we maintain the assumption of dynamic contracts with commitment, our model still generates these dynamics as well as cross-sectional price dispersion. However, always within the realm of Markov-perfect equilibria that we have narrowed our attention to, allowing for price discrimination opens the door to equilibrium multiplicity.

First, if firms were to use only prices (instead of full dynamic contracts with price-utility pairs) as their instrument for customer attraction, an equilibrium with price discrimination across customers of different tenures would look similar to that of Gourio and Rudanko (2014b): firms would attract customers by offering an instantaneous discount on the valuation v, and extract all surplus by charging v immediately after the customer joins the seller, and until separation. Assuming price discrimination in our setting with dynamic long-term contracts and commitment does *not* yield this result. This is because firms must still trade off static payoffs coming from the current price with dynamic ones coming from the promised utility. Importantly, in this case, tractability would be preserved along several important dimensions: (i) the equilibrium would still be block-recursive; and (ii) the optimal contract would still solve a joint surplus problem. More importantly, (iii) the joint surplus would be constant in the distribution of contracts across customers, so the equilibrium could still be characterized as a set of sequences solving a joint surplus problem. Qualitatively, allowing for price discrimination comes at only only one cost: (iv) price indeterminacy.

Let us discuss these results more formally. First, we must extend our baseline framework to allow discrimination across buyers. Let $\boldsymbol{\omega}_i = \{p_i, \mathbf{x}'_i(n'; \mathbf{s}')\}$ be the contract offered to the typical current customer $i = 1, \ldots, n$, which is composed of an individual-specific price p_i and a *personalized* menu of continuation utilities: $\mathbf{x}'_i(n'; \mathbf{s}')$ for $n' \in \{n-1, n, n+1\}$ and $\mathbf{s}' \in \{(z', \varphi), (z, \varphi')\}$. A seller is characterized by the collection $\{x_i\}_{i=1}^n$ of outstanding promises, and must choose: (i) a menu of contracts $\{\boldsymbol{\omega}_i\}_{i=1}^n$ for the n current customers; and (ii) a starting promised utility $x'_0 \in \mathbb{R}$ for the new incoming customer (if there is any). The HJB equation for the seller now reads:

$$
\begin{aligned}
rV^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big) = \max_{x'_0, \{\boldsymbol{\omega}_i\}_{i=1}^n} \Bigg\{ & \sum_{i=1}^n p_i - \mathcal{C}(n; z, \varphi) + \delta_f \Big(V_0^S(\varphi) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\Big) \\
& + \delta_c \sum_{j=1}^n \Big(V^S\Big(n-1, \{x'_i(n-1; z, \varphi)\}_{i=1}^n \backslash_- \{x'_j(n-1; z, \varphi)\}; z, \varphi\Big) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\Big) \\
& + \eta\big(\theta(x'_0; \varphi)\big) \Big(V^S\Big(n+1, \{x'_i(n+1; z, \varphi)\}_{i=1}^n \cup_+ \{x'_0\}; z, \varphi\Big) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\Big) \\
& + \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z) \Big(V^S\Big(n, \{x'_i(n; z', \varphi)\}_{i=1}^n; z', \varphi\Big) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\Big) \\
& + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi) \Big(V^S\Big(n, \{x'_i(n; z, \varphi')\}_{i=1}^n; z, \varphi'\Big) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\Big) \Bigg\}
\end{aligned}
$$

where $\backslash_-$ and $\cup_+$ are multiset difference and union operators.[58] The most important differences relative to the baseline model (equation (6)) have been highlighted in blue. Note that, now, when a customer $i = 1, \ldots, n$ separates, the vector of promised utilities shrinks in cardinality and the customers that remain obtain the new promise $x'_i(n-1; z, \varphi)$. The firm attracts new buyers by offering a starting utility x'_0 to the entering customer, while delivering the promised level $x'_i(n+1; z, \varphi)$ to each of the remaining n customers. The promise-keeping constraint now reads:

$$
\forall i = 1, \ldots, n: \quad x_i \leq V^B(n, \boldsymbol{\omega}_i; z, \varphi)
$$

for all $(z, \varphi) \in \mathcal{Z} \times \Phi$, establishing that the firm commits to each and every customer (and recognizes that each customer earns a different utility). As in the baseline model, we can solve for the optimal menu

[58] These operators are defined by $\{a, b, b\} \backslash_- \{b\} = \{a, b\}$ and $\{a, b\} \cup_+ \{b\} = \{a, b, b\}$, and they are needed here because the vector of promised utilities may contain more than one instance of the same element.

of contracts by solving for the joint surplus problem:

Proposition 5 *In the economy with price discrimination, the seller's and the joint surplus problems are equivalent:*

(i) Given a menu of contracts $\boldsymbol{\omega}_i = \{p_i, \mathbf{x}'_i(n'; \mathbf{s}')\}$ *for* $i = 1, \ldots, n$ *that maximize the seller's value subject to the promise-keeping constraint,* $\{\mathbf{x}'_i(n'; \mathbf{s}')\}_{i=1}^n$ *maximizes* $W(n, \{x_i\}_{i=1}^n; z, \varphi) := V^S(n, \{x_i\}_{i=1}^n; z, \varphi) + \sum_{i=1}^n x_i$*;*

(ii) Conversely, for every $\{\mathbf{x}'_i(n'; \mathbf{s}')\}_{i=1}^n$ *that maximizes* $W(n, \{x_i\}_{i=1}^n; z, \varphi)$*, there exists a menu of personalized price levels* $\{p_i\}_{i=1}^n$ *such that the collection* $\{p_i, \mathbf{x}'_i(n'; \mathbf{s}')\}_{i=1}^n$ *constitutes a solution to the seller's problem.*

The proof can be found in Appendix B.5. The characterization of the equilibrium is then very similar to that of the baseline model. First, we note that by utility-invariance of the joint surplus, we can write:

$$\forall (n, z, \varphi) \in \mathbb{N} \times \mathcal{Z} \times \Phi : \quad W_n(z, \varphi) = W\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big) \tag{33}$$

Letting $\{x'_0, \{x'_{i,n+1}(z, \varphi)\}_{i=1}^n\}$ be the set of optimal policies, the joint surplus solves the second-order difference equation:

$$(r+\delta_f) W_n(z, \varphi) = nv - \mathcal{C}(n; z, \varphi) + n(\delta_f + \delta_c) U^B(\varphi) + \eta(\theta(x'_0; \varphi)) \Big(W_{n+1}(z, \varphi) - W_n(z, \varphi) - \sum_{i=1}^n x'_{i,n+1}(z, \varphi)\Big)$$
$$+ n\delta_c\big(W_{n-1}(z, \varphi) - W_n(z, \varphi)\big) + \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z)\Big(W_n(z', \varphi) - W_n(z, \varphi)\Big) + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)\Big(W_n(z, \varphi') - W_n(z, \varphi)\Big)$$

Since only the aggregate utility $\sum_{i=1}^n x'_{i,n+1}(z, \varphi)$ is relevant from the joint surplus perspective, there is now a multiplicity of contracts that can be sustained in the optimal allocation.[59] This is stated formally in the following lemma:

Proposition 6 *In the economy with price discrimination, prices are not uniquely determined. There is a continuum of joint-surplus-maximizing contracts* $\{p_i^*, \mathbf{x}'^*_i(n'; \mathbf{s}')\}_{i=1}^n$ *that leave both the buyers and the seller indifferent.*

For the proof, see Appendix B.6. This multiplicity result did not emerge in the baseline model because of our inductive construction of the Markov-perfect equilibrium: the free entry problem delivered a unique choice of the promised utility, which by promise-keeping ensured a unique price. Given this allocation, and the fact that firms of the same size cannot propose different contracts, the problems of the firms of size n gave a unique allocation given the solution of that of firms of size $n-1$. With price discrimination, however, there is discretion in the way firms distribute promised utilities (and thus prices) across their different buyers, so long as the sum of utilities is maximized according to the joint surplus rule. In a sense, therefore, assuming no price discrimination can be seen as the price-discrimination equilibrium in which all customers are charged the same price. Our analysis above shows, however, that when looking at Markov strategies, other allocations can be sustained in equilibrium.

[59] We should note here that the baseline model delivers a unique equilibrium contract only within the class of stationary Markov perfect equilibria that we have restricted our attention to. However, this does *not* mean that other equilibria may not exist under broader equilibrium definitions. Our point in this section is to point out that multiplicity reemerges within our Markov environment after we introduce discrimination.

8 Conclusion

Recent studies indicate that a major source of variation in firm performance across a variety of industries stems from demand components that are idiosyncratic to the firm, and that price differences are key to explain revenue differences for firms of similar productivity levels. These observations can shed new light on the behavior of markups at the aggregate level. We have presented a dynamic search model of demand accumulation through firm pricing with aggregate and idiosyncratic shocks and a relevant scope for firm dynamics in order to study the connection between customer capital at the microeconomic level and macroeconomic dynamics. In the model, firms of different customer base sizes strategically use menus of prices and continuation promises in order to trade off two conflicting concerns: attracting new customers to increase future market share, and extracting surplus from incumbent customers to increase current profits. The model exhibits cross-sectional price dispersion, and offers a micro-founded interpretation for sluggish price dynamics at both the firm and the aggregate level.

We have analyzed a number of predictions on both pricing and firm dynamics dimensions, and found them to be qualitatively in line with data from a large panel of stores in the U.S. retail sector. We have found that stores with lower volumes of sales experience higher sales growth rates for their products, and set prices that are, on average, lower than those of their direct competitors. Using this evidence, we have estimated the model and conducted experiments on the response of the economy to aggregate shocks to productivity and demand. In these exercises, we have found both level and compositional effects: shocks that raise the marginal propensity to consume by buyers elicit more entry of small sellers, as the opportunities of capturing new customers become greater. Since these sellers are also the ones setting lower prices, the markup distribution experiences a left-ward shift, and output booms. We have also provided a novel channel for incomplete pass-through of shocks to prices: because firms internalize that their prices today determine not only current profits but also future ones, they are able to smooth out the negative impact of shocks by front-loading their contracts and charging higher prices today, giving rise to a countercyclical price response. In combination, these results can therefore rationalize the cyclicality of markups that we see in the data, both at the aggregate level as well as in the cross-section of firms.

Finally, we have used the model to account for some salient low-frequency features of the U.S. economy since the 1980s, namely the secular decline in business dynamism (Pugsley and Şahin (2015)), and the continued increase in the average markup and in market concentration (DeLoecker and Eeckhout (2017)). Seen through the lens of our model, a shock that depresses firm entry is associated with a shift in the firm distribution toward higher-markup firms, and thus a rise in market concentration. A continued decrease in the costs of buyer search or an increase in the entry cost of firms can explain these facts simultaneously. This suggests that incorporating micro-founded pricing environments into quantitative macro models is relevant to understand certain patterns in macroeconomic dynamics. Further investigating the scope of customer markets to explain these and other trends remains an interesting avenue for future work.

References

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

Alvarez, F., LeBihan, H. and Lippi, F. (2014). Small and large price changes and the propagation of monetary shocks. *NBER Working Paper 20155*.

Apel, M., Friberg, R. and Hallsten, K. (2005). Micro foundations of macroeconomic price adjustment: Survey evidence from Swedish firms. *Journal of Money, Credit and Banking*, **37** (2), 313 – 338.

Arkolakis, C. (2010). Market penetration costs and the new consumer margin in international trade. *Journal of Political Economy*, **118** (6), 1151 – 1199.

— (2016). A unified theory of firm selection and growth. *The Quarterly Journal of Economics*, **131** (1), 89–155.

Atkeson, A. and Kehoe, P. J. (2005). Modeling and measuring organization capital. *Journal of Political Economy*, **113** (5), 1026 – 1053.

Autor, D., Dorn, D., Katz, L. F., Patterson, C. and Reenen, J. V. (2017a). Concentrating on the fall of the labor share. *American Economic Review: Papers & Proceedings*, **107** (5), 180 – 185.

—, —, —, — and — (2017b). The fall of the labor share and the rise of superstar firms. *NBER Working Paper 23396*.

Azariadis, C. (1975). Implicit contracts and underemployment equilibria. *Journal of Political Economy*, **83** (6), 1183 – 1202.

Bai, Y., Rios-Rull, V. and Storesletten, K. (2012). Demand shocks as productivity shocks. *Working Paper, University of Pennsylvania*.

Baily, M. N. (1974). Wages and employment under uncertain demand. *The Review of Economic Studies*, **41** (1), 37 – 50.

Berman, N., Rebeyrol, V. and Vicard, V. (2017). Demand learning and firm dynamics: Evidence from exporters. *Working Paper, Aix Marseille School of Economics*.

Bhandari, A. and McGrattan, E. (2017). Sweat equity in U.S. private business. *Working Paper, University of Minnesota*.

Bils, M. (1989). Pricing in a customer market. *The Quarterly Journal of Economics*, **104** (4), 699 – 718.

Blinder, A. S., Canetti, E. R. D., Lebow, D. E. and Rudd, J. B. (1998). *Asking About Prices: A New Approach to Understanding Price Stickiness*. Russell Sage Foundation.

Bronnenberg, B. J. and Dubé, J.-P. (2017). The formation of consumer brand preferences. *The Annual Review of Economics*, **9**, 353 – 382.

—, Kruger, M. W. and Mela, C. F. (2008). The IRI marketing data set. *Marketing Science*, **27** (4), 745 – 748.

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

— and Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica*, **51** (4), 955 – 969.

—, Shi, S. and Wright, R. (2001). Pricing and matching with frictions. *Journal of Political Economy*, **109** (5), 1060 – 1085.

Butters, G. R. (1977). Equilibrium distributions of sales and advertising prices. *The Review of Economic Studies*, **44** (3), 465 – 491.

Cabral, L. and Mata, J. (2003). On the evolution of the firm size distribution: Facts and theory. *American Economic Review*, **93** (4), 1075 – 1090.

Carlsson, M., Messina, J. and Skans, O. N. (2014). Firm-level shocks and labor adjustments. *IFAU Working Paper*.

Cason, T. N. and Friedman, D. (2003). Buyer search and price dispersion: A laboratory study. *Journal of Economic Theory*, **112** (2), 232 – 260.

Caves, R. E. (1998). Industrial organization and new findings on the turnover and mobility of firms. *Journal of Economic Literature*, **36** (4), 1947 – 1982.

Christopoulou, R. and Vermeulen, P. (2008). Markups in the Euro Area and the U.S. over the period 1981-2004. *European Central Bank, Working Paper Series*, **856**.

Coibion, O., Gorodnichenko, Y. and Hong, G. H. (2015). The cyclicality of sales, regular, and effective prices: Business cycle and policy implications. *American Economic Review*, **105** (3), 993–1029.

DeLoecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, **79** (5), 1407 – 1451.

— and Eeckhout, J. (2017). The rise of market power and the macroeconomic implications. *NBER Working Paper 23687*.

—, Goldberg, P. K., Khandelwal, A. K. and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, **84** (2), 445 – 510.

— and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, **102** (6), 2437 – 2471.

Dinlersoz, E. M. and Yorukoglu, M. (2012). Information and industry dynamics. *American Economic Review*, **102** (2), 884 – 913.

Drozd, L. A. and Nosal, J. B. (2012). Understanding international prices: Customers as capital. *American Economic Review*, **102** (1), 364 – 395.

Dunne, T., Roberts, M. J. and Samuelson, L. (1988). Patterns of firm entry and exit in U.S. manufacturing industries. *The RAND Journal of Economics*, **19** (4), 495 – 515.

Eaton, J., Eslava, M., Jinkins, D., Krizan, C. J. and Tybout, J. (2014). A search and learning model of export dynamics. *Working Paper, Pennsylvania State University*.

Elsby, M. W. L. and Michaels, R. (2013). Marginal jobs, heterogeneous firms, and unemployment flows. *American Economic Journal: Macroeconomics*, **5** (1), 1 – 48.

Fabiani, S., Druant, M., Hernando, I., Kwapil, C., Landau, B., Loupias, C., Martins, F., Mathä, T. Y., Sabbatini, R., Stahl, H. and Stokman, A. C. (2004). The pricing behavior of firms in the Euro Area: New survey evidence. In *Inflation Persistence in the Euro Area (European Central Bank)*.

Faig, M. and Jerez, B. (2005). A theory of commerce. *Journal of Economic Theory*, **122** (1), 60 – 99.

Fishman, A. and Rob, R. (2003). Consumer inertia, firm growth and industry dynamics. *Journal of Economic Theory*, **109**, 24 – 38.

Fitzgerald, D., Haller, S. and Yedid-Levi, Y. (2017). How exporters grow. *Working Paper, Federal Reserve Bank of Minneapolis*.

Foster, L., Haltiwanger, J. and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, **98** (1), 394 – 425.

—, — and — (2016). The slow growth of new plants: Learning about demand? *Economica*, **83**, 91 – 129.

Gagnon, E. and López-Salido, D. (2014). Small price responses to large demand shocks. *Federal Reserve Board Discussion Paper*.

Galenianos, M. and Kircher, P. (2009). Directed search with multiple job applications. *Journal of Economic Theory*, **144** (2), 445 – 471.

— and — (2012). On the game-theoretic foundations of competitive search equilibrium. *International Economic Review*, **53** (1), 1 – 21.

Gilchrist, S., Schoenle, R., Sim, J. and Zakrajšek, E. (2016). Inflation dynamics during the financial crisis. *Working Paper, Boston University*.

Gourio, F. and Rudanko, L. (2014a). Can intangible capital explain cyclical movements in the labor wedge? *American Economic Review*, **104** (5), 183–88.

— and — (2014b). Customer capital. *Review of Economic Studies*, **81** (3), 1102 – 1136.

Gutiérrez, G. and Philippon, T. (2017). Declining competition and investment in the U.S. *NBER Working Paper 23583*.

Hall, R. E. (2008). General equilibrium with customer relationships: A dynamic analysis of rent-seeking. *Working Paper, Stanford University*.

Hall, S., Walsh, M. and Yates, A. (1997). How do U.K. companies set prices? *Bank of England Quarterly Bulletin*, pp. 180 – 192.

Hong, S. (2017). Customer capital, markup cyclicality, and amplification. *Working Paper, Princeton University*.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, **60** (5), 1127 – 1150.

Hottman, C. (2017). Retail markups, misallocation, and store variety across U.S. cities. *Board of Governors of the Federal Reserve System, Working Paper*.

—, Redding, S. J. and Weinstein, D. E. (2016). Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, **131** (3), 1291 – 1364.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics*, **124** (4), 1403 – 1448.

— and — (2014). The life cycle of plants in India and Mexico. *The Quarterly Journal of Economics*, **129** (3), 1035 – 1084.

Jovanovic, B. (1982). Selection and the evolution of the industry. *Econometrica*, **50** (3), 649 – 670.

Kaas, L. and Kimasa, B. (2016). Firm dynamics with frictional product and labor markets. *Working Paper, University of Konstanz*.

— and Kircher, P. (2015). Efficient firm dynamics in a frictional labor market. *American Economic Review*, **105** (10), 3030 – 3060.

Kaplan, G. and Menzio, G. (2015). The morphology of price dispersion. *International Economic Review*, **56** (4), 1165 – 1206.

—, —, Rudanko, L. and Trachter, N. (2016). Relative price dispersion: Evidence and theory. *Working Paper, University of Chicago*.

Klemperer, P. (1987). Markets with consumer switching costs. *The Quarterly Journal of Economics*, **102** (2), 375 – 394.

— (1995). Markets with consumer switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies*, **62**, 515 – 539.

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

Krusell, P. and Smith, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, **106** (5), 867 – 896.

Kugler, M. and Verhoogen, E. (2012). Prices, plant size, and product quality. *Review of Economic Studies*, **79**, 307 – 339.

Luttmer, E. G. (2006). Consumer search and firm growth. *Working Paper, Federal Reserve Bank of Minneapolis*.

— (2007). Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, **122** (3), 1103 – 1144.

McGrattan, E. (2015). Intangible capital and measured productivity. *Working Paper, University of Minnesota*.

— and Prescott, E. C. (2014). A reassessment of real business cycle theory. *American Economic Review*, **104** (5), 177–82.

Menzio, G. (2007). A search theory of rigid prices. *Working Paper, University of Pennsylvania*.

— and Shi, S. (2010). Block recursive equilibria for stochastic models of search on the job. *Journal of Economic Theory*, **145** (4), 1453 – 1494.

— and — (2011). Efficient search on the job and the business cycle. *Journal of Political Economy*, **119** (3), 468 – 510.

— and Trachter, N. (2015). Equilibrium price dispersion across and within stores. *Working Paper, University of Pennsylvania*.

Moscarini, G. and Postel-Vinay, F. (2013). Stochastic search equilibrium. *Review of Economic Studies*, **80** (4), 1545 – 1581.

Nakamura, E. and Steinsson, J. (2008). Five facts about prices: A reevaluation of menu cost models. *The Quarterly Journal of Economics*, **123** (4), 1415 – 1464.

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

Norris, J. (1997). *Markov Chains*. Cambridge University Press.

Paciello, L., Pozzi, A. and Trachter, N. (2016). Price dynamics with customer markets. *Working Paper, Federal Reserve Bank of Richmond*.

Perla, J. (2016). Product awareness, industry life cycles, and aggregate profits. *Working Paper, University of British Columbia*.

Peters, M. (2016). Heterogeneous markups, growth and endogenous misallocation. *Working Paper, Yale University*.

Petrosky-Nadeau, N. and Wasmer, E. (2015). Macroeconomic dynamics in a model of goods, labor, and credit market frictions. *Journal of Monetary Economics*, **72**, 97 – 113.

—, — and Zeng, S. (2016). Shopping time. *Economics Letters*, **143**, 52 – 60.

Phelps, E. S. and Winter, S. G. (1970). Optimal price policy under atomistic competition. *Microeconomic Foundations of Employment and Inflation Theory*, pp. 309 – 337.

Piveteau, P. (2017). An empirical dynamic model of trade with consumer accumulation. *Working Paper, Johns Hopkins University*.

Pozzi, A. and Schivardi, F. (2016). Demand or productivity: What determines firm growth? *RAND Journal of Economics*, **47** (3), 608 – 630.

Pugsley, B. and Şahin, A. (2015). Grown-up business cycles. *Federal Reserve Bank of New York Staff Reports 707*.

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

Renner, E. and Tyran, J.-R. (2004). Price rigidity in customer markets: an empirical study. *Journal of Economic Behavior and Organization*, **55** (4), 575 – 593.

Rossi-Hansberg, E. and Wright, M. L. J. (2007). Establishment size dynamics in the aggregate economy. *American Economic Review*, **97** (5), 1639 – 1664.

Rotemberg, J. J. and Woodford, M. (1991). Markups and the business cycle. *NBER Macroeconomics Annual*, **6**, 63 – 140.

— and — (1999). The cyclical behavior of prices and costs. *NBER Working Paper No. 6909*.

Rudanko, L. (2017). Pricing in a frictional product market. *Mimeo, Federal Reserve Bank of Philadelphia*.

Schaal, E. (2017). Uncertainty and unemployment. *Econometrica*, **85** (6), 1675 – 1721.

Shi, S. (2009). Directed search for equilibrium wage-tenure contracts. *Econometrica*, **77** (2), 561 – 584.

— (2016). Customer relationship and sales. *Journal of Economic Theory*, **166**, 483–516.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, **95** (1), 25 – 49.

Stokey, N. L. and Lucas, R. E. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature*, **35** (1), 40 – 59.

Tauchen, G. (1986). Finite state Markov-Chain approximations to univariate and vector autoregressions. *Economics Letters*, **20** (2), 177 – 181.

Varian, H. R. (1980). A model of sales. *The American Economic Review*, **70** (4), 651 – 659.

Wright, R., Kircher, P., Julien, B. and Guerrieri, V. (2017). Directed search: A guided tour. *NBER Working Paper 23884*.

Appendix

Contents. *The Appendix is structured as follows. Appendix A includes tables and figures. Appendix B presents all the proofs. Appendix C discusses the numerical implementation. Finally, Appendix D describes additional results, including the derivation of the HJB equations, distribution dynamics, price statistics at the stationary solution, and comparative statics exercises.*

A Tables and Figures

	Full Sample	**Sub-Sample**
Number of product categories	15	15
Number of chains	64	64
Number of stores	278	278
Number of UPCs	19,721	11,483
Stores per chain (average)	27	26
Stores per product (average)	59	88
Products per store (average)	4,180	3,638
Average price (USD)	7.75	8.32
Price dispersion	15.73%	10.55%
Total sales (Billion USD)	2.86	1.60
Number of weeks	365	365
MSAs considered	NY, LA	NY, LA
Number of observations	89,112,170	59,813,217

Table A.1: Descriptive statistics before and after restricting the sample. Source: IRI Symphony weekly data. Notes: *Price dispersion* is computed as the average standard deviation of log-standardized prices (equation (25)) across all time periods.

	Relative prices	**Normalized sales**	**Sales growth**
Mean	0	1	−.0008
Median	.0009	.9087	−.0003
Percentiles			
1st	−.3257	.2729	−.1905
10^{th}	−.1138	.4709	−.0772
25^{th}	−.0415	.6656	−.0378
75^{th}	.0486	1.2281	.0373
90^{th}	.1097	1.7029	.0765
99^{th}	.2809	2.3819	.1782
Dispersion			
St. Dev.	.1055	.4744	.0694
p90-p10 range	1.2504	3.6163	1.1661
p90-p50 range	1.1149	1.8740	1.0798
p50-p10 range	1.1215	1.9297	1.0799

Table A.2: Distribution of relative prices, normalized store sales (i.e. the ratio of total dollar sales within the store to average sales across products sold in the store), and annualized sales growth rates, across the whole 2001-2007 sample. Source: IRI Symphony weekly data.

	(1a) $\Delta\log(Sales_{us})$	(1b) $\Delta\log(Sales_{us})$	(2a) $\widehat{p}_{us}$	(2b) $\widehat{p}_{us}$	(3a) $\widehat{e}_s$	(3b) $\widehat{e}_s$
$\log(Sales_s)$	−.04036***	−.0392***	.0101***	.0063***	.0079***	.0045***
	(.0012)	(.0012)	(.0003)	(.0003)	(.00005)	(.00004)
Age FE	✗	✓	✗	✓	✗	✓
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
R^2	.1355	.1358	.1498	.1511	.9263	.9339
Obs. (millions)	21.15	21.15	28.95	28.95	28.95	28.95

Table A.3: Same as Table 1, except that the sample has been restricted to only those stores that belong to chains with a total number of stores below the median.

Year	**New stores**	**All stores**	**Entry rate**
2001	18	189	9.52%
2002	11	182	6.04%
2003	14	172	8.14%
2004	9	176	5.11%
2005	13	185	7.03%
2006	20	187	10.7%
2007	32	200	16%
Average	16.71	184.43	8.9%

Table A.4: Number of new stores (aged 52 weeks or less) and all existing stores, per year. The entry rate of stores is computed as the ratio of new stores to all stores. Source: IRI Symphony weekly data.

Moment	**Model**	**Data**	
A. Distribution of Relative Prices			
p90-p10 range	1.1994	1.2504	
p90-p50 range	1.0508	1.1149	
B. Distribution of Price Changes			
Average frequency	0.9639	0.9609	annualized
Median frequency	0.9814	0.9264	annualized
Average implied duration	0.2788	0.7817	years
Median implied duration	0.2503	0.3568	years
Average absolute change	0.0305	0.0313	log-points
Median absolute change	0.0312	0.0197	log-points
St. Dev. absolute change	0.0505	0.1415	%-points

Table A.5: Non-targeted moments: model vs. data. Notes: Data moments are taken from our IRI sample. See Appendix D.3 for the calculation and aggregation of firm-level price statistics in the model.

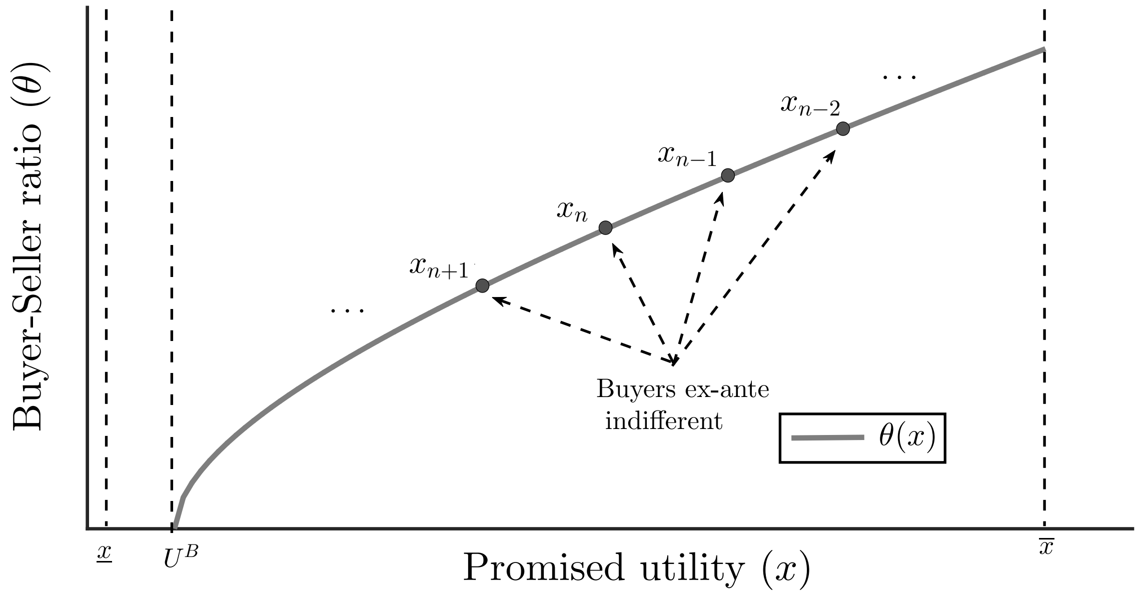

Figure A.1: Equilibrium tightness function $\theta : x \mapsto \mu^{-1}\left(\frac{\Gamma^B}{x-U^B}\right)$, and equilibrium markets across n, given $\mathbf{s} = (z, \varphi)$.

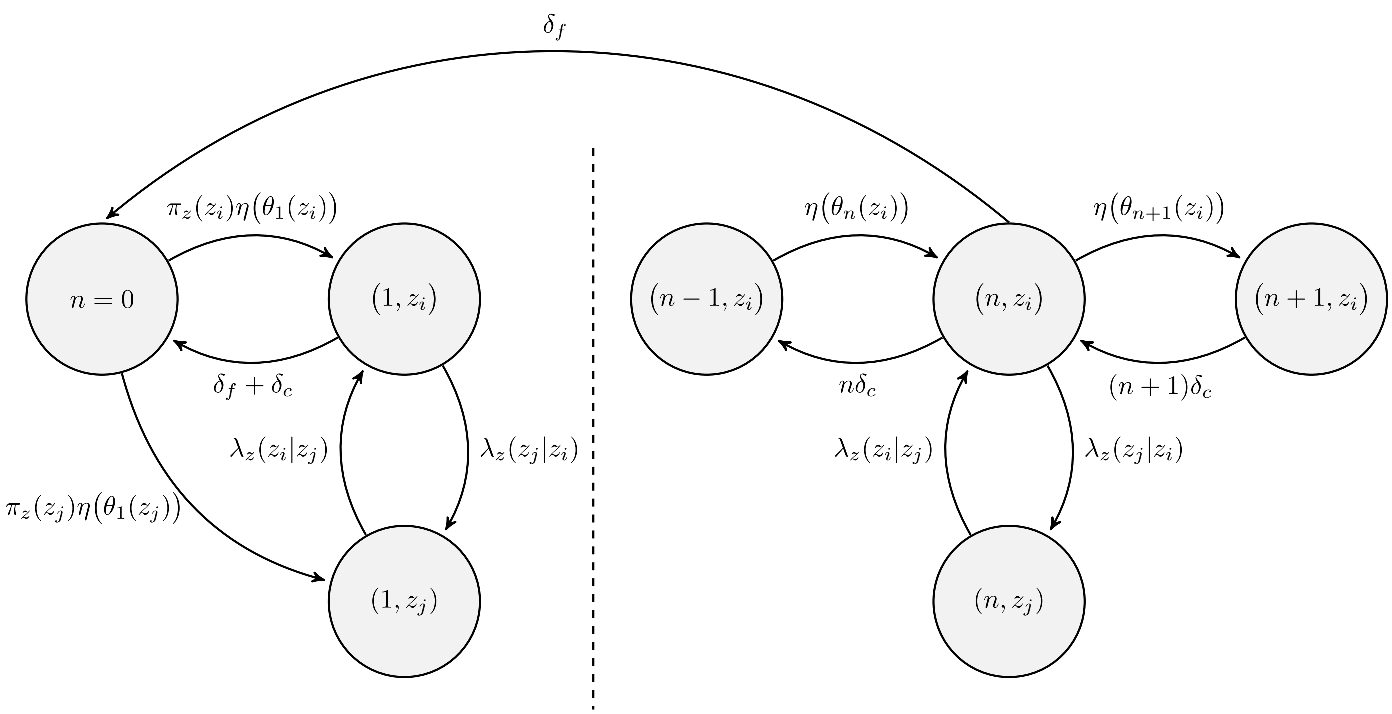

Figure A.2: Seller transitional dynamics in equilibrium, for a typical incumbent (right-hand side block) and for entrants (left-hand side block), where φ is fixed for expositional ease. Labels on arrows indicate flow rates.

Figure A.3: Distribution of normalized store sales (top), store sales growth rates (middle), and relative prices at the UPC level (bottom) in our final sample. Source: IRI Symphony weekly data.

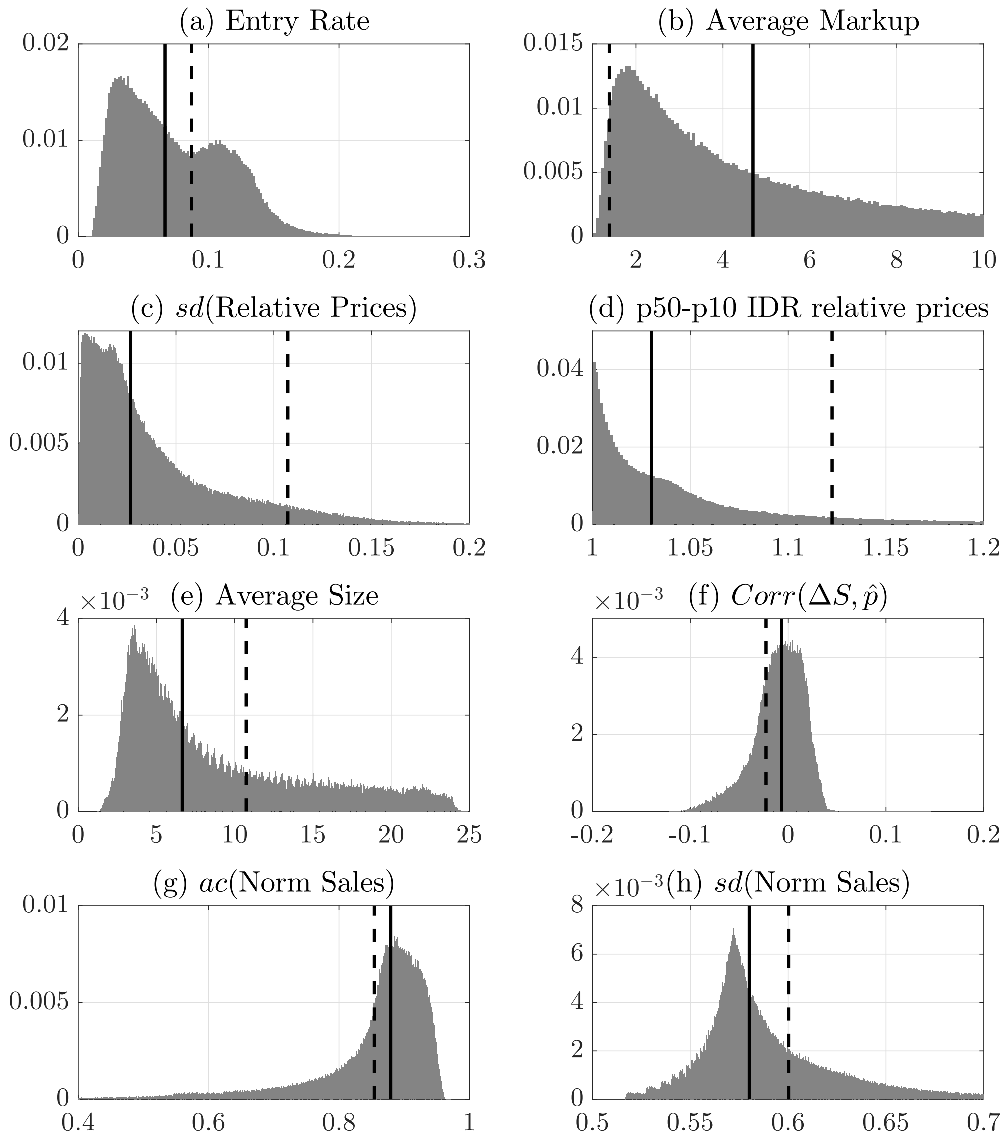

Figure A.4: Histograms of calibrated moments across different simulated economies in the parameter-search SMM algorithm. The dashed vertical line marks the calibrated value. The solid vertical line is the median of the distribution.

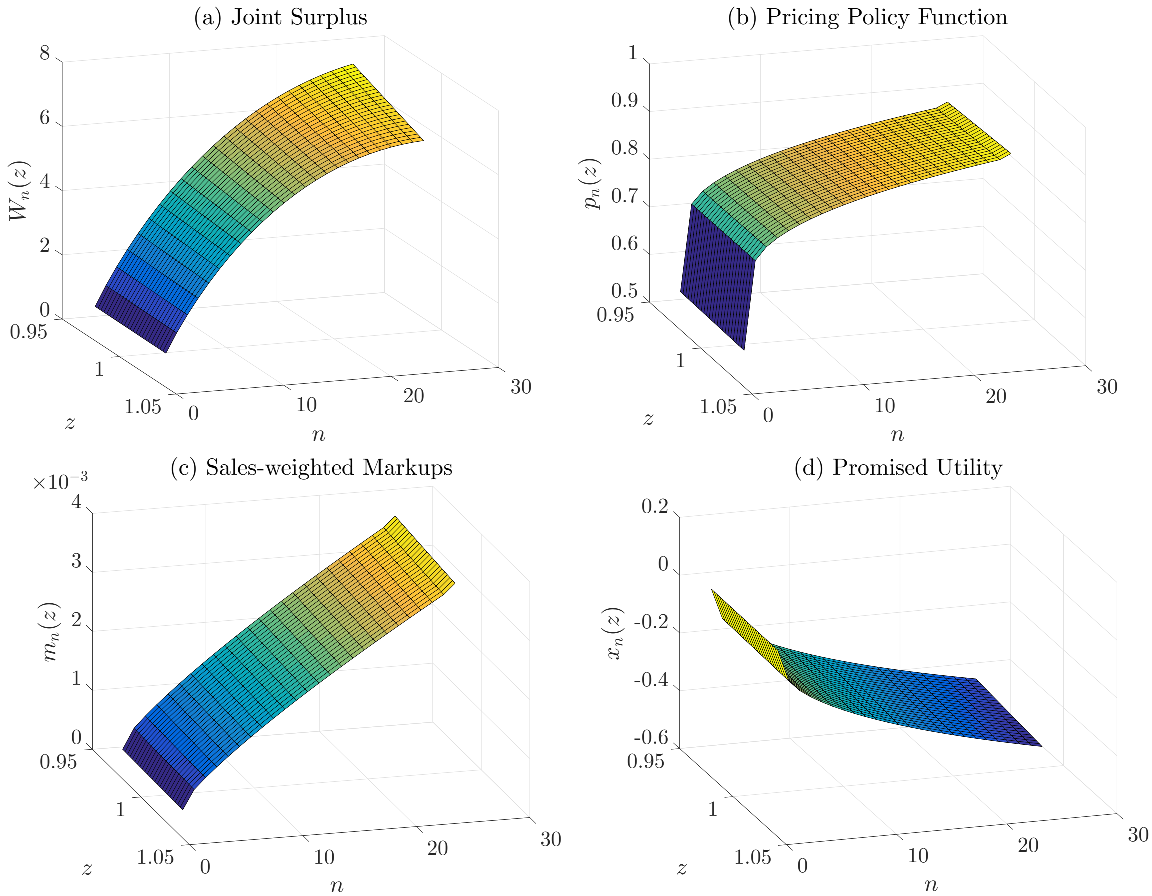

Figure A.5: Joint surplus function, pricing policy function, sales-weighted markups (equation (30)), and promised utility, in the (n, z) space, for the calibrated set of parameters. Higher values of z mean higher costs per customer (i.e. lower productivity).

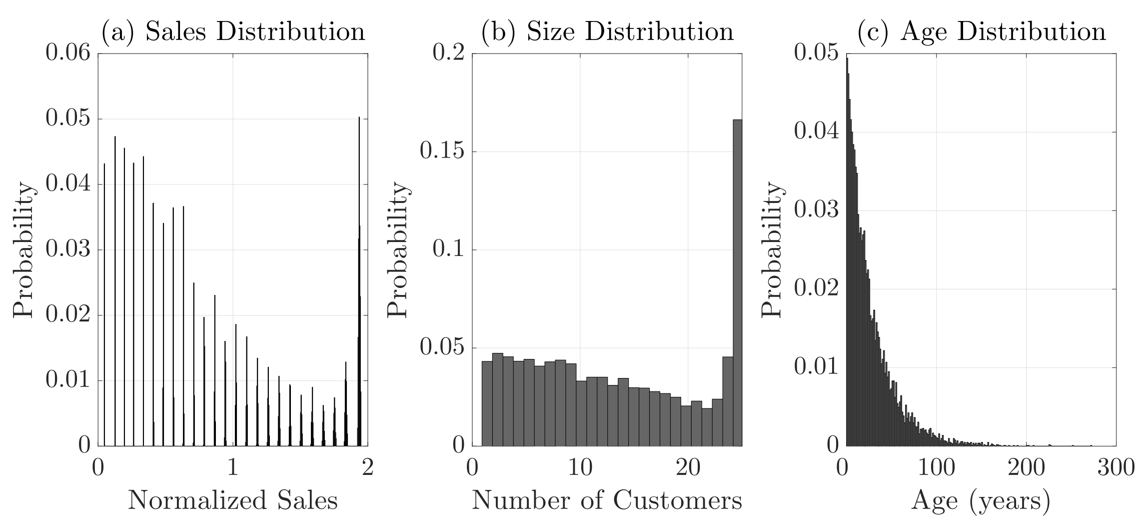

Figure A.6: Sales, customer, and seller age distributions, for the simulated economy under the calibrated set of parameters. Sales have been normalized by their mean.

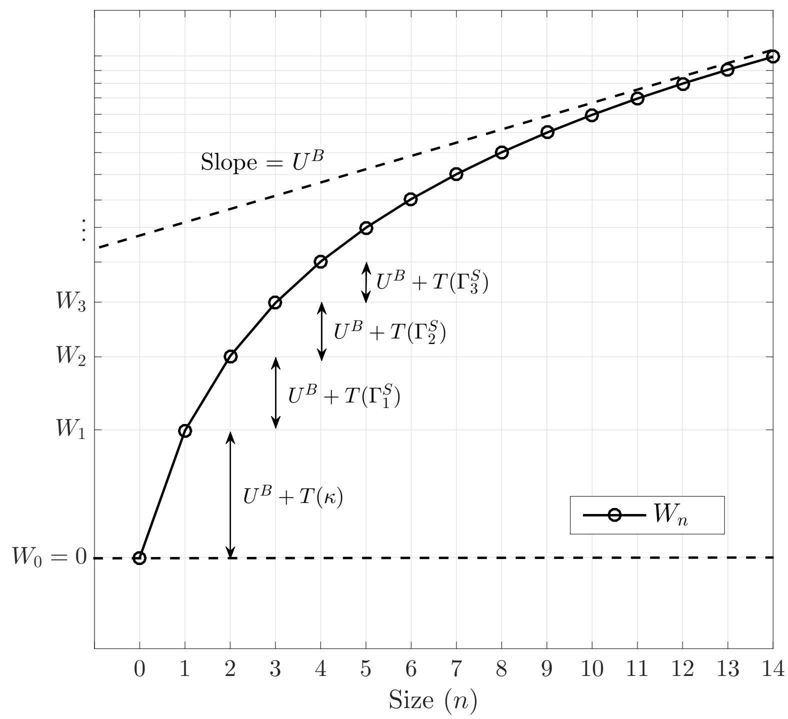

Figure A.7: Depiction of Proposition 4: Equilibrium joint surplus $\{W_n\}$ across n, for a fixed exogenous state (z, φ), where T is the mapping $T : x \mapsto \left(\frac{\Gamma^B}{\gamma}\right)^{\gamma} \left(\frac{x}{1-\gamma}\right)^{1-\gamma}$.

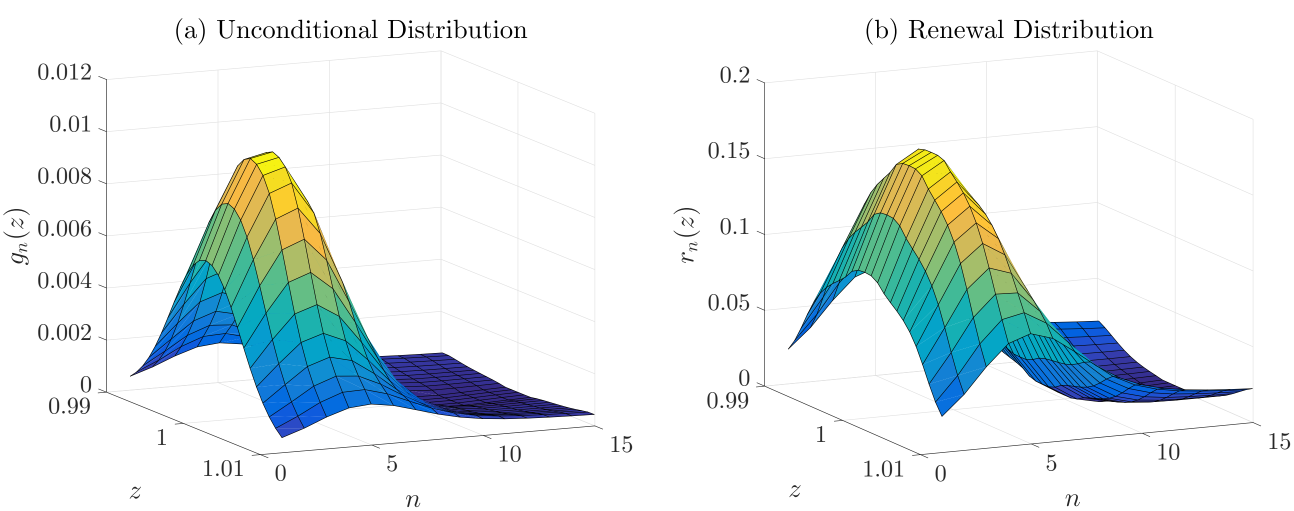

Figure A.8: Unconditional seller distribution (left) and renewal distribution (right), in the (n, z) space, for a numerical example. Note that the renewal density places more mass on smaller firms.

B Omitted Proofs

B.1 Proof of Proposition 1: Joint Surplus Problem

Proof. Denote by $\overline{\boldsymbol{\omega}} = \Big\{p, \{\mathbf{x}'(\mathbf{s}'), x'_+(z,\varphi), x'_-(z,\varphi)\}\Big\}$ a generic policy of the typical seller, where $\mathbf{x}'(\mathbf{s}') = \{x'(z',\varphi), x'(z,\varphi')\}$ stands for the promised utility under size n and a new state $\mathbf{s}' \neq (z,\varphi)$ (recall that $x'(z,\varphi) = x$ by stationarity), and $x'_+(z,\varphi)$ and $x'_-(z,\varphi)$ are the upsizing and downsizing choices, respectively. Then, the value of the seller in equilibrium, $V^S(n,x;\mathbf{s})$, can be written as the maximand on the right-hand side of (6), evaluated at $\overline{\boldsymbol{\omega}}$. That is:

$$V^S(n,x;z,\varphi) = \max_{\overline{\boldsymbol{\omega}}\in\Omega} \widetilde{V}^S(n;z,\varphi|\overline{\boldsymbol{\omega}}) \quad \text{s.t. } x \leq V^B(n,\overline{\boldsymbol{\omega}};z,\varphi)$$

where $\widetilde{V}^S(n;z,\varphi|\overline{\boldsymbol{\omega}})$ is given by:[60]

$$\begin{aligned}\widetilde{V}^S(n;z,\varphi|\overline{\boldsymbol{\omega}}) := \frac{1}{\rho(n;z,\varphi)}\Bigg[&pn - \mathcal{C}(n;z,\varphi) + \eta\Big(\theta\big(x'_+(z,\varphi);\varphi\big)\Big)V^S\Big(n+1,x'_+(z,\varphi);z,\varphi\Big) + n\delta_c V^S\Big(n-1,x'_-(z,\varphi);z,\varphi\Big)\\ &+ \sum_{z'\in\mathcal{Z}}\lambda_z(z'|z)V^S\Big(n,x'(z',\varphi);z',\varphi\Big) + \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)V^S\Big(n,x'(z,\varphi');z,\varphi'\Big)\Bigg] \quad \text{(B.1.1)}\end{aligned}$$

and we have defined $\rho(n;z,\varphi) := r + \delta_f + n\delta_c + \eta\big(\theta(x'_+(z,\varphi);\varphi)\big)$ as the *effective* discount rate of the firm, which adjusts the actual discount rate (r) for the death, separation, and growth rates faced by the agents.

From (B.1.1), it is clear that, for any given policy $\overline{\boldsymbol{\omega}}$, it is always optimal to offer the highest possible price that is consistent with promise-keeping. Indeed, the price has no bearing on the agents' incentives within the search market. Therefore, the PK constraint must bind with equality, and we can solve for the price p such that $x = V^B(n,\overline{\boldsymbol{\omega}};z,\varphi)$ using equation (5):

$$\begin{aligned}p^{PK}\Big(\big\{\mathbf{x}'(\mathbf{s}'),x'_+(z,\varphi),x'_-(z,\varphi)\big\}\Big) = {}& v(\varphi) - \rho(n;z,\varphi)x + \delta_f U^B(\varphi) + \eta\Big(\theta\big(x'_+(z,\varphi);\varphi\big)\Big)x'_+(z,\varphi) \quad \text{(B.1.2)}\\ &+ \delta_c\Big(U^B(\varphi) + (n-1)x'_-(z,\varphi)\Big) + \sum_{z'\in\mathcal{Z}}\lambda_z(z'|z)x'(z',\varphi) + \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)x'(z,\varphi')\end{aligned}$$

Intuitively, other things equal, the price level is higher when promised future utilities are higher and when the discounted value of the buyer is lower, since extracting surplus from the buyer today must be compensated in the future because of promise-keeping.

Using the above notation, we can now substitute the price (B.1.2) into the seller's value (B.1.1). After some straightforward algebra, we obtain:

$$\begin{aligned}\widetilde{W}(n,x;z,\varphi|\overline{\boldsymbol{\omega}}) = \frac{1}{\rho(n;z,\varphi)}\Bigg[&n\bigg(v(\varphi) + (\delta_f+\delta_c)U^B(\varphi)\bigg) - \bigg(\mathcal{C}(n;z,\varphi) + \eta\Big(\theta\big(x'_+(z,\varphi);\varphi\big)\Big)x'_+(z,\varphi)\bigg) \quad \text{(B.1.3)}\\ &+ \eta\Big(\theta\big(x'_+(z,\varphi);\varphi\big)\Big)W\Big(n+1,x'_+(z,\varphi);z,\varphi\Big) + n\delta_c W\Big(n-1,x'_-(z,\varphi);z,\varphi\Big)\\ &+ \sum_{z'\in\mathcal{Z}}\lambda_z(z'|z)W\Big(n,x'(z',\varphi);z',\varphi\Big) + \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)W\Big(n,x'(z,\varphi');z,\varphi'\Big)\Bigg]\end{aligned}$$

where we have defined:

$$\widetilde{W}(n,x;z,\varphi|\overline{\boldsymbol{\omega}}) := \widetilde{V}^S(n;z,\varphi|\overline{\boldsymbol{\omega}}) + nx \quad \text{and} \quad W(n,x;z,\varphi) := \max_{\overline{\boldsymbol{\omega}}\in\Omega}\widetilde{W}(n,x;z,\varphi|\overline{\boldsymbol{\omega}})$$

as the joint surplus under contract $\overline{\boldsymbol{\omega}}$, and the *maximized* joint surplus, respectively. Next, note that the right-

[60] Here we are arguing by free-entry that $V_0^S(\varphi) = 0$, $\forall\varphi\in\Phi$, to simplify the expression for $\widetilde{V}^S$. Moreover, we use that $\sum_{z'\in\mathcal{Z}}\lambda_z(z'|z) = \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi) = 0$.

hand side of equation (B.1.3) does not depend on x nor p, and so we can write the joint surplus under a given policy as :

$$\widetilde{W}(n,x;z,\varphi|\overline{\boldsymbol{\omega}}) = \widetilde{W}_n\Big(\big\{\mathbf{x}'(\mathbf{s}'), x'_+(z,\varphi), x'_-(z,\varphi)\big\}\Big)$$

This proves Part 2 of the proposition. Part 1 now readily follows. Since the joint surplus is invariant to the price level by construction, the optimal contract can be found by splitting the program into two separate stages. In the first stage, the seller chooses the continuation values $\big\{\mathbf{x}'(z,\varphi'), x'_+(z,\varphi), x'_-(z,\varphi)\big\}$ that maximize (B.1.3). In the second stage, once the surplus has been maximized, the seller chooses the promise-compatible price level (via equation (B.1.2)) that is consistent with the promised continuation utilities. Formally, the optimal contract $\boldsymbol{\omega}^*$ satisfies:

$$\boldsymbol{\omega}^* = \Big\{p^*, \big\{\mathbf{x}'^*(\mathbf{s}'), x'^*_+(z,\varphi), x'^*_-(z,\varphi)\big\}\Big\}$$

where

$$\big\{\mathbf{x}'^*(\mathbf{s}'), x'^*_+(z,\varphi), x'^*_-(z,\varphi)\big\} := \arg\max\ \widetilde{W}_n\Big(\big\{\mathbf{x}'(\mathbf{s}'), x'_+(z,\varphi), x'_-(z,\varphi)\big\}\Big) \tag{B.1.4a}$$

$$p^* := p^{PK}\Big(\big\{\mathbf{x}'^*(\mathbf{s}'), x'^*_+(z,\varphi), x'^*_-(z,\varphi)\big\}\Big) \tag{B.1.4b}$$

By expressing the problem of the seller in terms of $\widetilde{W}$, we have just shown that the contract that is optimally chosen by the firm, $\boldsymbol{\omega}^*$, must maximize the joint surplus. Conversely, for any set $\big\{\mathbf{x}'(\mathbf{s}'), x'_+(z,\varphi), x'_-(z,\varphi)\big\}$ of continuation values that maximizes the joint surplus, there is a price level, given by $p = p^{PK}\big(\big\{\mathbf{x}'(\mathbf{s}'), x'_+(z,\varphi), x'_-(z,\varphi)\big\}\big)$, that maximizes the seller's value subject to the PK constraint. Therefore, the seller's problem (equation (6)) and the joint surplus problem (equations (B.1.4a)-(B.1.4b)) are equivalent. □

B.2 Proof of Proposition 2: Invariant Distribution

Proof. Let $\big\{\theta_n(z,\varphi) : (n,z,\varphi) \in \mathcal{N} \times \mathcal{Z} \times \Phi\big\}$ be an equilibrium collection of market tightness levels, where $\mathcal{N} = \{1,\dots,\overline{n}\}$, and $\overline{n} < +\infty$ is a large integer. In matrix notation, for each aggregate state $\varphi \in \Phi$, the dynamical system can be written as:

$$\boldsymbol{\partial}_t \mathbf{S}_t(\varphi) = \mathbf{T}_\varphi \mathbf{S}_t(\varphi) \tag{B.2.1}$$

where $\mathbf{S}_t(\varphi) := \big(S_{0,t}(\varphi), \mathbf{S}_{1,t}^\top, \dots, \mathbf{S}_{\overline{n},t}^\top\big)^\top$, with $\mathbf{S}_{n,t} := \big(S_{n,t}(z_1), \dots, S_{n,t}(z_{k_z})\big)^\top$, and $\mathbf{T}_\varphi$ is the partitioned matrix:

$$\mathbf{T}_\varphi := \begin{pmatrix} t_{11} & \boldsymbol{\delta}_f^e + \boldsymbol{\delta}_c^e & \boldsymbol{\delta}_f^e & \boldsymbol{\delta}_f^e & \cdots & \boldsymbol{\delta}_f^e & \boldsymbol{\delta}_f^e & \boldsymbol{\delta}_f^e \\ \boldsymbol{\eta}_1^e(\varphi)^\top & \mathbf{D}_1(\varphi) & \boldsymbol{\delta}_{2,c} & \mathbf{0}_{k_z:k_z} & \cdots & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} \\ \mathbf{0}_{k_z:1} & \boldsymbol{\eta}_2(\varphi) & \mathbf{D}_2(\varphi) & \boldsymbol{\delta}_{3,c} & \cdots & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0}_{k_z:1} & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} & \cdots & \mathbf{D}_{\overline{n}-2}(\varphi) & \boldsymbol{\delta}_{\overline{n}-1,c} & \mathbf{0}_{k_z:k_z} \\ \mathbf{0}_{k_z:1} & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} & \cdots & \boldsymbol{\eta}_{\overline{n}-1}(\varphi) & \mathbf{D}_{\overline{n}-1}(\varphi) & \boldsymbol{\delta}_{\overline{n},c} \\ \mathbf{0}_{k_z:1} & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} & \mathbf{0}_{k_z:k_z} & \cdots & \mathbf{0}_{k_z:k_z} & \boldsymbol{\eta}_{\overline{n}}(\varphi) & \mathbf{D}_{\overline{n}}(\varphi) \end{pmatrix}$$

where $t_{11} := -\sum_z \pi_z(z)\eta(\theta_1(z,\varphi))$ is a scalar, $\mathbf{0}_{p:q}$ denotes a $p \times q$ matrix of zeros, and $\mathbf{T}_\varphi$ is a $K \times K$ square matrix, where $K = 1 + \overline{n}k_z$. Further, we have defined the following $1 \times k_z$ row vectors:

$$\boldsymbol{\delta}_f^e = \Big(\delta_f, \dots, \delta_f\Big); \quad \boldsymbol{\delta}_c^e = \Big(\delta_c, \dots, \delta_c\Big); \quad \boldsymbol{\eta}_1^e(\varphi) = \Big(\pi_z(z_1)\eta\big(\theta_1(z_1,\varphi)\big), \dots, \pi_z(z_{k_z})\eta\big(\theta_1(z_{k_z},\varphi)\big)\Big);$$

and the following $k_z \times k_z$ matrices:

$$\forall n = 2,\dots,\overline{n}: \qquad \boldsymbol{\delta}_{n,c} = \operatorname{diag}\Big(n\delta_c, \dots, n\delta_c\Big);$$

$$\boldsymbol{\eta}_n(\varphi) = \text{diag}\Big(\eta\big(\theta_n(z_1, \varphi)\big), \ldots, \eta\big(\theta_n(z_{k_z}, \varphi)\big)\Big);$$

$$\forall n = 1, \ldots, \overline{n}: \quad \mathbf{D}_n(\varphi) = \begin{pmatrix} d_n(z_1, \varphi) & \lambda_z(z_1|z_2) & \ldots & \lambda_z(z_1|z_{k_z}) \\ \lambda_z(z_2|z_1) & d_n(z_2, \varphi) & \ldots & \lambda_z(z_2|z_{k_z}) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_z(z_{k_z}|z_1) & \lambda_z(z_{k_z}|z_2) & \ldots & d_n(z_{k_z}, \varphi) \end{pmatrix}$$

where the diagonal elements of $\mathbf{D}_n(\varphi)$ are given by:

$$d_n(z_j, \varphi) = \begin{cases} -\Big(\delta_f + n\delta_c + \eta\big(\theta_{n+1}(z_j, \varphi)\big) + \sum_{\ell \neq j} \lambda_z(z_\ell|z_j)\Big) & \text{for } n = 1, \ldots, \overline{n} - 1 \\ -\Big(\delta_f + \overline{n}\delta_c + \sum_{\ell \neq j} \lambda_z(z_\ell|z_j)\Big) & \text{for } n = \overline{n} \end{cases}$$

(Recall that a graphical depiction of these transitions can be found in Figure A.2). Generally, system (B.2.1) describes an *irreducible* Markov chain, as any state $(n', z') \in \mathcal{N} \times \mathcal{Z}$ can be reached almost surely from some $(n, z) \neq (n', z')$. Moreover, the Markov chain is *aperiodic*. These properties, plus the fact that the state space is finite, guarantee that the Markov chain is *ergodic*. Therefore, by Theorem 11.2 of Stokey and Lucas (1989), the system converges to a unique steady-state distribution $\mathbf{S}^*(\varphi)$, for each $\varphi \in \Phi$.

More specifically, note that, thanks to the block-recursivity property, the equilibrium policies are not explicitly indexed by time (their time variation being fully encoded in the dependence to the aggregate state φ), so $\mathbf{T}_\varphi$ is constant. This means that we can solve the differential equation (B.2.1) directly. The solution is:

$$\mathbf{S}_t(\varphi) = e^{\mathbf{T}_\varphi t}\mathbf{S}_0(\varphi)$$

where the initial distribution $\mathbf{S}_0(\varphi) \in \mathbb{R}_+^K$ is given. To compute $e^{\mathbf{T}_\varphi t}$, consider the eigenvalue decomposition $\mathbf{T}_\varphi = \mathbf{E}_\varphi \boldsymbol{\Lambda}_\varphi \mathbf{E}_\varphi^{-1}$, where $\boldsymbol{\Lambda}_\varphi := \big(\lambda_1(\varphi), \ldots, \lambda_K(\varphi)\big)$ is the diagonal matrix of eigenvalues, and $\mathbf{E}_\varphi$ collects the corresponding eigenvectors. Defining $\mathbf{Z}_t(\varphi) := \mathbf{E}_\varphi^{-1}\mathbf{S}_t(\varphi)$, then $\boldsymbol{\partial}_t \mathbf{Z}_t(\varphi) = \boldsymbol{\Lambda}_\varphi \mathbf{Z}_t(\varphi)$, and because $\boldsymbol{\Lambda}_\varphi$ is a diagonal matrix, we can solve this differential equation element-wise, i.e. $\boldsymbol{\partial}_t Z_{i,t}(\varphi) = \lambda_i(\varphi) Z_{i,t}(\varphi)$ for each $i = 1, \ldots, K$. This is a simple system of ODEs with solution:

$$Z_{i,t}(\varphi) = c_i e^{\lambda_i(\varphi)t}, \quad i = 1, \ldots, K$$

where $c_i \in \mathbb{R}$ is a constant of integration. Since $\mathbf{S}_t(\varphi) = \mathbf{E}_\varphi \mathbf{Z}_t(\varphi)$, we have obtained:

$$\mathbf{S}_t(\varphi) = \sum_{i=1}^{K} c_i e^{\lambda_i(\varphi)t} \boldsymbol{v}_i \tag{B.2.2}$$

where $\boldsymbol{v}_i$ is the $K \times 1$ eigenvector associated to the i-th eigenvalue. Therefore, the stability of system (B.2.2) as $t \to +\infty$ depends on the sign of the eigenvalues of $\mathbf{T}_\varphi$. The trace of $\mathbf{T}_\varphi$ is:

$$tr(\mathbf{T}_\varphi) = \sum_{i=1}^{K} \lambda_i(\varphi) = -\sum_{j=1}^{k_z} \pi_z(z_j)\eta(\theta_1(z_j, \varphi)) + \sum_{n=1}^{\overline{n}} \sum_{j=1}^{k_z} d_n(z_j) < 0$$

The trace being unambiguously negative means that there is at least one negative eigenvalue, if not more. Letting $1 \leq \ell \leq K$ denote the number of negative eigenvalues, and re-ordering the eigenvalues from small to large with no loss of generality, we can then impose $c_j = 0, \forall j \in \{\ell+1, \ell+2, \ldots, K\}$, on equation (B.2.2), and let $t \to +\infty$ to find the stable solution. That is:

$$\mathbf{S}^*(\varphi) := \lim_{t \to +\infty} \sum_{j=1}^{\ell} c_j e^{\lambda_j(\varphi)t} \boldsymbol{v}_j \in \mathbb{R}_+^K$$

is the unique invariant distribution of sellers in state $\varphi \in \Phi$. □

B.3 Proof of Proposition 3: Efficiency

Proof. Consider a benevolent planner that is constrained by the search frictions of the economy and seeks to maximize aggregate welfare subject to the resource constraints of the economy. The planner can allocate resources freely, so the problem does not feature contracts or prices. Instead, we label each sub-market directly by its tightness, θ. To simplify notation, it is understood that time subscripts embody the entire history of aggregate shocks, which is taken as some arbitrary path $\varphi^t = (\varphi_j : j \leq t) \subseteq \Phi$.

The planner chooses time series for the tightness in each market segment, $\mathbf{\Theta}_t \equiv \{\theta_{n_t,t}(z_t) : (n_t, z_t) \in \mathbb{N} \times \mathcal{Z}\}$; distributions of inactive and active buyers across markets, $\mathbf{B}_t^I \equiv \{B_{n_t,t}^I(z_t) : (n_t, z_t) \in \mathbb{N} \times \mathcal{Z}\}$ and $\mathbf{B}_t^A \equiv \{B_{n_t,t}^A(z_t) : (n_t, z_t) \in \mathbb{N} \times \mathcal{Z}\}$; a measure of potential entrants $S_{0,t}$; and a distribution of firms across states, $\mathbf{S}_t \equiv \{S_{n_t,t}(z_t) : (n_t, z_t) \in \mathbb{N} \times \mathcal{Z}\}$, to maximize aggregate welfare in the economy:

$$\max_{\substack{\mathbf{\Theta}_t, \mathbf{B}_t^I, \mathbf{B}_t^A \\ S_{0,t}, \mathbf{S}_t}} \mathbb{E}_0 \int_0^{+\infty} e^{-rt} \Big\{ -\kappa(\varphi_t) S_{0,t} + \sum_{n_t=1}^{+\infty} \sum_{z_t \in \mathcal{Z}} \Big[v(\varphi_t) B_{n_t,t}^A(z_t) - \mathcal{C}(n_t; z_t, \varphi_t) S_{n_t,t}(z_t) - c(\varphi_t) B_{n_t,t}^I(z_t) \Big] \Big\} dt \tag{B.3.1}$$

In words, total welfare is equal to the present discounted value of aggregate consumption gains for active buyers, net of production operating costs for active sellers,[61] and net of the search and entry costs for inactive buyers and inactive sellers.

The planner is subject to three sets of constraints. The first one concerns the evolution of the firm distribution, which we described in Section 3.5 and reproduce again here for convenience:

$$\partial_t S_{0,t} = \delta_f \sum_{n_t=1}^{+\infty} \sum_{z_t \in \mathcal{Z}} S_{n_t,t}(z_t) + \delta_c \sum_{z_t \in \mathcal{Z}} S_{1,t}(z_t) - \sum_{z^e \in \mathcal{Z}} \pi_z(z^e) \eta\big(\theta_{1,t}(z^e)\big) S_{0,t} \tag{B.3.2a}$$

$$\begin{aligned} \partial_t S_{1,t}(z_t) = {} & \pi_z(z_t) \eta\big(\theta_{1,t}(z_t)\big) S_{0,t} + 2\delta_c S_{2,t}(z_t) + \sum_{\tilde{z} \neq z_t} \lambda_z(z_t|\tilde{z}) S_{1,t}(\tilde{z}) \\ & - \Big(\delta_f + \delta_c + \eta\big(\theta_{2,t}(z_t)\big) + \sum_{\tilde{z} \neq z_t} \lambda_z(\tilde{z}|z_t)\Big) S_{1,t}(z_t) \end{aligned} \tag{B.3.2b}$$

$$\begin{aligned} \forall n_t \geq 2 : \quad \partial_t S_{n_t,t}(z_t) = {} & \eta\big(\theta_{n_t,t}(z_t)\big) S_{n_t-1,t}(z_t) + (n_t+1)\delta_c S_{n_t+1,t}(z_t) + \sum_{\tilde{z} \neq z_t} \lambda_z(z_t|\tilde{z}) S_{n_t,t}(\tilde{z}) \\ & - \Big(\delta_f + n_t \delta_c + \eta\big(\theta_{n_t+1,t}(z_t)\big) + \sum_{\tilde{z} \neq z_t} \lambda_z(\tilde{z}|z_t)\Big) S_{n_t,t}(z_t); \end{aligned} \tag{B.3.2c}$$

for all $z_t \in \mathcal{Z}$, where z^e denotes the productivity draw upon entry. Note that implicit in these equations underlies the evolution of firms' size (i.e. the law of motion for n_t).

The second set of constraints describes the distribution of buyers across firms at any given time:

$$\forall (n_t, z_t) \in \mathbb{N} \times \mathcal{Z} : \quad B_{n_t,t}^A(z_t) = n_t S_{n_t,t}(z_t) \tag{B.3.3a}$$

$$\forall (n_t, z_t) \in \mathbb{N} \times \mathcal{Z} : \quad B_{n_t,t}^I(z_t) = \theta_{n_t,t}(z_t) S_{n_t-1,t}(z_t) \tag{B.3.3b}$$

$$1 = \sum_{n_t=1}^{+\infty} \sum_{z_t \in \mathcal{Z}} \Big(B_{n_t,t}^A(z_t) + B_{n_t,t}^I(z_t) \Big) \tag{B.3.3c}$$

Equation (B.3.3a) states that each customer consumes a single unit; equation (B.3.3b) states that each market segment is in equilibrium, in the sense that the measure of buyers who find a firm in any given market equals the measure of firms within that market who find a new customer; equation (B.3.3c) says that every buyer in the economy is in either the active or the inactive state.

Finally, the mass of potential entering firms needs to be non-negative in any aggregate state of the world:

$$S_{0,t} \geq 0 \tag{B.3.4}$$

The planning problem consists on maximizing (B.3.1) subject to the seven constraints listed above. We begin by simplifying the dimensionality of the problem. First, we use constraints (B.3.3a) and (B.3.3b) to rewrite (B.3.3c)

[61] The sum of these first two terms thus equals the total gains from trade.

as:

$$\sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}} n_t S_{n_t,t}(z_t) + \sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}} \theta_{n_t+1,t}(z_t) S_{n_t,t}(z_t) + S_{0,t}\sum_{z_t\in\mathcal{Z}}\theta_{1,t}(z_t) = 1 \tag{B.3.5}$$

Substituting constraints (B.3.3a) and (B.3.3b) into the objective function, we are left with the following problem:

$$\begin{aligned}\max_{\boldsymbol{\Theta}_t, S_{0,t}, \mathbf{S}_t} \mathbb{E}_0 \int_0^{+\infty} e^{-rt}\Bigg\{ &-\Bigg(\kappa(\varphi_t) + c(\varphi_t)\sum_{z_t\in\mathcal{Z}}\theta_1(z_t)\Bigg) S_{0,t} + v(\varphi_t)\sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}} n_t S_{n_t,t}(z_t) - \sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}}\mathcal{C}(n_t; z_t,\varphi_t) S_{n_t,t}(z_t) \\ &- c(\varphi_t)\sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}}\theta_{n_t+1,t}(z_t) S_{n_t,t}(z_t)\Bigg\}\mathrm{d}t\end{aligned}$$

subject to constraints (B.3.2a), (B.3.2b), (B.3.2c), (B.3.4), and (B.3.5). Conveniently, the variables $\mathbf{B}_t^I$ and $\mathbf{B}_t^A$ have disappeared from the problem, and the state vector has been reduced to the measures of firms: $\mathbb{S}_t \equiv [S_{0,t}, \mathbf{S}_t]$.

To solve the simplified planner's problem, we use standard tools from Optimal Control theory, where $\boldsymbol{\Theta}_t$ is the control variable. The current-value Hamiltonian of the simplified planning problem is:

$$\begin{aligned}\mathcal{H}_t(\boldsymbol{\Theta_t}; \mathbb{S}_t) := &-\Bigg(\kappa(\varphi_t) + c(\varphi_t)\sum_{z_t\in\mathcal{Z}}\theta_1(z_t)\Bigg) S_{0,t} + v(\varphi_t)\sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}} n_t S_{n_t,t}(z_t) - \sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}}\mathcal{C}(n_t; z_t,\varphi_t) S_{n_t,t}(z_t) \\ &- c(\varphi_t)\sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}}\theta_{n_t+1,t}(z_t) S_{n_t,t}(z_t) \\ &+ \phi_t\Bigg[1 - \sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}} n_t S_{n_t,t}(z_t) - \sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}}\theta_{n_t+1,t}(z_t) S_{n_t,t}(z_t) - S_{0,t}\sum_{z_t\in\mathcal{Z}}\theta_{1,t}(z_t)\Bigg] \\ &+ \psi_{0,t}\Bigg[\delta_f\sum_{n_t=1}^{+\infty}\sum_{z_t\in\mathcal{Z}} S_{n_t,t}(z_t) + \delta_c\sum_{z_t\in\mathcal{Z}} S_{1,t}(z_t) - \sum_{z^e\in\mathcal{Z}}\pi_z(z^e)\eta\big(\theta_{1,t}(z^e)\big) S_{0,t}\Bigg] \\ &+ \sum_{z_t\in\mathcal{Z}}\Bigg\{\psi_{1,t}(z_t)\Bigg[\pi_z(z_t)\eta\big(\theta_{1,t}(z_t)\big) S_{0,t} + 2\delta_c S_{2,t}(z_t) + \sum_{\tilde{z}\neq z_t}\lambda_z(z_t|\tilde{z}) S_{1,t}(\tilde{z}) \\ &\qquad - \Big(\delta_f + \delta_c + \eta\big(\theta_{2,t}(z_t)\big) + \sum_{\tilde{z}\neq z_t}\lambda_z(\tilde{z}|z_t)\Big) S_{1,t}(z_t)\Bigg] \\ &\quad + \sum_{n_t=2}^{+\infty}\psi_{n_t,t}(z_t)\Bigg[\eta\big(\theta_{n_t,t}(z_t)\big) S_{n_t-1,t}(z_t) + (n_t+1)\delta_c S_{n_t+1,t}(z_t) + \sum_{\tilde{z}\neq z_t}\lambda_z(z_t|\tilde{z}) S_{n_t,t}(\tilde{z}) \\ &\qquad - \Big(\delta_f + n_t\delta_c + \eta\big(\theta_{n_t+1,t}(z_t)\big) + \sum_{\tilde{z}\neq z_t}\lambda_z(\tilde{z}|z_t)\Big) S_{n_t,t}(z_t)\Bigg]\Bigg\} + \vartheta_t S_{0,t}\end{aligned}$$

where $\psi_{n,t}(z) \geq 0$, $n \geq 1$ (respectively, $\psi_{0,t} \geq 0$) is the co-state variable on the flow equation for $S_{n,t}(z)$ (respectively, $S_{0,t}$); $\phi_t \geq 0$ is the multiplier on (B.3.5); and $\vartheta_t \geq 0$ is the multiplier on the non-negative entry condition, where the corresponding complementary slackness hold.

In vector notation, the necessary conditions for optimality are:

$$\begin{aligned}\nabla_{\boldsymbol{\Theta}}\mathcal{H}_t(\boldsymbol{\Theta_t}; \mathbb{S}_t) &= \mathbf{0} \\ \nabla_{\mathbb{S}}\mathcal{H}_t(\boldsymbol{\Theta_t}; \mathbb{S}_t) &= -\nabla_t\boldsymbol{\psi}_t + r\boldsymbol{\psi}_t\end{aligned}$$

where ∇ denotes the gradient operator, and $\boldsymbol{\psi}_t$ is a stacked vector of co-state variables. These conditions are also sufficient because the Hamiltonian is quasi-concave. Indeed, the objective function is linear in both control and state variables, and because of Assumption 2 establishing concavity of η, all the constraints are concave in the control and linear in the states.

Regarding the first set of optimality conditions, for given $z_t \in \mathcal{Z}$ we have:

$$[\theta_1]: \qquad \phi_t + c(\varphi_t) = \Big(\psi_{1,t}(z_t) - \psi_{0,t}\Big)\pi_z(z_t)\frac{\partial \eta(\theta)}{\partial \theta}\Big|_{\theta=\theta_{1,t}(z_t)} \tag{B.3.6a}$$

$$[\theta_n : n \geq 2]: \qquad \phi_t + c(\varphi_t) = \Big(\psi_{n_t,t}(z_t) - \psi_{n_t-1,t}(z_t)\Big)\frac{\partial \eta(\theta)}{\partial \theta}\Big|_{\theta=\theta_{n_t,t}(z_t)} \tag{B.3.6b}$$

As for the second set of conditions, we have:

$$\begin{aligned}
[S_0]: \quad -\boldsymbol{\partial}_t\psi_{0,t} + r\psi_{0,t} =& -\kappa(\varphi_t) - \big(\phi_t + c(\varphi_t)\big)\sum_{z_t\in\mathcal{Z}}\theta_1(z_t) && \text{(B.3.7a)}\\
&+ \sum_{z^e\in\mathcal{Z}}\pi_z(z^e)\eta\big(\theta_{1,t}(z^e)\big)\psi_{1,t}(z^e) - \psi_{0,t}\sum_{z^e\in\mathcal{Z}}\pi_z(z^e)\eta\big(\theta_{1,t}(z^e)\big) + \vartheta_t\\
[S_{n_t}(z_t)]: \quad -\boldsymbol{\partial}_t\psi_{n_t,t}(z_t) + r\psi_{n_t,t}(z_t) =& \, n_t\big(v(\varphi_t) - \phi_t\big) - \big(\phi_t + c(\varphi_t)\big)\theta_{n_t+1,t}(z_t) - \mathcal{C}(n_t, z_t;\varphi_t) && \text{(B.3.7b)}\\
&+ \delta_f\Big(\psi_{0,t} - \psi_{n_t,t}(z_t)\Big) + n_t\delta_c\Big(\psi_{n_t-1,t}(z_t) - \psi_{n_t,t}(z_t)\Big)\\
&+ \eta\big(\theta_{n_t+1}(z_t)\big)\Big(\psi_{n_t+1,t}(z_t) - \psi_{n_t,t}(z_t)\Big) + \sum_{\tilde{z}\in\mathcal{Z}}\lambda_z(\tilde{z}|z_t)\Big(\psi_{n_t,t}(\tilde{z}) - \psi_{n_t,t}(z_t)\Big)
\end{aligned}$$

for given $z_t \in \mathcal{Z}$, where in the last line we have used that

$$\lambda_z(z|z) = -\sum_{\tilde{z}\neq z}\lambda_z(\tilde{z}|z)$$

for all $z \in \mathcal{Z}$, by the properties of the Markov chain. We will now show that a block-recursive equilibrium with non-negative entry of firms satisfies the optimality conditions of the planner by appropriately choosing the co-state variables of the planning problem. Note that equations (B.3.7a)-(B.3.7b) show that the co-state variables can be represented as HJB equations. Moreover, (B.3.6a)-(B.3.6b) are the corresponding first order conditions of those equations. Therefore, it suffices to find the values of the multipliers for which the HJB equations of the planner coincide with the joint surplus problem of the decentralized allocation.

Pick a decentralized equilibrium allocation $\big\{W_n(z,\varphi), x_n(z,\varphi), \theta_n(z,\varphi), U^B(\varphi) : (n,z,\varphi) \in \mathbb{N}\times\mathcal{Z}\times\Phi\big\}$, and consider the following realization for the multipliers:

$$\begin{aligned}
\phi_t(\varphi^t) &= rU^B(\varphi_t)\\
\psi_{0,t}(\varphi^t) &= 0\\
\forall n_t, z_t: \quad \psi_{n_t,t}(z_t,\varphi^t) &= W_{n_t}(z_t,\varphi_t) - n_tU^B(\varphi_t)
\end{aligned}$$

Under this guess, notice that $\boldsymbol{\partial}_t\psi_{0,t} = \boldsymbol{\partial}_t\psi_{n,t}(z_t) = 0$, $\forall n \geq 1$. Moreover, the multipliers depend only on the current realization of the aggregate state, and not on the entire history. Further, for a sufficiently low value of κ, we can impose strictly positive entry and therefore $\vartheta_t = 0$, $\forall t$.

Plugging these guesses into (B.3.7b), after some simple algebra we obtain:

$$\begin{aligned}
(r+\delta_f)W_{n_t}(z_t,\varphi_t) =& \, n_t\Big(v(\varphi_t) + (\delta_f+\delta_c)U^B(\varphi_t)\Big) - \mathcal{C}(n_t,z_t;\varphi_t)\\
&- \Big[\big(rU^B(\varphi_t) + c(\varphi_t)\big)\theta_{n_t+1,t}(z_t) + \eta\big(\theta_{n_t+1}(z_t)\big)U^B(\varphi_t)\Big]\\
&+ n_t\delta_c\Big(W_{n_t-1}(z_t,\varphi_t) - W_{n_t}(z_t,\varphi_t)\Big)\\
&+ \eta\big(\theta_{n_t+1}(z_t)\big)\Big(W_{n_t+1}(z_t,\varphi_t) - W_{n_t}(z_t,\varphi_t)\Big)\\
&+ \sum_{\tilde{z}\in\mathcal{Z}}\lambda_z(\tilde{z}|z_t)\Big(W_{n_t}(\tilde{z},\varphi_t) - W_{n_t}(z_t,\varphi_t)\Big)
\end{aligned}$$

Notice that the last equation resembles the maximized HJB equation for the joint surplus (equation (11)) except for the term highlighted in blue (second line). However, using that $\eta(\theta) = \theta\mu(\theta)$ and $x_{n+1}(z,\varphi) = U^B(\varphi) +$

$\frac{rU^B(\varphi)+c(\varphi)}{\mu(\theta_{n+1}(z,\varphi))}$ by inactive buyers' indifference, we obtain that this term is equal to:

$$\big(rU^B(\varphi_t)+c(\varphi_t)\big)\theta_{n_t+1,t}(z_t)+\eta\big(\theta_{n_t+1}(z_t)\big)U^B(\varphi_t)=\eta\left(\theta_{n_t+1,t}(z_t,\varphi_t)\right)x_{n_t+1,t}(z_t,\varphi_t) \tag{B.3.8}$$

Using this fact into the above equation and grouping terms, we then finally recognize the value of the joint surplus in the decentralized solution, equation (11).

Similarly, plugging the guess for the multipliers into (B.3.7a), we obtain:

$$\kappa(\varphi_t)=-\big(rU^B(\varphi_t)+c(\varphi_t)\big)\sum_{z_t\in\mathcal{Z}}\theta_1(z_t)+\sum_{z^e\in\mathcal{Z}}\pi_z(z^e)\eta\big(\theta_{1,t}(z^e)\big)\Big(W_1(z^e)-U^B(\varphi_t)\Big)$$

A final manipulation using (B.3.8) again then allows us to obtain the free entry condition in the decentralized allocation, equation (10).

Summing up, under an appropriate choice of the co-states, the planner's solution is equivalent to the problem of the decentralized economy. Therefore, the optimality conditions of the decentralized economy coincide with the first-order conditions of the planner (given by (B.3.6a)-(B.3.6b)), and the policies that we have obtained from the block-recursive equilibrium maximize aggregate welfare given our choice of the co-state multipliers. Hence, the equilibrium is efficient. □

B.4 Proof of Proposition 4: Joint Surplus Analytical Solution

Proof. The equilibrium allocation is composed of sequences $\big\{W_n(z,\varphi),x_n(z,\varphi),\theta_n(z,\varphi),p_n(z,\varphi):(n,z,\varphi)\in\mathbb{N}\times\mathcal{Z}\times\Phi\big\}$ satisfying equations (11), (12), (13), and (15), where the free entry condition:

$$\kappa(\varphi)=\sum_{z_0\in\mathcal{Z}}\pi_z(z_0)\eta\Big(\theta\big(x_1(z_0,\varphi);\varphi\big)\Big)\Big(W_1(z_0,\varphi)-x_1(z_0,\varphi)\Big) \tag{B.4.1}$$

must hold for all $\varphi\in\Phi$. Under the meeting rate $\mu(\theta)=\theta^{\gamma-1}$, equation (12) defines the following equilibrium mapping:

$$\theta:(x;\varphi)\mapsto\left(\frac{x-U^B(\varphi)}{\Gamma^B(\varphi)}\right)^{\frac{1}{1-\gamma}} \tag{B.4.2}$$

Some algebra shows that the joint surplus maximization rule (equation (13)) can be written as:

$$W_{n+1}(z,\varphi)-W_n(z,\varphi)-x_{n+1}(z,\varphi)=\frac{1-\gamma}{\gamma}\Big(x_{n+1}(z,\varphi)-U^B(\varphi)\Big) \tag{B.4.3}$$

for any $n\in\mathbb{N}$ and $(z,\varphi)\in\mathcal{Z}\times\Phi$. This equation reflects the relevant trade-offs in the equilibrium: when firms offer a higher value, they attract more buyers because the buyer's relative outside option worsens (right-hand side). Yet, the remaining value that accrues to the seller is also lower because part of the joint surplus is being transferred to the new customer (left-hand side). The joint-surplus maximizing rule splits the rents so that, for the marginal customer, these payoffs are equalized. Note that we can also write this condition as:

$$x_{n+1}(z,\varphi)-U^B(\varphi)=\gamma\underbrace{\Big(W_{n+1}(z,\varphi)-W_n(z,\varphi)-U^B(\varphi)\Big)}_{:=\Gamma_{n+1}(z,\varphi)}$$

showing that the buyer absorbs a fraction γ of the marginal gains from matching, $\Gamma_{n+1}(z,\varphi)$. Next, define:

$$\Gamma^S_n(z,\varphi):=(r+\delta_f)W_n(z,\varphi)-\pi_n(z,\varphi)+n\delta_c\Big(W_n(z,\varphi)-W_{n-1}(z,\varphi)\Big)-n(\delta_c+\delta_f)U^B(\varphi)-\Xi_n(z,\varphi) \tag{B.4.4}$$

where $\pi_n(z,\varphi):=nv(\varphi)-\mathcal{C}(n;z,\varphi)$ is the flow joint surplus, and

$$\Xi_n(z,\varphi):=\sum_{z'\in\mathcal{Z}}\lambda_z(z'|z)W_n(z',\varphi)+\sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)W_n(z,\varphi')$$

is the expected value of the joint surplus across exogenous states. Letting $\theta_{n+1}(z,\varphi):=\theta\big(x_{n+1}(z,\varphi);\varphi\big)$, note that:

$$
\begin{aligned}
\Gamma_n^S(z,\varphi) &= \eta\big(\theta_{n+1}(z,\varphi)\big)\Big(W_{n+1}(z,\varphi) - W_n(z,\varphi) - x_{n+1}(z,\varphi)\Big) \\
&= \left(\frac{1-\gamma}{\gamma}\right)\eta\big(\theta_{n+1}(z,\varphi)\big)\Big(x_{n+1}(z,\varphi) - U^B(\varphi)\Big) \\
&= \left(\frac{1-\gamma}{\gamma}\right)\theta_{n+1}(z,\varphi)\Gamma^B(\varphi)
\end{aligned}
$$

where the first line uses the HJB equation for the joint surplus (equation (11)), the second line uses (B.4.3), and the third line uses (B.4.2) and $\eta(\theta) = \theta\mu(\theta)$. The right-hand side of the first equality allows us to interpret Γ^S as the expected match surplus for the seller. Using the last equality, we have found the market tightness:

$$
\theta_{n+1}(z,\varphi) = \left(\frac{\gamma}{1-\gamma}\right)\frac{\Gamma_n^S(z,\varphi)}{\Gamma^B(\varphi)} \tag{B.4.5}
$$

for all $n \geq 1$. Finally, we can write (B.4.3) as $W_{n+1}(z,\varphi) - W_n(z,\varphi) = U^B(\varphi) + \gamma^{-1}\big(x_{n+1}(z,\varphi) - U^B(\varphi)\big) = U^B(\varphi) + \gamma^{-1}\Gamma^B(\varphi)\theta_{n+1}(z,\varphi)^{1-\gamma}$. Using (B.4.5) and rearranging terms, we obtain our desired result:

$$
W_{n+1}(z,\varphi) = W_n(z,\varphi) + U^B(\varphi) + \left(\frac{\Gamma^B(\varphi)}{\gamma}\right)^{\gamma}\left(\frac{\Gamma_n^S(z,\varphi)}{1-\gamma}\right)^{1-\gamma} \tag{B.4.6}
$$

This is a second-order difference equation in n. The boundary conditions are $W_0 = 0$ (as the joint value is nil when the seller has no customers), and W_1 set to satisfy the free entry condition (B.4.1). By (B.4.3), we know that $W_1 - x_1 = (1-\gamma)(W_1 - U^B)$ and $x_1 - U^B = \gamma(W_1 - U^B)$, and thus we can write (B.4.1) as:

$$
\kappa(\varphi) = (1-\gamma)\left(\frac{\Gamma^B(\varphi)}{\gamma}\right)^{\frac{\gamma}{\gamma-1}}\sum_{z_0\in\mathcal{Z}}\pi_z(z_0)\Big(W_1(z_0,\varphi) - U^B(\varphi)\Big)^{\frac{1}{1-\gamma}}
$$

our desired result. □

Figure A.7 shows the solution to the second-order difference equation graphically. In equilibrium, the joint surplus value for each additional customer of the match is an increasing and concave sequence in size, with the gain for the first customer determined by the expected value at entry (equal to κ). As the seller accumulates more customers, the joint surplus flattens as the promised utility of each additional customer converges to the outside option, $x_n \searrow U^B$, thereby making the rate of attraction for new customers shrink down to zero, $\theta_n \searrow 0$.

B.5 Proof of Proposition 5: Optimal Contracts under Price Discrimination

Proof. The argument is conceptually similar to that of the proof to Proposition 1 (see Appendix B.1). Letting $\{\overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n\}$, with $\overline{\boldsymbol{\omega}}_i := \{\overline{p}_i, \overline{\mathbf{x}}_i'(n';\mathrm{s}')\}$ and $\overline{\mathbf{x}}_i'(n';\mathrm{s}') = \{\overline{x}_i'(n+1;z,\varphi), \overline{x}_i'(n-1;z,\varphi), \{\overline{x}_i'(n;z',\varphi) : z' \in \mathcal{Z}\}, \{\overline{x}_i'(n;z,\varphi') : \varphi' \in \Phi\}\}$, form a generic policy for the firm, the firm's problem can be written as:

$$
V^S\Big(n, \{x_i\}_{i=1}^n; z,\varphi\Big) := \max_{\overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n} \widetilde{V}^S\Big(n; \overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n; z,\varphi\Big) \quad \text{s.t. } x_i \leq V^B(n, \overline{\boldsymbol{\omega}}_i; z,\varphi), \ \ \forall i = 1,\ldots,n
$$

where:

$$
\begin{aligned}
\widetilde{V}^S\Big(n; \overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n; z,\varphi\Big) :=& \frac{1}{\rho(n;z,\varphi)}\Bigg[\sum_{i=1}^n \overline{p}_i - \mathcal{C}(n;z,\varphi) + \delta_c\sum_{j=1}^n V^S\Big(n-1, \{\overline{x}_i'(n-1;z,\varphi)\}_{i=1}^n \setminus_- \{\overline{x}_j'(n-1;z,\varphi)\}; z,\varphi\Big) \\
&+ \eta\big(\theta(\overline{x}_0';\varphi)\big)V^S\Big(n+1, \{\overline{x}_i'(n+1;z,\varphi)\}_{i=1}^n \cup_+ \{\overline{x}_0'\}; z,\varphi\Big) \\
&+ \sum_{z'\in\mathcal{Z}}\lambda_z(z'|z)V^S\Big(n, \{\overline{x}_i'(n;z',\varphi)\}_{i=1}^n; z',\varphi\Big) + \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)V^S\Big(n, \{\overline{x}_i'(n;z,\varphi')\}_{i=1}^n; z,\varphi'\Big)\Bigg]
\end{aligned} \tag{B.5.1}
$$

is the value of the seller, with $\rho(n;z,\varphi) := r + \delta_f + n\delta_c + \eta\big(\theta(\overline{x}_0';\varphi)\big)$ being the effective discount rate. The value of buyer $i = 1, \ldots, n$ under this policy is:

$$rV^B(n, \overline{\boldsymbol{\omega}}_i; z, \varphi) = v(\varphi) - p_i + (\delta_f + \delta_c)\Big(U^B(\varphi) - V^B(n, \overline{\boldsymbol{\omega}}_i; z, \varphi)\Big) + (n-1)\delta_c\Big(\overline{x}_i'(n-1; z, \varphi) - V^B(n, \overline{\boldsymbol{\omega}}_i; z, \varphi)\Big)$$
$$+ \eta\big(\theta(\overline{x}_0'; \varphi)\big)\Big(\overline{x}_i'(n+1; z, \varphi) - V^B(n, \overline{\boldsymbol{\omega}}_i; z, \varphi)\Big) + \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z)\Big(\overline{x}_i'(n; z', \varphi) - V^B(n, \overline{\boldsymbol{\omega}}_i; z, \varphi)\Big)$$
$$+ \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)\Big(\overline{x}_i'(n; z, \varphi') - V^B(n, \overline{\boldsymbol{\omega}}_i; z, \varphi)\Big)$$

Notice that the firm is re-optimizing after changing size. By monotonicity of preferences, the promise-keeping constraint will bind for each customer:

$$x_i = V^B(n, \overline{\boldsymbol{\omega}}_i; z, \varphi), \qquad \forall i = 1, \ldots, n$$

From this equation, we can solve for the promise-compatible price level to be charged to each customer under the policy $\left\{\overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n\right\}$:

$$p_i^{PK}\Big(\big\{\overline{x}_0', \{\mathbf{x}_j'(n'; \mathbf{s}')\}_{j=1}^n\big\}\Big) = v(\varphi) - \rho(n; z, \varphi)x_i + \delta_f U^B(\varphi) + \delta_c\Big(U^B(\varphi) + (n-1)\overline{x}_i'(n-1; z, \varphi)\Big)$$
$$+ \eta\big(\theta(\overline{x}_0'); \varphi\big)\overline{x}_i'(n+1; z, \varphi) + \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z)\overline{x}_i'(n; z', \varphi) + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)\overline{x}_i'(n; z, \varphi') \tag{B.5.2}$$

Importantly, note that the price level for a specific customer i is independent of the distribution of utilities for all the *other* customers, that is:

$$p_i^{PK}\Big(\big\{\overline{x}_0', \{\mathbf{x}_j'(n'; \mathbf{s}')\}_{j \neq i}\big\} \cup_+ \{\mathbf{x}_i'(n'; \mathbf{s}')\}\Big) = p_i^{PK}\Big(\big\{\overline{x}_0', \{\mathbf{x}_{\phi(j)}'(n'; \mathbf{s}')\}_{\phi(j) \neq i}\big\} \cup_+ \{\mathbf{x}_i'(n'; \mathbf{s}')\}\Big)$$

for any arbitrary bisection $\phi : \{1, \ldots, n\} \to \{1, \ldots, n\}$. Therefore, since the firm's problem internalizes the price level, the resulting maximization should be independent of the distribution of utilities. Indeed, plugging (B.5.2) into (B.5.1) and rearranging terms yields:

$$\widetilde{W}\Big(n, \{x_i\}_{i=1}^n; \overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n; z, \varphi\Big) := \frac{1}{\rho(n; z, \varphi)}\Bigg[n\Big(v(\varphi) + (\delta_f + \delta_c)U^B(\varphi)\Big) - \Big(\mathcal{C}(n; z, \varphi) + \eta\big(\theta(\overline{x}_0'; \varphi)\big)\sum_{i=1}^n \overline{x}_i'(n+1; z, \varphi)\Big)$$
$$+ \delta_c \sum_{j=1}^n W\Big(n-1, \{\overline{x}_i'(n-1; z, \varphi)\}_{i=1}^n \setminus_- \{\overline{x}_j'(n-1; z, \varphi)\}; z, \varphi\Big) \tag{B.5.3}$$
$$+ \eta\big(\theta(\overline{x}_0'; \varphi)\big)W\Big(n+1, \{\overline{x}_i'(n+1)\}_{i=1}^n \cup_+ \{\overline{x}_0'\}; z, \varphi\Big)$$
$$+ \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z)W\Big(n, \{\overline{x}_i'(n; z', \varphi)\}_{i=1}^n; z', \varphi\Big) + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)W\Big(n, \{\overline{x}_i'(n; z, \varphi')\}_{i=1}^n; z, \varphi'\Big)\Bigg]$$

where we have defined:

$$\widetilde{W}\Big(n, \{x_i\}_{i=1}^n; \overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n; z, \varphi\Big) := \widetilde{V}^S\Big(n; \overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n; z, \varphi\Big) + \sum_{i=1}^n x_i$$

and:

$$W\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big) := \max_{\overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n} \widetilde{W}\Big(n, \{x_i\}_{i=1}^n; \overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n; z, \varphi\Big)$$

as the joint surplus under policy $\{\overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n\}$, and the *maximized* joint surplus, respectively. Finally, noting that the right-hand side of (B.5.3) does not depend on the initial distribution of utilities $\{x_i\}_{i=1}^n$ nor the price level, we can write the joint surplus under a given policy as:

$$\widetilde{W}\Big(n, \{x_i\}_{i=1}^n; \overline{x}_0', \{\overline{\boldsymbol{\omega}}_i\}_{i=1}^n; z, \varphi\Big) = \widetilde{W}_n\Big(\overline{x}_0', \{\overline{\mathbf{x}}_i'(n'; \mathbf{s}')\}_{i=1}^n; z, \varphi\Big)$$

This allows us to break up the optimal contracting problem into two separate stages. Where $\{x_0'^*, \{p_i^*, \mathbf{x}_i'^*(n';\mathbf{s}')\}_{i=1}^n\}$ denotes an optimal contract, we have:

$$\{x_0'^*, \{\mathbf{x}_i'^*(n';\mathbf{s}')\}_{i=1}^n\} = \arg\max\ \widetilde{W}_n\Big(\overline{x}_0', \{\overline{\mathbf{x}}_i'(n';\mathbf{s}')\}_{i=1}^n; z, \varphi\Big)$$
$$p_i^* = p_i^{PK}\Big(\Big\{\overline{x}_0'^*, \{\mathbf{x}_j'^*(n';\mathbf{s}')\}_{j=1}^n\Big\}\Big), \ \ \forall i = 1, \dots, n$$

It thus follows that the joint surplus problem and the seller's problem are equivalent. □.

B.6 Proof of Proposition 6: Price Indeterminacy under Discrimination

Proof. Let $\varepsilon \in \mathbb{R}$ be an arbitrary number. The goal of the proof is to show that there is some $\beta_n(\varphi) > 0$ (possibly a function of size and the aggregate state) for which, if a given contract with $\boldsymbol{\omega}^b = \{p_i + \varepsilon\beta_n(\varphi), \mathbf{x}_i'(n';\mathbf{s}') + \varepsilon\}_{i=1}^n$ is optimal, then each customer and the seller maximize their value under contract $\boldsymbol{\omega}^a = \{p_i, \mathbf{x}_i'(n';\mathbf{s}')\}_{i=1}^n$. The value of contract $\boldsymbol{\omega}_i^b$ for customer $i = 1, \dots, n$ is:

$$\begin{aligned}
rV^B(n, \boldsymbol{\omega}_i^b; z, \varphi) &= v(\varphi) - p_i - \varepsilon\beta_n(\varphi) + (\delta_f + \delta_c)\Big(U^B(\varphi) - V^B(n, \boldsymbol{\omega}_i; z, \varphi)\Big) + (n-1)\delta_c\Big(x_i'(n-1; z, \varphi) + \varepsilon - V^B(n, \boldsymbol{\omega}_i^b; z, \varphi)\Big) \\
&\quad + \eta\big(\theta(x_0'(\varphi); \varphi)\big)\Big(x_i'(n+1; z, \varphi) + \varepsilon - V^B(n, \boldsymbol{\omega}_i^b; z, \varphi)\Big) + \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z)\Big(x_i'(n; z', \varphi) + \varepsilon - V^B(n, \boldsymbol{\omega}_i; z, \varphi)\Big) \\
&\quad + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)\Big(x_i'(n; z, \varphi') + \varepsilon - V^B(n, \boldsymbol{\omega}_i; z, \varphi)\Big) \\
&= v(\varphi) - p_i - \varepsilon\Big(\beta_n(\varphi) - (n-1)\delta_c - \eta\big(\theta(x_0'(\varphi); \varphi)\big)\Big) + (\delta_f + \delta_c)\Big(U^B(\varphi) - V^B(n, \boldsymbol{\omega}_i^b; z, \varphi)\Big) \\
&\quad + (n-1)\delta_c\Big(x_i'(n-1; z, \varphi) - V^B(n, \boldsymbol{\omega}_i^b; z, \varphi)\Big) \\
&\quad + \eta\big(\theta(x_0'(\varphi); \varphi)\big)\Big(x_i'(n+1; z, \varphi) - V^B(n, \boldsymbol{\omega}_i^b; z, \varphi)\Big) + \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z)\Big(x_i'(n; z', \varphi) - V^B(n, \boldsymbol{\omega}_i^b; z, \varphi)\Big) \\
&\quad + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)\Big(x_i'(n; z, \varphi') - V^B(n, \boldsymbol{\omega}_i^b; z, \varphi)\Big) \\
&= rV^B(n, \boldsymbol{\omega}_i^a; z, \varphi) + \varepsilon\Big(\beta_n(\varphi) - (n-1)\delta_c - \eta\big(\theta(x_0'(\varphi); \varphi)\big)\Big)
\end{aligned}$$

where we have used $\sum_{z' \in \mathcal{Z}} \lambda_z(z'|z)\varepsilon = \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)\varepsilon = 0$ in the second equality. Thus, $V^B(n, \boldsymbol{\omega}_i^a) = V^B(n, \boldsymbol{\omega}_i^b)$ if, and only if,

$$\beta_n(\varphi) = (n-1)\delta_c + \eta\big(\theta(x_0'(\varphi); \varphi)\big) \tag{B.6.1}$$

As for the seller's value, note that:

$$\begin{aligned}
rV^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big) = \max_{x_0'(\varphi), \{\boldsymbol{\omega}_i\}_{i=1}^n} \Bigg\{ &\sum_{i=1}^n p_i + n\varepsilon\beta_n(\varphi) - \mathcal{C}(n; z, \varphi) + \delta_f\left(V_0^S(\varphi) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\right) \\
&+ \delta_c \sum_{j=1}^n \left(V^S\Big(n-1, \{x_i'(n-1; z, \varphi) + \varepsilon\}_{i=1}^n \setminus_- \{x_j'(n-1; z, \varphi) + \varepsilon\}; z, \varphi\Big) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\right) \\
&+ \eta\big(\theta(x_0'(\varphi); \varphi)\big)\left(V^S\Big(n+1, \{x_i'(n+1; z, \varphi) + \varepsilon\}_{i=1}^n \cup_+ \{x_0'(\varphi)\}; z, \varphi\Big) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\right) \\
&+ \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z)\left(V^S\Big(n, \{x_i'(n; z', \varphi) + \varepsilon\}_{i=1}^n; z', \varphi\Big) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\right) \\
&+ \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)\left(V^S\Big(n, \{x_i'(n; z, \varphi') + \varepsilon\}_{i=1}^n; z, \varphi'\Big) - V^S\Big(n, \{x_i\}_{i=1}^n; z, \varphi\Big)\right) \Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&= \max_{x_0'(\varphi),\{\boldsymbol{\omega}_i\}_{i=1}^n} \Bigg\{ \sum_{i=1}^n p_i + n\varepsilon\beta_n(\varphi) - \mathcal{C}(n;z,\varphi) + \delta_f\bigg(V_0^S(\varphi) - V^S\Big(n,\{x_i\}_{i=1}^n;z,\varphi\Big)\bigg) \\
&\qquad + \delta_c \sum_{j=1}^n \bigg(W_{n-1}(z,\varphi) - \sum_{i\neq j} x_i'(n-1;z,\varphi) - (n-1)\varepsilon - V^S\Big(n,\{x_i\}_{i=1}^n;z,\varphi\Big)\bigg) \\
&\qquad + \eta\big(\theta(x_0'(\varphi);\varphi)\big)\bigg(W_{n+1}(z,\varphi) - \sum_{i=1}^n x_i'(n+1;z,\varphi) - x_0' - n\varepsilon - V^S\Big(n,\{x_i\}_{i=1}^n;z,\varphi\Big)\bigg) \\
&\qquad + \sum_{z'\in\mathcal{Z}} \lambda_z(z'|z)\bigg(W_n(z',\varphi) - \sum_{i=1}^n x_i'(n;z',\varphi) - n\varepsilon - V^S\Big(n,\{x_i\}_{i=1}^n;z,\varphi\Big)\bigg) \\
&\qquad + \sum_{\varphi'\in\Phi} \lambda_\varphi(\varphi'|\varphi)\bigg(W_n(z,\varphi') - \sum_{i=1}^n x_i'(n;z,\varphi') - n\varepsilon - V^S\Big(n,\{x_i\}_{i=1}^n;z,\varphi\Big)\bigg)\Bigg\} \\
&= rV^S\Big(n,\{x_i\}_{i=1}^n;z,\varphi\Big) + n\varepsilon\Big(\beta_n(\varphi) - (n-1)\delta_c - \eta\big(\theta(x_0'(\varphi);\varphi)\big)\Big)
\end{aligned}
$$

where we have used the definition of W in the second equality. Thus, under our choice for $\beta_n(\varphi)$ in (B.6.1), the value of the seller does not change, either. In sum, both buyers and seller are indifferent between contracts $\{\boldsymbol{\omega}_i^a\}_{i=1}^n$ and $\{\boldsymbol{\omega}_i^b\}_{i=1}^n$. Consequently, the joint surplus does not change by definition, and therefore contract $\{\boldsymbol{\omega}_i^a\}_{i=1}^n$ is optimal if, and only if, $\{\boldsymbol{\omega}_i^b\}_{i=1}^n$ is optimal. Generally, these is a continuum of optimal contracts, indexed by ε. $\square$

C Numerical Appendix

C.1 Stationary Solution Algorithm

To solve for the stationary equilibrium, we solve for two nested fixed-point problems. The innermost problem is the maximization of the joint surplus function, for a given value of inactivity, U^B. Since W defines a contraction, we use a value function iteration algorithm for this step. The outermost fixed point problem is on $U^B(\varphi)$ which, for each and every aggregate state, must satisfy the free entry condition. For this step, we use a bisection method, whereby $U^B(\varphi)$ is updated depending on whether there is too much, or not enough, entry. Throughout, the state space grid is fixed at $\mathcal{N}\times\mathcal{Z}\times\Phi$, where $\mathcal{N}=\{1,\dots,\bar{n}\}$, with $\bar{n}\in\mathbb{N}$ a sufficiently large bound on firm size, and $\mathcal{Z}=\{z_i\}_{i=1}^{k_z}$ and $\Phi=\{\varphi_j\}_{j=1}^{k_\varphi}$. The (z,φ) processes are parametrized according to the description in Section C.2. In the calibration, we set $k_\varphi=1$, $k_z=25$, and $\bar{n}=50$.

The following describes the steps of the algorithm:

Step 1. Set the counter to $k=0$. Choose guesses $\underline{U}^{(0)}(\varphi)$ and $\overline{U}^{(0)}(\varphi)\gg\underline{U}^{(0)}(\varphi)$ for each $\varphi\in\Phi$. Set the value of inactivity to:

$$
U^{B(0)}(\varphi) = \frac{1}{2}\left(\underline{U}^{(0)}(\varphi) + \overline{U}^{(0)}(\varphi)\right)
$$

Step 2. For any given $k\in\mathbb{N}$ and $n\in\mathcal{N}$, use value function iteration to find the fixed point $W_n^{(k)}(z,\varphi)$ of:

$$
\begin{aligned}
(r+\delta_f)W_n^{(k)}(z,\varphi) = \; & n\Big(v(\varphi) + (\delta_f+\delta_c)U^{B(k)}(\varphi)\Big) - \mathcal{C}(n;z,\varphi) + n\delta_c\Big(W_{n-1}^{(k)}(z,\varphi) - W_n^{(k)}(z,\varphi)\Big) \\
& + \max_{x_{n+1}'(z,\varphi)} \left\{ \eta\circ\mu^{-1}\left(\frac{\Gamma^{B(k)}(\varphi)}{x_{n+1}'(z,\varphi) - U^{B(k)}(\varphi)}\right)\Big(W_{n+1}^{(k)}(z,\varphi) - W_n^{(k)}(z,\varphi) - x_{n+1}'(z,\varphi)\Big)\right\} \\
& + \sum_{z'\in\mathcal{Z}}\lambda_z(z'|z)\Big(W_n^{(k)}(z',\varphi) - W_n^{(k)}(z,\varphi)\Big) + \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)\Big(W_n^{(k)}(z,\varphi') - W_n^{(k)}(z,\varphi)\Big)
\end{aligned}
$$

where $\Gamma^{B(k)} = c(\varphi) + rU^{B(k)}(\varphi) - \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)\Big(U^{B(k)}(\varphi') - U^{B(k)}(\varphi)\Big)$. Store the corresponding policy functions: $\Big\{x_{n+1}'^{(k)}(z,\varphi) : (n,z,\varphi)\in\mathcal{N}\times\mathcal{Z}\times\Phi\Big\}$.

Step 3. For each $\varphi\in\Phi$, compute the object:

$$\Delta^{(k)}(\varphi) := \kappa(\varphi) - \sum_{z_0 \in \mathcal{Z}} \pi_z(z_0) \left\{ \eta \circ \mu^{-1} \left(\frac{\Gamma^{B(k)}(\varphi)}{x_1'^{(k)}(z_0,\varphi) - U^{B(k)}(\varphi)} \right) \left(W_1^{(k)}(z_0,\varphi) - x_1'^{(k)}(z_0,\varphi) \right) \right\}$$

Stop if $\Delta^{(k)}(\varphi) \in [-\varepsilon, \varepsilon]$ for all $\varphi \in \Phi$ and some small tolerance $\varepsilon > 0$. Otherwise, set

$$U^{B(k+1)}(\varphi) = \frac{1}{2} \left(\underline{U}^{(k+1)}(\varphi) + \overline{U}^{(k+1)}(\varphi) \right)$$

for each $\varphi \in \Phi$, where:

(a) If $\Delta^{(k)}(\varphi) > \varepsilon$, then $\underline{U}^{(k+1)}(\varphi) = \underline{U}^{(k)}(\varphi)$ and $\overline{U}^{(k+1)}(\varphi) = U^{B(k)}(\varphi)$;

(b) If $\Delta^{(k)}(\varphi) < -\varepsilon$, then $\underline{U}^{(k+1)}(\varphi) = U^{B(k)}(\varphi)$ and $\overline{U}^{(k+1)}(\varphi) = \overline{U}^{(k)}(\varphi)$;

and go back to Step 2. with $[k] \hookleftarrow [k+1]$.

The advantage of this approach is that the policy function in Step 2. can be expressed as a function of only W and U^B (recall Proposition B.4), the two functions over which we iterate. Both fixed-point algorithms are fast and converge within only a few iterations. After convergence, we have the full equilibrium sequences for W, defined on every point of the state space grid $\mathcal{N} \times \mathcal{Z} \times \Phi$, from which market tightness, prices, and distributions can be readily computed using our analytical results described above.

C.2 Numerical Approximation of the Exogenous State Processes

This appendix shows how to parametrize and estimate continuous-time Markov chain (CTMC) processes. In the context of our model, we have two such exogenous processes: z and φ. Consider the idiosyncratic shock, for instance (the same structure applies to the aggregate shock). Firstly, the $k_z \times k_z$ infinitesimal generator matrix $\mathbf{\Lambda}_z$ has the usual form, i.e. the elements of each row vector add up to zero:

$$\mathbf{\Lambda}_z = \begin{pmatrix} -\sum_{j\neq 1}\lambda_{1j} & \lambda_{12} & \dots & \lambda_{1k_z} \\ \lambda_{21} & -\sum_{j\neq 2}\lambda_{2j} & \dots & \lambda_{2k_z} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{k_z 1} & \lambda_{k_z 2} & \dots & -\sum_{j\neq k_z}\lambda_{k_z j} \end{pmatrix}$$

where λ_{ij} is short-hand for $\lambda_z(z_j|z_i)$, $z_i, z_j \in \mathcal{Z}$. Since this level of generality would require the estimation of $k_z(k_z-1)$ transition rates, we reduce the parameter space by specializing the CTMC as follows:

- First, we assume z follows a driftless Ornstein-Uhlenbeck (OU) process in logs. An OU process is a type of mean-reverting and autoregressive CTMC which can be loosely viewed as the continuous-time analogue of an AR(1).[62] Formally:

  $$\mathrm{d}\log z_t = -\rho_z \log z_t \mathrm{d}t + \sigma_z \mathrm{d}B_t$$

  where B_t is a standard Brownian motion, and $\rho_z, \sigma_z > 0$ are parameters.

- Operationally, in the numerical version of the model in which time is partitioned and takes values in $\mathbb{T} = \{\Delta, 2\Delta, 3\Delta, \dots\}$, we implement this process by using the *Euler-Maruyama method*, that is:

  $$\log z_k = (1-\rho_z\Delta)\log z_{k-1} + \sigma_z\sqrt{\Delta}\varepsilon_k^z, \qquad \varepsilon_k^z \sim iid\,\mathcal{N}(0,1) \tag{C.2.1}$$

  for each $k \in \mathbb{T}$. Notice that this is an AR(1) processes with autocorrelation $\widetilde{\rho}_z := 1 - \rho_z\Delta$ and variance $\dfrac{\sigma_z^2}{\rho_z(1+\widetilde{\rho}_z)}$. Thus, $\rho_z > 0$ can be seen as a measure of mean-reversion, with *lower* values corresponding to *higher* persistence.

[62] For another example of a continuous-time search-and-matching model with shocks that uses Ornstein-Uhlenbeck processes, see Shimer (2005).

- The discrete-time process above is defined on a continuous state space, so we then employ standard techniques (Tauchen (1986)) to approximate the AR(1) processes using a discrete-time, discrete-state Markov chain that we define on the theoretical grid, $\mathcal{Z}$. The outcome of this method is a transition probability matrix $\mathbf{\Pi}_z = \left(\pi_{ij}\right)$, where π_{ij} denotes the probability of a z_i-to-z_j transition in the $\mathbb{T}$ space.
- Finally, to map this specification back into continuous time, we use the fact that, for small enough $\Delta > 0$, transition probabilities are well approximated by transition rates in the following sense:[63]

$$\forall i = 1, \ldots, k_z : \quad \pi_{ij} \approx \lambda_{ij}\Delta, \forall j \neq i \quad \text{and} \quad \pi_{ii} \approx 1 - \sum_{j \neq i} \lambda_{ij}\Delta$$

when $\Delta > 0$ is small enough.

This methodology therefore has the obvious advantage that, instead of having to calibrate the whole collection $\left\{\lambda_{ij}^z\right\}$ of transition rates individually, these can be backed out directly from estimating only two parameters: ρ_z and σ_z.

D Additional Theoretical Results

D.1 HJB Equations

In this appendix, we show how to derive the HJB equations for inactive buyers (equation (2)), active buyers (equation (5)), incumbent sellers (equation (6)), and potential entrant sellers (equation (8)), from their discrete-time counterparts. Throughout, we assume a discrete-time stationary environment in which time intervals are equidistant and of some (short) length $\Delta > 0$.

Inactive Buyers

The value of an inactive buyers in state $\varphi \in \Phi$ is:

$$U^B(\varphi) = \max_{\widehat{x}(\varphi) \in \mathcal{X}} u^B\left(\widehat{x}(\varphi); \varphi\right)$$

where $u^B\left(\widehat{x}(\varphi); \varphi\right)$ is given by the Bellman equation:

$$\begin{aligned} u^B\left(\widehat{x}(\varphi); \varphi\right) = &- c(\varphi)\Delta + e^{-r\Delta}\Bigg(\Big[\eta\big(\theta(\widehat{x}(\varphi); \varphi)\big)\Delta + o(\Delta)\Big] \max\Big\{\widehat{x}(\varphi), u^B\left(\widehat{x}(\varphi); \varphi\right)\Big\} + \sum_{\varphi' \in \Phi}\Big[\lambda_\varphi(\varphi'|\varphi)\Delta + o(\Delta)\Big] u^B\left(\widehat{x}(\varphi); \varphi'\right) \\ &+ \Big[1 - \eta\big(\theta(\widehat{x}(\varphi); \varphi)\big)\Delta - \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi)\Delta + o(\Delta)\Big] u^B\left(\widehat{x}(\varphi); \varphi\right)\Bigg) + o(\Delta) \end{aligned} \tag{D.1.1}$$

where $o(\Delta)$ has the property $\lim_{\Delta \to 0} \frac{o(\Delta)}{\Delta} = 0$. The intuitive interpretation is standard. Here, and in every value function to follow, we use that, for a given Poisson arrival rate $k \geq 0$, the term $k\Delta + o(\Delta)$ (respectively, $1 - k\Delta + o(\Delta)$) approximates the probability of exactly one arrival (respectively, no arrivals) of the Poisson shock within some time interval $[t, t + \Delta]$. This approximation is valid for $\Delta > 0$ small enough. The probability of two or more arrivals is approximately equal to $o(\Delta)$, and accounted for in the object $o(\Delta)$. Moreover, the discount factor $e^{-r\Delta}$ approximates the usual discrete-time discounting of $\frac{1}{1+r\Delta}$ when $\Delta > 0$ is small enough.

It should be noted that, if $\theta(x; \varphi) > 0$, then it must be that $\max\{x - U^B(\varphi), 0\} > 0$, or else no buyer would visit the market. Thus, any matched customer is ex-post better off than any inactive buyer. On the one hand, this means that, so long as the size of shocks is sufficiently restricted, we may ignore the possibility of voluntary transition

[63] Formally, in this step we are using the result that any CTMC with fixed *transition rate* matrix $\mathbf{\Lambda}$ maps into a discrete-time chain with *transition probability* matrix $\mathbf{\Pi}(t)$ at time $t \in \mathbb{Z}_+$ in which holding times between arrivals are independently and exponentially distributed. In particular, $\boldsymbol{\partial}_t\mathbf{\Pi}(t) = \mathbf{\Lambda}\mathbf{\Pi}(t)$, and thus $\mathbf{\Pi}(t) = e^{\mathbf{\Lambda}t}$ (when $\mathbf{\Pi}_0 = \mathbf{I}$). Hence, the probability of a z_i-to-z_j transition after time Δ is given by $\frac{\lambda_{ij}}{\sum_{j \neq i} \lambda_{ij}}\left(1 - e^{-\sum_{j \neq i} \lambda_{ij}\Delta}\right)$ when $j \neq i$, and by $e^{-\sum_{j \neq i} \lambda_{ij}\Delta}$ otherwise. When $\Delta > 0$ is sufficiently small, these probabilities are well approximated by $\lambda_{ij}\Delta$ and $\left(1 - \sum_{j \neq i} \lambda_{ij}\Delta\right)$, respectively.

to inactivity from our HJB equations for buyers, without loss of generality. Moreover, since inactive buyers are always willing to enter into the product market, the seller's optimal design of the contract $\boldsymbol{\omega}$ need not incorporate a participation constraint.

To obtain equation (2), subtract $e^{-r\Delta}u^B\big(\widehat{x}(\varphi);\varphi\big)$ from both sides, divide every term by Δ, and take the continuous-time limit as $\Delta\to 0$ (using that $\lim_{\Delta\to 0}\frac{1-e^{-r\Delta}}{\Delta}=r$).

Active Buyers

Now consider the value of being the customer of a firm of size n that promises continuation utility x in state $\mathbf{s}=(z,\varphi)\in\mathcal{Z}\times\Phi$, and offers contract $\boldsymbol{\omega}=\big\{p,\mathbf{x}'(n';\mathbf{s}')\big\}$. In the discrete-time approximation, we have:

$$
\begin{aligned}
V^B(n,\boldsymbol{\omega};\mathbf{s}) = \big(v(\varphi)-p\big)\Delta + e^{-r\Delta}\Bigg(&\Big[(\delta_f+\delta_c)\Delta+o(\Delta)\Big]U^B(\varphi)+\Big[(n-1)\delta_c\Delta+o(\Delta)\Big]\max\Big\{x'(n-1;\mathbf{s}),U^B(\varphi)\Big\}\\
&+\Big[\eta\Big(\theta\big(x'(n+1;\mathbf{s});\varphi\big)\Big)\Delta+o(\Delta)\Big]\max\Big\{x'(n+1;\mathbf{s}),U^B(\varphi)\Big\}\\
&+\sum_{z'\in\mathcal{Z}}\Big[\lambda_z(z'|z)\Delta+o(\Delta)\Big]\max\Big\{x'(n;z',\varphi),U^B(\varphi)\Big\}+\sum_{\varphi'\in\Phi}\Big[\lambda_\varphi(\varphi'|\varphi)\Delta+o(\Delta)\Big]\max\Big\{x'(n;z,\varphi'),U^B(\varphi')\Big\}\\
&+\Big[1-\delta_f\Delta-n\delta_c\Delta-\eta\Big(\theta\big(x'(n+1;\mathbf{s});\varphi\big)\Big)\Delta-\sum_{z'\in\mathcal{Z}}\lambda_z(z'|z)\Delta-\sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)\Delta+o(\Delta)\Big]\max\Big\{V^B(n,\boldsymbol{\omega};\mathbf{s}),U^B(\varphi)\Big\}\Bigg)\\
&+o(\Delta)
\end{aligned}
$$

This equation represents the buyer's Bellman equation in full generality. Particularly, we recognize the possibility that, as indicated by the $\max$ operators, the customer is free to opt out of the match at the beginning of period $t+\Delta$, in which case she stops buying from the firm and must become inactive for at least one period. The buyer may also transition to inactivity if either the firm dies or the match is destroyed exogenously. If she remains within the firm, she acknowledges that the match might change in value once the firm shrinks (second line), grows (third line), or is hit by an exogenous shock (fourth line). The fifth line is the complementary case in which none of the Poisson shocks hit, in which case the customer is still given the chance to transition to inactivity.

To obtain our final HJB equation, we first recall that transiting to inactivity is not optimal under any contingency. Since, in equilibrium, the seller offers continuation values in the set of equilibrium markets, we may eliminate the $\max$ operators above. Finally, we can subtract $e^{-r\Delta}V^B(n,\boldsymbol{\omega};\mathbf{s})$ from both sides, divide every term by Δ, and take the continuous-time limit as $\Delta\to 0$ to obtain equation (5).

Incumbent Sellers

Consider now a seller with $n\in\mathbb{N}$ customers who is currently promising a value x in state $\mathbf{s}=(z,\varphi)$. The seller must choose contract $\boldsymbol{\omega}=\big\{p,\mathbf{x}'(n';\mathbf{s}')\big\}$ to maximize:

$$
\begin{aligned}
V^S(n,x;\mathbf{s}) = \max_{\boldsymbol{\omega}\in\Omega}\Bigg\{\big(np-\mathcal{C}(n;\mathbf{s})\big)\Delta+e^{-r\Delta}\Bigg(&\Big[\delta_f\Delta+o(\Delta)\Big]V_0^S(\varphi)+\Big[n\delta_c\Delta+o(\Delta)\Big]\max\Big\{V^S\big(n-1,x'(n-1;\mathbf{s});\mathbf{s}\big),V_0^S(\varphi)\Big\}\\
&+\Big[\eta\Big(\theta\big(x'(n+1;\mathbf{s})\big);\varphi\Big)\Delta+o(\Delta)\Big]\max\Big\{V^S\big(n+1,x'(n+1;\mathbf{s});\mathbf{s}\big),V_0^S(\varphi)\Big\}\\
&+\sum_{z'\in\mathcal{Z}}\Big[\lambda_z(z'|z)\Delta+o(\Delta)\Big]\max\Big\{V^S\big(n,x'(n;z',\varphi);z',\varphi\big),V_0^S(\varphi)\Big\}\\
&+\sum_{\varphi'\in\Phi}\Big[\lambda_\varphi(\varphi'|\varphi)\Delta+o(\Delta)\Big]\max\Big\{V^S\big(n,x'(n;z,\varphi');z,\varphi'\big),V_0^S(\varphi')\Big\}\\
&+\Big[1-\delta_f\Delta-n\delta_c\Delta-\eta\Big(\theta\big(x'(n+1;\mathbf{s});\varphi\big)\Big)\Delta-\sum_{z'\in\mathcal{Z}}\lambda_z(z'|z)\Delta-\sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)\Delta+o(\Delta)\Big]\max\Big\{V^S(n,x;\mathbf{s}),V_0^S(\varphi)\Big\}\Bigg)\Bigg\}\\
&+o(\Delta)
\end{aligned}
$$

subject to:

$$V^B(n, \boldsymbol{\omega}; \mathbf{s}) \geq x$$

As in the case of buyers above, we write the seller's Bellman equation in full generality. In particular, at the beginning of period $t+\Delta$, the seller can choose to cease operations and transition to $n=0$, as indicated by the $\max$ operators. Otherwise, the interpretation of the above equation is very similar to the one given in the main text. To arrive at equation (6), we anticipate that $V_0^S(\varphi)=0$, $\forall t$, by free entry, and therefore $V^S(n,x;\mathbf{s}) \geq V_0^S(\varphi)=0$, $\forall (n,x;\mathbf{s})$, without loss of generality. Then, we can proceed by taking the usual continuous-time limit, as described above.

Entrant Sellers

Finally, we characterize the problem of potential entrants. In the discrete-time approximation, potential entrants draw their initial productivity z_0 from the distribution π_z, after which they pay the set-up cost $\kappa(\varphi)>0$ allowing them to post a contract. Let $v_0^S(\mathbf{s}_0)$ denote the value of drawing productivity z_0, where $\mathbf{s}_0 := (z_0, \varphi)$. Then, the entrant chooses the initial contract $\mathbf{x}'(1;\mathbf{s}_0) \in \mathcal{X}$, specifying the promised utility for the first customer of the firm, to maximize:

$$v_0^S(\mathbf{s}_0) = \max_{\mathbf{x}'(1;\mathbf{s}_0)\in\mathcal{X}} \Bigg\{ e^{-r\Delta}\Bigg(\Big[\eta\Big(\theta\big(x'(1;\mathbf{s}_0);\varphi\big)\Big)\Delta + o(\Delta)\Big] \max\Big\{V^S\big(1, x'(1;\mathbf{s}_0);\mathbf{s}_0\big), V_0^S(\varphi)\Big\} + \sum_{\varphi'\in\Phi}\Big[\lambda_\varphi(\varphi'|\varphi)\Delta + o(\Delta)\Big] V_0^S(\varphi') \\ + \Big[1 - \eta\Big(\theta\big(x'(1;\mathbf{s});\varphi\big)\Big)\Delta - \sum_{\varphi'\in\Phi}\lambda_\varphi(\varphi'|\varphi)\Delta + o(\Delta)\Big] V_0^S(\varphi)\Bigg)\Bigg\} + o(\Delta)$$

where $V_0^S(\varphi)$ is the expected value of entry, defined by:

$$V_0^S(\varphi) = -\kappa(\varphi)\Delta + \sum_{z_0\in\mathcal{Z}} \pi_z(z_0) v_0^S(z_0,\varphi)$$

To obtain equation (8), subtract $e^{-r\Delta}V_0^S(\varphi)$ from both sides of the second equation by making use of the fact that $\sum_{z_0\in\mathcal{Z}}\pi_z(z_0)=1$, and take the continuous-time limit in the usual fashion, noting that, by free entry, we can drop the possibility of voluntary separation, and thereby eliminate the $\max$ operator from the first equation.

D.2 Derivations of Section 3.5

Kolmogorov Forward Equations

To derive the KFEs, we work with the equilibrium *shares* of agent types, defined by $g_{n,t}(z) := \frac{S_{n,t}(z)}{\mathcal{S}_t}$, for each $n\in\mathbb{N}$ and z, where $\mathcal{S}_t := \sum_{n\geq 1}\sum_z S_{n,t}(z)$ is the total measure of incumbents. Note that g is a probability mass function (p.m.f.), with $g_{n,t}(z)\leq 1$ and $\sum_{n\geq1}\sum_z g_{n,t}(z)=1$, $\forall t\geq 0$, for each given φ.

After a period of size $\Delta>0$, the share of firms of type (n,z) when $n=1$ becomes:

$$g_{1,t+\Delta}(z) = \Big[\pi_z(z)\eta\big(\theta_{1,t+\Delta}(z,\varphi)\big)\Delta + o(\Delta)\Big]\frac{S_{0,t}(\varphi)}{\mathcal{S}_t} + 2\Big[\delta_c\Delta + o(\Delta)\Big] g_{2,t}(z) + \sum_{\tilde z\neq z}\Big[\lambda_z(z|\tilde z)\Delta + o(\Delta)\Big] g_{1,t}(\tilde z) \\ + \Big[1 - \delta_f\Delta - \delta_c\Delta - \eta\big(\theta_{2,t+\Delta}(z,\varphi)\big)\Delta - \sum_{\tilde z\neq z}\lambda_z(\tilde z|z)\Delta + o(\Delta)\Big] g_{1,t}(z) \tag{D.2.1}$$

Similarly, for $n\geq 2$, we have:

$$g_{n,t+\Delta}(z) = \Big[\eta\big(\theta_{n,t+\Delta}(z,\varphi)\big)\Delta + o(\Delta)\Big] g_{n-1,t}(z) + (n+1)\Big[\delta_c\Delta + o(\Delta)\Big] g_{n+1,t}(z) + \sum_{\tilde z\neq z}\Big[\lambda_z(z|\tilde z)\Delta + o(\Delta)\Big] g_{n,t}(\tilde z) \\ + \Big[1 - \delta_f\Delta - n\delta_c\Delta - \eta\big(\theta_{n+1,t+\Delta}(z,\varphi)\big)\Delta - \sum_{\tilde z\neq z}\lambda_z(\tilde z|z)\Delta + o(\Delta)\Big] g_{n,t}(z) \tag{D.2.2}$$

These equations describe the law of motion for $g_{n,t}(\cdot)$: due to customer acquisition, attrition, or an exogenous shock, the first line shows the shares of firms transitioning into state (n, z); the second line shows the share of firms of type (n, z) that are not hit by any shock, and thereby remain type (n, z). Subtracting $g_{n,t}(z)$ from both sides of equation (D.2.2) and dividing by Δ gives:

$$\begin{aligned}\frac{g_{n,t+\Delta}(z) - g_{n,t}(z)}{\Delta} &= \Big[\eta\big(\theta_{n,t+\Delta}(z,\varphi)\big) + \frac{o(\Delta)}{\Delta}\Big]g_{n-1,t}(z) + (n+1)\Big[\delta_c + \frac{o(\Delta)}{\Delta}\Big]g_{n+1,t}(z) \\ &+ \sum_{\tilde{z}\neq z}\Big[\lambda_z(z|\tilde{z}) + \frac{o(\Delta)}{\Delta}\Big]g_{n,t}(\tilde{z}) - \Big[\delta_f + n\delta_c + \eta\big(\theta_{n+1,t+\Delta}(z,\varphi)\big) + \sum_{\tilde{z}\neq z}\lambda_z(\tilde{z}|z) + \frac{o(\Delta)}{\Delta}\Big]g_{n,t}(z)\end{aligned}$$

Taking the limit as $\Delta \to 0$,

$$\boldsymbol{\partial}_t g_{n,t}(z) = \eta\big(\theta_{n,t}(z,\varphi)\big)g_{n-1,t}(z) + (n+1)\delta_c g_{n+1,t}(z) + \sum_{\tilde{z}\neq z}\lambda_z(z|\tilde{z})g_{n,t}(\tilde{z}) - \Big(\delta_f + n\delta_c + \eta\big(\theta_{n+1,t}(z,\varphi)\big) + \sum_{\tilde{z}\neq z}\lambda_z(\tilde{z}|z)\Big)g_{n,t}(z)$$

A similar derivation on (D.2.1) shows that, for $n = 1$,

$$\boldsymbol{\partial}_t g_{1,t}(z) = \pi_z(z)\eta\big(\theta_{1,t}(z,\varphi)\big)\frac{S_{0,t}(\varphi)}{\mathcal{S}_t} + 2\delta_c g_{2,t}(z) + \sum_{\tilde{z}\neq z}\lambda_z(z|\tilde{z})g_{1,t}(\tilde{z}) - \Big(\delta_f + \delta_c + \eta\big(\theta_{2,t}(z,\varphi)\big) + \sum_{\tilde{z}\neq z}\lambda_z(\tilde{z}|z)\Big)g_{1,t}(z)$$

It remains to show the law of motion for the measure of potential entrants, $S_{0,t}(\varphi)$. In this case, for given φ, we have:

$$S_{0,t+\Delta}(\varphi) = \Big[\delta_f\Delta + o(\Delta)\Big]\mathcal{S}_t + \Big[\delta_c\Delta + o(\Delta)\Big]\sum_z S_{1,t}(z) + \Big[1 - \sum_{z_0}\pi_z(z_0)\eta\big(\theta_{1,t+\Delta}(z_0,\varphi)\big)\Delta + o(\Delta)\Big]S_{0,t}(\varphi)$$

Taking the continuous-time limit in the usual way, we arrive at:

$$\boldsymbol{\partial}_t S_{0,t}(\varphi) = \Big(\delta_f + \delta_c\sum_z g_{1,t}(z)\Big)\mathcal{S}_t - \sum_{z_0}\pi_z(z_0)\eta\big(\theta_{1,t}(z_0,\varphi)\big)S_{0,t}(\varphi)$$

In the stationary solution, inflows and outflows are equated for every pair of idiosyncratic states (n, z), so $\boldsymbol{\partial}_t g_{n,t}(z) = 0$ and $\boldsymbol{\partial}_t S_{0,t}(\varphi) = 0$. The system of KF equations then becomes a system of second-order equations which can be solved numerically on the state-space grid. In particular, we find a solution for the matrix $\{g_n(z)\}_{n,z}$, and the share of potential entrants per incumbent firm, $h_0(\varphi) := S_0(\varphi)/\mathcal{S}$.

Computing the Aggregate Stationary Measures of Agents

Once we have found a solution $\{g_n(z)\}_{n,z}$ for the invariant size distribution, we can compute aggregate measures as follows. Firstly, using (17) we obtain the object $b_n^A(z) := \frac{B_n^A(z)}{\mathcal{S}}$ by:

$$b_n^A(z) = n g_n(z)$$

Then, $b^A := B^A/\mathcal{S} = \sum_{n=1}^{+\infty}\sum_{z\in\mathcal{Z}} n g_n(z)$. On the other hand, from equation (16) we know that $B_n^I(z,\varphi) = \mathcal{S}\theta_n(z,\varphi)g_{n-1}(z)$. Therefore, adding across states (from $n = 2$ onward) yields:

$$\begin{aligned}\mathcal{S}\sum_{n=2}^{+\infty}\sum_{z\in\mathcal{Z}}\theta_n(z,\varphi)g_{n-1}(z) &= \sum_{n=2}^{+\infty}\sum_{z\in\mathcal{Z}}B_n^I(z,\varphi) \\ &= B^I - \sum_{z\in\mathcal{Z}}B_1^I(z,\varphi) \\ &= 1 - B^A - \sum_{z\in\mathcal{Z}}\theta_1(z,\varphi)S_0(\varphi)\end{aligned}$$

Using the definitions above, the last line can be written as $\mathcal{S} = \dfrac{1 - \Big(b^A + h_0(\varphi)\sum_z \theta_1(z,\varphi)\Big)\mathcal{S}}{\sum_{n\geq 2}\sum_z \theta_n(z,\varphi)g_{n-1}(z)}$. Solving for $\mathcal{S}$, we obtain the stationary measure of active sellers:

$$\mathcal{S} = \left(b^A + h_0(\varphi)\sum_{z\in\mathcal{Z}}\theta_1(z,\varphi) + \sum_{n=1}^{+\infty}\sum_{z\in\mathcal{Z}}\theta_{n+1}(z,\varphi)g_n(z)\right)^{-1}$$

Once the total measure of sellers is known, all other aggregate measures can be easily obtained. For instance, the mass of potential entrants is $S_0(\varphi) = \mathcal{S}h_0(\varphi)$, and the different measures of incumbent sellers can be obtained by $S_n = \mathcal{S}g_n$. The measure of active buyers is $B^A = \mathcal{S}b^A$, and the measure of inactive buyers is $B^I = 1 - B^A$.

Special Case: Analytical Solution for the Invariant Size Distribution

Assume an environment without exogenous (z,φ) shocks, and let us introduce the object $\sigma_n = S_n/(S_0+\mathcal{S})$, defined for any $n = 0, 1, 2, \ldots$ Then, when $\delta_f = 0$, we can re-write the flow equations of Section 3.5 at steady state as:

$$\eta(\theta_n)\sigma_{n-1} + (n+1)\delta_c\sigma_{n+1} - \big(\eta(\theta_{n+1}) + n\delta_c\big)\sigma_n = 0$$

for any $n \geq 1$, and $\delta_c\sigma_1 - \eta(\theta_1)\sigma_0 = 0$. Since $\sum_{n=0}^{+\infty}\sigma_n = 1$ by construction, $\{\sigma_n\}$ follows a stationary birth-death process, with Markov transition rates $\eta(\theta_{n+1})$ and $n\delta_c$ for transitions $n \to (n+1)$ and $n \to (n-1)$, respectively. If we solve the difference equation on $n \geq 1$ recursively, we find:

$$\sigma_n = \frac{1}{n!}\frac{\prod_{i=0}^{n-1}\eta(\theta_{i+1})}{(\delta_c)^n}\sigma_0 \tag{D.2.3}$$

Imposing that $\sum_{n=0}^{+\infty}\sigma_n = 1$ in equation (D.2.3) yields:

$$\sigma_0 = \left(1 + \sum_{n=1}^{+\infty}\frac{1}{n!}\frac{\prod_{i=0}^{n-1}\eta(\theta_{i+1})}{(\delta_c)^n}\right)^{-1} \tag{D.2.4}$$

From the last expression, it is clear that $\{\sigma_n\}$ admits an ergodic representation if, and only if:

$$\sum_{n=1}^{+\infty}\frac{1}{n!}\frac{\prod_{i=0}^{n-1}\eta(\theta_{i+1})}{(\delta_c)^n} < +\infty \tag{D.2.5}$$

Under necessary condition (D.2.5), the stationary solution of the birth-death process $\{\sigma_n\}$ is given by (D.2.3)-(D.2.4). Using that $g_n = \sigma_n(1 + S_0/\mathcal{S})$ for $n \geq 1$, and $S_0/\mathcal{S} = \sigma_0/(1-\sigma_0)$, we then have:

$$g_n = \frac{S_0}{\mathcal{S}}\frac{1}{n!}\frac{\prod_{i=0}^{n-1}\eta(\theta_{i+1})}{(\delta_c)^n}, \qquad \text{with } \frac{S_0}{\mathcal{S}} = \left[\sum_{n=1}^{+\infty}\frac{1}{n!}\frac{\prod_{i=0}^{n-1}\eta(\theta_{i+1})}{(\delta_c)^n}\right]^{-1},$$

Note that for as long as there is a fat right-tail in the distribution of market tightness, this is a realistic description of the fat-tailed, Pareto-like size distributions that we see in the data (e.g. Luttmer (2007)).

D.3 Conditional and Aggregate Price Statistics

This section shows how to calculate, using the model's stationary solution, the conditional and aggregate price statistics that we use in the validation exercise.

0. Background and Definitions

Consider a price spell whose starting date is normalized to $\underline{t} = 0$ and which lasts until some unknown time $\bar{t} \geq 0$. Let T denote the total duration of the price spell (a continuous, non-negative random variable), and let $\mathcal{F} : \mathbb{R}_+ \to [0,1]$ be the c.d.f. of T. We define the *survival function* associated to duration T, denoted $\mathcal{S}^T$, as the probability that the price spell lasts at least $t \leq \bar{t}$ periods, i.e. $\mathcal{S}_t^T := \mathbf{Pr}[T \geq t] = 1 - \mathcal{F}_t$. Consequently, the probability that the price spell will end in the $[t, t+\Delta]$ interval is:

$$\mathbf{Pr}[t < T \leq t+\Delta] = \mathcal{S}_t^T - \mathcal{S}_{t+\Delta}^T$$

The *hazard function* is defined as the probability that a spell ends within the $[t, t+\Delta]$ interval, conditional on having lasted until time t, i.e. the object $\mathbf{Pr}[t < T \leq t+\Delta|T > t]$. Using Bayes' rule, we can write the hazard function in terms of the survival function as follows: $\mathbf{Pr}[t < T \leq t+\Delta|T > t] = 1 - \mathcal{S}_{t+\Delta}^T / \mathcal{S}_t^T$. The *instantaneous hazard rate* is then defined by the continuous-time limit: $h_t := \lim_{\Delta \searrow 0} \frac{1}{\Delta}\left(1 - \mathcal{S}_{t+\Delta}^T / \mathcal{S}_t^T\right)$. Using L'Hôpital's rule, we have:

$$h_t = -\boldsymbol{\partial}_t \log \mathcal{S}_t^T \tag{D.3.1}$$

Hence, defining the *cumulative hazard* as $\mathcal{H}_t := \int_0^t h_s \mathrm{d}s$, the cumulative hazard and the survival functions are related by $\mathcal{S}_t^T = \exp\{-\mathcal{H}_t\}$ (as $\mathcal{S}_0^T = 1 - \mathcal{H}_0 = 1$). Using this result, note that we can write the discrete-time hazard function in terms of the instantaneous hazard as follows:

$$\begin{aligned}
\mathbf{Pr}[t < T \leq t+\Delta|T > t] &= 1 - \frac{\mathcal{S}_{t+\Delta}^T}{\mathcal{S}_t^T} \\
&= 1 - \exp\Big\{\mathcal{H}_{t+\Delta} - \mathcal{H}_t\Big\} \\
&= 1 - \exp\left\{-\int_t^{t+\Delta} h_s \mathrm{d}s\right\}
\end{aligned} \tag{D.3.2}$$

Finally, the expected duration of price spells is given by $\mathbb{E}\{T\} = \int_0^{+\infty} t\mathrm{d}\mathcal{F}_t$. Integrating by parts and using that $\mathcal{S}_t^T = 1 - \mathcal{F}_t$, we obtain:

$$\mathbb{E}\{T\} = \int_0^{+\infty} \mathcal{S}_t^T \mathrm{d}t \tag{D.3.3}$$

Let us now compute these objects at the stationary solution of the model. Throughout, we consider a typical firm of fixed type $(n_t, z_t) = (n, z) \in \mathbb{N} \times \mathcal{Z}$ at time t, and let the random variable $T_n(z, \varphi)$ denote the duration of price spells of the firm in aggregate state $\varphi \in \Phi$.

1. Instantaneous Hazard Rate

Conditional on firm survival, the probability that a firm of type (n, z) changes its price at some time within the interval $[t, t+\Delta]$, given that the price spell was still ongoing at date t, is:

$$\begin{aligned}
\mathbf{Pr}\Big[t < T_n(z,\varphi) \leq t+\Delta \Big| T_n(z,\varphi) > t\Big] = &\Big[\eta\big(\theta_{n+1,t+\Delta}(z,\varphi)\big)\Delta + o(\Delta)\Big] + n\Big[\delta_c\Delta + o(\Delta)\Big] \\
&+ \sum_{\tilde{z}\neq z}\Big[\lambda_z(\tilde{z}|z)\Delta + o(\Delta)\Big] + \sum_{\tilde{\varphi}\neq\varphi}\Big[\lambda_\varphi(\tilde{\varphi}|\varphi)\Delta + o(\Delta)\Big]
\end{aligned}$$

where $o(\Delta)$ collects higher-order terms. In words, the hazard of a price change is the joint probability of an increase in size, a decrease in size, and an exogenous shock out of the current state, respectively. The *instantaneous hazard rate* (as defined in (D.3.1)) is, therefore:

$$h_n(z,\varphi) = \eta\big(\theta_{n+1}(z,\varphi)\big) + n\delta_c + \sum_{\tilde{z}\neq z}\lambda_z(\tilde{z}|z) + \sum_{\tilde{\varphi}\neq\varphi}\lambda_\varphi(\tilde{\varphi}|\varphi)$$

Note the absence of time subscripts in the above expression. This is a convenient implication of our block-recursive structure: as described in Section 3.5, $\theta_{n,t}(z,\varphi)$ is a jump variable that need not explicitly be indexed by time t. That is, all time variation in the instantaneous hazard rate is encapsulated in its explicit dependence on the aggregate state φ. Thus, for a given φ, the hazard of price changes is constant in time.

Two relevant implications of this result follow:

- Since the instantaneous hazard is flat at the firm level, the firm-level cumulative hazard is linear in time (though non-linear in the aggregate state):

$$\mathcal{H}_{n,t}(z,\varphi) = h_n(z,\varphi)t \tag{D.3.4}$$

The survival function, in turn, takes the simple form $\mathcal{S}^T_{n,t}(z,\varphi) = \exp\big\{-h_n(z,\varphi)t\big\}$.

- This result does not mean, however, that hazard rates are not time-dependent at higher levels of aggregation. Indeed, the measured aggregate hazard rate of price changes is time-varying because aggregate shocks generate slow-moving dynamics in the distribution of firms across states. In particular, the cross-sectional average hazard of price changes is equal to:

$$H_t(\varphi) := \sum_{n\in\mathbb{N}}\sum_{z\in\mathcal{Z}} g_{n,t}(z)h_n(z,\varphi)$$

where $\{g_{n,t}(z,\varphi)\}$ is the firm-size probability mass function (p.m.f.), $g_n(z) = S_{n,t}(z)\Big/\sum_n\sum_z S_{n,t}(z)$. For instance, in periods of high firm entry, the size distribution shifts to the left, so the aggregate hazard puts more weight on the hazard rates of small firms.

2. Frequency of Price Changes

We define the frequency of price changes over a time window of length one (i.e. $1/\Delta$ sub-periods) as the cumulative probability of a price change after a spell of such length. Using equation (D.3.2) and the fact that the instantaneous hazard rate is flat at the firm level (equation (D.3.4)), we can now easily write this probability as:

$$f_n(z,\varphi) = 1 - \exp\Big\{-h_n(z,\varphi)\Big\} \tag{D.3.5}$$

The frequency of price changes at the (n,z)-level is a jump variable that is time-independent for as long as there are no transitions in the aggregate state φ. The average frequency of price adjustment in the cross-section of firms is:

$$F_t(\varphi) := \sum_{n\in\mathbb{N}}\sum_{z\in\mathcal{Z}} g_{n,t}(z)f_n(z,\varphi)$$

Hence, the aggregate frequency of price changes evolves over time according to the underlying distribution dynamics.

3. Expected Duration of Price Spells

From equation (D.3.4), it is readily seen that the price duration $T_n(z,\varphi)$ follows an exponential distribution with parameter $h_n(z,\varphi)$. The average duration (equation (D.3.3)) then reduces to the reciprocal of the instantaneous hazard. Expressed in terms of frequency, this means:

$$\mathbb{E}\big\{T_n(z,\varphi)\big\} = -\frac{1}{\log\big(1-f_n(z,\varphi)\big)} \tag{D.3.6}$$

At the population level, once again expected durations are affected by the slow-moving distributional dynamics. Then:

$$D_t(\varphi) := \sum_{n\in\mathbb{N}}\sum_{z\in\mathcal{Z}} \frac{g_{n,t}(z)}{h_n(z,\varphi)}$$

is average expected duration of prices at time t.

4. Moments of the Distribution of Price Changes

Finally, we report moments of the distribution of (non-zero) price log-changes.

- The *expected absolute price change* in market (n,z) is defined as the average log change in prices. Denoting $\widehat{p} \equiv \log p$, we have:

$$
\begin{aligned}
\mathbb{E}_t\Big\{\Big|\widehat{p}_{n,t+\Delta}(z,\varphi)-\widehat{p}_{n,t}(z,\varphi)\Big|\Big\} &= \Big(\eta\big(\theta_{n+1,t+\Delta}(z,\varphi)\big)\Delta+o(\Delta)\Big)\times\Big|\widehat{p}_{n+1,t+\Delta}(z,\varphi)-\widehat{p}_{n,t}(z,\varphi)\Big| \\
&\quad + n\Big(\delta_c\Delta+o(\Delta)\Big)\times\Big|\widehat{p}_{n-1,t+\Delta}(z,\varphi)-\widehat{p}_{n,t}(z,\varphi)\Big| \\
&\quad + \sum_{\tilde{z}\neq z}\Big(\lambda_z(\tilde{z}|z)\Delta+o(\Delta)\Big)\times\Big|\widehat{p}_{n,t+\Delta}(\tilde{z},\varphi)-\widehat{p}_{n,t}(z,\varphi)\Big| \\
&\quad + \sum_{\tilde{\varphi}\neq\varphi}\Big(\lambda_\varphi(\tilde{\varphi}|\varphi)\Delta+o(\Delta)\Big)\times\Big|\widehat{p}_{n,t+\Delta}(z,\tilde{\varphi})-\widehat{p}_{n,t}(z,\varphi)\Big|
\end{aligned}
$$

where $|.|$ denotes the absolute value. Therefore, letting $\mu_n^\Delta(z,\varphi)$ denote the expected absolute price log-change, taking the continuous-time limit we obtain:

$$
\begin{aligned}
\mu_n^\Delta(z,\varphi) &= \eta\big(\theta_{n+1}(z,\varphi)\big)\Big|\widehat{p}_{n+1}(z,\varphi)-\widehat{p}_n(z,\varphi)\Big| + n\delta_c\Big|\widehat{p}_n(z,\varphi)-\widehat{p}_{n-1}(z,\varphi)\Big| \\
&\quad + \sum_{\tilde{z}\neq z}\lambda_z(\tilde{z}|z)\Big|\widehat{p}_n(\tilde{z},\varphi)-\widehat{p}_n(z,\varphi)\Big| + \sum_{\tilde{\varphi}\neq\varphi}\lambda_\varphi(\tilde{\varphi}|\varphi)\Big|\widehat{p}_n(z,\tilde{\varphi})-\widehat{p}_n(z,\varphi)\Big|
\end{aligned} \tag{D.3.7}
$$

- The *variance* of the distribution of price changes is given by:

$$
\mathbb{V}_t\Big\{\Big|\widehat{p}_{n,t+\Delta}(z,\varphi)-\widehat{p}_{n,t}(z,\varphi)\Big|\Big\} = \mathbb{E}_t\Bigg\{\bigg(\Big|\widehat{p}_{n,t+\Delta}(z,\varphi)-\widehat{p}_{n,t}(z,\varphi)\Big| - \mathbb{E}_t\Big\{\Big|\widehat{p}_{n,t+\Delta}(z,\varphi)-\widehat{p}_{n,t}(z,\varphi)\Big|\Big\}\bigg)^2\Bigg\}
$$

Following the usual derivation, in the continuous-time limit we obtain:

$$
\begin{aligned}
\sigma_n^\Delta(z,\varphi) &= \eta\big(\theta_{n+1}(z,\varphi)\big)\Big(\Big|\widehat{p}_{n+1}(z,\varphi)-\widehat{p}_n(z,\varphi)\Big|-\mu_n^\Delta(z,\varphi)\Big)^2 + n\delta_c\Big(\Big|\widehat{p}_n(z,\varphi)-\widehat{p}_{n-1}(z,\varphi)\Big|-\mu_n^\Delta(z,\varphi)\Big)^2 \\
&\quad + \sum_{\tilde{z}\neq z}\lambda_z(\tilde{z}|z)\Big(\Big|\widehat{p}_n(\tilde{z},\varphi)-\widehat{p}_n(z,\varphi)\Big|-\mu_n^\Delta(z,\varphi)\Big)^2 + \sum_{\tilde{\varphi}\neq\varphi}\lambda_\varphi(\tilde{\varphi}|\varphi)\Big(\Big|\widehat{p}_n(z,\tilde{\varphi})-\widehat{p}_n(z,\varphi)\Big|-\mu_n^\Delta(z,\varphi)\Big)^2
\end{aligned} \tag{D.3.8}
$$

where $\sigma_n^\Delta(z,\varphi)$ denotes the variance of price changes.

Time subscripts have again been dropped from the firm-level statistics by the block recursivity argument: pricing policies, when conditioned on the realization of the aggregate state, are time-invariant. At the population level, these moments now cannot be aggregated using g (the unconditional firm distribution), for not all firms change prices every period. Instead, we use the so-called *renewal distribution* of firms, that is, the distribution of firms conditional on a price adjustment. Since the probability that a firm of type (n,z) changes prices is given by the frequency $f_n(n,z)$ (equation (D.3.5)), the renewal distribution is given by:

$$
r_{n,t}(z,\varphi) := \frac{g_{n,t}(z)f_n(z,\varphi)}{\sum_{n\in\mathbb{N}}\sum_{z\in\mathcal{Z}} g_{n,t}(z)f_n(z,\varphi)}
$$

Figure A.8 compares the unconditional and the renewal p.m.f.'s in the calibrated economy. We note that there is more mass on smaller firms relative to the unconditional distribution, as these firms change prices more frequently.

Then, the average expected size of a price change and the average standard deviation of price changes are given by:

$$
\mathcal{M}_t^\Delta(\varphi) := \sum_{n\in\mathbb{N}}\sum_{z\in\mathcal{Z}} r_{n,t}(z,\varphi)\mu_n^\Delta(z,\varphi) \quad \text{and} \quad \Sigma_t^\Delta(\varphi) := \sum_{n\in\mathbb{N}}\sum_{z\in\mathcal{Z}} r_{n,t}(z,\varphi)\sqrt{\sigma_n^\Delta(z,\varphi)}
$$

respectively.

D.4 Extension: Business-Stealing

We now present an alternative way to model that, on top of inactive buyers, firms can use their pricing contracts to attract customers that are already buying from other firms. One way to incorporate such possibility is to assume that, when posting a contract $\boldsymbol{\omega}$, a firm attracts a buyer that is randomly drawn from the pool of *all* buyers, whether matched or unmatched. Then, firms now face an aggregate risk $\tau(\varphi) > 0$ of losing a customer to another seller, where $\tau(\varphi)$ is determined in equilibrium. One complication now is that τ is an equilibrium object which will depend on the aggregate state through both market tightness and the firm distribution. This means that sellers would need to forecast τ in an economy with aggregate shocks, thereby breaking the block recursivity property. A model with no aggregate shocks can still be solved, however, by using the solution methods developed above.

In the latter case, the HJB equation of a seller of size n promising utility x changes only slightly. In particular, V^S is again given by (6), except now the "effective" customer separation rate is:

$$\widehat{\delta}_c(n) := \delta_c + \frac{\tau}{n}$$

Here, $\delta_c > 0$ is the exogenous separation rate, and τ/n is the per-customer *endogenous* separation rate. Since meeting rates are independent across firms, the aggregate customer attrition rate that is due to firm pricing is $\tau := \int_{x\in\mathcal{X}} \eta\big(\theta(x)\big)\mathrm{d}S(x)$, where $S(x)$ is the measure of firms promising utility x. The HJB equation of the buyer (equation (5)) will be modified similarly, using $\widehat{\delta}_c(n;\varphi)$ for δ_c.

Since firms are atomistic and their contracts cannot affect the aggregate attrition rate τ, this variable is taken as given by sellers. Thus, the equivalence between seller's and joint surplus problems allows us to write the equilibrium aggregate separation rate more simply as:

$$\tau(\varphi) = \sum_{n=1}^{+\infty} \eta\big(\theta_n(z)\big) S_{n-1}(z) \tag{D.4.1}$$

where $S_n(z)$ is the measure of firms of type (n, z), and $\eta\big(\theta_n(z)\big)$ is the rate at which firms of type (n, z) attract customers. Moreover, since both inactive and active buyers can now form new matches, every buyer (matched or otherwise) is effectively in the pool of potential new customers for firms, and thus market tightness is defined by:

$$\theta_n(z) = \frac{1}{S_{n-1}(z)} \left(B_n^I(z) + \sum_{m=1}^{+\infty} \sum_{\tilde{z}\in\mathcal{Z}} B_m^A(\tilde{z}) \right)$$

for any $n \geq 1$ and $z \in \mathcal{Z}$.

D.5 Comparative Statics

In this section, we present a set of comparative statics exercises on the stationary recursive equilibrium to illustrate the behavior of the joint surplus, firm growth rates, price levels, and promised utilities, across parameter values. For illustration, we focus on a simplified version of the model in which the exogenous shocks (z, φ) are turned off. This simplification makes the intuition simpler and does not change the qualitative features of the equilibrium.

Figure D.1 shows, as a function of seller's size n, the equilibrium market tightness, promised utility, and price level for a numerical example.[64] We observe that, when the seller enters with $n = 1$ customer, she attracts a higher measure of buyers than any other type of firm (left panel). This is because these firms promise a high continuation utility to their buyers (middle panel) by setting relatively low prices (right panel). As growth unfolds, the promised utility decreases as the seller must compromise across more buyers, causing market tightness, and thus firm growth rates, to decline. In the stationary solution, this implies that a relatively larger measure of firms is of small size, which translates into a right-skewed stationary firm distribution (left panel in Figure D.2). The distribution of customers (middle panel) is similarly skewed by construction, whereas the distribution of inactive buyers looking to be matched with a firm has a larger mass on the left, as it is smaller firms who make the most attractive promises from an ex-post perspective.

In order to understand how sellers of different sizes implement the decline in promised utilities through adjustments in the price level, we can decompose the latter into the different components described in equation (15).

[64] For these examples, we use the parameter values: $v = 1$; $c = 0.63$; $\kappa = 0.98$; $\delta_f = 0.09$; $w = 0.15$; $\delta_c = 0.20$; and $\gamma = 0.546$.

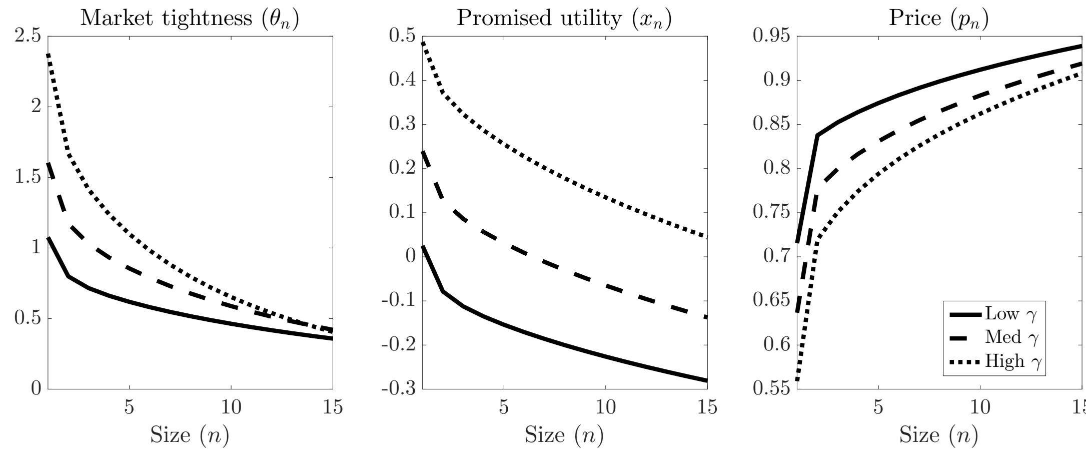

Figure D.1: Numerical Example: Market tightness, promised utility, and prices for the simple model with no exogenous (z, φ) shocks. Comparative statics with matching elasticity (γ).

In the model without shocks, the price decomposition simplifies to:

$$p_n = \underbrace{v - rx_n}_{\text{Baseline } \geq 0} + \underbrace{\delta_f(U^B - x_n)}_{\text{Exit } \leq 0} + \underbrace{\psi(x_{n+1})(x_{n+1} - x_n)}_{\text{Growth } \leq 0} + \underbrace{n\delta_c\left(\frac{U^B + (n-1)x_{n-1}}{n} - x_n\right)}_{\text{Separation } \leq 0} \tag{D.5.1}$$

where we have defined $\psi : x \mapsto \eta \circ \mu^{-1}\left(\frac{\Gamma^B}{x - U^B}\right)$ by equation (12). As the seller grows, the baseline price that the buyer is charged increases with size (first panel in Figure D.3). The intuition is standard: since sellers cannot price discriminate, they increasingly prefer to extract rents from the current base as the base increases in size. The overall price level is then adjusted for the different state transitions. Both the exit and the growth components (second and third panels) are negative, putting downward pressure on prices. As sellers grow, the promised utility for buyers approaches the outside option U^B, and the adjustment on price due to firm exit decreases in absolute value. Similarly, the growth compensation is relatively larger when the seller is small, as increases in the customer base lead to larger revisions of the continuation utility. As the seller grows, however, the opportunity cost of firm growth for the matched buyer (i.e. $x_{n+1} - x_n$) declines, and the overall price level needs to be compensated less for the event of a growth shock. Finally, the separation component (fourth panel) is placing downward pressure on prices as well, since the expected value that each customer expects to obtain after a separation (the object $\frac{U^B + (n-1)x_{n-1}}{n}$) exceeds her current value (x_n). Both of these are decreasing in size, and thus the net effect depends on the elasticity of the promised utility x_n with respect to size.

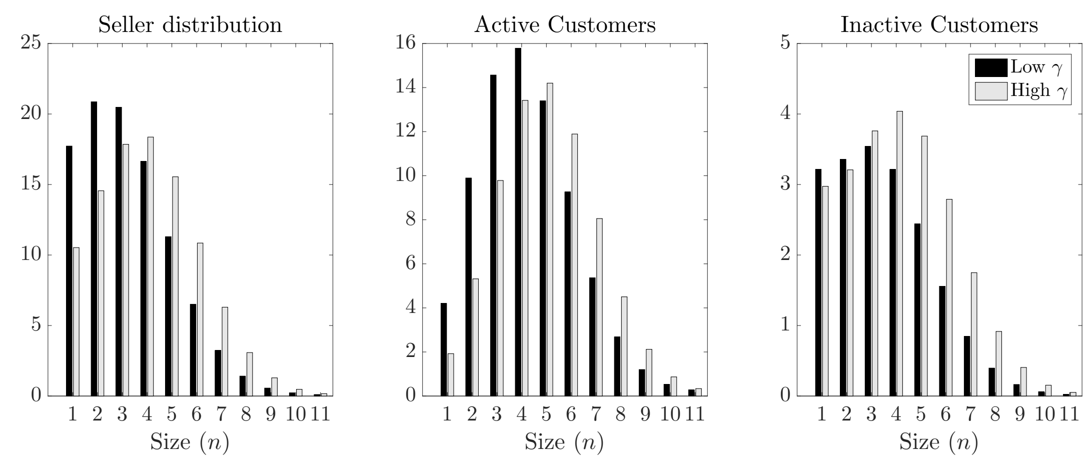

Figure D.2: Numerical Example: Stationary distributions. Comparative statics with matching elasticity (γ). The diagrams plot the stationary distribution of sellers ($\{S_n\}$), active customers (equation (17)), and inactive buyers (equation (16)).

Let us now compare across different levels of the matching elasticity. According to our intuition above, a higher value for γ means that buyers absorb a higher share of the joint surplus. Indeed, this implies that market tightness, promised utilities and, therefore, the growth rate of sellers, are all higher when rent shares are more favorable to the customer (see Figure D.1). Sellers implement this with lower prices, and therefore manage to grow larger (see Figure D.2). Besides changes in levels, we also observe interesting slope effects: when incoming customers extract larger rent shares, the increase in market tightness is higher for smaller firms, and therefore growth rates are tilted toward them. This generates relatively faster growth for smaller firms, which accounts for the relative decrease in the measure of small firms that we see in Figure D.2. Interestingly, these slope effects are implemented through the price level (p_n) rather than the continuation values (x_n): as γ increases, the price schedule becomes steeper, but the promised utility does not.

Figures D.4, D.5, and D.6 below show comparative statics for the entry cost (κ), the separation rate (δ_c), and the search cost (c), respectively, with similar intuitions to the ones used above.

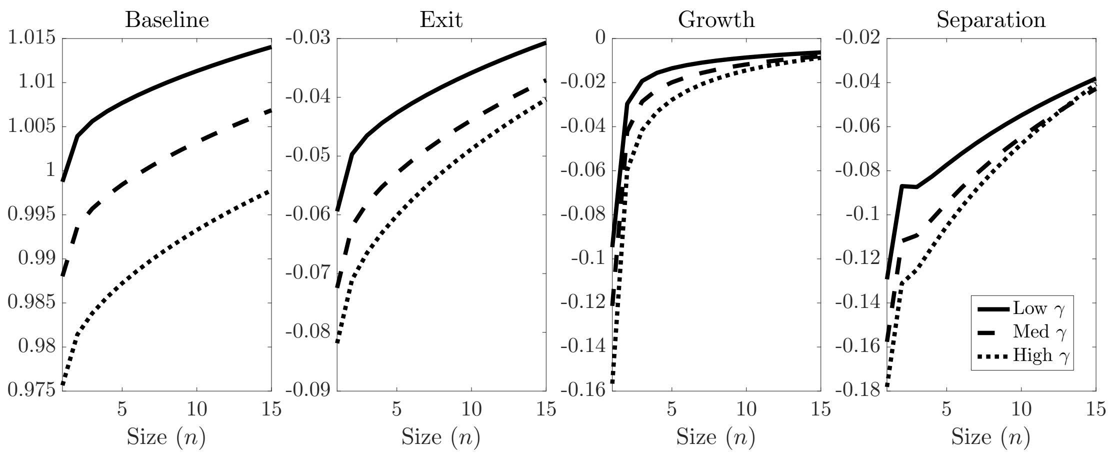

Figure D.3: Numerical Example: Price decomposition. Comparative statics with matching elasticity (γ). The diagrams plot the different price components of equation (D.5.1).

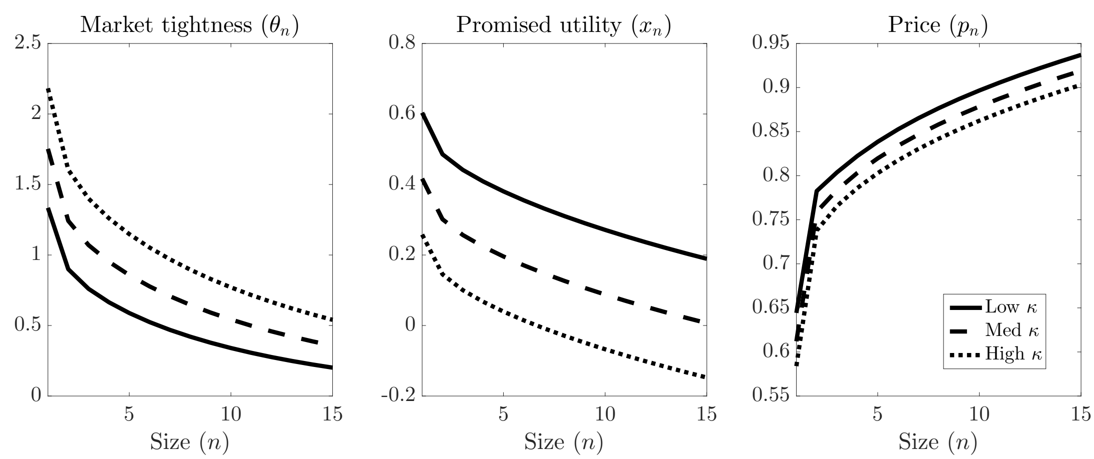

Figure D.4: Comparative statics with entry cost (κ).

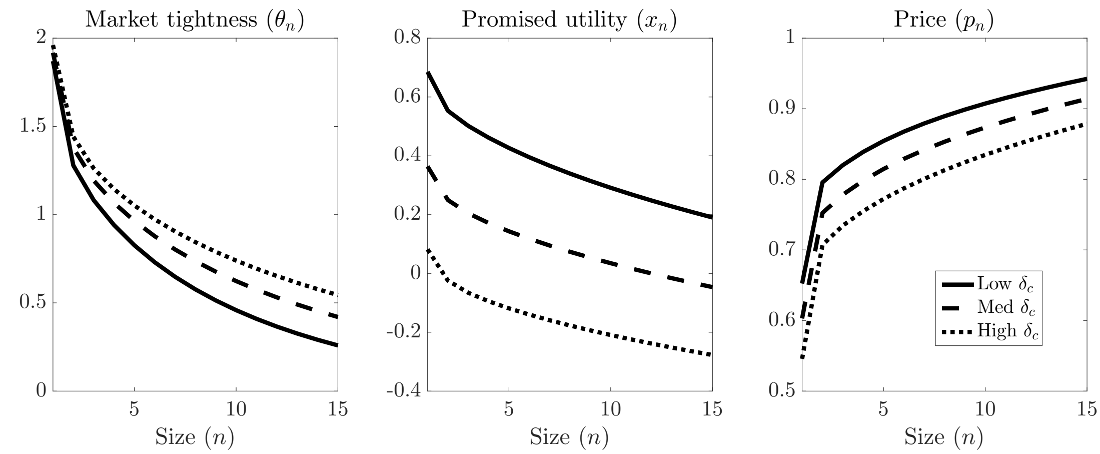

Figure D.5: Comparative statics with separation rate (δ_c).

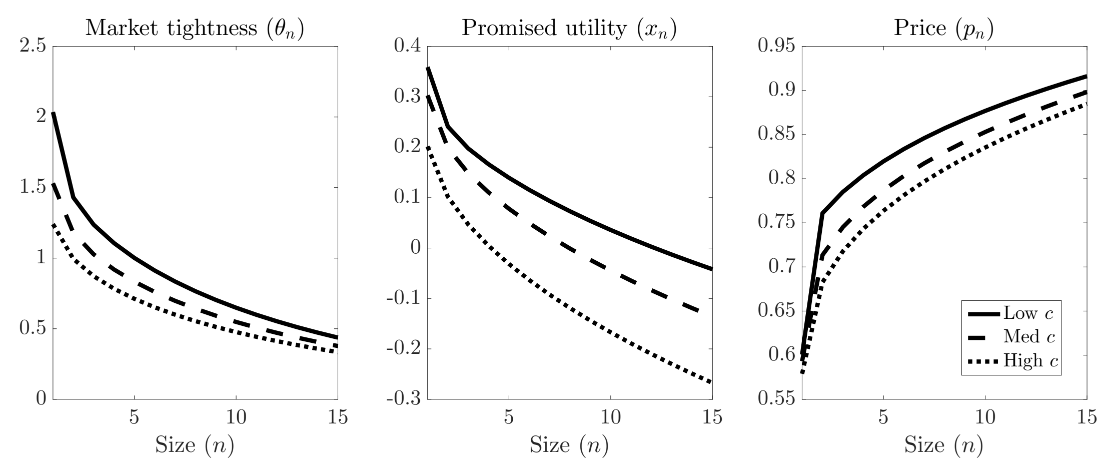

Figure D.6: Comparative statics with search cost (c).