Firm Dynamics and Pricing under Customer Capital Accumulation

Pau Roldan† Sonia Gilbukh‡
New York University NYU Stern

First draft: December 31, 2016
This draft: March 15, 2018

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 calibrate the model to match 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 response of markups to aggregate shocks. We find that there is incomplete price pass-through leading to procyclicality in the average markup, with smaller firms being more responsive to the shock. Finally, we show that our estimated model can offer an interpretation for two secular trends experienced in the U.S. since the early 1980s: the decline in business dynamism and the rise in the average markup.

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, Philipp Kircher, Julian Kozlowski, Virgiliu Midrigan, Espen Moen, Simon Mongey, Joshua O’Donnell, Jesse Perla, Michael Peters, ThomasPhilippon, Tom Schmitz, Venky Venkateswaran, Ewout Verriest, and Basil Williams; from seminar audiences at Bank of Spain, BI Norwegian Business School, Carlos III, CEMFI, European University Institute, IESE, NYU, Federal Reserve Bank of St. Louis, Universitat Autònoma de Barcelona, University of Essex, and University of Toronto Rotman; and from conference participants at SED 2017, SAe 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.

†New York University, 19 W 4th St., New York, NY 10012. Corresponding author: pau.roldan@nyu.edu.

‡New York University Stern School of Business, 44 W 4th Street, New York, NY 10012.
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, in the retail sector 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. Moreover, Foster et al. (2008, 2016) have documented that, even though new and well-established firms exhibit very different behavior within a variety of commodity-like markets, the productivity advantage of entrants is only small and it dissipates within the first few years of operation. It is instead differences in idiosyncratic demand components which account for the bulk of the observed heterogeneity in performance. Hence, the evidence suggests a persistent demand-side channel of variation: firm investment in demand accumulation could account for differences in the life-cycle of businesses of similar productivities but different sizes.

In this paper, we formalize these ideas by developing an equilibrium theory of firm dynamics in 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. We then use an estimated version of the model to explore two macroeconomic questions. First, we explore the dynamics of prices and markups in response to aggregate supply and demand shocks, with an emphasis on how the response is

---

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. (2017) (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.
heterogeneous in the cross section of firms. Second, we use the model to connect two secular
trends that have occurred in the U.S. since the early 1980s: the decline in the firm entry rate,
and the increase in the average markup.

We model 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 het-
erogeneous in the number of buyers that they sell the product to, since their choice of the
contract endogenously determines the rate at which demand accumulates over time. Out-
side of the market, idle sellers must pay a fixed entry (or “market penetration”) cost 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 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 endoge-
nously, 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 a balance between instantaneous revenues (via high prices)
and future market shares (via high continuation promises). 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

\(^2\) 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.
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 exist constant returns in technology. Therefore, on top of price dispersion, the model generates a well-defined and right-skewed firm distribution.

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. To further obtain analytical tractability, we exploit the fact that the search equilibrium is block-recursive, a common property of models of directed search (e.g. Shi (2009), Menzio and Shi (2010, 2011)). This property implies that, in order to evaluate payoffs, buyers and sellers need not keep track of the distribution of agents across aggregate 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. 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.

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 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 estimation, we calibrate the model to moments of the distribution of relative prices and sales, which we take from our sample of micro-pricing data. 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 conduct two quantitative exercises. In the first exercise, we analyze the response of the economy to both aggregate demand and aggregate supply shocks. In this exercise, we find both level and distributional effects. First, we show that the price pass-through of aggregate temporary supply shocks (e.g. marginal costs) is incomplete: in the wake of a negative shock, firms choose to front-load their contracts by charging slightly 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 the shock, firms choose to trade-off immediate losses to future market shares, which they achieve by inter-temporally transferring the burden of shocks onto their buyers. Since the price level reacts less than one for one to the increase in marginal costs, the markup is procyclical. In addition, we describe the effects of aggregate demand shocks on firm pricing. Shocks that lower the marginal propensity to consume by buyers generate a bust in demand and lower prices instantaneously. Since the shock mean-reverts, firms depress their promises on impact but increase prices in the transition. Overall, the markup response in this case is also procyclical. Thus, the model provides a micro-founded explanation for why price-cost markups could be procyclical in the data in response to both aggregate supply and demand shocks (e.g. Nekarda and Ramey (2013)).

These level effects are accompanied by important distributional changes in the transition. Through a decrease in the continuation promise of firms, both negative supply and negative demand shocks lead to a decrease in the number of new matches, and firms temporarily shrink in size. This implies that the pass-through is less incomplete for larger firms. The reason is that the slow-moving, left-ward shift in the size distribution, combined with the fact that small firm’s pricing policies are relatively more sensitive to size changes in the estimated version of the model, implies that a stronger markup response for smaller firms. Moreover, the latter type of firms also experience a more persistent response, because during the transition the measure of lower-price, smaller-size firms relatively increases.

In our second and final quantitative exercise, 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 the average markup of the economy. 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 trend coupled with a slow 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 (e.g. 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 by introducing a decline in the buyer’s search cost (e.g. the introduction of new shopping technologies such as the Internet). In the estimated model, a decline in the search cost implies an increase in the average size of firms and a subsequent increase in the average markup in the economy. The reason is that, when firms seek to retain buyers whose outside options have improved, they must raise their promises to new customers. The increase in the promise is paid for by raising prices on the current customers. This, in turn, increases the meeting rate of firms, who start to grow faster. The entry rate falls because, even though the probability of finding the first customer has increased, there is now less scope for new buyers as a higher share of them are finding incumbents. These effects thus lead to a shift of the firm size distribution toward higher-markup firms.
Related Literature  There is a large amount of survey evidence suggesting 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. More recently, using the same pricing dataset that we use in Section 4.1, Paciello et al. (2017) 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 sticky).

Our paper is primarily related to a long tradition of building customer capital into macroeconomic models of firm pricing. Work 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. 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

---

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)).

4 As an application of this approach, Gilchrist et al. (2017) 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.
future.\footnote{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 renge 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 above a certain threshold.}

While our model shares these features with the literature, a contribution 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.\footnote{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 (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.} 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 (2017)), and household behavior (Hall (2008)). A number of papers have further analyzed how expenditures on intangibles may shape the evolution of firms and industries. 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 plantspecific 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.\footnote{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 (2017) studies the implications of product sorting by uninformed consumers on the industry life cycle and the degree of market competition; Bai et al. (2017) 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.}

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 firms use discounts to attract 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 3, 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 in our framework 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 4.3).\(^8\) Luttmer (2006) and Fishman and Rob (2003) also study the implications of customer acquisition for the firm size distribution, but in those papers there is no meaningful role for prices. In contrast, like us, Paciello et al. (2017) 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 of different productivities with retention concerns, we offer a complementary view whereby firms of different sizes use prices to attract customers.\(^9\)

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-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 stylized patterns of growth, which we will document for our sample of retail firms in Section 4.1, 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.

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. (2017) 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 set lower markups relative to

---

\(^{8}\) 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. For a full discussion on this issue, see Section 5.2.

\(^{9}\) Methodologically, another innovation of our framework relative to Paciello et al. (2017) 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.
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 are important contributors 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 4.1. 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 (2017) 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) or Kaas and Kircher (2015).10 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 dynamics. Secondly, we make use of dynamic long-term 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.

Outline The remainder of the paper proceeds as follows. In Section 2 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 3 discusses the main mechanism, shows that search frictions can deliver different profiles for prices, and explains the role of each central assumption. Section 4 describes our application to the U.S. retail sector, and proceeds to the calibration of the model and its quantitative results, including

10 For a recent survey of directed search theory, see Wright et al. (2017).
the response of the economy to aggregate shocks. Section 4.4 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 5 presents extensions to the baseline model, and Section 6 concludes. The Appendix includes supplementary tables and figures, all the proofs, and some additional theoretical results.

2 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.

2.1 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 \( \varphi \) taking values in a discrete and finite support \( \Phi := \{ \varphi < \cdots < \varphi_\varphi \} \), 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 \( \varphi \) follows a homogenous continuous-time Markov chain with generator matrix \( \Lambda_\varphi := [\lambda_\varphi(\varphi' | \varphi)] \), where \( \lambda_\varphi(\varphi' | \varphi) \) denotes the intensity rate of a \( \varphi \)-to-\( \varphi' \) transition.

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 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 \( \varphi \) is visited, the economy always spends a non-zero measure of time in it).

12 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.
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 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 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.\footnote{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 (2017) provides a micro-foundation for this view based on the evolution of buyers’ consideration sets.}

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, \( \varphi \). Since \( v \) and \( c \) relate directly to buyers’ preferences, this state-dependence incorporates the possibility of aggregate and exogenous demand fluctuations into the model.\footnote{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.}

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 \( n_t \in \mathbb{N} := \{1, 2, 3, \ldots\} \) 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 \( Z := \{z, \ldots, z\} \) of cardinality \( |Z| = k_z \geq 2 \). Like the aggregate state, the idiosyncratic state follows a continuous-time Markov chain with generator matrix \( \Lambda_z := [\lambda_z(z'|z)] \), where \( \lambda_z(z'|z) \) denotes the transition rate from \( z \) to \( z' \).\footnote{The usual conditions apply. For all \( z \in Z \): \( \lambda_z(z|z) \leq 0; \lambda_z(z'|z) \geq 0, \forall z' \neq z; \sum_{z' \in Z} \lambda_z(z'|z) = 0; \sum_{z' \neq z} \lambda_z(z'|z) < +\infty. \)} 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 \( C(n; z, \varphi) \), which depend on the idiosyncratic state \( (n, z) \), as well as possibly the aggregate state \( \varphi \). Further, we make the following assumptions:

**Assumption 1** For all \( (z, \varphi) \in Z \times \Phi \):

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

(ii) \( C(n; z, \varphi) \) 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 \( C \) 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 \( C \) 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 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 \( \delta_c > 0 \).\(^{16}\) 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, \( \varphi \). A firm must pay this cost every time it has lost all its customers and intends to re-enter the market, so \( \kappa \) 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, \( \kappa \) 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 \( \pi_z \), 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

\(^{16}\) In Section 5.1 we present ways to endogenize this customer separation rate.
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).\(^{17}\) 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.\(^ {18}\) 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.\(^ {19}\) 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.

**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 generically denote this value by $x$, let $\mathcal{X} = [x, \bar{x}] \subseteq \mathbb{R}_+$ 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 seek to be matched under promised utility $x$, which is posted by a mass $S(x; \varphi) \geq 0$ of sellers. 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 mar-

\(^{17}\) 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.

\(^{18}\) 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 3.

\(^{19}\) Intuitively, this is meant to capture the idea that, in largely populated markets where implicit relationships develop, buyers are anonymous to the seller. In Section 5.2 we discuss the implications of relaxing the no discrimination assumption.
ket’s tightness. Importantly, agents take the mapping \( \theta : X \times \Phi \rightarrow [0, +\infty) \) 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 \( x \in X \), a single applicant obtains offer \( x \) at the endogenous Poisson arrival rate \( \mu(\theta(x; \varphi)) \geq 0 \), while a seller successfully finds an applicant for offer \( x \) at the Poisson arrival rate \( \eta(\theta(x; \varphi)) \geq 0 \), where \( \eta(\theta) = \theta \mu(\theta) \).

Further, we impose the following regularity conditions:

**Assumption 2** The meeting rates satisfy:

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

(ii) \( \eta \) is increasing and concave; \( \mu \) is decreasing and convex;

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

(iv) \( \eta(0) = \lim_{\theta \uparrow +\infty} \mu(\theta) = 0 \), and \( \lim_{\theta \uparrow +\infty} \eta(\theta) = \lim_{\theta \downarrow 0} \mu(\theta) = +\infty \).

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 \( x \in X \) 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 \( (n, x; s) \), where henceforth we use \( s = (z, \varphi) \) 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.\textsuperscript{20}

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 \((n, x; s)\) as the object:\textsuperscript{21}

\[ \omega := \{ p, x'(n'; s') \} \]

The elements of a recursive contract \(\omega\) 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 \(x'(n'; s') \subseteq X\) 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 \(s' \in \{(z', \phi), (z, \phi')\}\). Hence, by stationarity, conditional on a fixed exogenous state \(s\) (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. \(x'(n'; s') = x\) when \(n' = n\) and \(s' = s\)).\textsuperscript{22} 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 2.5.

\section{2.2 Buyer’s Problem}

\textbf{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 \(U^B(\phi)\) in state \(\phi \in \Phi\). The buyer enters the sub-market that offers the highest valuation, and therefore:

\[ U^B(\phi) = \max_{\hat{x}(\phi) \in X} u^B(\hat{x}(\phi); \phi) \quad (1) \]

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

\[ ru^B(x; \varphi) = -c(\varphi) + \mu(\theta(x; \varphi)) \left( x - u^B(x; \varphi) \right) + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi) \left( u^B(x; \varphi') - u^B(x; \varphi) \right) \] (2)

for any \( x \in \mathcal{X} \). \(^{23}\) Equation (1) states that the inactive buyer searches in the product market that promises the highest expected value, \( \hat{x}(\varphi) \). 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 \( \varphi \) to some \( \varphi' \), occurring at rate \( \lambda_\varphi(\varphi'|\varphi) \).\(^{24}\)

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

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

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 \( \mathcal{X}^*(\varphi) := \{ x \in \mathcal{X} : \theta(x; \varphi) > 0 \} \subseteq \mathcal{X} \) be the equilibrium set of markets in state \( \varphi \in \Phi \). Hence, for any given aggregate state \( \varphi \in \Phi \), we have:

\[ \mu(\theta(x; \varphi)) \left( x - U^B(\varphi) \right) = \Gamma^B(\varphi) \] (3)

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

\[ \Gamma^B(\varphi) := c(\varphi) + ru^B(\varphi) - \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi) \left( U^B(\varphi') - U^B(\varphi) \right) \] (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.

\(^{23}\) For a derivation of all the HJB equations in the main text, see Appendix D.1.

\(^{24}\) 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, \( u^B(x; \varphi) \) is actually shorthand notation for \( u^B(x; \varphi, \theta) \), where \( \theta \) here is a function mapping from \( \mathcal{X} \times \Phi \) to \( \mathbb{R}_+ \). Similar concise notation will be used throughout the paper.
For a given value of inactivity $U^B(\varphi)$, this ex-ante revenue-equalization condition uniquely pins down 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.\footnote{For an in-depth discussion of the game-theoretical foundations of this assumption, see Galenianos and Kircher (2009, 2012).}

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 $\mu$ 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 buyer who is currently consuming the homogeneous good from a firm of size $n \in \mathbb{N}$ and idiosyncratic productivity $z \in \mathbb{Z}$, under contract $\omega = \{p, x'(n'; s')\}$. The contract delivers the promised value $x$ to the customer, and it specifies the current price $p$ and the new continuation promises $x'(n'; s')$, to be delivered by the seller after a $n$-to-$n'$ or $s$-to-$s'$ transition, where $n' \in \{n - 1, n + 1\}$ and $s' \in \{(z', \varphi), (z, \varphi')\}$.

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

$$rV^B(n, \omega; s) = v(\varphi) - p + (\delta_f + \delta_c)(U^B(\varphi) - V^B(n, \omega; s))$$

$$+ (n - 1)\delta_c\left(x'(n - 1; s) - V^B(n, \omega; s)\right)$$

$$+ \eta\left(\theta\left(x'(n + 1; s); \varphi\right)\right)\left(x'(n + 1; s) - V^B(n, \omega; s)\right)$$

$$+ \sum_{z' \in \mathbb{Z}} \lambda_z(z'|x)\left(x'(n; z', \varphi) - V^B(n, \omega; s)\right)$$

$$+ \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi'|x)\left(x'(n; z, \varphi') - V^B(n, \omega; s)\right)$$

For an in-depth discussion of the game-theoretical foundations of this assumption, see Galenianos and Kircher (2009, 2012).
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 $p \in \omega$; 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 $x'_{(n-1);s} \in \omega$ 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 value $x'_{(n+1);s} \in \omega$. 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 $x'_{(n+1);s}$ 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.

### 2.3 Seller’s Problem

#### Incumbent Sellers

Consider a seller with idiosyncratic productivity $z \in Z$ 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 $V^S(n, x; s)$, is to select a new contract $\omega = \{p, x'(n'; s')\}$ for all of its $n$ customers so as to maximize the life-time value:

$$rV^S(n, x; s) = \max_{\omega \in \Omega} \left\{ pm - C(n; s) + \delta_f \left( V^S_0(\varphi) - V^S(n, x; s) \right) \ight. \right.
\right.
\right.
\right.
\right.$$
\right.
\right.$$
\right.
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.$$
\right.
where \( V_0^S(\varphi) \) denotes the value of having no customers (which we derive below). The right side of equation (6) has the following components. The term \([pn - C(n; s)]\) 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 \( x'(n - 1; s) \in \omega \). The second line shows that, by posting a new offer \( x'(n + 1; s) \in \omega \), 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 \( \theta \) schedule. In the event of a successful match, the seller would implement the new continuation value, and its state vector would transition from \((n, x; s)\) into \((n + 1, x'(n + 1; s); s)\). Finally, the value of the firm could change exogenously because of a state transition from \( s = (z, \varphi) \) to \( s' \in \{(z', \varphi), (z, \varphi')\} \), as captured by the last two terms in equation (6).

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

\[
V^B(n, \omega; s) \geq x
\]  

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 \( \pi_z \) distribution. For each possible realization \( z_0 \in Z \), the contract is the object \( \{x'(1; z_0, \varphi)\} \), specifying the utility promised to the first customer of the firm under state \((z_0, \varphi)\). Thus, the potential entrant chooses amongst a menu of contracts, \( \omega_0(\varphi) := \{x'(1; z_0, \varphi)\}_{z_0 \in Z} \), contingent on each realization of productivity at entry.

\[26\] The object \( \Omega := \mathbb{R} \times [x, x]^k \) denotes the set of admissible contracts, and \( k = 3k_z k_{\varphi} - 1 \). For \( n = 1 \), we note that \( x'(n - 1; s) = \emptyset, \forall s \in Z \times \Phi \), and denote \( V^S(n - 1, x'(n - 1; s); s) \) by \( V_0^S(\varphi) \).
The ex-ante value of the potential entrant in aggregate state $\varphi$ is, therefore:

$$ rV_0^S(\varphi) = -\kappa(\varphi) + \sum_{z_0 \in Z} \pi_z(z_0)v_0^S(z_0, \varphi) + \sum_{\varphi' \in \Phi} \lambda_{\varphi'}(\varphi') \left( V_0^S(\varphi') - V_0^S(\varphi) \right) \quad (8) $$

This value is composed of the set-up flow cost $\kappa$ (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:

$$ v_0^S(z_0, \varphi) := \max_{x' \in X} \eta(\theta(x'; \varphi)) \left( V^S(1, x'; z_0, \varphi) - V_0^S(\varphi) \right) \quad (9) $$

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, the 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:

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

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.

### 2.4 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 simplifies the state space and renders the equilibrium computationally tractable.

**Joint Surplus Problem**

To start, consider a typical firm whose state vector is $(n, x; s)$, where recall that $s := (z, \varphi)$ collects the exogenous states. As seen in the last section, the optimal contract $\omega =$
\{p, x'(n'; s')\} 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 \(V^B(n, \omega; s)\).

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

\[
W(n, x; s) := V^S(n, x; s) + nx
\]

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

\[
(r + \delta_f)W(n, x; s) = \max_{x'(n'; s')} \left\{ n \left( v(\varphi) + (\delta_f + \delta_c)U^B(\varphi) \right) - \left( \mathcal{C}(n; s) + \eta \left( \theta(x'(n + 1; s); \varphi) \right)x'(n + 1; s) \right) + \eta \left( \theta(x'(n + 1; s); \varphi) \right) \left( W(n + 1, x'(n + 1; s); s) - W(n, x; s) \right) + n\delta_c \left( W(n - 1, x'(n - 1; s); s) - W(n, x; s) \right) + \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z) \left( W(n, x(z', \varphi); z', \varphi') - W(n, x; s) \right) + \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi'|\varphi) \left( W(n, x(z, \varphi'); z, \varphi') - W(n, x; s) \right) \right\} \tag{10}
\]

Intuitively, equation (10) 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 \(n \left( v(\varphi) + (\delta_f + \delta_c)U^B(\varphi) \right)\) 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, \(\mathcal{C}(n; s)\), and the expected costs of offering a life-time continuation value of \(x'(n + 1; s)\) 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 \(\varphi\).

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

**Proposition 1 (Joint Surplus Problem)**
i. The firm’s and the joint surplus problems are equivalent:

(a) Given a contract $\omega^* = \{p, x'(n'; s')\}$ that maximizes (6), $x'(n'; s')$ is a solution to (10).
(b) Conversely, for every vector $x'(n'; s')$ that solves (10), there exists a unique $p$ for which $\{p, x'(n'; s')\}$ is a solution to (6).

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

Proof. See 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 (10). In this problem, the contract is designed 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 fashion.

Part ii. of the proposition thus follows immediately from the first one, and clarifies why problem (10) 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 $x'(n'; s')$ that maximizes the size of the surplus under every possible combination of future states. In the second stage, the price level implements such an allocation, thereby splitting and distributing rents among the $(n + 1)$ agents involved, 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 $s = (z, \varphi)$ serve as sufficient statistics to determine the current surplus-maximizing policies. Thus, given $s$, there exists a sequence $\{W_n(s)\}_{n=1}^{\infty}$ of positive real numbers such that the joint surplus can be expressed as:

$$\forall (n, x) \in \mathbb{N} \times \mathcal{X} : \quad W_n(s) = W(n, x; s)$$

As a result, the policy that solves problem (10) 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.\footnote{For an application of this idea to a rich firm-dynamics search model of the labor market, see Schaal (2017).}
Characterization

Let us characterize the equilibrium policies that result from problem (10). 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)). Therefore, the tightness of market $x$ is:

$$\theta(x; \varphi) = \mu^{-1} \left( \frac{\Gamma^B(\varphi)}{x - U^B(\varphi)} \right)$$  \hspace{1cm} (11)

By Assumption 2.1 and continuity of $\theta$ on $x$, equation (10) describes the maximization of a continuous function over a compact support, so there exist promises $\{x'(n+1; s), x'(n-1; s), x'(n; s')\}$ and a price level $p(n; s)$ that solve the joint surplus problem. Once again, note that we index these policies by $n$, but not $x$.

Stage 1. Continuation promises  Let us begin with the upsizing choice. First, using equation (11) and differentiability of $\eta$, the following first-order condition is sufficient for optimality:

$$\frac{\partial \eta(\theta(x'; \varphi))}{\partial x'} (W_{n+1}(s) - W_n(s)) = \frac{\partial \eta(\theta(x'; \varphi))}{\partial x'} x' + \eta(\theta(x'; \varphi))$$  \hspace{1cm} (12)

Let $x'(n+1; s)$ denote the solution. Intuitively, the optimal continuation value $x'(n+1; s)$ 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 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 the raised promised value on the likelihood that the firm meets a new customer. On the other hand, for every dollar spent the continuation promise, 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 of all currently captive buyers of the firms, by $\eta(\theta(x'(n+1; s); \varphi))$ utils, which is required by promise-keeping.

As for the choices of $x'(n; s')$ and $x'(n-1; s)$, note that these do not feature anywhere in equation (10) once we impose that the joint surplus is invariant to promised utilities (Proposition 1, part ii.). Therefore, these objects cannot be determined by a surplus-maximizing condition similar to (12). Instead, these 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 (11), the firm’s choice must be consistent with the sorting behavior of inactive buyers. By symmetry, we have $x'(n-1; s) = x'(m+1; s)$ for two different sellers of size $n \geq 2$ and $m = n - 2$. In words, the optimal downsizing choice for a size-$n$ seller (left side of the

---

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

equality) must be consistent with the optimal upsizing choice for a firm of size $n - 2$ (right side). Similarly, when transitioning to another state, equation (11) and symmetry require that $x'(n; s') = x'(m + 1; s')$ for two sellers of sizes $n \geq 2$ and $m = n - 1$. Finally, for $n = 1$, the free entry condition must be satisfied, implying:

$$
k(\varphi) = \sum_{z_0 \in \mathcal{Z}} \pi(z_0) \eta\left(\theta(x'(1; z_0, \varphi); \varphi)\right) \left(W_1(z_0, \varphi) - x'(1; z_0, \varphi)\right)$$  \hspace{1cm} (13)

In sum, the set of active market segments in equilibrium is $X^* := \{x'(n; z, \varphi) : (n, z, \varphi) \in \mathbb{N} \times \mathcal{Z} \times \Phi\}$, where $x'(n; z, \varphi)$ solves (12) for all $n \geq 2$, and (13) for $n = 1$.

Once the equilibrium markets are pinned down, the remaining equilibrium objects readily follow. First, equilibrium market tightness levels are given by $\theta_n(z, \varphi) := \theta(x'(n; z, \varphi); \varphi)$ via equation (11). Since $\theta(x; \varphi)$ is an increasing and continuous function of $x$ (equation (3)), then $\theta_n$ inherits the size-dependence in $x'$. For instance, when $x'$ is decreasing in $n$, 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. Figure A.1 in Appendix A depicts the different markets in equilibrium for this case, where $s = (z, \varphi)$ is fixed to ease the visualization. All equilibrium markets are distributed on the $\theta$ 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. Indeed, we can write the law of motion of the seller’s customer base as:

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

where $\Delta > 0$ is small, and $o(\Delta)$ satisfies $\lim_{\Delta \to 0} \frac{o(\Delta)}{\Delta} = 0$. In the example of Figure A.1, the probability of attracting a new customer (first line) is relatively higher for smaller sellers. As the seller grows, the attrition probability (second line) increases proportionally to the seller’s size, while the the attraction probability decays, leading to slower seller growth. Eventually, conditional on survival, these differences in growth rates ensure that sellers converge to a stationary size (see Section 2.5).

**Stage 2. Prices** Finally, the equilibrium price is given implicitly by the PK constraint, which binds with equality. First, we replace $V^B(n, \omega_n; z, \varphi) = x'(n; z, \varphi)$ and $\omega = \{p; x'(n'; z', \varphi') : (n', z', \varphi') \in \{n - 1, n, n + 1\} \times \mathcal{Z} \times \Phi\}$ in equation (5). Then, solving for $p$ we obtain the equilibrium price level:
The optimal price level for a firm of type \((n, z)\) can be decomposed into the following additive terms. The first one 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. 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 again the parametrization under which \(x'\) decreases in \(n\). First, the firm offers a price reduction of \(\delta_f(U^B - x'(n)) \leq 0\) 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 eventuality that the firm grows. This compensation is thus labeled as the growth component.

Third, the firm adjusts the price for the event 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 separation 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 Section 3 we will discuss the different price effects that may be present in equilibrium and provide some intuition for the direction of the dependence on size.

\[ p_n(z, \varphi) = \underbrace{v(\varphi) - rx'(n; z, \varphi)}_{\geq 0} + \underbrace{\delta_f(U^B(\varphi) - x'(n; z, \varphi))}_{\leq 0} + \underbrace{\eta(\theta_{n+1}(z, \varphi))(x'(n+1; z, \varphi) - x'(n; z, \varphi))}_{\leq 0} + \underbrace{\sum_{z' \in Z} \lambda_z(z' | z)(x'(n; z', \varphi) - x'(n; z, \varphi))}_{\text{Separation component}} + \underbrace{\sum_{\varphi' \in \Phi} \lambda_{\varphi'}(\varphi' | \varphi)(x'(n; z, \varphi') - x'(n; z, \varphi))}_{\text{Aggregate-shock component}} \]

\[ \text{Baseline component} \quad \text{Exit component} \quad \text{Growth component} \quad \text{Separation component} \quad \text{Idiosyncratic-shock component} \quad \text{Aggregate-shock component} \]
2.5 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)) implies that the equilibrium tightness on each market adjusts to be consistent with agents’ beliefs. 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 derive the aggregate dynamics of the model (on and off steady state). Let \( S_{n,t}(z) \geq 0 \) be the total measure of firms of size \( n \) with idiosyncratic productivity \( z \in Z \) 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) buyers within market \( x'(n + 1; z, \varphi) \), market tightness must guarantee that:

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

at every \( t \geq 0 \) for all \( n \in \mathbb{N} \). Using that \( \eta(\theta) = \theta \mu(\theta) \), equation (16) can be written

\[
\mu(\theta_{n}(z, \varphi)) B^I_{n,t}(z, \varphi) = \eta(\theta_{n}(z, \varphi)) S_{n-1,t}(z),
\]

stating that the measure of inactive buyers who

\[\text{[Footnote]}\]

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-buyer firms). This is the approach followed, for example, by the bulk of 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.
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:

\[
\forall \varphi \in \Phi, \forall t \geq 0 : \sum_{n=1}^{+\infty} \sum_{z \in \mathcal{Z}} B^A_{n,t}(z) + \sum_{n=1}^{+\infty} \sum_{z \in \mathcal{Z}} B^I_{n,t}(z, \varphi) = 1 \tag{18}
\]

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 \(\varphi\) 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 \(\varphi\). Through equation (17), \(B^A_{n,t}(z)\) changes smoothly over time, while through equation (16) market tightness \(\theta_{n,t}(z, \varphi)\) 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 \(S_0(z, \varphi)\), jumps following a \(\varphi\)-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 \(\varphi\)-switch in such a way for the aggregate measure of inactive buyers to adjust only smoothly over time.

In short, while the 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 \(\mathbb{N} \times \mathcal{Z}\).

First, for \(n = 1\) we have:

---

\[\text{The derivation of the KF equations to follow can be found in Appendix D.2.}\]
\[ \partial_t S_{1,t}(z) = \pi_z(z)\eta(\theta_1(z, \varphi))S_{0,t}(\varphi) + 2\delta_c S_{2,t}(z) + \sum_{\tilde{z} \neq z} \lambda_z(z|\tilde{z})S_{1,t}(\tilde{z}) \]
\[ - \left( \delta_f + \delta_c + \eta(\theta_2(z, \varphi)) + \sum_{\tilde{z} \neq z} \lambda_z(\tilde{z}|z) \right)S_{1,t}(z) \]  

(19)

for any \( z \in \mathcal{Z} \), where \( \partial_t \) 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 \( \theta_1 \). Therefore, the dynamics of \( S_1 \) are smooth (i.e. not indexed by \( \varphi \)).

Similarly, for any \( n \geq 2 \):

\[ \partial_t S_{n,t}(z) = \eta(\theta_n(z, \varphi))S_{n-1,t}(z) + (n+1)\delta_c S_{n+1,t}(z) + \sum_{\tilde{z} \neq z} \lambda_z(z|\tilde{z})S_{n,t}(\tilde{z}) \]
\[ - \left( \delta_f + n\delta_c + \eta(\theta_{n+1}(z, \varphi)) + \sum_{\tilde{z} \neq z} \lambda_z(\tilde{z}|z) \right)S_{n,t}(z) \]  

(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:

\[ \partial_t S_{0,t}(\varphi) = \delta_f S_t + \delta_c \sum_{z \in \mathcal{Z}} S_{1,t}(z) - \sum_{z_0 \in \mathcal{Z}} \pi_z(z_0)\eta(\theta_{1,t}(z_0, \varphi))S_{0,t}(\varphi) \]  

(21)

where \( S_t := \sum_{n=1}^{+\infty} \sum_{z \in \mathcal{Z}} S_{n,t}(z) \) is the total measure of incumbent firms (i.e. firms with one or more customers). The usual interpretation 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 the model’s dynamics. We can now specialize these equations to obtain the time-invariant distribution of firms by equating flows in and out of every possible state: \( \partial_t S_{n,t}(z) = 0, \forall (n, z) \in \mathbb{N} \times \mathcal{Z}. \)32 The following result

---

32 In general, an analytical solution for the stationary distribution does not exist. One exception is for the economy without shocks and with \( \delta_f = 0 \). In this case, the flow equations amount to a continuous-time
ensures that the dynamics are convergent:

**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 $\varphi \in \Phi$.

*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.

### 2.6 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 $V^S(\cdot, \varphi) : \mathbb{N} \times \mathcal{X} \times \mathbb{Z} \to \mathbb{R}$ and $V^B(\cdot, \varphi) : \mathbb{N} \times \mathcal{O} \times \mathbb{Z} \to \mathbb{R}$; a value of inactivity $U^B(\varphi) \in \mathbb{R}$; a joint surplus $W_n(z, \varphi)$, price $p_n(z, \varphi)$, and continuation promises $x'(n' + 1; z, \varphi), x'(n - 1; z, \varphi), \{x'(n; z', \varphi) : z' \in \mathbb{Z}\}, \{x'(n; z, \varphi') : \varphi' \in \Phi\}$ for firms of type $(n, z) \in \mathbb{N} \times \mathbb{Z}$; a decision rule $\hat{x}(\varphi)$ for inactive buyers; a market tightness function $\theta(\cdot, \varphi) : \mathcal{X} \to \mathbb{R}^+$; aggregate measures of agents: $\{S_0(\varphi), S(\varphi), B^A(\varphi), B^B(\varphi)\}$; and a distribution of sellers and buyers: $\{S_n(z), B^A_n(z), B^B_n(z, \varphi) : (n, z) \in \mathbb{N} \times \mathbb{Z}\}$; such that: [i] the value functions solve (5) and (6), $U^B(\varphi)$ satisfies the free-entry condition (13), and the joint surplus $W_n(z, \varphi)$ solves (10); [ii] price and continuation promises satisfy (15) and (12), respectively; [iii] $\hat{x}(\varphi)$ solves the inactive buyer’s problem, (1)-(2); [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 2.5.

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. This insight is instructive for the numerical implementation, which exploits the nested fixed-point nature of the problem to solve for equation (10) via value function iteration on $W$ and a bisection step on $U^B$ (see Appendix C.1 for the details). 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 size of the exogenous state so that the free entry condition, and therefore ex-ante revenue equalization across markets, can be met in all states of nature.

---

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 case.

---

A preliminary proof of this result is available upon request.
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 2.5. In our environment, the planner chooses distributions of buyers and sellers, as well as market tightness levels, in order to solve:

\[
\max E_0 \left\{ \int_0^{+\infty} e^{-rt} W_t(\varphi_t) dt \right\} \tag{22}
\]

where

\[
W_t(\varphi_t) \equiv -\kappa(\varphi_t) S_{0,t} + \sum_{n_t=1}^{+\infty} \sum_{z_t \in Z} \left( v(\varphi_t) B_{n_t,t}^A(z_t) - C(n_t; z_t, \varphi_t) S_{n_t,t}(z_t) - c(\varphi_t) B_{n_t,t}^I(z_t) \right)
\]

Aggregate welfare in this economy 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, we then establish:

**Proposition 3 (Efficiency)** A Recursive Equilibrium is efficient.

*Proof.* See Appendix B.3.

This result implies that, given the search frictions, 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 out other interpretations, our setting demonstrates that this type of dispersion may also be generated through efficient pricing.

## 3 Understanding the Mechanism

In the previous section, we have described a model in line with the original intuition behind Rotemberg and Woodford (1991): sellers use prices as a way to invest into larger market shares in the future. 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 key assumptions and describe the role that they play in the model.
3.1 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 $(z, \varphi) \in Z \times \Phi$:

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

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

(23)

where $\Gamma^B(\varphi)$ is given by (4), and $\{\Gamma^S_n(z, \varphi)\}_{n=1}^{+\infty}$ is given in equation (B.4.3) of Appendix B.4.

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

**Proof.** See 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. The result has an intuitive interpretation. First, in the Appendix we show that:

$$\Gamma^B = \mu(\theta_{n+1}) \left( x_{n+1} - U^B \right)$$

**Ex-post**

buyer net gains from *new* match

and

$$\Gamma^S_n = \eta(\theta_{n+1}) \left( W_{n+1} - W_n - x_{n+1} \right)$$

**Ex-post**

seller net gains from *new* match

For each equation, the right hand-side of the equality gives the product of the *ex-post* net gain from a *new* match times the probability that a match occurs (from the perspective of buyer and seller, respectively). Thus, $\Gamma^B$ and $\Gamma^S_n$ can be interpreted as the *ex-ante* net gains from matching for each agent, respectively. The former is constant in $n$ because of ex-ante buyer indifference (equation (3)). The latter depends on seller size. In particular, what the seller extracts from a new match ex-post is the total gain in joint surplus $(W_{n+1} - W_n) > 0$, net of the value $x_{n+1}$ that was promised to the new consumer in case of a successful meeting. Thus, Proposition 4 says that, with a Cobb-Douglas matching function, the equilibrium

---

34 For the remainder of this section, we suppress state dependence to alleviate notation. This does not affect the intuition in any significant way.
marginal net gain in joint surplus for each additional customer is a convex combination of the \textit{ex-ante} net match gains that accrue to the new customer and her seller.

The matching elasticity parameter $\gamma$ governs how the surplus is shared \textit{ex-post} between the seller and the new customer. In particular, the new customer’s \textit{ex-post} net gains are given by:

$$x_{n+1} - U^B = \gamma \left( W_{n+1} - W_n - U^B \right)$$

showing that a fraction $\gamma$ of the total gains in joint surplus are absorbed by the new incoming buyer. Interestingly, however, the \textit{ex-post} gains for the seller must incorporate rents shared with the new buyer \textit{as well as} all those shared with the pre-existing buyers. To see this, we can use $W_n = V^S_n + nx_n$ by definition to obtain that:

$$V^S_{n+1} - V^S_n = \left( W_{n+1} - W_n - x_{n+1} \right) + n(x_n - x_{n+1})$$

By Proposition 4, we have:

$$W_{n+1} - W_n - x_{n+1} = (1 - \gamma) \left( W_{n+1} - W_n - U^B \right)$$

In words, after a $n \to (n+1)$ transition, the seller absorbs a share $(1 - \gamma)$ of the total net gain in joint surplus from the \textit{new} customer (term [A]). In addition, some surplus is transferred between the seller and the $n$ pre-existing customers (term [B]), as each pre-existing buyer sees their valuation change from $x_n$ to $x_{n+1} \neq x_n$ as a result of the size increase.

For example, when buyer valuation is decreasing in $n$ (i.e. $x_n > x_{n+1}$), term [B] is a positive transfer from all $n$ pre-existing customers to their seller. Figure 1 shows, in this case, how net gains for buyers and sellers change with size (from both \textit{ex-ante} and \textit{ex-post} perspectives). We see in the right-most panel that, as the seller grows, he increasingly prefers to
extract rents from the current customer base (term [B] increases), and is less concerned with extracting surplus from the new consumer (term [A] decreases).

But why does the seller’s size affect the incentives to build a customer base in the first place? To explain this, we proceed in two steps: first, we show how, given 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 the price level for a size-\( n \) firm (equation (15)) around the \((x_n, x_{n+1})\) equilibrium promises. Respectively, this yields:

\[
\frac{\partial p_n}{\partial x} \bigg|_{x=x_n} = -\left(r + \delta_f + \eta(\theta_{n+1}) + n\delta_c \right) < 0 \tag{25a}
\]

\[
\frac{\partial p_n}{\partial x} \bigg|_{x=x_{n+1}} = \eta(\theta_{n+1}) + \eta(\theta_{n+1}) \frac{\partial \eta(\theta_{n+1})}{\partial x} \bigg|_{x=x_{n+1}} \frac{x_{n+1} - x_n}{\eta(\theta_{n+1})} \leq 0 \tag{25b}
\]

Equation (25a) shows the relationship between the utility that is currently being paid to each incumbent buyer (a pre-determined state for the seller) and the price that each buyer is charged (a choice variable). This relationship is always negative because the PK constraint binds in equilibrium: higher prices are detrimental to each customer’s valuation.

More interesting is the effect of a seller’s promise to a potentially new customer (a choice variable for the seller) on the price that the seller charges to its currently incumbent customers (equation (25b)). The marginal effect has two additive terms of opposite sign. The first part (term [C]) says that each additional util that the seller offers to the new potential customer must come from revenue raised through an increase in the price that is currently being charged to each one of the pre-existing customers of the seller. In expectation, the additional util is worth \( \eta(\theta_{n+1}) \), as this is the effective probability with which a new customer will actually join the group. The second term (the product of [C] and [D]) shows that a further adjustment is necessary. Offering an additional util to the new potential customer raises the probability with which a new customer is successfully attracted. This change in the likelihood of matching is term [D], which has a positive sign by equation (3). Importantly, as we have argued above (term [B]), when a new customer arrives there is change in the value of all incumbent customers, equal to \((x_{n+1} - x_n) \neq 0\), which is a transfer between buyers and seller. By commitment, the price must be adjusted accordingly. For instance, when customer utility decreases as the firm grows \((x_{n+1} - x_n < 0)\), the price must be adjusted downward (as the product of [C] and [D] is negative). Intuitively, the buyers receive a price compensation through this channel. In this case, the overall effect of the new promise \( x_{n+1} \) on the
current price $p_n$ is ambiguous, reflecting that prices may increase or decrease, depending on parameters, when the sellers raise their future promises. In contrast, when incumbent customers benefit from firm growth ($x_{n+1} - x_n > 0$), raising the promise must necessarily be accompanied by a price increase.

**Intuition for the size dependence** It remains to argue what the dependence between $x_n$ and size $n$ is in equilibrium. To make the argument transparent, consider that marginal costs are constant in $n$ (i.e. $C(n) \propto n$), so that all size dependence emerges only from the product-market frictions.

Figure 2: Numerical Example: Promised utility, price, market tightness, and firm growth, as a function of size, for the simple model with no exogenous $(z, \varphi)$ shocks, and a constant marginal cost (i.e. $C(n) \propto n$). Firm growth has been decomposed between the rate of customer attraction (solid line) and that of customer attrition (dashed line).

Figure 2 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 (i) satisfy $x_1 > U^B$, or else the customer would prefer to remain unmatched, and (ii) ensure that the expected value at entry pays for the entry cost, or $\kappa = \eta(\theta(x_1))V^S_1$.

After the seller has acquired its first customer, $\kappa$ becomes sunk and the seller posts 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 $W_1 = V^S_1 + x_1$ to $W_2 = V^S_2 + 2x_2$. Since the number of customers has grown proportionally, but customers must still absorb part of the added value to agree to remain matched (Proposition 4), 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. To implement the strictly decreasing sequence of promise, the seller chooses an increasing price
path (by equation (25a)).

How does this argument depend on the degree of frictions? An important parameter for
the size dependence is the size of entry costs $\kappa$. Figure 3 uses the same parameter values as
Figure 2, but with a much lower value for $\kappa$. The path of promised utilities and prices is now
different. First, the seller must still enter with a high promise for 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 constraint to raise front-load resources to make up for the costs of
market penetration. Instead, promises can be back-loaded, and prices may start decreasing
with size.

![Graph showing promised utilities, prices, tightness, and firm growth against size](image)

**Figure 3:** Same as Figure 2, but with a lower value for $\kappa$.

Therefore, importantly, the positive correlation between prices and sizes is not built into
the model, and the model can be used to fit different price-size dependences. 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 decreasing for export markets. In contrast, Fitzgerald et al. (2017) find
no dynamics of prices after adjusting for selection. Through the lens of our model, these
observations could be rationalized by different entry costs.

### 3.2 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.
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 dynamic rent-extraction trade-off between incoming customers and the current base, the firm would now have to 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 5.1, and discuss the technical challenges it presents.

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 4.1), 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 collapse to this form of pricing when we allow for discrimination. Allowing for discrimination is our environment not only preserves the block-recursivity property, which is key for providing tractability, but it also preserves efficient firm dynamics and price dispersion. Importantly, however, assuming discriminatory contracting results into a new feature: equilibrium multiplicity in the form of price indeterminacy. We explain these results in detail in Section 5.2.

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 will not price gouge ex-post, and they remain loyal to their seller in order to avoid going through the costly inactive state. As discussed briefly in Section 2.2, 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, 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 under appropriately designed “implicit contracts”, paired with trigger strategies.
on the buyer side (see Nakamura and Steinsson (2011) for a discussion in customer market settings). The implicit contracts literature (going back to Baily (1974) and Azariadis (1975)) shows that these type of contracts typically exhibit history-dependence, which in our framework would break the recursive structure. 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 no longer hold. For these reasons, seller’s commitment is a key aspect for our set-up.

4 Quantitative Analysis

Let us now turn to the quantitative part of the paper. First, we present the data that will be used to calibrate our model. Then, we proceed to the computational implementation of the equilibrium, present the estimation 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.

4.1 Data

Estimating our model requires the use of disaggregated data which 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. We use micro-pricing data on the U.S. retail sector. 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. In this interpretation of the theory, sellers are stores, and buyers are private consumers. We will use data of unique 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 the observed store transactions. Finally, under this framing of the model, we interpret the buyer’s search cost as a proxy for the transport, information and/or utility costs associated to finding and/or switching away from trusted suppliers.

We use weekly micro-level data from the IRI Symphony scanner data set.35 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 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.\textsuperscript{36} 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.\textsuperscript{37}

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 within each \( UPC \times \text{market} \times \text{week} \) 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.\textsuperscript{38} 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.\textsuperscript{39} Table A.1 in Appendix A shows some descriptive statistics of our data before and after applying these restrictions.

To capture 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

\textsuperscript{36} 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.

\textsuperscript{37} 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.

\textsuperscript{38} 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.

\textsuperscript{39} A “promotion” is defined by the IRI as a temporary price reduction of 5% or greater. Sales are quite unresponsive to the business cycle, as documented by Coibion et al. (2015) for the IRI data, and therefore excluding should not change our life-cycle results significantly.
relative measure of prices. That is, we define:

\[ \hat{p}_{usm,t} = \log P_{usm,t} - \frac{1}{N_{usm,t}^S} \sum_{s=1}^{N_{usm,t}^S} \log P_{usm,t} \] (26)

where \( N_{usm,t}^S \) is the number of stores selling good \( u \) in market \( m \) and week \( t \). In words, \( \hat{p}_{usm,t} \) 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 \( \hat{p}_{usm,t} \) across products, 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.

4.2 Estimation

Parametrization

Let us proceed with the estimation of the model using these micro data. 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:

\[ C(n; z, \varphi) = \tilde{c}(z, \varphi) \cdot n^\psi \] (27)

where \( \tilde{c}(z, \varphi) > 0 \) is a size-invariant 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 seller size for each state, given by

\[ n^*(z, \varphi) \equiv \left( \frac{\psi \tilde{c}(z, \varphi)}{v} \right)^{\frac{1}{1-\psi}} \]

beyond which the static flow surplus \( \pi_n(z, \varphi) = nv - C(n; z, \varphi) \) is strictly decreasing and the seller does not want to grow further.\(^{40}\) The scale parameter changes across sellers and aggregate states, with \( \tilde{c}(z_i, \varphi_j) = w e^{z_i+\varphi_j} \), 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 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 \( \varphi \). 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

\(^{40}\) Even though, as we saw in Section 3, the existence of a stationary size does not hinge on the curvature in the cost function, the parameter \( \psi \) will help us pin down the size dependence in prices more easily.
s ∈ \{z, ϕ\}, this would require the estimation of \(k_s(k_s-1)\) additional parameters, a potentially very 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 in logs. These, in turn, are approximated on finite grids using the Tauchen (1986) method.\(^{41}\) 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 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: \((v, r, δ_f, δ_c, w, ψ, κ, γ, c)\), corresponding to the value of consumption, the time discount rate, the separation rates of sellers 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 economy in its long-run equilibrium. These shocks will be re-introduced in Section 4.3, when we study the response of markups to supply and demand shocks.

**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. The exogenous separation probability is set to match a 0.044% weekly customer turnover rate (corresponding to \(δ_c ≈ 0.2041\) 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. (2017), who estimated it using the same database that we use here. 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 seller exit rate \(δ_f\); (ii) the cost scale \(w\); (iii) the seller 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

\(^{41}\) Full details can be found in Section C.2 of the numerical appendix.
model, identifying each parameter separately is hard in our environment, though we can provide some intuition for how each one is informative about specific moments. Methodologically, we estimate the parameters jointly by matching a combination of aggregate and seller-level 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 as well as 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 sellers over $T = 100$ years of data which we discretize with time steps of equidistant length $\Delta = 0.01$ each. All sellers are drawn from the stationary distribution $\{S_n(z) : (n, z) \in \mathbb{N} \times \mathcal{Z}\}$ at time $t = 0$ and evolve endogenously through simulated Markov chains that replicate the state dynamics described in Section 2.5. To allow for convergence in the distribution, we drop the first half of the time sample when computing the average simulated moments. For the productivity distribution $\pi_z$, 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) store-level moments related to the 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.3 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 sellers who either die or lose their last remaining customer. By equation (21), this means:

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

(28)

These formulas hold in and out of steady state, but are equal to each other in the absence of $\varphi$-shocks, so the entry rate in the data helps us identify the exogenous exit rate $\delta_f$ 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, in the literature estimates vary substantially depending on the empirical methodology used, 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.\footnote{Using a model-free approach and manufacturing data at the plant level from Slovenia, \textit{DeLoecker and Warzynski} (2012) find that median markups range from 10% to 28%, depending on the specification, while \textit{DeLoecker et al.} (2016) find even more variation, from 15% up to 43%, using similar methods for India. In a study spanning 1981-2004, \textit{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.} We
choose to target a markup of 39%, a number that we impute from the average ratio over our sample period (2001-2007) of gross margins to sales in the retail sector. We obtain this number from the latest Annual Retail Trade Report of the U.S. Census Bureau. 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:

\[
\bar{m} = \sum_{n \in N} \sum_{z \in \mathcal{Z}} m_n(z); \quad \text{with } m_n(z) := s_n(z) \frac{p_n(z)}{mc_n(z)}
\]  

(29)

where \(s_n(z) = \sum_{n, z} np_n(z)\) is the sales share of type \((n, z)\) firms, and \(mc_n(z) := C(n; z) - C(n - 1; z)\) is the marginal cost of this type of firms. Though many parameters affect the average markup, \(\gamma\) 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 3).

At the store level, we target several moments from the distribution of prices and sales that we obtain from our IRI sample. The cost parameters \((\psi, w)\) 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 (26)), our 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 sellers 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 implications of pricing when sellers grow and shrink through their pricing decisions.

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 \(\rho_z\)), and the dispersion in the distribution of normalized sales (pinning down the volatility \(\sigma_z\)). Finally, we need to calibrate the search cost for buyers, \(c\), and the market penetration cost for sellers, \(\kappa\). As we discussed in Section 3, these parameters are important to pin down the dependence between seller size and seller price, which determines two key aspects of firm dynamics: (i) the growth rate of sellers across sizes; and (ii) the stationary size. For the former, we target the correlation between store product-level sales growth rates and the relative price of those products. Regarding (ii), we target the stationary size of sellers in the data to make average size comparable between

\[43\] 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%.
data and model. In the model, we measure the average size of firms as the mean number of units sold per firm. Since each customer consumes only one unit, the average size is (see e.g. Luttmer (2006)):

$$L = \left( \sum_{n \in \mathbb{N}} \sum_{z \in \mathbb{Z}} \frac{1}{n} L_n(z) \right)^{-1}; \quad \text{with } L_n(z) := \frac{nS_n(z)}{\sum_{n,z} nS_n(z)} \quad (30)$$

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, so we target this number in the estimation.

**Estimation Results**

The full set of calibrated parameter values is presented in Table 1, and the result of the calibration exercise in terms of moment-matching is presented in Table 2. 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 that cannot be captured by the model.

Figure A.5 in Appendix A plots the joint surplus, the pricing policy function, the measured markups (equation (29)) and the promised utility, in the space of seller sizes $(n)$ and productivities $(z)$, for the calibrated set of parameters. 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 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.7 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 size distribution with a fat right-tail in both seller revenues and 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.

---

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

45 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.
<table>
<thead>
<tr>
<th>Param.</th>
<th>Value</th>
<th>Description</th>
<th>Source / Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v$</td>
<td>1</td>
<td>Value of consumption</td>
<td>Normalization</td>
</tr>
<tr>
<td>$r$</td>
<td>0.05</td>
<td>Discount rate</td>
<td>5% annual risk-free rate</td>
</tr>
<tr>
<td>$\delta_c$</td>
<td>0.2041</td>
<td>Separation rate</td>
<td>Paciello et al. (2017)</td>
</tr>
</tbody>
</table>

**Estimated internally**

<table>
<thead>
<tr>
<th>Param.</th>
<th>Value</th>
<th>Description</th>
<th>Source / Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_f$</td>
<td>0.0738</td>
<td>Firm exit rate</td>
<td>Annual store entry rate</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.5339</td>
<td>Matching elasticity</td>
<td>Average markup</td>
</tr>
<tr>
<td>$\psi$</td>
<td>1.4044</td>
<td>Cost curvature</td>
<td>Standard deviation of relative prices</td>
</tr>
<tr>
<td>$w$</td>
<td>0.1510</td>
<td>Cost scale</td>
<td>p50-p10 inter-decile range in relative prices</td>
</tr>
<tr>
<td>$c$</td>
<td>0.5457</td>
<td>Buyer search cost</td>
<td>Average store size</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1.6214</td>
<td>Firm entry cost</td>
<td>Sales growth and relative price correlation</td>
</tr>
<tr>
<td>$\rho_z$</td>
<td>0.0751</td>
<td>Persistence of $z$</td>
<td>Autocorrelation in normalized store sales</td>
</tr>
<tr>
<td>$\sigma_z$</td>
<td>0.1034</td>
<td>Volatility of $z$</td>
<td>Dispersion in normalized store sales</td>
</tr>
</tbody>
</table>

**Frequency** Annual

**Table 1:** Full set of calibrated parameters in the baseline estimation. Notes: The parameters $(\rho_z, \sigma_z)$ correspond to the Euler-Maruyama equation (C.2.1) of the Ornstein-Uhlenbeck process for $z$. See Appendix C.2 for details.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Model</th>
<th>Data</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Aggregate moments</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual entry rate</td>
<td>0.087</td>
<td>0.089</td>
<td>IRI (Table A.3)</td>
</tr>
<tr>
<td>Average markup (2001-07)</td>
<td>1.388</td>
<td>1.383</td>
<td>U.S. Census</td>
</tr>
<tr>
<td><strong>B. Store-level moments</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sd(Relative prices)</td>
<td>0.1072</td>
<td>0.1055</td>
<td>IRI (Table A.1)</td>
</tr>
<tr>
<td>p50-p10 IDR relative prices</td>
<td>1.1224</td>
<td>1.1215</td>
<td>IRI (Table A.1)</td>
</tr>
<tr>
<td>Average store size</td>
<td>10.73</td>
<td>12.44</td>
<td>IRI</td>
</tr>
<tr>
<td>corr(Sales growth, Relative price)</td>
<td>-0.023</td>
<td>-0.007</td>
<td>IRI</td>
</tr>
<tr>
<td>ac(Normalizes sales)</td>
<td>0.854</td>
<td>0.828</td>
<td>IRI</td>
</tr>
<tr>
<td>sd(Normalized sales)</td>
<td>0.6</td>
<td>0.474</td>
<td>IRI</td>
</tr>
</tbody>
</table>

**Table 2:** 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.4 in Appendix A. 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 or other forms of rigidity), 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 formulas 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.\textsuperscript{46} 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.4 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.\textsuperscript{47} 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. Figure A.6 in Appendix A shows how these measures of price changes vary across firm size, with smaller firms experiencing more frequent and larger price changes.

\textsuperscript{46} We focus only on \textit{regular} price changes, which we define (following Coibion \textit{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.

\textsuperscript{47} 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.
4.3 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 plays on macroeconomic transmission. In particular, we seek to identify how sellers’ incentives to accumulate customers can generate amplification and persistence through both level and distributional effects, as well as the degree of pass-through on prices and the cyclicality of markups.\(^{48}\)

Supply Shocks

Starting from the stationary equilibrium of the calibrated economy, we first study the response to a negative and transitory 1\% supply shock to the marginal cost, i.e. a 1\% increase in \(C(n; z, \varphi)\) parametrized as in equation (27). We assume that the aggregate state \(\varphi\) follows a mean-reverting process in logs (details in Appendix C.2). The shock hits at some time \(t_0\), and the process continues without any further shocks for all \(t > t_0\).\(^{49}\)

Figure 4 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)). Thus, active buyers are hit harder than sellers by the shock. As a result of the lower promises, 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, while prices increase in response to the shock, 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 higher price). Yet, in order to ensure their recovery, they choose an increasing path of promised utilities in the transition. As the shock is smoothed out inter-temporally via these two contracting instruments, the price response is muted. In the calibrated economy, in particular, the price response is above 12\% of the size of the shock. 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). More importantly, since prices react less than one for one, the average measured markup (panel (l), solid line) is procyclical.

\(^{48}\) Henceforth, our measure of cyclicality is the comovement between the response of the variable in question and that of the measure of active buyers. The latter equals total output in the economy because each buyer consumes one unit.

\(^{49}\) 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.
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).

Notes: All responses expressed in %-deviations from steady-state. The shock hits at date $t_0 = 0$. Paths are smoothed out with 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 (28). Panels (i) and (j) are computed using equation (28). Panel (l) is computed using equation (29).
When analyzing the markup response by seller size, we find that smaller sellers (dashed line) respond stronger on impact, while the largest sellers’ response (dotted line) is weaker than the average. 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. In the model, this occurs because smaller sellers experience more frequent and larger price changes per unit of time (Figure A.6 in Appendix A), since the optimal pricing policy is concave in seller size.

To explain the behavior of measured markups in the transition and in the cross section, we must first understand the distributional consequences 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)). These trends continue for a few periods, that is until eventually reversed by the continued increase in promised utilities, and thus seller growth rates. In the first phase of the transition, therefore, the size distribution is gradually shifting to the left. 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. Therefore, the persistence of the shock on markups is higher for smaller sellers, as reflected in the fact that markups take longer to mean-revert for these type of sellers relative to larger ones.

To further illustrate that sellers smooth out the effects of the adverse shock inter-temporally by raising prices imperfectly and depressing promised utilities, Figure 5 shows how the responses discussed above vary with the average duration of customer relationships, as measured by $\delta_c$. In particular, we compare the baseline economy (solid), with an economy in which the duration of customer relationships is one-third shorter (dashed) and the remaining parameters remain fixed at their original calibrated values. 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 $\delta_c$ 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 needs to decrease less on impact (panel (a)) as these captive buyers are not expected to remain matched for long anyway. Accordingly, the seller can charge higher prices on impact (panel (c)), thereby alleviating the adverse effects of the shock on their

---

50 Here is how to understand the paths for entry and exit rates. Since the risk of exiting has become 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)). In turn, the reason why the entry rate spikes up is that tightness in the entry market (i.e. for $n = 1$) has actually increased, unlike the tightness in all other markets, making the ex-post rate of acquiring the first customer higher. Finally, the reason why $\theta_1$ behaves differently than in other markets is that, in order to enter into the economy, potential sellers must raise enough resources to pay for the fixed market penetration cost, $\kappa$. While these costs have remained 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.
own value (panel (b)). Moreover, the effects of the shock on prices and continuation utilities become naturally less persistent. Finally, the fact that the price pass-through becomes less incomplete as $\delta_c$ increases means that the absolute response of the average markup (panel (d)) is weaker. In the limit as $\delta_c$ 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.

![Figure 5: 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 $\delta_c$. Notes: See Figure 4.]

**Demand Shocks**

Recent research has emphasized the relevance that consumer shopping behavior may have on macroeconomic dynamics. For instance, Bai et al. (2017) 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 to aggregate shocks, and Paciello et al. (2017) 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 $v$. 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 $\varphi$ process implying a half life of about three years, following Paciello et al. (2017) and in line with estimates by Bai et al. (2017).\footnote{Paciello et al. (2017) use a quarterly autocorrelation of 0.98 for their demand shock. Using the Euler-Maruyama method described in 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 $\rho_\varphi = \frac{1-0.98^{114}}{\Delta} \approx 0.0808$.}
Figure 6: 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. See Figure 4.
Figure 6 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 to them. 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 this initial drop in the matching rate leads sellers to progressively shrink in size (panel (h)), and more buyers to remain inactive (panel (g)).

In this respect, the demand shock is akin to a productivity shock (Figure 4), in line with the intuition in Bai et al. (2017). In particular, the compositional effects are similar as in our previous analysis, with a left-ward shift in the firm size distribution accounting for the increase in the exit rate and in the increase in the relative contribution of low-markup firms to the average markup response. 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’ 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.

Once again, sellers understand the temporary nature of the shock ex-ante, so in the transition they increase their promises, and seller growth picks up until sellers eventually go back to positive growth on average. That is, as before, the size distribution shifts left in a first phase of the transition, and returns back to its original position in the long-run. Thus, the level effects of the demand shock 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 seller, 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. At the other end of the distribution, in contrast, the shock decreases the relative mass of large sellers, which mutes the markup response for these type of sellers. Thus, the cyclical 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.

In sum, while prices are procyclical after demand shocks and countercyclical after supply shocks, markups are always procyclical to both supply and demand shocks. The latter is a feature that standard models of price rigidities, such as the New Keynesian model, have some trouble generating, particularly for aggregate demand fluctuations. Yet, recent empir-
ical studies tend to rule against countercyclical markup variation conditional on either type of shock (e.g. Nekarda and Ramey (2013)). We have proposed a micro-founded mechanism through which this empirical observation can be rationalized.

The Margins of Price Adjustment

To conclude this section, we seek to further understand the main margins along which prices adjust in response to aggregate shocks. For this, we decompose the price response into the response of the different price components 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 7](image)

**Figure 7:** Impulse responses to negative and temporary 1% supply and demand shocks (same as Figures 4-6). 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 7 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 procyclical 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 a nearly one-to-one 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.

We also observe that, in both cases, the “growth” component is procyclical, while the “exit” and “separation” components are countercyclical. 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.

### 4.4 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 and experienced a 40% increase in the 1980-2014 period (see Figure 8, right). In levels, the markup has increased from about 20% in 1980 to about 65% in 2014. 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.\(^{52}\) As in our model, firm size and markups in their data are positively correlated within narrowly defined industries.

![Entry Rate](https://example.com/entry_rate.png)

![Average Markup](https://example.com/average_markup.png)

**Figure 8:** Left: Entry rates across all sectors for the U.S. (1980 – 2014), in changes relative to 1980. Entry rate 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](http://www.census.gov/ces/dataproducts/bds/data.html). Right: Average markup in the firm cross-section across all industries for the U.S. (1980 – 2014), in changes relative to 1980. 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 these data with us.

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

---

\(^{52}\) 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 \(Q\).
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 2014, a 40% decrease (see Figure 8, left).

We interpret this evidence through the lens of our model by implementing a decrease in buyer search costs. This follows the idea that, since the early 1980s, new shopping technologies (e.g. the Internet) have emerged which have made searching for product-supplier pairs cheaper for buyers. Empirically, it has been argued that, in certain product markets, the emergence of these technologies has not had the price-lowering effect that standard theories of price dispersion would predict, for example through an increase in competition (see e.g. Ellison and Ellison (2018) and references therein). In this section, we show that our model can shed some light upon this discussion.

Figure 9 plots the entry rate, average markup, and welfare (equation (22)) across different stationary solutions for different values of \( c \) (with the remaining parameters fixed at their calibrated values). Note that we express these variables in changes relative to the highest possible value of \( c \) for which a solution exists. We see that the model can overall generate nearly up to a 10% change in the entry rate and a 14% change in the average markup, i.e. about one-fourth of the actual long-run changes in the data. Interestingly, a decline in search cost also raises welfare substantially in spite of the increase in markups. This is because, as we shall now discuss, not only do resources increase through lower costs directly, but also aggregate consumption increases as a higher share of buyers become matched.

![Figure 9: Entry rate, average markup, and welfare, across steady states for different values of the buyer search cost, with all values normalized to unity for the highest possible \( c \). The vertical dashed line marks the calibrated value for \( c \).](image)

To understand these effects, Figure 10 shows the effects of changes in \( c \) on the equilibrium policies for promised utility and prices, as well as the effects on the stationary size distribution. First, when search costs decrease, a matched buyer now faces a better outside option \( (U^B \text{ goes up}) \). Thus, in order to keep her customers from voluntarily separating (for which they would now incur a lower cost \( c \)), the seller needs to increase her continuation promises.
ex-post ($x_n$ goes up for all $n$). To see how sellers implement these changes through prices, we must return to our partial analysis in Section 3, equation (25b). There, we argued that whenever prices are strictly increasing in seller size (as is the case in the estimated model), raising the promise on the new buyer leads to both upward (term $[C]>0$) and downward (term $[E]<0$) pressure on the price charged to the current buyers. For the calibrated set of parameters, the former effect dominates: though buyers are ex-post being offered a higher value, they now consume at a higher price ($p_n$ increases). Yet, their outside option has improved sufficiently for them to be willing to remain matched with their seller in spite of the price increase. Intuitively, the seller is now squeezing a higher share of the surplus out of their current buyers. This is reflected on the left-most panel of Figure 10, showing that the ex-post gain from matching for new buyers net of their outside option, $(x_n - U^B)$, is actually lower for lower search costs.

![Figure 10: Policy functions for buyers' net ex-post gains from matching and prices as a function of seller size, for a low $c$ (solid line) and a high $c$ (dashed line), and comparison of the stationary size distribution between both cases.](image)

We have thus argued that prices are higher for a lower $c$. Since sellers are now making better promises, they find new matches faster and experience higher rates of growth. Hence, the size distribution shifts to the right (third panel in Figure 10). All in all, both because prices are higher and because the average firm of the economy is now relatively larger and more expensive, the average markup of the economy increases when search costs decline.

The effect of $c$ on the entry rate is a priori also ambiguous. By formula (28), the entry rate is the product of the effective rate of attracting the first customer $\eta(\theta(x_1))$, and the measure of inactive sellers relative to incumbents $\frac{S_0}{\sum_{n \geq 1} s_n}$. Along the intensive margin, when the search cost of buyers decreases, the seller-specific probability of finding the first customer increases for the outside seller (as $x_1$ goes up). Along the extensive margin, however, when incumbent sellers grow faster and more matches are made, the relative measure of inactive sellers declines, as the total measure of buyers is fixed exogenously. In other words, as $c$ decreases, there is less scope for entering because less buyers are unsuccessful in finding active sellers, and the extensive margin of demand in the entrants’ market diminishes. In
the calibrated set of parameters, this extensive-margin effect is stronger, and the entry rate goes down.

5 Extensions

In this section, we relax some central assumptions of the baseline model of Section 2. First, we outline how to endogenize customer separations in Section 5.1. Then, in Section 5.2 we show how to relax the assumption of no price discrimination across customers, and show that the model preserves its main structure though the predictions may change substantially.

5.1 Endogenous Customer Separations

To introduce customer seller-to-seller transitions, we can model customer search explicitly. 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 other.

Introducing this additional dimension into our full model with aggregate shocks 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. This would change the environment substantially, so we leave it for future work.

To show how to incorporate endogenous customer transitions within the model, we thus must turn off the aggregate shocks. The search decision of a customer currently obtaining value $V^B$ is then:

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

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 $\hat{x}(n, \omega; z)$ for a customer matched with firm of type $(n, z)$ under contract $\omega$, the first-order condition reads:

$$\left( \hat{x}(n, \omega; z) - V^B(n, \omega; z) \right) \frac{\partial \mu(\theta(x))}{\partial x} \bigg|_{x = \hat{x}(n, \omega; z)} = -\mu(\theta(\hat{x}(n, \omega; z)))$$ (31)

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 $\hat{x}(n, \omega; z)$ is increasing in $V^B(n, \omega; z)$. 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 sellers) 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 sellers. The buyers’ and seller’s HJB equations are then identical to (5) and (6), respectively, except that we now must replace $\delta_c$ by an “effective” customer separation rate, given by:

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

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:

$$\theta_n(z) = \frac{1}{S_{n-1}(z)} \left( B_{n1}^n(z) + B_{n1}^A(z) \right)$$

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

### 5.2 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.

The first observation to be noted is that, if sellers 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. However, assuming price discrimination in a model with dynamic long-term contracts and commitment does not yield this result. This is because sellers must still trade off static payoffs coming from the current price with dynamic ones coming from the promised utilities. 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 by a set of promised utility sequences solving a joint surplus problem. Allowing for price discrimination comes at 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 \( \omega_i = \{ p_i, x_i'(n'; s') \} \) be the contract offered to the typical customer \( i = 1, \ldots, n \), which is composed of an individual-specific price level \( p_i \), and a personalized menu of continuation utilities \( x_i'(n'; s') \), one for each \( n' \in \{ n - 1, n, n + 1 \} \) and \( s' \in \{(z', \varphi), (z, \varphi')\} \). A seller is characterized by the collection \( \{ x_i \}_{i=1}^n \) of outstanding promises, and must choose: (i) a menu of contracts \( \{ \omega_i \}_{i=1}^n \) 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:

\[
\begin{align*}
rv^S(n, \{ x_i \}_{i=1}^n; z, \varphi) &= \max_{x_0', (\omega_i)_{i=1}^n} \left\{ \sum_{i=1}^n p_i - C(n; z, \varphi) + \delta f \left( V_0^S(\varphi) - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \right. \\
&+ \delta \sum_{j=1}^n \left( V^S(n-1, \{ x_j(n-1; z, \varphi) \}_{j=1}^n \cup \{ x_j(n-1; z, \varphi) \}; z, \varphi) - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \\
&+ \eta(\theta(x_0'; \varphi)) \left( V^S(n+1, \{ x_i(n+1; z, \varphi) \}_{i=1}^n \cup \{ x_0'; z, \varphi \}; z, \varphi) - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \\
&+ \sum_{z' \in \mathcal{Z}} \lambda_z(z'|z) \left( V^S(n, \{ x_i'(n; z', \varphi) \}_{i=1}^n; z', \varphi) - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \\
&+ \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi'|\varphi) \left( V^S(n, \{ x_i'(n; z, \varphi') \}_{i=1}^n; z, \varphi') - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \right. \\
&\left. \left. \right\} \right. \\
\end{align*}
\]

where \( \setminus \) and \( \cup_+ \) are multiset difference and union operators.\(^{53}\) 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 \( x_0'(n-1; z, \varphi) \). The firm promises, and must choose: (i) a menu of contracts \( \{ \omega_i \}_{i=1}^n \) 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 promise-keeping constraint now reads:

\[
\forall i = 1, \ldots, n : \quad x_i \leq V^B(n, \omega_i; z, \varphi)
\]

for all \((z, \varphi) \in \mathcal{Z} \times \Phi\), establishing that the firm commits to each and every customer (and recognizes that each customer may be enjoying a different utility). As in the baseline model, we can solve for the optimal menu of contracts by solving for the joint surplus problem:

\(^{53}\) These operators are defined by \( \{a, b, b\} \setminus \{b\} = \{a, b\} \) 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.
Proposition 5 In the economy with price discrimination, the seller’s and the joint surplus problems are equivalent:

(i) Given a menu of contracts $\omega_i = \{p_i, x'_i(n'; s')\}$ for $i = 1, \ldots, n$ that maximize the seller’s value subject to the promise-keeping constraint, $\{x'_i(n'; s')\}_{i=1}^n$ 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 $\{x'_i(n'; s')\}_{i=1}^n$ 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 $\{p_i, x'_i(n'; s')\}_{i=1}^n$ constitutes a solution to the seller’s problem.

Proof. See 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:

$$\forall (n, z, \varphi) \in \mathbb{N} \times Z \times \Phi : W_n(z, \varphi) = W(n, \{x_i\}_{i=1}^n; z, \varphi) \quad (32)$$

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:

$$(r + \delta f)W_n(z, \varphi) = n v(\varphi) - C(n; z, \varphi) + n(\delta f + \delta c)U^B(\varphi)$$

$$+ \eta(\theta(x'_0; \varphi))\left(W_{n+1}(z, \varphi) - W_n(z, \varphi) - \sum_{i=1}^n x'_{i,n+1}(z, \varphi)\right)$$

$$+ n\delta_c(W_{n-1}(z, \varphi) - W_n(z, \varphi)) + \sum_{z' \in Z} \lambda_z(z'|z)\left(W_n(z', \varphi) - W_n(z, \varphi)\right)$$

$$+ \sum_{\varphi' \in \Phi} \lambda_{\varphi'}(\varphi'|\varphi)\left(W_n(z, \varphi') - W_n(z, \varphi)\right)$$

Importantly, 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.\(^{54}\) 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 $\{p^*_i, x^*_i(n'; s')\}_{i=1}^n$ that leave both the buyers and the seller indifferent.

\(^{54}\) 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 goal in this section is only to point out that multiplicity reemerges within the Markov environment after we introduce discrimination.
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 sellers of the same size cannot propose different contracts, the problems of the sellers 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 sellers 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.

6 Conclusion

Recent studies indicate that a major source of variation in the performance of businesses across a variety of industries stems from demand components that are idiosyncratic to the firm. Further, price differences are key to explain revenue differences for firms of similar productivity levels. These observations shed new light on the behavior of markups at the aggregate level. We have presented a dynamic search model of demand accumulation through 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 the macroeconomic dynamics of prices and markups. In the model, sellers of different 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 dynamics at both the firm and the aggregate level.

We have analyzed a number of predictions on both pricing and firm dynamics dimensions. Using product-level data from the U.S. retail sector, 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. In response to adverse and mean-reverting aggregate shocks, sellers inter-temporally smooth out the effects of the shock on prices by transferring the burden onto their future buyers, giving rise to an incomplete price response. This gives a rationale for procyclicality in the markup response to both supply and demand shocks, as documented empirically (e.g. Nekarda and Ramey (2013)). Moreover, we have also shown that smaller sellers experience stronger and more persistent responses, due to the slow adjustment in the size distribution during the
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 the model, a shock that depresses firm entry is associated with a shift in the firm distribution toward higher-markup firms. A continued decrease in the costs of buyer search can explain these facts simultaneously.

Overall, these results suggest that incorporating micro-founded pricing environments into quantitative macro models is important to understand certain patterns of macroeconomic dynamics and firm heterogeneity. Further investigating the ability of customer markets to explain these and other dimensions remains an exciting avenue for future work.
References


Appendix

Contents. Appendix A includes additional figures and tables. 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, and price statistics along the stationary solution.

A Additional Figures and Tables

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

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

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 (26)) across all time periods.
\[ n = 0 (1, z_i) \]
\[ \delta_f + \delta_c \]
\[ \lambda_z(z_i|z_j) \]
\[ \lambda_z(z_j|z_i) \]
\[ \pi_z(z_i)\eta(\theta_1(z_i)) \]
\[ \pi_z(z_j)\eta(\theta_1(z_j)) \]

\[ (n-1, z_i) \]
\[ n\delta_c \]
\[ \lambda_z(z_i|z_j) \]
\[ \lambda_z(z_j|z_i) \]
\[ \eta(\theta_n(z_i)) \]

\[ (n, z_i) \]
\[ (n+1, z_i) \]

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

<table>
<thead>
<tr>
<th>Relative prices</th>
<th>Normalized sales</th>
<th>Sales growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Median</td>
<td>0.0009</td>
<td>0.9087</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentiles</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>-0.3257</td>
<td>0.2729</td>
</tr>
<tr>
<td>10(^{th})</td>
<td>-0.1138</td>
<td>0.4709</td>
</tr>
<tr>
<td>25(^{th})</td>
<td>-0.0415</td>
<td>0.6656</td>
</tr>
<tr>
<td>75(^{th})</td>
<td>0.0486</td>
<td>1.2281</td>
</tr>
<tr>
<td>90(^{th})</td>
<td>0.1097</td>
<td>1.7029</td>
</tr>
<tr>
<td>99(^{th})</td>
<td>0.2809</td>
<td>2.3819</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dispersion</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>St. Dev.</td>
<td>0.1055</td>
<td>0.4744</td>
</tr>
<tr>
<td>p90-p10 range</td>
<td>1.2504</td>
<td>3.6163</td>
</tr>
<tr>
<td>p90-p50 range</td>
<td>1.1149</td>
<td>1.8740</td>
</tr>
<tr>
<td>p50-p10 range</td>
<td>1.1215</td>
<td>1.9297</td>
</tr>
</tbody>
</table>

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.
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 (29)), 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: Hazard rate, frequency, duration, expected price change, and price change dispersion, as a function of size $(n)$, in the calibrated economy.
Figure A.7: 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.8: Depiction of Proposition 4: Equilibrium joint surplus \( \{W_n\} \) across \( n \), for a fixed exogenous state \((z, \varphi)\), where \( T \) is the mapping \( T : x \mapsto \left( \frac{\kappa}{z} \right)^{\gamma} \left( \frac{x}{z} \right)^{1-\gamma} \).
Figure A.9: 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.

---

Table A.3: 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.

<table>
<thead>
<tr>
<th>Year</th>
<th>New stores</th>
<th>All stores</th>
<th>Entry rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>18</td>
<td>189</td>
<td>9.52%</td>
</tr>
<tr>
<td>2002</td>
<td>11</td>
<td>182</td>
<td>6.04%</td>
</tr>
<tr>
<td>2003</td>
<td>14</td>
<td>172</td>
<td>8.14%</td>
</tr>
<tr>
<td>2004</td>
<td>9</td>
<td>176</td>
<td>5.11%</td>
</tr>
<tr>
<td>2005</td>
<td>13</td>
<td>185</td>
<td>7.03%</td>
</tr>
<tr>
<td>2006</td>
<td>20</td>
<td>187</td>
<td>10.7%</td>
</tr>
<tr>
<td>2007</td>
<td>32</td>
<td>200</td>
<td>16%</td>
</tr>
<tr>
<td>Average</td>
<td>16.71</td>
<td>184.43</td>
<td>8.9%</td>
</tr>
</tbody>
</table>

---

Table A.4: 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.
B Omitted Proofs

B.1 Proof of Proposition 1: Joint Surplus Problem

Proof. Denote by \(\overline{\omega} = \{\overline{p}, \overline{x}'(n'; s')\}\) a generic policy of the typical seller in state \((n; z, \varphi)\), where \(\overline{p}\) is the price level,

\[\overline{x}'(n'; s') = \{\overline{x}'(n + 1; z, \varphi), \overline{x}'(n - 1; z, \varphi), \{\overline{x}'(n; z', \varphi) : z' \in \mathbb{Z}\}, \{\overline{x}'(n; z, \varphi') : \varphi' \in \Phi\}\}\]

is the set of promised utilities, and \(\overline{x}'(n + 1; z, \varphi)\) and \(\overline{x}'(n - 1; z, \varphi)\) are the upsizing and downsizing choices, respectively. Recall that \(\overline{x}'(n; z, \varphi) = x\) by stationarity.

The value of the seller in equilibrium, \(V^S(n, x; z, \varphi)\), can be written as the maximand on the right-hand side of (6), evaluated at \(\overline{\omega}\). That is:

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

where \(\overline{V}^S(n, z, \varphi|\overline{\omega})\) is given by:\(^{55}\)

\[
\overline{V}^S(n, z, \varphi|\overline{\omega}) := \frac{1}{\rho(n; z, \varphi)} \left[ p_n - C(n, z, \varphi) + \eta \left( \theta(\overline{x}'(n + 1; z, \varphi); \varphi) \right) V^S(n + 1, \overline{x}'(n + 1; z, \varphi); z, \varphi) \right. \\
+ n \delta_c V^S(n - 1, \overline{x}'(n - 1; z, \varphi); z, \varphi) \\
+ \left. \sum_{z' \in \mathbb{Z}} \lambda_z(z'|z) V^S(n, \overline{x}'(n; z', \varphi); z', \varphi) + \sum_{\varphi' \in \Phi} \lambda_{\varphi'}(\varphi'|\varphi) V^S(n, \overline{x}'(n; z, \varphi'); z, \varphi') \right]
\]

and we have defined \(\rho(n; z, \varphi) := r + \delta_f + n \delta_c + \eta(\theta(\overline{x}'(n + 1; z, \varphi); \varphi))\) 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 \(\overline{\omega}\), 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 \(\overline{p}\) such that \(x = V^B(n, \overline{\omega}; z, \varphi)\) using equation (5). The resulting price is a function of the promised values denoted by \(p^{PK}\) and given by:

\[p^{PK} : \overline{x}'(n'; s') \mapsto \{v(\varphi) - \rho(n; z, \varphi)x + \delta_f U^B(\varphi) + \eta(\theta(\overline{x}'(n + 1; z, \varphi); \varphi))\overline{x}'(n + 1; z, \varphi) \]

\[+ \delta_c \left( U^B(\varphi) + (n - 1)\overline{x}'(n - 1; z, \varphi) \right) + \sum_{z' \in \mathbb{Z}} \lambda_z(z'|z) \overline{x}'(n; z', \varphi) + \sum_{\varphi' \in \Phi} \lambda_{\varphi'}(\varphi'|\varphi) \overline{x}'(n; z, \varphi') \}
\]

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 by commitment.

Using the above notation, we can now substitute the price level \(p^{PK} (\overline{x}'(n'; s'))\) from (B.1.2) into the seller’s value (B.1.1). After some straightforward algebra, we obtain:

\[\overline{W}(n, x; z, \varphi|\overline{\omega}) = \frac{1}{\rho(n; z, \varphi)} \left[ n \left( v(\varphi) + (\delta_f + \delta_c) U^B(\varphi) \right) \right]
\]

\(^{55}\) Here we are arguing by free-entry that \(\overline{V}^S(0; 0, \varphi) = 0\), \(\forall \varphi \in \Phi\), to simplify the expression for \(\overline{V}^S\). Moreover, we use that \(\sum_{z' \in \mathbb{Z}} \lambda_z(z'|z) = \sum_{\varphi' \in \Phi} \lambda_{\varphi'}(\varphi'|\varphi) = 0\).
\[-C(n; z, \varphi) + \eta\left(\theta\left(\overline{x}'(n + 1; z, \varphi)\right)\right)\overline{x}'(n + 1; z, \varphi)\]

\[+ \eta\left(\theta\left(\overline{x}'(n + 1; z, \varphi)\right)\right)W\left(n + 1, \overline{x}'(n + 1; z, \varphi)\right) + n\delta_xW\left(n - 1, \overline{x}'(n - 1; z, \varphi)\right)\]

\[+ \sum_{z' \in \Omega} \lambda_x(z'|z)W\left(n, \overline{x}'(n; z', \varphi); z', \varphi\right) + \sum_{\varphi' \in \Phi} \lambda_x(\varphi'|\varphi)W\left(n, \overline{x}'(n; z, \varphi'); z, \varphi'\right)\]

(B.1.3)

where we have defined:

\[\tilde{W}(n, x; z, \varphi|\omega) := \overline{V}_n^S(n; z, \varphi|\omega) + nx\text{ and } W(n, x; z, \varphi) := \max_{\omega \in \Omega} \tilde{W}(n, x; z, \varphi|\omega)\]

as the joint surplus under contract \(\omega\), and the maximized joint surplus, respectively. Next, note that the right-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:

\[\tilde{W}(n, x; z, \varphi|\omega) = \tilde{W}_n\left(\overline{x}'(n'; s'); z, \varphi\right)\]

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 vector of continuation values \(\overline{x}'(n'; s')\) that maximizes (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).

Formally, the optimal contract is \(\omega^* = \left\{p^*, \overline{x}^*(n'; s')\right\}\), where:

\[\overline{x}^*(n'; s') := \arg \max_{x} \tilde{W}_n(x; z, \varphi)\]  

(B.1.4a)

\[p^* := p^{PK}\left(\overline{x}^*(n'; s')\right)\]  

(B.1.4b)

By expressing the problem of the seller in terms of \(\tilde{W}\), we have just shown that the contract that is optimally chosen by the firm, \(\omega^*\), must maximize the joint surplus. Conversely, for any vector \(\overline{x}'(n'; s')\) of continuation values that maximizes the joint surplus, there is a price level, given by \(p^* = p^{PK}\left(\overline{x}^*(n'; s')\right)\), 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. \(\square\)

### B.2 Proof of Proposition 2: Invariant Distribution

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

\[\partial_t S_t(\varphi) = T_\varphi S_t(\varphi)\]  

(B.2.1)

where \(S_t(\varphi) := (S_{0,t}(\varphi), S_{1,t}^\top, \ldots, S_{\pi,t}^\top)^\top\), with \(S_{n,t}(z) := (S_{n,t}(z_1), \ldots, S_{n,t}(z_k))^\top\), and \(T_\varphi\) is the partitioned matrix.
\[
T_\varphi := \begin{pmatrix}
t_{11} & \delta_{f} & \delta_{c} & \delta_{f} & \delta_{f} & \cdots & \delta_{f} & \delta_{f} & \delta_{c} \\
\eta_{f}(\varphi)^T & D_1(\varphi) & \delta_{2,c} & 0_{k_z,k_z} & \cdots & 0_{k_z,k_z} & 0_{k_z,k_z} & 0_{k_z,k_z} & 0_{k_z,k_z} \\
0_{k_z,1} & \eta_{2}(\varphi) & D_2(\varphi) & \delta_{3,c} & \cdots & 0_{k_z,k_z} & 0_{k_z,k_z} & 0_{k_z,k_z} & 0_{k_z,k_z} \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0_{k_z,1} & 0_{k_z,k_z} & 0_{k_z,k_z} & \cdots & D_{\pi-2}(\varphi) & \delta_{\pi-1,c} & 0_{k_z,k_z} \\
0_{k_z,1} & 0_{k_z,k_z} & 0_{k_z,k_z} & \cdots & \eta_{\pi-1}(\varphi) & D_{\pi-1}(\varphi) & \delta_{\pi,c} \\
0_{k_z,1} & 0_{k_z,k_z} & 0_{k_z,k_z} & \cdots & 0_{k_z,k_z} & \eta_{\pi}(\varphi) & D_{\pi}(\varphi) \\
\end{pmatrix}
\]

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

\[
\delta_f = (\delta_f, \ldots, \delta_f); \quad \delta_c = (\delta_c, \ldots, \delta_c);
\]

and the following \( k_z \times k_z \) matrices:

\[
\forall n = 2, \ldots, \pi: \quad \eta_n(\varphi) = \text{diag}(\eta(\theta_n(z_1, \varphi)), \ldots, \eta(\theta_n(z_{k_z}, \varphi)));
\]

\[
\forall n = 1, \ldots, \pi: \quad D_n(\varphi) = \begin{pmatrix}
d_n(z_1, \varphi) & \lambda_z(z_1|z_2) & \cdots & \lambda_z(z_1|z_{k_z}) \\
\lambda_z(z_2|z_1) & d_n(z_2, \varphi) & \cdots & \lambda_z(z_2|z_{k_z}) \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_z(z_{k_z}|z_1) & \lambda_z(z_{k_z}|z_2) & \cdots & d_n(z_{k_z}, \varphi)
\end{pmatrix}
\]

where the diagonal elements of \( D_n(\varphi) \) are given by:

\[
d_n(z_j, \varphi) = \begin{cases}
-\left(\delta_f + n \delta_c + \eta(\theta_{n+1}(z_j, \varphi)) + \sum_{\ell \neq j} \lambda_z(z_{\ell}|z_j)\right) & \text{for } n = 1, \ldots, \pi - 1 \\
-\left(\delta_f + \pi \delta_c + \sum_{\ell \neq j} \lambda_z(z_{\ell}|z_j)\right) & \text{for } n = \pi
\end{cases}
\]

(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 \((n', z') \in N \times Z\) 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 \( S^*(\varphi) \), for each \( \varphi \in \Phi \).

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

\[
S_t(\varphi) = e^{T_\varphi t} S_0(\varphi)
\]

where the initial distribution \( S_0(\varphi) \in \mathbb{R}_+^K \) is given. To compute \( e^{T_\varphi t} \), consider the eigenvalue decomposition \( T_\varphi = E_\varphi \Lambda_\varphi E_\varphi^{-1} \), where \( \Lambda_\varphi := (\lambda_1(\varphi), \ldots, \lambda_K(\varphi)) \) is the diagonal matrix of eigenvalues, and \( E_\varphi \) collects the corresponding eigenvectors. Defining \( Z_t(\varphi) := E_\varphi^{-1} S_t(\varphi) \), then \( \partial_t Z_t(\varphi) = \Lambda_\varphi Z_t(\varphi) \), and because \( \Lambda_\varphi \) is a diagonal matrix, we can solve this differential equation element-by-element, i.e. \( \partial_t Z_{i,t}(\varphi) = \lambda_i(\varphi) Z_{i,t}(\varphi) \) 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 the constant of integration. Since \( S_t(\varphi) = E_{\varphi}Z_t(\varphi) \), we have obtained:

\[
S_t(\varphi) = \sum_{i=1}^{K} c_i e^{\lambda_i(\varphi)t} v_i
\] (B.2.2)

where \( v_i \) 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 \( T_\varphi \). The trace of \( T_\varphi \) is:

\[
\text{tr}(T_\varphi) = \sum_{i=1}^{K} \lambda_i(\varphi) = -\sum_{j=1}^{K} \pi_z(z_j)\eta(\theta_1(z_j, \varphi)) + \sum_{n=1}^{\pi} \sum_{j=1}^{K} d_n(z_j) < 0
\]

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:

\[
S^*(\varphi) := \lim_{t \to +\infty} \sum_{j=1}^{\ell} c_j e^{\lambda_j(\varphi)t} v_j \in \mathbb{R}_+^K
\]

is the unique invariant distribution of sellers in state \( \varphi \in \Phi \). \( \square \)

### 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, \( \theta \). To simplify notation, it is understood that time subscripts embody the entire history of aggregate shocks, which is taken to be some arbitrary path \( \varphi^t = (\varphi_j : j \leq t) \subseteq \Phi \).

The planner chooses:

- The tightness in each market segment, \( \Theta_t \equiv \{\theta_{n,t}(z_t) : (n_t, z_t) \in \mathbb{N} \times \mathcal{Z}\} \);
- Distributions of inactive and active buyers across markets, \( B^I_t \equiv \{B^I_{n,t}(z_t) : (n_t, z_t) \in \mathbb{N} \times \mathcal{Z}\} \) and \( B^A_t \equiv \{B^A_{n,t}(z_t) : (n_t, z_t) \in \mathbb{N} \times \mathcal{Z}\} \);
- A measure of potential entrants \( S_{0,t} \);
- A distribution of firms across states \( S_t \equiv \{S_{n,t}(z_t) : (n_t, z_t) \in \mathbb{N} \times \mathcal{Z}\} \).

The planner’s objective is:

\[
\max_{\Theta_t, B^I_t, B^A_t, S_{0,t}, S_t} E_0 \int_0^{+\infty} e^{-rt} W_t(\varphi_t) dt
\]

(B.3.1)

where

\[
W_t(\varphi_t) = -\kappa(\varphi_t)S_{0,t} + \sum_{n_t=1}^{+\infty} \sum_{z_t \in \mathcal{Z}} \left[ v(\varphi_t)B^A_{n,t}(z_t) - C(n_t; z_t, \varphi_t)S_{n,t}(z_t) - c(\varphi_t)B^I_{n,t}(z_t) \right]
\]

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,\(^{56}\) and net of the search and entry costs for inactive buyers and inactive sellers.

---

\(^{56}\) The sum of these first two terms equals the total gains from trade.
The planner is subject to three sets of constraints. The first one concerns the evolution of the firm distribution, which we described in Section 2.5 and reproduce again here for convenience:

\[
\begin{align*}
\partial_t S_{0,t} &= \delta_f \sum_{n_t=1}^{\infty} \sum_{z_t \in \mathcal{Z}} S_{n_t,t}(z_t) + \delta_c \sum_{z_t \in \mathcal{Z}} S_{1,t}(z_t) - \sum_{z_t' \in \mathcal{Z}} \pi_{z}(z') \eta(\theta_{1,t}(z')) S_{0,t} \\
\theta_t S_{1,t}(z_t) &= \pi_{z}(z_t) \eta(\theta_{1,t}(z_t)) S_{0,t} + 2 \delta_c S_{2,t}(z_t) + \sum_{\tilde{z} \neq z_t} \lambda_{z}(z_t | \tilde{z}) S_{1,t}(\tilde{z}) \\
&- \left( \delta_f + \delta_c + \eta(\theta_{2,t}(z_t)) + \sum_{\tilde{z} \neq z_t} \lambda_{z}(\tilde{z} | z_t) \right) S_{1,t}(z_t)
\end{align*}
\]  
\(\forall n_t \geq 2: \quad \theta_t S_{n_t,t}(z_t) = \eta(\theta_{n_t,t}(z_t)) 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}) \\
&- \left( \delta_f + n_t \delta_c + \eta(\theta_{n_t+1,t}(z_t)) + \sum_{\tilde{z} \neq z_t} \lambda_{z}(\tilde{z} | z_t) \right) S_{n_t,t}(z_t);
\]

for all \(z_t \in \mathcal{Z}\), where \(\zeta\) denotes the productivity draw upon entry. The second set of constraints describes the distribution of buyers across firms at any given time:

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

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

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

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) 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:

\[
\max_{\Theta_t, S_{0,t}, S_t} \mathbb{E}_0 \int_0^{+\infty} e^{-rt} \left\{ - \left( \kappa(\varphi_t) + c(\varphi_t) \sum_{z_t \in \mathcal{Z}} \theta_1(z_t) \right) S_{0,t} + v(\varphi_t) \sum_{n_t=1}^{+\infty} \sum_{z_t \in \mathcal{Z}} n_t S_{n_t,t}(z_t) \right.
\]

\[
- \sum_{n_t=1}^{+\infty} \sum_{z_t \in \mathcal{Z}} c(n_t; z_t, \varphi_t) S_{n_t,t}(z_t) - c(\varphi_t) \sum_{n_t=1}^{+\infty} \sum_{z_t \in \mathcal{Z}} \theta_{n_t+1,t}(z_t) S_{n_t,t}(z_t) \right\} dt
\]

subject to (B.3.2a), (B.3.2b), (B.3.2c), (B.3.4), and (B.3.5). Conveniently, the variables \(B_t^I\) and \(B_t^A\) have disappeared from the problem, and the state vector has been reduced to the measures of firms: \(S_t \equiv [S_{0,t}, S_t]\).
To solve the simplified planner’s problem, we use standard tools from Optimal Control theory, where \(\Theta_t\) is the control variable. The current-value Hamiltonian of the simplified planning problem is:

\[
\mathcal{H}_t(\Theta_t; S_t) := - \left( \kappa(\varphi_t) + c(\varphi_t) \sum_{z_t \in Z} \theta_t(z_t) \right) S_{0,t} + v(\varphi_t) \sum_{n_t = 1}^{\infty} \sum_{z_t \in Z} n_t S_{n_t,t}(z_t) - \sum_{n_t = 1}^{\infty} \sum_{z_t \in Z} C(n_t; z_t, \varphi_t) S_{n_t,t}(z_t) - c(\varphi_t) \sum_{n_t = 1}^{\infty} \sum_{z_t \in Z} \theta_{n_t+1,t}(z_t) S_{n_t,t}(z_t) + \phi_t \left[ 1 - \sum_{n_t = 1}^{\infty} \sum_{z_t \in Z} n_t S_{n_t,t}(z_t) - \sum_{n_t = 1}^{\infty} \sum_{z_t \in Z} \theta_{n_t+1,t}(z_t) S_{n_t,t}(z_t) - S_{0,t} \sum_{z_t \in Z} \theta_{1,t}(z_t) \right] + \psi_0, t \left[ \delta f \sum_{n_t = 1}^{\infty} \sum_{z_t \in Z} S_{n_t,t}(z_t) + \delta_c \sum_{z_t \in Z} S_{1,t}(z_t) - \sum_{z^c \in Z} \pi_z(z^c) \eta(\theta_{1,t}(z^c)) S_{0,t} \right] + \sum_{z_t \in Z} \left\{ \psi_{1,t}(z_t) \left[ \pi_z(z_t) \eta(\theta_{1,t}(z_t)) S_{0,t} + 2 \delta_c S_{2,t}(z_t) + \sum_{\tilde{z} \not= z_t} \lambda_{z_t} S_{1,t}(\tilde{z}) \right] \right. \\
- \left( \delta f + \delta_c + \eta(\theta_{2,t}(z_t)) + \sum_{\tilde{z} \not= z_t} \lambda_z(z_t) \right) S_{1,t}(z_t) \left. \right\} + \psi_{0,t} S_{0,t}
\]

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 \(\psi_{0,t} \geq 0\) is the multiplier on the non-negative entry condition, where the corresponding complementary slackness hold.

In vector notation, the necessary conditions for optimality are:

\[
\nabla_{\Theta} \mathcal{H}_t(\Theta_t; S_t) = 0 \quad \nabla_{S_t} \mathcal{H}_t(\Theta_t; S_t) = -\nabla_t \psi_t + r \psi_t
\]

where \(\nabla\) denotes the gradient operator, and \(\psi_t\) 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 \(\eta_t\), all the constraints are concave in the control and linear in the states.

Regarding the first set of optimality conditions, for given \(z_t \in Z\) we have:

\[
[\theta_t] : \quad \phi_t + c(\varphi_t) = \left( \psi_{1,t}(z_t) - \psi_{0,t} \right) \pi_z(z_t) \frac{\partial \eta(\theta)}{\partial \theta} \bigg|_{\theta = \theta_{1,t}(z_t)} \tag{B.3.6a}
\]

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

As for the second set of conditions, we have:

\[
[S_0] : \quad -\nabla_t \psi_{0,t} + r \psi_{0,t} = -\kappa(\varphi_t) - \left( \phi_t + c(\varphi_t) \right) \sum_{z_t \in Z} \theta_t(z_t) + \sum_{z^c \in Z} \pi_z(z^c) \eta(\theta_{1,t}(z^c)) \psi_{1,t}(z^c) \tag{B.3.7a}
\]

80
of the joint surplus in the decentralized solution, equation (10). Using that allocation.

First order conditions of those equations. Therefore, it suffices to find the values of the multipliers for state variables can be represented as HJB equations. Moreover, (B.3.6a)-(B.3.6b) are the corresponding co-state variables of the planning problem. Note that equations (B.3.7a)-(B.3.7b) show that the co-

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 \( \{ W_n(z, \varphi), x_n(z, \varphi), \theta_n(z, \varphi), U^B(\varphi) : (n, z, \varphi) \in \mathbb{N} \times Z \times \Phi \} \), and consider the following realization for the multipliers:

\[
\begin{align*}
\phi_t(\varphi_t) & = rU^B(\varphi_t) \\
\psi_0(t, \varphi_t) & = 0 \\
\forall n_t, z_t : \quad \psi_{n_t,t}(z_t, \varphi_t) & = W_{n_t}(z_t, \varphi_t) - n_tU^B(\varphi_t)
\end{align*}
\]

Under this guess, notice that \( \partial_t \psi_{0,t} = \partial_t \psi_{n,t}(z_t) = 0, \forall n \geq 1 \). 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 \( \kappa \), 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:

\[
(r + \delta_f)W_{n_t}(z_t, \varphi_t) = n_t \left( v(\varphi_t) + (\delta_f + \delta_c)U^B(\varphi_t) \right) - C(n_t, z_t; \varphi_t)
\]

\[
- \left[ \left( rU^B(\varphi_t) + c(\varphi_t) \right) \theta_{n_t+1,t}(z_t) + \eta(\theta_{n_t+1}(z_t))U^B(\varphi_t) \right]
\]

\[
+ n_t \delta_c \left( W_{n_t-1}(z_t, \varphi_t) - W_{n_t}(z_t, \varphi_t) \right)
\]

\[
+ \eta(\theta_{n_t+1}(z_t)) \left( W_{n_t+1}(z_t, \varphi_t) - W_{n_t}(z_t, \varphi_t) \right)
\]

\[
+ \sum_{\tilde{z} \in Z} \lambda_{\tilde{z}}(\tilde{z}|z_t) \left( W_{n_t}(\tilde{z}, \varphi_t) - W_{n_t}(z_t, \varphi_t) \right)
\]

Notice that the last equation resembles the maximized HJB equation for the joint surplus (equation (10)) except for the second line. 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:

\[
(rU^B(\varphi_t) + c(\varphi_t)) \theta_{n_t+1,t}(z_t) + \eta(\theta_{n_t+1}(z_t))U^B(\varphi_t) = \eta(\theta_{n_t+1,t}(z_t, \varphi_t)) x_{n_t+1,t}(z_t, \varphi_t) \quad (B.3.8)
\]

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 (10).

Similarly, plugging the guess for the multipliers into (B.3.7a), we obtain:
\[\kappa(\varphi_t) = -(rU^B(\varphi_t) + c(\varphi_t)) \sum_{z_t \in Z} \theta_1(z_t) + \sum_{z^* \in Z} \pi(z^*) \eta(\theta_1,t(z^*)) \left(W_1(z^*) - U^B(\varphi_t)\right)\]

A final manipulation using (B.3.8) again then allows us to obtain the free entry condition in the decentralized allocation, equation (13).

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 constrained-efficient. □

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

**Proof.** The equilibrium allocation is composed of sequences

\[\{W_n(z, \varphi), x_n(z, \varphi), \theta_n(z, \varphi), p_n(z, \varphi) : (n, z, \varphi) \in \mathbb{N} \times Z \times \Phi\}\]

satisfying equations (10), (11), and (15), where the free entry condition (13) pins down \(x_1\) and the first-order condition (12) pins down \(x_n\) given \(x_{n-1}\), \(n \geq 2\), for any \(\varphi \in \Phi\).

For \(\mu(\theta) = \theta^{-1}, \gamma \in (0, 1)\), equation (11) defines the following equilibrium mapping:

\[\theta : (x; \varphi) \mapsto \left(\frac{x - U^B(\varphi)}{\Gamma^B(\varphi)}\right)^{1/\gamma} \]  

(B.4.1)

Some algebra shows that the joint surplus maximization rule (equation (12)) can be written as:

\[W_{n+1}(z, \varphi) - W_n(z, \varphi) - x_{n+1}(z, \varphi) = \frac{1-\gamma}{\gamma} \left(x_{n+1}(z, \varphi) - U^B(\varphi)\right)\]

(B.4.2)

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:

\[x_{n+1}(z, \varphi) - U^B(\varphi) = \gamma \left(W_{n+1}(z, \varphi) - W_n(z, \varphi) - U^B(\varphi)\right)\]

\[:= \Gamma_{n+1}(z, \varphi)\]

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

Next, define:

\[\Gamma^S_n(z, \varphi) := (r + \delta_f)W_n(z, \varphi) - \pi_n(z, \varphi) + n\delta_c\left(W_n(z, \varphi) - W_{n-1}(z, \varphi)\right) - n(\delta_c + \delta_f)U^B(\varphi) - \Xi_n(z, \varphi)\]

(B.4.3)

where \(\pi_n(z, \varphi) := n\nu(\varphi) - C(n; z, \varphi)\) is the flow joint surplus, and

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

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

\[\Gamma^S_n(z, \varphi) = \eta(\theta_{n+1}(z, \varphi)) \left(W_{n+1}(z, \varphi) - W_n(z, \varphi) - x_{n+1}(z, \varphi)\right)\]
where the first line uses the HJB equation for the joint surplus (equation (10)), the second line uses (B.4.2), and the third line uses (B.4.1) and \( \eta(\theta) = \theta \mu(\theta) \). The right-hand side of the first equality allows us to interpret \( \Gamma^S \) as the expected match surplus for the seller. Using the last equality, we have found the market tightness:

\[
\theta_{n+1}(z, \varphi) = \left( \frac{\gamma}{1 - \gamma} \right) \frac{\Gamma^S_n(z, \varphi)}{\Gamma^B(\varphi)} 
\]

for all \( n \geq 1 \). Finally, we can write (B.4.2) as

\[
W_{n+1}(z, \varphi) - W_n(z, \varphi) = U^B(\varphi) + \gamma^{-1}(x_{n+1}(z, \varphi) - U^B(\varphi)) = U^B(\varphi) + \gamma^{-1}(W_1(z, \varphi))^1 - \gamma
\]

Using (B.4.4) and rearranging terms, we obtain our desired result:

\[
W_{n+1}(z, \varphi) = W_n(z, \varphi) + U^B(\varphi) + \left( \frac{\Gamma^B(\varphi)}{\gamma} \right) \left( \frac{\Gamma^S_n(z, \varphi)}{1 - \gamma} \right)^{1 - \gamma} \tag{B.4.5}
\]

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 (13). By (B.4.2), 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 (13) as:

\[
\eta(\varphi) = (1 - \gamma) \left( \frac{\Gamma^B(\varphi)}{\gamma} \right)^{\frac{\gamma}{\gamma - 1}} \sum_{z_0 \in Z} \pi(z_0) \left( W_1(z_0, \varphi) - U^B(\varphi) \right)^{\frac{1}{1 - \gamma}}
\]

our desired result. \( \square \)

Figure A.8 shows the solution to the second-order difference equation graphically when \( x_n \) is a weakly decreasing sequence in \( n \). 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 \( \kappa \)). 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 \( \{\pi_0, \{\omega_i\}_{i=1}^n\} \), with \( \omega_i := \{\pi_i, \pi_i'(n'; s')\} \) and \( \pi_i'(n; z, \varphi) = \{\pi_i(n + 1; z, \varphi), \pi_i(n - 1; z, \varphi), \pi_i(n; z', \varphi) : z' \in Z\}, \{\pi_i'(n; z, \varphi) : \varphi' \in \Phi\} \), form a generic policy for the firm, the firm’s problem can be written as:

\[
V^S\left(n, \{x_i\}_{i=1}^n; z, \varphi\right) := \max_{\pi_0, \{\omega_i\}_{i=1}^n} \tilde{V}^S\left(n, \pi_0', \{\omega_i\}_{i=1}^n; z, \varphi\right) \text{ s.t. } x_i \leq V^B(n, \omega_i; z, \varphi), \forall i = 1, \ldots, n
\]

where:

\[
\tilde{V}^S\left(n, \pi_0', \{\omega_i\}_{i=1}^n; z, \varphi\right) := \frac{1}{\rho(n; z, \varphi)} \left[ \sum_{i=1}^n p_i - C(n; z, \varphi) \right] \tag{B.5.1}
\]
+ δ_c \sum_{j=1}^{n} V^S \left( n-1, \{ \mathcal{P}_i(n-1; z, \varphi) \}_{i=1}^{n} \right) \cup \{ \mathcal{P}_f(n-1; z, \varphi) \}_{i=1}^{n} \right) 
+ \eta(\theta(\mathcal{P}_0'; \varphi)) V^S \left( n+1, \{ \mathcal{P}_i(n+1; z, \varphi) \}_{i=1}^{n} \right) 
+ \sum_{z' \in Z} \lambda_z(z'|z) V^S \left( n, \{ \mathcal{P}_{i}(n; z', \varphi) \}_{i=1}^{n} \right) 
+ \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi'|\varphi) V^S \left( n, \{ \mathcal{P}_{i}(n; z, \varphi') \}_{i=1}^{n} \right)

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

\[
\nu^S(n, \mathcal{W}_i; z, \varphi) = v(\varphi) - p_i + (\delta_f + \delta_c) \left( U^B(\varphi) - V^B(n, \mathcal{W}_i; z, \varphi) \right) 
+ (n-1)\delta_c \left( \mathcal{P}_{i}(n-1; z, \varphi) - V^B(n, \mathcal{W}_i; z, \varphi) \right) 
+ \eta(\theta(\mathcal{P}_0'; \varphi)) \left( \mathcal{P}_{i}(n+1; z, \varphi) - V^B(n, \mathcal{W}_i; z, \varphi) \right) 
+ \sum_{z' \in Z} \lambda_z(z'|z) \left( \mathcal{P}_{i}(n; z', \varphi) - V^B(n, \mathcal{W}_i; z, \varphi) \right) 
+ \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi'|\varphi) \left( \mathcal{P}_{i}(n; z, \varphi') - V^B(n, \mathcal{W}_i; z, \varphi) \right)
\]

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

\[
x_i = V^B(n, \mathcal{W}_i; z, \varphi), \quad \forall i = 1, \ldots, n
\]

From this equation, we can solve for the promise-compatible price level to be charged to each customer under the policy \( \{ \mathcal{P}_0', \{ \mathcal{W}_i \}_{i=1}^{n} \} \):

\[
p_i^{PK} \left( \{ \mathcal{P}_0', \{ \mathcal{X}_j(n'; s') \}_{j \neq i} \}_{i=1}^{n} \right) = v(\varphi) - \rho(n; z, \varphi)x_i + \delta_f U^B(\varphi) 
+ \delta_c \left( U^B(\varphi) + (n-1)\mathcal{P}_{i}(n-1; z, \varphi) \right) 
+ \eta(\theta(\mathcal{P}_0'; \varphi)) \mathcal{P}_{i}(n+1; z, \varphi) 
+ \sum_{z' \in Z} \lambda_z(z'|z) \mathcal{P}_{i}(n; z', \varphi) 
+ \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi'|\varphi) \mathcal{P}_{i}(n; z, \varphi')
\]  

(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:

\[
p_i^{PK} \left( \{ \mathcal{P}_0', \{ \mathcal{X}_j(n'; s') \}_{j \neq i} \}_{i=1}^{n} \right) = p_i^{PK} \left( \{ \mathcal{P}_0', \{ \mathcal{X}_j(n'; s') \}_{j \neq i} \}_{i=1}^{n} \right)
\]

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:

\[
\nu \left( n, \{ x_i \}_{i=1}^{n}, \mathcal{P}_0', \{ \mathcal{W}_i \}_{i=1}^{n}; z, \varphi \right) := \frac{1}{\rho(n; z, \varphi)} \left[ n \left( v(\varphi) + (\delta_f + \delta_c) U^B(\varphi) \right) \right]
\]

(B.5.3)
\[- \left( C(n; z, \varphi) + \eta(\theta(x_0^n; \varphi)) \sum_{i=1}^n x_i'(n+1; z, \varphi) \right) \]

\[+ \delta \epsilon \sum_{j=1}^n W(n-1, \{x_j'(n-1; z, \varphi)\}_{i=1}^n \setminus \{x_j'(n-1; z, \varphi)\}; z, \varphi) \]

\[+ \eta(\theta(x_0^n; \varphi)) W(n+1, \{x_j'(n+1)\}_{i=1}^n \cup \{x_0^n\}; z, \varphi) \]

\[+ \sum_{z' \in \mathbb{Z}} \lambda_z(z'|z) W(n, \{x_j'(n; z', \varphi)\}_{i=1}^n; z, \varphi) \]

\[+ \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi'|\varphi) W(n, \{x_j'(n; z, \varphi')\}_{i=1}^n; z, \varphi') \]

where we have defined:

\[\tilde{W}(n, \{x_i\}_{i=1}^n; x_0^n, \{\omega_i\}_{i=1}^n; z, \varphi) := \tilde{V}(n; x_0^n, \{\omega_i\}_{i=1}^n; z, \varphi) + \sum_{i=1}^n x_i\]

and:

\[W(n, \{x_i\}_{i=1}^n; z, \varphi) := \max_{x_0^n, \{\omega_i\}_{i=1}^n} \tilde{W}(n, \{x_i\}_{i=1}^n; x_0^n, \{\omega_i\}_{i=1}^n; z, \varphi)\]

as the joint surplus under policy \{x_0^n, \{\omega_i\}_{i=1}^n\}, and the \textit{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:

\[\tilde{W}(n, \{x_i\}_{i=1}^n; x_0^n, \{\omega_i\}_{i=1}^n; z, \varphi) = \tilde{W}(n; x_0^n, \{x_i^n(n'; s')\}_{i=1}^n; z, \varphi)\]

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

\[\{x_0^n, \{x_i^n(n'; s')\}_{i=1}^n\} = \arg \max_{x_0^n, \{x_i^n\}_{i=1}^n} \tilde{W}(n; x_0^n, \{x_i^n\}_{i=1}^n; z, \varphi)\]

\[p_i^* = p_i^{PK}(\{x_0^n, \{x_i^n(n'; s')\}_{i=1}^n\}), \quad \forall i = 1, \ldots, n\]

Thus, the joint surplus and the seller’s problems are equivalent. \(\square\)

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

\textbf{Proof.} Let \(\epsilon \in \mathbb{R}\) 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 \(\omega^b = \{p_i + \epsilon \beta_n(\varphi), x_i^n(n'; s') + \epsilon\}_{i=1}^n\) is optimal, then each customer and the seller maximize their value under contract \(\omega^a = \{p_i, x_i^n(n'; s')\}_{i=1}^n\). The value of contract \(\omega_i^b\) for customer \(i = 1, \ldots, n\) is:

\[r V^B(n, \omega_i^b; z, \varphi) = v(\varphi) - p_i - \epsilon \beta_n(\varphi) + (\delta + \delta_c) \left(U^B(\varphi) - V^B(n, \omega_i; z, \varphi)\right)\]

\[+ (n-1) \delta \epsilon \left(x_i'(n-1; z, \varphi) + \epsilon - V^B(n, \omega_i^b; z, \varphi)\right)\]

\[+ \eta(\theta(x_0^b; \varphi)) \left(x_i^b(n+1; z, \varphi) + \epsilon - V^B(n, \omega_i^b; z, \varphi)\right)\]

\[+ \sum_{z' \in \mathbb{Z}} \lambda_z(z'|z) \left(x_i^b(n; z', \varphi) + \epsilon - V^B(n, \omega_i^b; z, \varphi)\right)\]
\[ + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi' | \varphi) \left( x_i'(n; z, \varphi') + \varepsilon - V^B(n, \omega_i; z, \varphi') \right) \]

\[ = v(\varphi) - p_i - \varepsilon \left( \beta_n(\varphi) - (n - 1)\delta_c - \eta(\theta(x_0'; \varphi)) \right) \]

\[ + (\delta_f + \delta_c) \left( U^B(\varphi) - V^B(n, \omega_i; z, \varphi) \right) \]

\[ + (n - 1)\delta_c \left( x_i'(n - 1; z, \varphi) - V^B(n, \omega_i; z, \varphi) \right) \]

\[ + \eta(\theta(x_0'; \varphi)) \left( x_i'(n + 1; z, \varphi) - V^B(n, \omega_i; z, \varphi) \right) \]

\[ + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi' | \varphi) \left( x_i'(n; z, \varphi') - V^B(n, \omega_i; z, \varphi') \right) \]

\[ = rV^B(n, \omega_i; z, \varphi) + \varepsilon \left( \beta_n(\varphi) - (n - 1)\delta_c - \eta(\theta(x_0'; \varphi)) \right) \]

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

\[ \beta_n(\varphi) = (n - 1)\delta_c + \eta(\theta(x_0'; \varphi)) \quad \text{(B.6.1)} \]

As for the seller’s value, note that:

\[ rV^S(n, \{ x_i \}_{i=1}^n; z, \varphi) = \max_{x_0'} \left\{ \sum_{i=1}^n p_i + n\varepsilon \beta_n(\varphi) - C(n; z, \varphi) \right\} \]

\[ + \delta_f \left( V^S_0(\varphi) - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \]

\[ + \delta_c \sum_{j=1}^n \left( V^S(n - 1, \{ x_i'(n - 1; z, \varphi) + \varepsilon \}_{i=1}^n \setminus \{ x_i'(n-1; z, \varphi) + \varepsilon \}; z, \varphi) \right. \]

\[ - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \]

\[ + \eta(\theta(x_0'; \varphi)) \left( V^S(n + 1, \{ x_i'(n + 1; z, \varphi) + \varepsilon \}_{i=1}^n \cup \{ x_0'(n) \}; z, \varphi) \right. \]

\[ - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \]

\[ + \sum_{z' \in Z} \lambda_z(z'|z) \left( V^S(n, \{ x_i'(n; z', \varphi) + \varepsilon \}_{i=1}^n; z', \varphi) - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \]

\[ + \sum_{\varphi' \in \Phi} \lambda_\varphi(\varphi' | \varphi) \left( V^S(n, \{ x_i'(n; z, \varphi') + \varepsilon \}_{i=1}^n; z, \varphi') - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \}

\[ = \max_{x_0'(n), \{ \omega_i \}_{i=1}^n} \left\{ \sum_{i=1}^n p_i + n\varepsilon \beta_n(\varphi) - C(n; z, \varphi) + \delta_f \left( V^S_0(\varphi) - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \right\} \]

\[ + \delta_c \sum_{j=1}^n \left( W_{n-1}(z, \varphi) - \sum_{i \neq j} x_i'(n-1; z, \varphi) - (n - 1)\varepsilon - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \right) \]

\[ + \eta(\theta(x_0'; \varphi)) \left( W_{n+1}(z, \varphi) - \sum_{i=1}^n x_i'(n + 1; z, \varphi) - x_0' - n\varepsilon \right) \]

\[ - V^S(n, \{ x_i \}_{i=1}^n; z, \varphi) \]
+ \sum_{z' \in Z} \lambda(z'|z) \left( W_n(z', \varphi) - \sum_{i=1}^{n} x'_i(n; z', \varphi) - n \varepsilon - V^S \left( n, \left\{ x_i \right\}_{i=1}^{n}; z, \varphi \right) \right) \\
+ \sum_{\varphi' \in \Phi} \lambda(\varphi'|\varphi) \left( W_n(z, \varphi') - \sum_{i=1}^{n} x'_i(n; z, \varphi') - n \varepsilon - V^S \left( n, \left\{ x_i \right\}_{i=1}^{n}; z, \varphi \right) \right) \right\} \\
= rV^S \left( n, \left\{ x_i \right\}_{i=1}^{n}; z, \varphi \right) + n \varepsilon \left( \beta_n(\varphi) - (n - 1) \delta_n - \eta(\theta(x_0^b(\varphi); \varphi)) \right)

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 $\{\omega^a_i\}_{i=1}^{n}$ and $\{\omega^b_i\}_{i=1}^{n}$. Consequently, the joint surplus does not change by definition, and therefore contract $\{\omega^a_i\}_{i=1}^{n}$ is optimal if, and only if, $\{\omega^b_i\}_{i=1}^{n}$ is optimal. Generally, these is a continuum of optimal contracts, indexed by $\varepsilon$. □

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 $\{U^B(\varphi) : \varphi \in \Phi\}$, which 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 $N \times Z \times \Phi$, where $N = \{1, \ldots, \bar{n}\}$, with $\bar{n} \in \mathbb{N}$ a sufficiently large bound on firm size, and $Z = \{z_i\}_{i=1}^{k_z}$ and $\Phi = \{\varphi_j\}_{j=1}^{k_\varphi}$. The $(z, \varphi)$ 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 $U^{(0)}(\varphi)$ and $\overline{U}^{(0)}(\varphi) \gg U^{(0)}(\varphi)$ for each $\varphi \in \Phi$. Set the value of inactivity to:

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

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

$$(r + \delta_f)W_n^{(k)}(z, \varphi) = n \left( v(\varphi) + (\delta_f + \delta_c)U^{B(k)}(\varphi) \right) - C(n; z, \varphi)$$

$$+ n \delta_c \left( W_{n-1}^{(k)}(z, \varphi) - W_n^{(k)}(z, \varphi) \right)$$

$$+ \max_{z_{n+1}} \left\{ \eta \circ \mu^{-1} \left( \frac{\Gamma^{B(k)}(\varphi)}{x_{n+1}^{(k)}(z, \varphi) - U^{B(k)}(\varphi)} \right) \left( W_{n+1}^{(k)}(z, \varphi) - W_n^{(k)}(z, \varphi) - x_{n+1}^{(k)}(z, \varphi) \right) \right\}$$

$$+ \sum_{z' \in Z} \lambda(z'|z) \left( W_{n}^{(k)}(z', \varphi) - W_{n}^{(k)}(z, \varphi) \right)$$

$$+ \sum_{\varphi' \in \Phi} \lambda(\varphi'|\varphi) \left( W_{n}^{(k)}(z, \varphi') - W_{n}^{(k)}(z, \varphi) \right)$$

87
where $\Gamma^{B(k)} = c(\varphi) + r U^{B(k)}(\varphi) - \sum_{\varphi' \in \Phi} \lambda_{\varphi'(\varphi')} \left( U^{B(k)}(\varphi') - U^{B(k)}(\varphi) \right)$. Store the corresponding policy functions: \( \left\{ x_{n+1}^{(k)}(z, \varphi) : (n, z, \varphi) \in \mathcal{N} \times \mathcal{Z} \times \Phi \right\} \).

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

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

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

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

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

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

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

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

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: the idiosyncratic state $z$, and the aggregate state $\varphi$. Consider the idiosyncratic shock, for instance (the same structure applies to the aggregate shock). Firstly, the $k_z \times k_z$ infinitesimal generator matrix $\Lambda_z$ has the usual form, i.e. the elements of each row vector add up to zero:

$$\Lambda_z = \begin{pmatrix}
-\sum_{j \neq 1} \lambda_{1j} & \lambda_{12} & \cdots & \lambda_{1k_z} \\
\lambda_{21} & -\sum_{j \neq 2} \lambda_{2j} & \cdots & \lambda_{2k_z} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_{k_z1} & \lambda_{k_z2} & \cdots & -\sum_{j \neq k_z} \lambda_{k_zj}
\end{pmatrix}$$

where $\lambda_{ij}$ 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 a large number $k_z(k_z - 1)$ of 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).\(^{57}\) Formally:

$$d \log z_t = -\rho_z \log z_t dt + \sigma_z dB_t$$

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

---

\(^{57}\) For another example of a continuous-time search-and-matching model with shocks that uses Ornstein-Uhlenbeck processes, see Shimer (2005).
• Operationally, in the numerical version of the model in which time is partitioned and takes values in $\mathbb{T} = \{\Delta, 2\Delta, 3\Delta, \ldots\}$, we implement this process by using the Euler-Maruyama method, that is:

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

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

• The discrete-time process (C.2.1) is then estimated using the Tauchen (1986) with a discrete-state Markov chain that we define on the theoretical grid, $Z$. The outcome of this method are estimates for $(\rho_z, \sigma_z)$, and a transition probability matrix $\Pi_z = (\pi_{ij})$, where $\pi_{ij}$ 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:

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

when $\Delta > 0$ is small enough.

## 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:

$$U^B(\varphi) = \max_{\tilde{\varphi}(\varphi) \in \mathcal{X}} u^B(\tilde{\varphi}(\varphi); \varphi)$$

where $u^B(\tilde{\varphi}(\varphi); \varphi)$ is given by the Bellman equation:

$$u^B(\tilde{\varphi}(\varphi); \varphi) = -c(\varphi) \Delta + e^{-r\Delta} \left[ \eta(\theta(\tilde{\varphi}(\varphi); \varphi)) \Delta + o(\Delta) \right] \max \left\{ \tilde{\varphi}(\varphi), u^B(\tilde{\varphi}(\varphi); \varphi) \right\}$$

$$+ \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi'|\varphi) \Delta + o(\Delta) \right] u^B(\tilde{\varphi}(\varphi); \varphi')$$

---

58 Formally, in this step we are using the result that any CTMC with fixed transition rate matrix $\Lambda$ maps into a discrete-time chain with transition probability matrix $\Pi(t)$ at time $t \in \mathbb{Z}_+$ in which holding times between arrivals are independently and exponentially distributed. In particular, $\theta_t \Pi(t) = \Lambda \Pi(t)$, and thus $\Pi(t) = e^{\Lambda t}$ (when $\Pi_0 = I$). Hence, the probability of a $z_i$-to-$z_j$ transition after time $\Delta$ is given by $\frac{\lambda_{ij}}{\sum_{j \neq i} \lambda_{ij}} (1 - e^{-\sum_{j \neq i} \lambda_{ij} \Delta})$ 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 $(1 - \sum_{j \neq i} \lambda_{ij} \Delta)$, respectively.
\[ + \left[ 1 - \eta(\theta(\bar{\varphi}; \varphi)) \Delta - \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi'|\varphi) \Delta + o(\Delta) \right] u^B(\bar{\varphi}(\varphi); \varphi) + o(\Delta) \]

where \( o(\Delta) \) has the property \( \lim_{\Delta \to 0} \frac{o(\Delta)}{\Delta} = 0 \). The usual interpretation applies. 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 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 \( \omega \) need not incorporate a participation constraint.

To obtain equation (2), subtract \( e^{-r\Delta} u^B(\bar{\varphi}(\varphi); \varphi) \) from both sides, divide every term by \( \Delta \), 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

The value of the customer is:

\[
V^B(n, \omega; s) = (v(\varphi) - p)\Delta + e^{-r\Delta} \left\{ (\delta_f + \delta_c)\Delta + o(\Delta) \right\} U^B(\varphi) \\
+ \left[ (n - 1)\delta_c\Delta + o(\Delta) \right] \max \left\{ x'(n - 1; s), U^B(\varphi) \right\} \\
+ \left[ \eta \left( \theta(x'(n + 1; s); \varphi) \right) \Delta + o(\Delta) \right] \max \left\{ x'(n + 1; s), U^B(\varphi) \right\} \\
+ \sum_{z' \in \mathbb{Z}} \left[ \lambda_{z}(z'|z)\Delta + o(\Delta) \right] \max \left\{ x'(n; z', \varphi), U^B(\varphi') \right\} \\
+ \left[ 1 - \delta_f \Delta - n\delta_c\Delta - \eta \left( \theta(x'(n + 1; s); \varphi) \right) \Delta \\
- \sum_{z' \in \mathbb{Z}} \lambda_{z}(z'|z)\Delta - \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi'|\varphi)\Delta + o(\Delta) \right] \max \left\{ V^B(n, \omega; s), U^B(\varphi) \right\} \bigg] \\
+ o(\Delta) 
\]

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 \( e^{-r\Delta} V^B(n, \omega; s) \) from both sides, divide every term by \( \Delta \), and take the continuous-time limit as \( \Delta \to 0 \) to obtain equation (5).

### Incumbent Sellers

The typical seller must choose contract \( \omega = \{ p, x'(n'; s') \} \) to maximize:

\[
V^S(n, x; s) = \max_{\omega \in \Omega} \left\{ (np - C(n; s))\Delta + e^{-r\Delta} \left[ \delta_f \Delta + o(\Delta) \right] V^S_0(\varphi) \right\} 
\]
Entrant Sellers

After this, they pay the set-up cost \( \kappa(\varphi) > 0 \) allowing them to post a contract. Let \( v_0^S(s_0) \) denote the value of drawing productivity \( z_0 \), where \( s_0 := (z_0, \varphi) \). Then, the entrant chooses the initial contract \( x'(1; s_0) \in \mathcal{X} \), specifying the promised utility for the first customer of the firm, to maximize:

\[
v_0^S(s_0) = \max_{x'(1; s_0) \in \mathcal{X}} \left\{ e^{-\delta \Delta} \left[ \eta \left( \theta(x'(1; s_0); \varphi) \right) \Delta + o(\Delta) \right] \max \left\{ V^S(1, x'(1; s_0); s_0), V_0^S(\varphi) \right\} + \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi') \Delta + o(\Delta) \right\}
\]

subject to:

\[
V^B(n, \omega; s) \geq x
\]

To arrive at equation (6), we anticipate that \( V_0^S(\varphi) = 0, \forall t \), by free entry, and therefore \( V^S(n, x; s) \geq V_0^S(\varphi) = 0, \forall (n, x; s) \), without loss of generality. Then, we can proceed by taking the usual continuous-time limit, as described above.

Entrant Sellers

In the discrete-time approximation, potential entrants draw their initial productivity \( z_0 \) from \( \pi_z \). After this, they pay the set-up cost \( \kappa(\varphi) > 0 \) allowing them to post a contract. Let \( v_0^S(s_0) \) denote the value of drawing productivity \( z_0 \), where \( s_0 := (z_0, \varphi) \). Then, the entrant chooses the initial contract \( x'(1; s_0) \in \mathcal{X} \), specifying the promised utility for the first customer of the firm, to maximize:

\[
v_0^S(s_0) = \max_{x'(1; s_0) \in \mathcal{X}} \left\{ e^{-\delta \Delta} \left[ \eta \left( \theta(x'(1; s_0); \varphi) \right) \Delta + o(\Delta) \right] \max \left\{ V^S(1, x'(1; s_0); s_0), V_0^S(\varphi) \right\} + \sum_{\varphi' \in \Phi} \lambda_{\varphi}(\varphi') \Delta + o(\Delta) \right\}
\]

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 \( e^{-\delta \Delta} V_0^S(\varphi) \) 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 2.5

Kolmogorov Forward Equations

To derive the KFEs, we work with the equilibrium shares of agent types, defined by \( g_{n,t}(z) := \frac{S_{n,t}(z)}{S_t} \), for each \( n \in \mathbb{N} \) and \( z \), where \( S_t := \sum_{n \geq 1} \sum_z S_{n,t}(z) \) 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 \( \sum_{n \geq 1} \sum_z g_{n,t}(z) = 1 \), \( \forall t \geq 0 \), for each given \( \varphi \).

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

\[
g_{1,t+\Delta}(z) = \left[ \pi_z(z) \eta(\theta_{1,t+\Delta}(z, \varphi)) \Delta + o(\Delta) \right] \frac{S_{0,t}(\varphi)}{S_t} + 2 \left[ \delta_c \Delta + o(\Delta) \right] g_{2,t}(z) + \sum_{\tilde{\varphi} \neq \varphi} \left[ \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) \Delta + o(\Delta) \right] g_{1,t}(\tilde{\varphi})
\]

\[
+ \left[ 1 - \delta_f \Delta - \delta_c - \eta(\theta_{2,t+\Delta}(z, \varphi)) \Delta - \sum_{\tilde{\varphi} \neq \varphi} \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) \Delta + o(\Delta) \right] g_{1,t}(z)
\]

Similarly, for \( n \geq 2 \), we have:

\[
g_{n,t+\Delta}(z) = \left[ \eta(\theta_{n,t+\Delta}(z, \varphi)) \Delta + o(\Delta) \right] g_{n-1,t}(z) + (n + 1) \left[ \delta_c \Delta + o(\Delta) \right] g_{n+1,t}(z) + \sum_{\tilde{\varphi} \neq \varphi} \left[ \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) \Delta + o(\Delta) \right] g_{n,t}(\tilde{\varphi})
\]

\[
+ \left[ 1 - \delta_f \Delta - n\delta_c - \eta(\theta_{n+1,t+\Delta}(z, \varphi)) \Delta - \sum_{\tilde{\varphi} \neq \varphi} \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) \Delta + o(\Delta) \right] g_{n,t}(z)
\]

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 \( \Delta \) gives:

\[
\frac{g_{n,t+\Delta}(z) - g_{n,t}(z)}{\Delta} = \left[ \eta(\theta_{n,t+\Delta}(z, \varphi)) + \frac{o(\Delta)}{\Delta} \right] g_{n-1,t}(z) + (n + 1) \left[ \delta_c + \frac{o(\Delta)}{\Delta} \right] g_{n+1,t}(z) + \sum_{\tilde{\varphi} \neq \varphi} \left[ \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) + \frac{o(\Delta)}{\Delta} \right] g_{n,t}(\tilde{\varphi})
\]

\[
- \left[ \delta_f + n\delta_c + \eta(\theta_{n+1,t+\Delta}(z, \varphi)) + \sum_{\tilde{\varphi} \neq \varphi} \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) \right] g_{n,t}(z)
\]

Taking the limit as \( \Delta \to 0 \),

\[
\partial_t g_{n,t}(z) = \eta(\theta_{n,t}(z, \varphi)) g_{n-1,t}(z) + (n + 1) \delta_c g_{n+1,t}(z) + \sum_{\tilde{\varphi} \neq \varphi} \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) g_{n,t}(\tilde{\varphi}) - \left( \delta_f + n\delta_c + \eta(\theta_{n+1,t}(z, \varphi)) + \sum_{\tilde{\varphi} \neq \varphi} \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) \right) g_{n,t}(z)
\]

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

\[
\partial_t g_{1,t}(z) = \pi_z(z) \eta(\theta_{1,t}(z, \varphi)) \frac{S_{0,t}(\varphi)}{S_t} + 2 \delta_c g_{2,t}(z) + \sum_{\tilde{\varphi} \neq \varphi} \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) g_{1,t}(\tilde{\varphi}) - \left( \delta_f + \delta_c + \eta(\theta_{2,t}(z, \varphi)) + \sum_{\tilde{\varphi} \neq \varphi} \lambda_{\tilde{\varphi}}(z|\tilde{\varphi}) \right) g_{1,t}(z)
\]

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

\[
S_{0,t+\Delta}(\varphi) = \left[ \delta_f \Delta + o(\Delta) \right] S_t + \left[ \delta_c \Delta + o(\Delta) \right] \sum_z S_{1,t}(z) \\
+ \left[ 1 - \sum_{z_0} \pi_z(z_0) \eta(\theta_{1,t+\Delta}(z_0, \varphi)) \Delta + o(\Delta) \right] S_{0,t}(\varphi)
\]

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

\[
\partial_t S_{0,t}(\varphi) = \left( \delta_f + \delta_c \sum_z g_{1,t}(z) \right) S_t - \sum_{z_0} \pi_z(z_0) \eta(\theta_{1,t}(z_0, \varphi)) S_{0,t}(\varphi)
\]

In the stationary solution, inflows and outflows are equated for every pair of idiosyncratic states \((n, z)\), so \(\partial_t g_{n,t}(z) = 0\) and \(\partial_t S_{0,t}(\varphi) = 0\). 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, \(h_0(\varphi) := S_0(\varphi)/S\).

**Computing the Aggregate Stationary Measures of Agents**

To compute aggregates, we first use (17) to obtain the object \(b_n^A(z) := \frac{B_n^A(z)}{S}\) by:

\[
b_n^A(z) = n g_{n}(z)
\]

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

\[
S = \sum_{n=2}^{+\infty} \sum_{z \in Z} \theta_n(z, \varphi) g_{n-1}(z) = \sum_{n=2}^{+\infty} \sum_{z \in Z} B_n^I(z, \varphi) = 1 - B^I - \sum_{z \in Z} \theta_1(z, \varphi) S_0(\varphi)
\]

Using the definitions above, the last line can be written as

\[
S = \frac{1 - \left( b^A + h_0(\varphi) \sum_z \theta_1(z, \varphi) \right) S}{\sum_{n \geq 2} \sum_z \theta_n(z, \varphi) g_{n-1}(z)}
\]

Solving for \(S\), we obtain the stationary measure of active sellers:

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

Once the total measure of sellers is known, all other aggregate measures can be easily obtained. For instance, the mass of potential entrants is \(S_0(\varphi) = S h_0(\varphi)\), and the different measures of incumbent sellers can be obtained by \(S_n = S g_n\). The measure of active buyers is \(B^A = S b^A\), 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 \((z, \varphi)\) shocks, and let us introduce the object \(\sigma_n = S_n/(S_0 + S)\), defined for any \(n = 0, 1, 2, \ldots\) Then, when \(\delta_f = 0\), we can re-write the flow equations of Section 2.5 at steady state as:

93
\[ \eta(\theta_n) \sigma_{n-1} + (n + 1) \delta_c \sigma_{n+1} - (\eta(\theta_{n+1}) + n \delta_c) \sigma_n = 0 \]

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:

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

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

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

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

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

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 \( g_n = \sigma_n (1 + S_0/S) \) for \( n \geq 1 \), and \( S_0/S = \sigma_0/(1 - \sigma_0) \), we then have:

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

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

\[ \Pr[t \leq T \leq t + \Delta] = S_{t+\Delta}^t - S_t^t \]

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. \( \Pr[t \leq T \leq t + \Delta | T > t] \). Using Bayes’ rule, we can write the hazard function in terms of the survival function as follows: \( \Pr[t \leq T \leq t + \Delta | T > t] = 1 - S_{t+\Delta}^t / S_t^T \). The instantaneous hazard rate is then defined by the continuous-time limit, \( h_t := \lim_{\Delta \to 0} \frac{1}{\Delta} \left( 1 - S_{t+\Delta}^t / S_t^T \right) \). Using L’Hôpital’s rule:

\[ h_t = -\partial_t \log S_t^T \quad (D.3.1) \]

Hence, defining the cumulative hazard as \( H_t := \int_0^t h_s ds \), the cumulative hazard and the survival functions are related by \( S_T^t = \exp\{-H_t\} \) (as \( S_0^t = 1 - H_0 = 1 \)). Using this result, note that we can
write the discrete-time hazard function in terms of the instantaneous hazard as follows:

\[
\Pr[t < T \leq t + \Delta | T > t] = 1 - \exp\left\{- \int_{t}^{t+\Delta} h_s ds\right\}
\]  

(D.3.2)

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

\[
\mathbb{E}\{T\} = \int_0^{+\infty} S_T^T dt
\]  

(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 \((n_t, z_t) = (n, z) \in \mathbb{N} \times \mathbb{Z}\) 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:

\[
\Pr\left[ t < T_{n}(z, \varphi) \leq t + \Delta \bigg| T_{n}(z, \varphi) > t \right] = \eta(\theta_{n+1,t+\Delta}(z, \varphi)) \Delta + o(\Delta) + n(\delta_c \Delta + o(\Delta)) + \sum_{z \neq z} \lambda_z(z | z) \Delta + o(\Delta) + \sum_{\varphi \neq \varphi} \lambda_{\varphi}(\varphi | \varphi) \Delta + o(\Delta)
\]

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:

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

Note the absence of time subscripts in the above expression. This is a convenient implication of our block-recursive structure. 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):

\[
H_{n,t}(z, \varphi) = h_n(z, \varphi)t
\]  

(D.3.4)

The survival function, in turn, takes the simple form \( S_{n,t}^T(z, \varphi) = \exp\left\{ -h_n(z, \varphi)t \right\} \).

- Hazard rates are, however, time-dependent at higher levels of aggregation, because aggregate shocks generate slow-moving dynamics in the distribution of firms across states. The cross-sectional average hazard of price changes is equal to:

\[
H_t(\varphi) := \sum_{n \in \mathbb{N}} \sum_{z \in \mathbb{Z}} g_n(z) h_n(z, \varphi)
\]

where \( g_n(z) = S_{n,t}(z) / \sum_n \sum_z S_{n,t}(z) \) is the firm-size probability mass function (p.m.f.). 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:

$$f_n(z, \varphi) = 1 - \exp \left\{ - h_n(z, \varphi) \right\}$$  \hspace{1cm} (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 \(\varphi\). The average frequency of price adjustment in the cross-section of firms is:

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

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:

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

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

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

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 \(\hat{p} \equiv \log p\), we have:

$$\mathbb{E}_t \left\{ \left| \hat{p}_{n,t+\Delta}(z, \varphi) - \hat{p}_{n,t}(z, \varphi) \right| \right\} = \left( \eta(\theta_{n+1,t+\Delta}(z, \varphi) + o(\Delta)) \right) \left| \hat{p}_{n+1,t+\Delta}(z, \varphi) - \hat{p}_{n,t}(z, \varphi) \right|$$

$$+ n \left( \delta_c \Delta + o(\Delta) \right) \left| \hat{p}_{n-1,t+\Delta}(z, \varphi) - \hat{p}_{n,t}(z, \varphi) \right|$$

$$+ \sum_{\tilde{z} \neq z} \left( \lambda_z(\tilde{z}|z) \Delta + o(\Delta) \right) \left| \hat{p}_{n,t+\Delta}(\tilde{z}, \varphi) - \hat{p}_{n,t}(z, \varphi) \right|$$

$$+ \sum_{\tilde{\varphi} \neq \varphi} \left( \lambda_{\varphi}(\tilde{\varphi}|\varphi) \Delta + o(\Delta) \right) \left| \hat{p}_{n,t+\Delta}(z, \tilde{\varphi}) - \hat{p}_{n,t}(z, \varphi) \right|$$

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:

$$\mu_{n}^\Delta(z, \varphi) = \eta(\theta_{n+1}(z, \varphi)) \left| \hat{p}_{n+1}(z, \varphi) - \hat{p}_{n}(z, \varphi) \right| + n \delta_c \left| \hat{p}_{n}(z, \varphi) - \hat{p}_{n-1}(z, \varphi) \right|$$

$$+ \sum_{\tilde{z} \neq z} \lambda_z(\tilde{z}|z) \left| \hat{p}_{n}(\tilde{z}, \varphi) - \hat{p}_{n}(z, \varphi) \right| + \sum_{\tilde{\varphi} \neq \varphi} \lambda_{\varphi}(\tilde{\varphi}|\varphi) \left| \hat{p}_{n}(z, \tilde{\varphi}) - \hat{p}_{n}(z, \varphi) \right|$$

• The variance of the distribution of price changes is given by:
\[ \forall t \left\{ \left| \hat{p}_{n,t+\Delta}(z, \varphi) - \hat{p}_{n,t}(z, \varphi) \right| \right\} = \] 
\[ E_t \left\{ \left( \left| \hat{p}_{n,t+\Delta}(z, \varphi) - \hat{p}_{n,t}(z, \varphi) \right| - E_t \left\{ \left| \hat{p}_{n,t+\Delta}(z, \varphi) - \hat{p}_{n,t}(z, \varphi) \right| \right\} \right)^2 \right\} \]

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

\[ \sigma_n^\Delta(z, \varphi) = \eta(\theta_{n+1}(z, \varphi)) \left( \left| \hat{p}_{n+1}(z, \varphi) - \hat{p}_n(z, \varphi) \right| - \mu_n^\Delta(z, \varphi) \right)^2 \]
\[ + n \delta_z \left( \left| \hat{p}_n(z, \varphi) - \hat{p}_{n-1}(z, \varphi) \right| - \mu_n^\Delta(z, \varphi) \right)^2 \]
\[ + \sum_{\tilde{z} \neq z} \lambda_z(\tilde{z}|z) \left( \left| \hat{p}_n(\tilde{z}, \varphi) - \hat{p}_n(z, \varphi) \right| - \mu_n^\Delta(z, \varphi) \right)^2 \]
\[ + \sum_{\tilde{\varphi} \neq \varphi} \lambda_{\varphi}(\tilde{\varphi}|\varphi) \left( \left| \hat{p}_n(z, \tilde{\varphi}) - \hat{p}_n(z, \varphi) \right| - \mu_n^\Delta(z, \varphi) \right)^2 \]

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:

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

Figure A.9 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:

\[ M^\Delta_t(\varphi) := \sum_{n \in N} \sum_{z \in Z} r_{n,t}(z, \varphi) \mu_n^\Delta(z, \varphi) \quad \text{and} \quad \Sigma^\Delta_t(\varphi) := \sum_{n \in N} \sum_{z \in Z} r_{n,t}(z, \varphi) \sqrt{\sigma_n^\Delta(z, \varphi)} \]

respectively.