The Matching between Demand and Productivity

Jeremy Pearce[*] Liangjie Wu[†] [‡§]

March 20, 2025

Abstract

Aggregate productivity depends on how firm-level productivity matches with demand. We study this matching by developing a dynamic model of oligopolistic firms with heterogeneous productivity that grow their demand through marketing. Due to the oligopolistic competition, firms may market more due to high productivity or if they have accumulated significant demand from the past; this can drive path dependency and misallocation. Calibrating the model to firm-level price and sales data, we find the growth of residual demand drives the majority of firm size growth, and there is a sizable mismatch between demand and productivity. Quantitatively, policies that aim to correct static distortions due to markups can exacerbate dynamic misallocation between demand and productivity.

Key Words: Firm Dynamics, Productivity, Customer Capital, Market Concentration, Competition, Innovation.

JEL Code: O31, O32, O34, O41, D22, D43, L11, L13, L22

[*]Federal Reserve Bank of New York, jeremy.pearce@ny.frb.org;

[†]Einaudi Institute for Economics and Finance, liangjie.wu@eief.it;

[‡]Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[§]We thank Tincho Almuzara for helpful comments. The views expressed here are the authors' and do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

Aggregate productivity depends on how firm-level productivity matches with demand. Recent evidence suggests that demand and customer base rather is more central than cost productivity in driving firm size differences (Hottman et al., 2016; Foster et al., 2016; Einav et al., 2021; Afrouzi et al., 2023). This has implications for both aggregate productivity and the dynamic incentives of firms, as customer base is persistent and requires significant firm-level investment (Gourio et al., 2014). We study this interaction in oligopolistic markets, where firms have heterogeneous productivity, market power, and build demand through marketing. In this environment, we ask: what is the primary driver of market share, and how does this interact with marketing, aggregate productivity, and efficiency?

We answer this question empirically, theoretically, and quantitatively. We first focus on the empirical relationship between demand, productivity, and firm size. In the cross-section and over time, there is an imperfect correlation between productivity and market share, primarily due to variation driven by demand. We then introduce the dynamic customer base investment decision into an otherwise standard oligopolistic competition model (Atkeson and Burstein, 2008). Firms invest in customer capital to build out demand. Due to oligopolistic competition, firms market disproportionately more when larger, leading to path dependency. We quantify the model and find a sizable mismatch of demand and productivity, among which 21% is due to endogenous marketing decisions and 79% is due to the exogenous persistence of demand. Our model provides new positive and normative insights. On the positive side, the connection between productivity shocks at the firm level and aggregate productivity depends on the existing distribution of demand and the ability to build customer capital. On the normative side, a size-based policy that aims to correct the static distortions due to markups can distort the dynamic matching between demand and productivity.

Starting empirically, we establish the importance of demand in driving firm market share dispersion using detailed scanner data. We follow Hottman et al. (2016) by decomposing the observed firm size according to a nested constant elasticity of substitution demand system. Our measure of firm-level demand is the *residual* demand, which is the sales in excess of what is predicted by prices adjusted by the demand elasticity. By decomposing firm market shares into productivity, markup, and residual demand

components, we document that residual demand explains approximately four times as much as productivity differences in the cross-sectional variation and growth in firm size. Our variance decomposition reveals that residual demand accounts for 94% of market share variation, while productivity accounts for 19%, with productivity slightly more important for larger firms (27%).

Demand and productivity are imperfectly correlated at the firm level. In our empirical specification, we find a negative correlation of -0.26; this correlation disconnects productivity from market share. Highly productive firms do not capture market share proportional to their productivity, yet they do build demand in response to positive productivity shocks. We understand this through a structural VAR. We find that current residual demand depends on past productivity and past demand itself, with a coefficient of 0.20 and 0.70, respectively, while current productivity does not depend on past demand but solely on past productivity itself. However, firms that get positive productivity shocks must play "catch-up" by building demand.

We build a theory that connects market share at the firm level to the dynamic interaction of demand and productivity. The model takes the oligopolistic competition framework into a dynamic setting, where firms endogenously invest in their customer base. In the model, the productivity of firms fluctuates exogenously. Firms can invest in their customer base by paying a marketing cost. In addition to endogenous marketing, the customer base also has exogenous persistence. Two parameters, marketing cost and exogenous persistence, allow us to decompose the observed mismatch.

There are two core features of customer base that differ from productivity in our framework. First, while productivity changes are discrete and can be quite large, customer base takes time to build and moves in a continuous fashion. Another core feature of the customer base that differs from productivity is in the direct zero-sum element of business stealing. Motivated by the marketing literature, we assume that firms' investment in customer capital has direct spillovers to other firms depending on how rivalrous advertising is, governed by a single parameter of marketing rivalry. As long as advertising captures customers, this negative spillover generates costs for other firms. In our baseline model, we take a rivalry parameter such that marketing is purely reallocative, in that the level of customer base per se does not bring utility, but its matching with productivity is welfare-relevant. This leads to overall overinvestment in marketing, as it does not generate the positive gains productivity changes do. However, most of our

focus is on the allocative role of marketing.

Two competing economic forces emerge from the model. First, a complementarity between productivity and residual demand encourages efficient investments in customer capital. Second, a strategic incentive from variable markups sustains less productive firms with strong brand positions. This distorted incentive can create path dependency and long-run inefficiencies. A social planner who only observes market share and assumes this results from one-dimensional productivity can miss in terms of policy and exacerbate long-run inefficiencies by subsidizing inefficient firms. Further, a social planner focused only on static misallocation through variable markups may also induce inefficient marketing choices at the firm level. This distortion emerges from the interaction between customer capital and variable markups, but it speaks to more general distortions that emerge through increasing returns in customer acquisition to firm size.

This model is challenging to analyze, as firms are granular; firm-level shocks become aggregate shocks. Part of our contribution is to provide a tractable algorithm to compute a model using a quadratic approximation. We approximate the equilibrium profits from an oligopolistic competition game with quadratic functions that depend on the entire vector of firm-level demand and productivity in the market, with the criteria of (a) minimizing the distance between actual profits and approximated profits and (b) preserving the monotonicity of the approximated profits in the relevant domain. This approximation performs well, especially for large firms that matter disproportionally more. With the quadratic profit functions, our equilibrium becomes a quadratic dynamic game among firms. The results from game theory help us establish that the firm's value functions are also quadratic and their optimal marketing decisions are linear in the states. This result has two important outcomes for our setting. First, we now solve for a tractable equilibrium. Second, our model exactly mimics the AR(1) regressions of firm-level productivity that are popular in the firm dynamics literature.

We estimate the model to match the VAR results from the data and the dispersion of demand and productivity. Our model can replicate this VAR result almost exactly and offers a decomposition of the path dependency into exogenous and endogenous forces. We direct our attention to the efficiency implications of our quantified model. We utilize the method from Baqaee and Farhi (2019) to show that the imperfect correlation between demand and productivity can lead to significant aggregate welfare losses. How much of this misallocation is due to exogenous or endogenous forces leads to different policy

conclusions. We then turn to the dynamics and focus on the nature of size-dependent policies in the literature. Standard size-dependent policies correct for markups by subsidizing large firms with high markups. This policy can backfire in our world, where market share is jointly determined by productivity and endogenous demand. A subsidy to alleviate the static misallocation can lead to long-run welfare *losses* of 21%, whereas the efficient benchmark can improve welfare by 41%. The welfare losses come from policymakers subsidizing firms with low productivity and high customer capital (statically) and subsidizing wasteful advertising to generate size (dynamically).

Our quantitative model highlights the importance of independently understanding the determinants of firm market share as driven by demand and productivity. Most results in the literature (such as Boar and Midrigan, 2019 Berger et al., 2019, and Edmond et al., 2023) focus on markets where only productivity determines market share. These papers find that specific size-dependent policies (subsidizing large firms with high markups) restore efficiency in markets with market power. We recast the same set of policy instruments in the literature in our environment, where productivity and customer capital jointly determine market share and compare them to the planner's solution. In our environment, a size-dependent policy corrects static distortions due to markups but can exacerbate the dynamic mismatch between customer capital and productivity. This exacerbation is because firms may be significant due to customer accumulation, not productivity, and these subsidies promote more brand consolidation in those firms. This dynamic distortion can undo the welfare gains from correcting markups.

Our results point to several directions. Our paper provides an important message on the measurement of aggregate productivity. Aggregate productivity could move due to structural changes in the customer base or productivity shocks, and the nature of the transmission will have important implications for the persistence of productivity growth. Normatively, our paper points to demand formation as having important dynamic and policy implications. Firm-level investments may be suboptimal, and standard policies to eliminate markups may backfire. Understanding this interaction is essential, as it varies substantially by market. Policymakers need attunement to the joint distribution of demand and productivity to speak to specific policies when engaging in innovation or antitrust policies.

The remainder of this section reviews the literature, while the rest of the paper is structured as follows. Section 3 introduces the empirical framework for our study, the

decomposition of demand and productivity, and the empirical dynamics of both. Section 2 introduces the theoretical model with oligopolistic firms with heterogeneous productivity and customer capital. Section 4 estimates the model by uniting theory and empirics and studying the positive implications of our quantified model. Section 5 discusses policy implications and how the quantified model changes our understanding of size-based policies, advertising, and antitrust. We start with a review of the literature.

Related Literature. The core contribution of this paper is to connect the literature on productivity to customer capital in a dynamic setting. In doing so, we speak to two main strands of literature. First, we build on the growing literature of demand-driven firm size differences and the interaction with productivity. Second, we contribute to the literature on oligopolistic firm competition and bring it to a dynamic setting to study policies and dynamic misallocation. Third, our paper focusing on demand and productivity as two central intangibles of the firm contributes also to the study of intangible assets.

Aggregate productivity is classically a supply-side concept. There is a well-established literature on the origins of economic growth, heterogenous firm productivity, and the transmission from firm-level changes to the aggregate economy (e.g., Jovanovic, 1982, Hopenhayn, 1992, Aghion and Howitt, 1992, Klette and Kortum, 2004, Akcigit and Kerr, 2018). We build on this endogenous growth literature by connecting the firm-level productivity to changes in aggregate productivity through heterogeneity in firm's demand side. This heterogeneity can drive misallocation, which builds on a rich theoretical and empirical literature on factor misallocation and reallocation (Aghion et al., 2001, Haltiwanger et al., 2014, Acemoglu et al., 2018, Peters, 2020, and Liu et al., 2022). There is rising interest in this literature in how growth and firm heterogeneity interact with market power. Akcigit and Ates (2021, 2023) focus on the knowledge diffusion gaps between leaders and followers driving rising concentration and falling business dynamism. Large firms may also leverage their market share to increase markups.

While productivity remains a central literature in macroeconomics, there has been growing interest in the importance of demand in driving market share amongst economists (Dinlersoz and Yorukoglu, 2012; Gourio and Rudanko, 2014). This naturally connects to firms' choices of how to build demand, depending on their productivity and the nature of competition (Cavenaile and Roldan-Blanco, 2021; Cavenaile et al., 2022; Greenwood et al., 2021; Ignaszak and Sedlácek, 2022; Cavenaile et al., 2023). The interaction of

customer capital and productivity can generate endogenous customer misallocation, as firms' investment in customers depends on productivity and their markups (Bornstein and Peter, 2022; Afrouzi et al., 2023). We extend this concept of firms that hold demand and supply into a dynamic setting with oligopolistic firms. This generates new model features, such as path dependency of marketing and strategic interactions of large firms.

Hottman et al. (2016) study multi-product firms and find that the "appeal" of firms, or residual demand, explains the largest share of sales variation across firms. More recently, Eslava et al. (2024) found the same result when evaluating drivers of firm plant share. This demand takes time to build, as Argente et al. (2018, 2020) and Jaravel (2018) explore how product creation and destruction are ubiquitous in product markets. Argente et al. (2021) and Einav et al. (2021) document that product sales expansion is primarily due to an expanding customer base. Bhandari and McGrattan (2020) focus on this customer base or "sweat equity" as an endogenous asset. These channels can interact and may affect the measurement of productivity (Foster et al., 2008; Syverson, 2011) and misallocation (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2017; David and Venkateswaran, 2019). We study this relationship empirically and theoretically.

Our setting in this paper is oligopolistic competition among firms. We extend the insight from Atkeson and Burstein (2008) that variable markups may generate misallocation into a dynamic setting. This builds on a host of papers that focus on static misallocation and markups (Boar and Midrigan, 2019; Edmond et al., 2018; Mongey, 2021; Berger et al., 2019), which we also connect to the interaction between productivity and demand. The core mechanism is the role of strategic complementarities, which Amiti et al. (2019) find to be a strong force in international markets that drive variable markups. This setting connects to an extensive literature on markups and the negative and positive effects of concentration (such as De Loecker and Eeckhout, 2018; De Loecker et al., 2020; Gutiérrez and Philippon, 2017; Eggertsson et al., 2018; Hall, 2018; Autor et al., 2020; Akcigit and Ates, 2021). In extending the study of firm competition with markups to a dynamic setting, we are close to Peters (2020), who studies endogenous dynamic misallocation with limit pricing. We consider a more general framework of competition where multiple firms carry demand and productivity and compete for market share, with a particular focus on the role of endogenous demand.

Finally, we contribute to the literature on the importance of intangible assets at the firm level. Firm productivity and demand are two central intangible assets to firm

success and will serve a central role as the importance of intangibles continues to rise (Haskel and Westlake, 2017, Crouzet and Eberly, 2019, Olmstead-Rumsey, 2022, Syverson, 2019, De Ridder, 2024). Demand is a naturally central intangible asset. Bain (1956) classically noted that "(t)he advantage to established sellers accruing from buyer preferences for their products as opposed to potential-entrant products is on the average larger and more frequent in occurrence at large values than any other barrier to entry." Theoretically, brand value can generate persistent profits in markets with imperfect information (Schmalensee, 1978, Schmalensee, 1982, and Shapiro, 1983). The power of branding has been detailed empirically as consumer brand preferences are quite persistent (e.g., in Bronnenberg et al., 2009, 2012), making it an intangible asset at the center of firm value. This paper integrates it with the intangible of productivity empirically, theoretically, and quantitatively.

2 Empirical Model and Firm Size Decomposition

We start with a simple empirical model that can be informed by Nielsen Retail Measurement Services (RMS) scanner data on the determinants of firm market share and market concentration. We use this framework to decompose the main forces driving firm market share, primarily focusing on productivity and demand components. This section discusses the data construction and the drivers of cross-sectional and time-series dispersion in firm market share.

2.1 Model: Static Demand System

We start by presenting a framework of the decomposition of firm market share, which we will then take to the data. Following Hottman et al. (2016), we first decompose firms' revenues into four static components: demand differences (customer capital), production costs (productivity), scale of production (marginal cost), and markups. This static decomposition reveals which factors matter most for explaining cross-sectional differences in firm size. These components become crucial for understanding firms' dynamic incentives to invest in customer capital, which we study in Section 3.

Our frequency of analysis is yearly. We assume there is a representative household who spends 1 unit of expenditure on a measure 1 of product groups, indexed by $k \in$

$[0,1]$. Within each product group k, there are J_k firms, indexed by $j = 1, ..., J_k$. The real consumption, denoted by $\mathbf{C}_t$, is aggregated across product groups via a Cobb-Douglas aggregator:

$$\mathbf{C}_t = \exp\left(\int_0^1 \phi_k \log C_{kt} dk\right), \tag{1}$$

where $\phi_k > 0$ is the share of product groups in the representative consumption bundle. In each product group k, we assume that firm (j,k) faces a downward-sloping demand curve, with σ_k as the demand elasticity for each firm in group k. Firms use labor as the only input and the wage is normalized to 1,

$$C_{kt} = \left[\frac{\sum_{j=1}^{J} e^{\frac{b_{jkt}}{\sigma_k}} c_{jkt}^{\frac{\sigma_k - 1}{\sigma_k}}}{\left(\sum_{j=1}^{J} e^{b_{jkt}}\right)^{\theta}}\right]^{\frac{\sigma_k}{\sigma_k - 1}}, \ \sigma_k > 1. \tag{2}$$

The sales for each firm j emerges jointly from productivity a_{jkt}, where $c_{jkt} = e^{\frac{a_{jkt}}{\sigma_k}} l_{jkt}^{\omega}$, and customer capital b_{jkt}. The nature of b_{jkt} has a zero-sum element: as all firms get more customer capital, there is a business stealing effect, as represented by the denominator in equation (7). The degree of business stealing depends on the nature of advertising. θ governs the degree of this spillover. $\theta = 0$ is the world of no negative spillover: advertising enters the utility directly. $\theta = 1$ is the case of no net gain from advertising. Consumers only have one unit of attention, which is fixed and allocated across firms. These forces connect to the optimal pricing decision of firm (j,k), which is the following:

$$p_{jkt} = \mu_{jkt} \times \exp\left(\frac{1}{1-\sigma_k} a_{jkt}\right),$$

where μ_{jkt} is the markup and $\mu_{jkt} = \frac{\epsilon_{jkt}}{\epsilon_{jkt}-1}$ and a_{jkt} is the labor productivity. In addition to productivity, we assume firms also differ in their demand-side factors, which we refer to as customer capital. More specifically, we assume the real consumption at firm (j,k) is $b_{jkt}c_{jkt}$, where c_{jkt} is the quantity consumed and b_{jkt} is the customer capital.

This parsimonious model allows us to write the logarithm of firm sales as an additive function of the three factors (demand, productivity, and markups) as well as group-level factors. By taking a difference of the terms with the geometric averages, we write:

$$\Delta_k \log s_{jkt} = \Delta_k \log b_{jkt} + \omega(1-\epsilon_{jk})\Delta_k \log a_{jkt} - (1-\epsilon_{jk})\mu_{jkt}$$

The above equation decomposes log market share differences into customer capital differences, $\Delta_k \log b_{jkt}$, productivity differences scaled by demand elasticity, $(1-\epsilon_{jk})\Delta_k \log a_{jkt}$, and markup differences, $(1-\epsilon_{jk})\mu_{jkt}$, which negatively affect market share. σ is the elasticity of substitution across firms within a group and ω is the coefficient that governs diminishing returns at the firm level. We take these parameters from Hottman et al. (2016) in our empirical exercises.

One discrepancy in our exercises from Hottman et al. (2016) is that firms are multi-product, which requires a firm-level price aggregator. We follow Hottman et al. (2016) and assume products are aggregated through a CES. Further, firms face an upward supply curve. The mechanics of both of these are discussed in the Appendix as the specifics are not important for the economics discussed in this section.

2.2 Data Construction

With our empirical framework in hand, we turn to the data to construct each empirical object. The data enables us to separate the demand and supply components of firms' brand holdings (following Hottman et al., 2016). The data enables a breakdown at the product, brand, or firm level.

Our empirical analysis requires detailed data on prices, market shares, and firm characteristics over time to identify the joint evolution of customer capital and productivity. To enable the study of price and sales data, we employ detailed bar-code level data from Kilts-NielsenIQ Retail Measurement Services Data from the University of Chicago Booth School of Business. The data are large and comprehensive in the consumer product space. This dataset delivers significant coverage for products, brands, and firms, which we detail in Appendix B (see Argente et al., 2020 for more detail on this merge). We now turn to the summary statistics across firms in our sample in Table 1.

For each product group k, we observe sales and prices at the firm-year level. Following Hottman et al. (2016), we focus on price and market share variation within product groups to control for differences across categories. Table 1 displays summary statistics.

Our final sample includes 432 firms per product group on average, with substantial variation in firm size - the 90th percentile firm has sales of \$3.6 million compared to \$452,000 for the 10th percentile firm. The largest firm in a typical product group captures 31% market share, while the second largest firm captures 13%. Market share drops off

Table 1: Sample Statistics

	Mean	Median	Std. dev.	10th Pctile	90th Pctile	Max
Firms per Group	432	309	347	117	966	1740
Firm Sales ($1000)	1775	1072	3021	452	3569	54430
Top Firm Share	0.30	0.25	0.17	0.12	0.57	1.00
Top 3 Firm Share	0.51	0.48	0.18	0.27	0.76	1.00
Top 10 Firm Share	0.73	0.73	0.15	0.51	0.91	1.00

Notes: Weighted by product group-quarter sales. Firm and UPC sales in thousands.

quickly beyond the top firms - the 10th largest firm captures only 2% share.

The panel structure of the Nielsen data enables us to estimate the persistence of customer capital across the firm size distribution. Our framework exploits the continuous evolution of market shares to identify how brand advantages persist and accumulate over time. This is particularly important for understanding the dynamics of large firms, which maintain stable outsized leadership despite facing continual competition from more productive entrants.

2.3 Contemporaneous Interaction between Demand and Productivity

We combine our data and our framework. We ask about the correlation between the main objects of interest in our empirical framework. We then turn to firm market share decompositions to ask which force is most likely to drive market share. Finally, we focus on the nature of firm growth as a function of demand and productivity. We now turn to the correlations in Table 2.

Table 2: Correlations at the Firm Level

	Group-Adj. Sales	Demand	Productivity	Marginal Cost	Markup
Group-Adj. Sales	1				
Demand	0.669***	1			
Productivity	0.547***	-0.255***	1		
Marginal Cost	0.664***	-0.109***	0.989***	1	
Markup	0.247***	0.144***	0.169***	0.196***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The correlations in Table 2 reveal several important patterns. First, demand has the strongest correlation with sales (0.669), even higher than productivity (0.547). Second,

while both productivity and scale are highly correlated with sales, they are negatively correlated with customer capital, suggesting these advantages may substitute rather than complement each other. Third, markups show weaker correlations with all variables, indicating market power may play a smaller role in explaining sales variation.

While correlations show the relationships between components, they don't quantify each factor's contribution to overall market share variation. To measure these contributions precisely, we perform a β-variance decomposition where all terms sum to one, allowing us to interpret each coefficient as the share of variance explained. We introduce four components into this decomposition. As earlier, we have our appeal (or customer capital), which is residual demand for the firm conditional on price. We then also have the firm-level productivity z_j, and the firm-level cost scale, which comes from diminishing returns on the production function. Finally, we include the markups. This can be seen in Table 3.

Table 3: β-Variance Decompositions

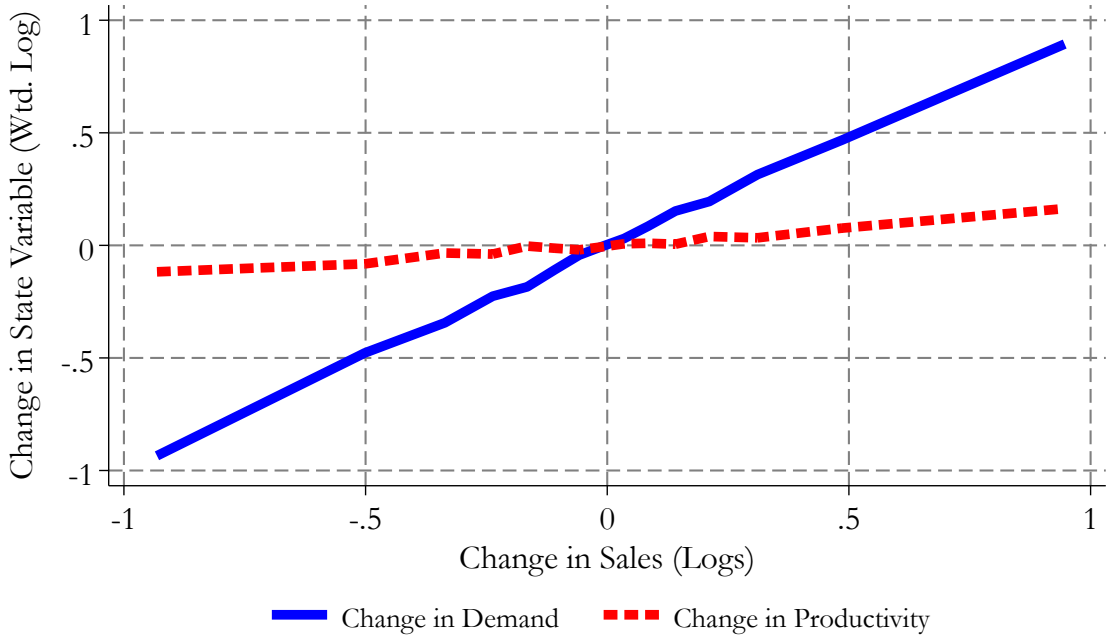
	Demand	Productivity	Marginal Cost	Markup
Group-Adj. Sales	0.942	0.191	-0.133	-0.001
	(0.002)	(0.001)	(0.000)	(0.000)
Top 10 Firms				
Group-Adj. Sales	0.920	0.273	-0.176	-0.020
	(0.016)	(0.013)	(0.004)	(0.000)

We see from the β-variance decomposition a couple of interesting results. First, appeal is again the most significant factor driving market share. Second, the productivity channel is also important. In the cross-section, a productivity alone explains 19% of market share. Overall, residual is about four-times as strong as productivity in explaining changes in market share in the cross-section. The decreasing returns to scale from the cost curve ("Cost DRS") makes it such that when appeal and productivity each increase, the cost effect will diminish their overall expansion in market share due to the convexity of the cost function ($\omega < 1$).

2.4 Dynamic Interaction between Demand and Productivity

We now look to Figure 1 to see how changes in firm market share can be attributed to changes in productivity and appeal. We see that the relationship looks quite symmetric.

Figure 1: The Evolution of Firm Size



Notes: Top two firms determined by average market share, logged and initialized at period 0. The blue line represents the leader and red line represents the follower. The dotted lines remove all net inflows/outflows from each firm contributing to market share. Shares are defined within each product group and weighted by product group size in sales. Source: USPTO/RMS NielsenIQ.

We now group our variables into four components: customer capital (or "total appeal"), productivity a which takes into account diminishing returns (ω), the "scale" which also takes into account diminishing returns, and markups. As noted earlier, the residual demand component is the most essential.

We now turn to the dynamics of demand and productivity more structurally. We estimate the joint evolution of productivity and customer capital using a dynamic panel framework, which has a similar structure to Heathcote et al. (2010). Our baseline specification consists of two equations that relate the dynamic evolution of demand and productivity:

$$a_{jt} = \rho_{aa} a_{jt-1} + \gamma_j^a + \nu_{ajt} \tag{3}$$

$$b_{jt} = \rho_{ba} a_{jt-1} + \rho_{bb} b_{jt-1} + \gamma_j^b + \nu_{bjt} \tag{4}$$

where a_{jt} represents log productivity and b_{jt} represents log customer capital for firm j at time t. The terms γ_j and ξ_j capture firm-specific fixed effects, while ν_{ajt} and ν_{bjt} represent idiosyncratic shocks.

We estimate these regressions employing Arellano and Bond (1991) where we use the second lag as an instrument for the first lag, discussed in detail in Appendix D. We now present two main sets of regressions that focus on the dynamic evolution of demand and cost productivity. Table 4 reports the evolution of firm-level productivity and demand.

Table 4: Dynamic Panel Estimations

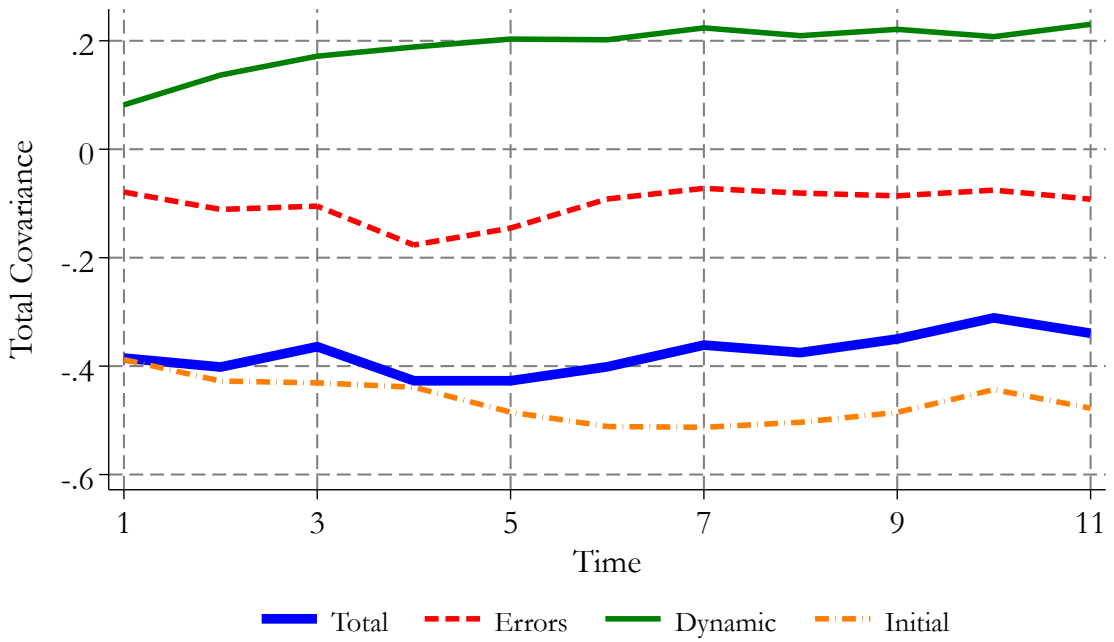
	Panel A: Arellano-Bond GMM				Panel B: IV	
	(1) a_t	(2) a_t	(3) b_t	(4) b_t	(5) a_t	(6) b_t
$a_t - 1$	0.846***	0.852***		0.185***	0.589***	0.144**
	(0.032)	(0.032)		(0.047)	(0.042)	(0.070)
$b_t - 1$		0.007	0.927***	0.976***		0.746***
		(0.015)	(0.037)	(0.039)		(0.032)
Observations	10920	10920	10920	10920	10010	10010
AR(1) p-value	0.000	0.000	0.000	0.000		
AR(2) p-value	0.041	0.042	0.368	0.487		
Hansen p-value	0.000	0.000	0.000	0.000		
F-statistic					194.52	270.64
Weak ID F-stat					224.02	109.80

There are a few key takeaways from Table 4. Cost productivity is fairly responsive to itself (as seen in columns 1 and 2), but is unresponsive to demand. This matches our theoretical framework where cost productivity has an immediate impact on demand through marketing, but there is no immediate feedback.[1] One important characteristic of our model is that productivity can lead to shifts in demand but demand does not directly shift firm productivity. We also show demand responds to both cost productivity and demand, consistent with our model of firm advertising and persistent demand as customer capital. Together our identification will rely on column 1 column 4 of Table 4. Both of these statistical models passed the standard tests for Arellano-Bond panel dynamics, the AR(1) test (low p-value), AR(2) test (high p-value), and the Hansen test (high p-value). We now turn to studying the dynamics in greater detail.

[1]There may indeed be dynamic feedback between customer capital and endogenous innovation. We leave this for later work.

Drivers of the Dynamic Covariance. In this exercise, we take the correlation at the beginning of the sample and track the correlation of productivity and residual demand over time. In the maroon line, we residualize the shocks that hit a and b, which are highly correlated. This can be seen in Figure 2.

Figure 2: Covariance Decomposition



Notes: This figure disaggregates the change in welfare (yellow) due to change in customer capital (blue) productivity (dashed red) and their interaction (dotted green). The decomposition is available in equation (15). Authors' calculations.

We see that the steady state point reached is a strongly positive correlation between productivity and residual demand, which comes from the positive dynamic relationship between the two. In the two flat lines, we plot the theoretical correlation under the two cases. We find that the overall fit matches the distribution in the data.

The importance of customer capital in both the cross section and growth of firms illustrates a fundamental challenge for market efficiency. While standard models predict market share should flow to the most productive firms, our evidence suggests customer capital may be equally or more important in determining firm success. This raises crucial questions about the dynamics of customer capital accumulation. In particular, how customer capital and productivity interact across firms will have important implications for the firm size distribution, firm-level incentives, and aggregate output. we turn to this

next in a dynamic model.

2.5 Endogenous Demand

This section has presented evidence that demand is an important driver of market share along with productivity. We now turn to the dynamics of endogenous demand, where firms build customer capital through the introduction of new products.

We focus on three channels by which demand changes. First, firms introduce new products within the same group to boost attention for their basket of goods. Second, existing products exhibit churn as a result of market dynamics. Third, products exit the market. We ask how these three processes contribute to change in sales and change in overall customer capital in this section. For sales, we are interested in the evolution of firm sales from these three terms,

$$\Delta \ln \text{Sales}_{jt} = \underbrace{2\frac{\text{Cont}_t - \text{Cont}_{t-1}}{\text{Sales}_t + \text{Sales}_{t-1}}}_{\text{Continuing products}} + \underbrace{2\frac{\text{Entry Sales}}{\text{Sales}_t + \text{Sales}_{t-1}}}_{\text{Entry}} \underbrace{-2\frac{\text{Exit Sales}}{\text{Sales}_t + \text{Sales}_{t-1}}}_{\text{Exit}} \tag{5}$$

This delivers the sales decomposition, which tells us how the activity of each margin affects overall sales at the firm level. We also present the decomposition of appeal or residual demand, which has a slightly different structure. This is presented in Equation (6),

$$\Delta \ln b_{jkt} = \frac{N_{cont}}{N_{jkt}} \cdot (\ln b_{cont,t} - \ln b_{cont,t-1}) + \frac{N_{entry}}{N_{jkt}} \cdot \ln b_{entry} - \frac{N_{exit}}{N_{jk,t-1}} \cdot \ln b_{exit}. \tag{6}$$

We now turn to the decomposition of sales and appeal growth at the firm level in Table 5.

The decomposition of sales and appeal suggest some important margins. First, when it comes to new product entry, its role for expanding firm appeal is central, in both level and in how it varies with total appeal. New products explain approximately 42% of the variation in aggregate appeal. Theoretically, this aligns with the endogenous force as firms introduce new products in order to maintain high appeal. This is consistent with Argente et al. (2020), who find constant product entry is a central driver to firm's maintaining appeal.

Table 5: Beta Decomposition of Changes in Firm-Level Sales and Appeal: Top 10 Firms by Sales (2007-2017)

Component	Beta Coefficient	Std. Error	Mean Level
	Panel A: Sales Decomposition		
Continuing products	0.739	0.005	-0.061
New product entry	0.202	0.004	0.059
Product exit	0.059	0.002	-0.008
Total	—	—	-0.010
	Panel B: Appeal Decomposition		
Continuing products	0.290	0.004	-0.069
New product entry	0.421	0.007	0.226
Product exit	0.289	0.007	-0.127

Notes: This table presents the beta decomposition of changes in firm sales and appeal into three components: continuing products, new product entry, and product exit. The sample is restricted to the top 10 firms by sales within each product group for the period 2007-2017. Beta coefficients represent each component's contribution to the total variance of changes. All estimates are weighted by product group sales.

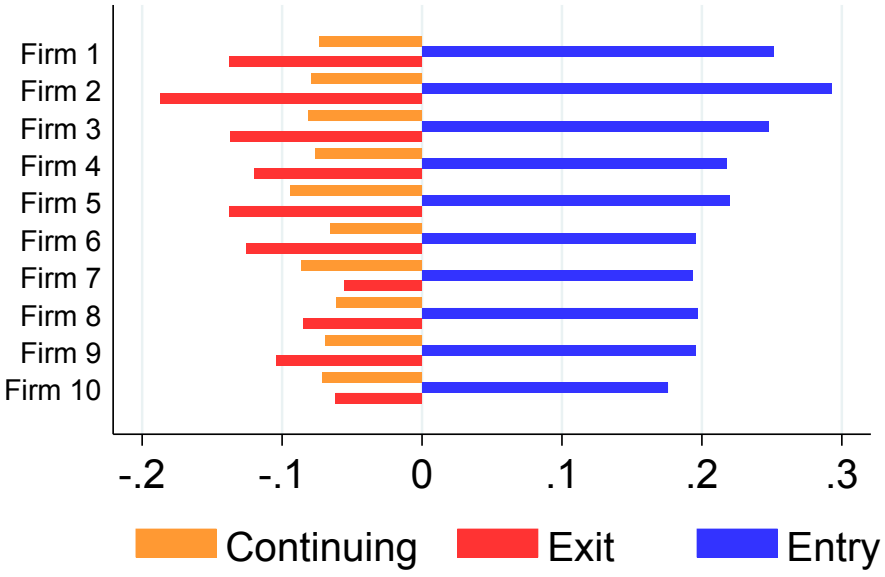
Heterogeneous Appeal by Firm Size. Large firms tend to be large due to their residual demand, as we've shown here and as noted in the literature. Figure 3 details the changes in firm appeal coming from continuing products, exiting products, and entering products. We document this by firm rank within a product group, where the group is weighted by total average sales in the group.

Discussion of Empirical Results. Our empirical exercises extend standard workhorse empirical models for measuring misallocation and productivity. We find that demand differences are indeed important drivers of firm productivity, and they are persistent through the introduction of new products with high appeal. However, from this exercise, it is not clear the endogenous versus exogenous drivers of this misallocation. This question is central for policy analysis and normative understanding of the drivers of market share and aggregate productivity. We turn to this next through the lens of the model where we focus on the dynamic manifestation of the static distortions.

3 A Model of Endogenous Demand

We set up an equilibrium model of customer capital investment with oligopolistic firms. To highlight the new mechanisms of the model, we start by presenting the environment with one single market and introduce other general equilibrium elements and aggrega-

Figure 3: Drivers of Residual Demand



Notes: This figure disaggregates the change in appeal into continuing products (orange), exiting products (red) and entering products (blue). The decomposition is available in equation (6). Authors' calculations.

tion at the end of this section.

In this single market, a finite number of firms engage in Cournot competition, internalizing their impacts on the market, as in Atkeson and Burstein (2008). Dynamically, they can invest in customer capital by costly marketing activities, while their productivities fluctuates over time.

3.1 Dynamic Environment

We now extend the demand model from Section 2.1 into a dynamic setting, by explicitly modelling the marketing decisions and accumulation of customer capital. In our baseline model, we assume product groups (k) are homogeneous for ease of exposition. This extension can be performed using the same analytical tool we discussed below. We start with households and then extend the discussion to firms, pricing, and customer capital dynamics. Our model is written in continuous time.

Households. A representative household endogenously supplies L_t units of labor and consumes products produced by J firms. These J firms differ in both their customer capital and productivity. More specifically, we model the customer capital as the appeal b_{jt} and labor productivity as a_{jt} discussed below. The products from firms are imperfect substitutes across firms within group k, with σ as the substitution elasticity.

$$C_t = \left[\frac{\sum_{j=1}^{J} e^{b_{jt}/\sigma} c_{jt}^{\frac{\sigma-1}{\sigma}}}{\left(\sum_{j=1}^{J} e^{b_{jt}}\right)^{\theta}} \right]^{\frac{\sigma}{\sigma-1}}, \sigma > 1. \tag{7}$$

In equation (7), C_t is the real consumption from this market, and c_{jt} is the consumption from firm j. Everything else in this demand system is standard, other than the inclusion of a term that models the rivalry of customer capital among firms. More precisely, we assume the real consumption is scaled by the term $\left(\sum_{j=1}^{J} e^{b_{jt}}\right)^{\theta}$, where $\theta \in [0,1]$ is the rivalry in marketing. This parameter allows the model to span between two alternative theories of marketing. When $\theta = 0$, we can treat customer capital as another form of productivity; When $\theta = 0$, marketing per se does not create utility but a reallocative activity. Building on the theory of fixed customer attention, we assume $\theta = 1$ and normalize the total customer capital in each product group to be $\sum_{j \in J} e^{b_{jt}} = 1$. This normalization takes the a zero-sum view marketing in the sense that marketing at the firm level does not directly create utility.[2] The change of customer capital across firms reallocates customer attention; the efficiency of this churn depends on firm productivity. As a result of this process, customer capital and productivity at the firm level are complementary. Due to limited attention of consumers, the economy faces a sorting problem between customer capital and productivity. When one firm gains customer capital another firm must lose customer capital, which makes the joint allocation of productivity and customer capital crucial for efficiency and aggregate productivity. At the heart of this model is the dynamics of this process.

Evolution of Demand and Productivity. In addition to their customer capital, firms differ in labor productivity. We assume firm j produces using a production function $c_{jt} = \exp\left(\frac{a_{jt}}{\sigma-1}\right) l_{jt}$, where l_{jt} is the labor input.

[2] In Section 4.3, we relax this assumption and study the welfare implications. This decision is irrelevant for firm decisions but relevant for welfare.

Both customer capital and firm productivity evolve over time. Our model assumes that firms endogenously invest in demand while the productivity of firms fluctuates exogenously. We discuss the validity of this assumption in Section 4. Our goal of this quantitative model is to mimic closely the empirical framework. To make the model flexible and tractable, we assume that the productivity of firm j evolves according to the following continuous Markov process. Firm productivity can take value from the N different values: $a_{jt} \in \{\lambda_n\}_{n=1}^{N}$ where $\lambda_n < \lambda_{n+1}$ for any n. All firms start from the lowest possible productivity λ_1 and can switch between two adjacent productivity levels with a Poisson arrival rate α. This jump process closely mimics the AR(1) process we used in the empirical analysis. The productivity switching rate α governs the persistence of the productivity process and the values of λ_n govern its volatility.

The customer capital of firm j evolves both endogenously and exogenously. Endogenously, firm j can grow its customer capital with marketing investment η_{jt}. Exogenously, the customer capital depreciates, which we model as the mean-reversion rate $\rho_b > 0$. We emphasize that in the primitives, our model does not impose any dependence of customer capital on productivity. Instead, we show the dependence arises endogenously through firms' optimization. Mathematically, the law of motion for b_{jt} is

$$\dot{b}_{jt} = \eta_{jt} - \rho_b b_{jt},\ b_{j0} = 0. \tag{8}$$

Marketing investment is costly. We model this as a quadratic cost in advertising, $\gamma\frac{\eta_{jt}^2}{2}$.

Oligopolistic Competition. The structure of competition has important implications for brand reallocation across firms. We assume that the duopolists take as given the aggregate price index across all product groups, **P**, yet are large enough to internalize their impact on their own product group (as in Atkeson and Burstein, 2008).[3] As a result, they charge a variable markup in the equilibrium, which is a function of their market shares, $\mu(s_j) = \frac{\epsilon(s_j)}{\epsilon(s_j)-1}$, and thus the ability of consumers to substitute away from the firm, or perceived elasticity, $\epsilon(s)$. We assume firms engage in Cournot competition, which implies that the perceived elasticity, as a function of their market share s, is $\epsilon(s) = \left(\frac{1}{\sigma}(1-s)+s\right)^{-1}$.

[3]This can be micro-founded as a multi-product firm in a given category, as in Pearce and Wu (2022). This captures the fact that firms in given product groups provide many products and brands within the same group.

3.2 Characterization

Firms make static pricing decisions and benefit from more productivity and customer capital due to two forces. First, more composite productivity leads to higher market share and a higher level of profits. Second, due to variable markups, firms can extract more profit per unit when market share is higher. In the dynamics, this informs their marketing decisions and the value of their joint customer capital and productivity. As a result, we proceed in this fashion to characterize the equilibrium.

First, we derive the outcomes from the static pricing equilibrium and write the equilibrium market shares and profits as functions of the distribution of productivity and customer capital. Second, we embed the static profits into the dynamic setting to define a symmetric Markov Perfect Equilibrium. Characterizing this equilibrium involves solving a fixed point problem in a functional space. Our last step is to provide an analytical solution to an approximated game. This analytical solution provides tractability and an intuitive link between firm choices and aggregate variables, which we bring to the quantitative analysis.

Static Pricing Equilibrium. As firms engage in oligopolistic quantity competition, their equilibrium markups are given by the formula: $p_{jt} = \mu_{jt} \exp\left(\frac{a_{jt}}{\sigma-1}\right)$, where the markup $\mu_{jt} = \frac{\sigma}{\sigma-1}\frac{1}{1-s_{jt}}$, where s_{jt} is the equilibrium market share for the firm. Coupled with the demand curves from the CES preference, the vector of market shares $(s_{jt})_{j=1}^{J_{k(j)}}$ is the solution to:

$$s_{jt} = \frac{\exp(a_{jt} + b_{jt})(1 - s_{jt})^{\sigma-1}}{\sum_{\ell=1}^{J} \exp(a_{\ell t} + b_{\ell t})(1 - s_{\ell t})^{\sigma-1}}. \tag{9}$$

Given the market shares, firm j's profit, π_{jt} is given by:

$$\pi_{jt} = \frac{1}{\sigma} s_{jt} + \frac{\sigma - 1}{\sigma} s_{jt}^2.$$

There are a few observations that are useful to state here. First, we note that the system is homogeneous of degree zero in $(a_{jt}, b_{jt})_{j=1}^{J}$. This homogeneity also applies to profits. Economically, only the *relative* productivity and customer capital matter for profits. Noting this property, we instead write the shares and profits in terms of the relative productivity a_{jt} and the relative customer capital b_{jt}: Secondly, only the sum of the demand and supply factors matters for the static profits. More precisely, if we define

$z_{jt} \equiv a_{jt} + b_{jt}$, we can write the share equations as

$$s_{jt} = \frac{\exp(z_{jt})(1 - s_{jt})^{\sigma-1}}{\sum_{\ell=1}^{J} \exp(z_{\ell t})(1 - s_{\ell t})^{\sigma-1}}.$$

Economically, whether market shares come from demand-side or supply-side factors of firms does not matter for profits. We will refer to z_{jt} as the composite productivity of firm j.

We use $\boldsymbol{\pi} : \mathbb{R}^J \mapsto \mathbb{R}^J_{++}$ to denote the mapping from the vector of compounded productivity $\mathbf{z}_t = (z_{jt})_{j=1}^J$ to the profits $\boldsymbol{\pi}_t \equiv (\pi_{jt})_{j=1}^J$. We use $\pi_j(\mathbf{z})$ to denote firm j's profit. A similar mapping could be defined for the market shares $\mathbf{s} : \mathbb{R} \mapsto \Delta^J$, where Δ^J is the J-dimensional simplex. With this notation, we now define and characterize the dynamic marketing equilibrium.

Dynamic Marketing Equilibrium. We are focused on firms of sufficient size that impact the overall product group in which they operate. In this case, the proper equilibrium concept we focus on is the Markov Perfect Equilibrium. As discussed in the static pricing equilibrium, the static profits only depend on the composite productivity $\mathbf{z}$. Dynamically, due to the differential dynamics of productivity and customer capital, the full payoff relevant state of firms must separate a and b. For notational simplicity, we denote $x_j \equiv (a_j, b_j)$. The vector notation is $\mathbf{x} = (x_j)_{j=1}^J$.

To prepare the notation for the equilibrium, we denote $\eta_j(\mathbf{x})$ as the marketing strategy of firm j. That is, given firm j's state x and the collection of its competitors' states $\mathbf{x}_-$, η_j is the optimal marketing strategy of firm j. In a Markov perfect equilibrium, firm j takes as given the strategies of other firms and maximizes:

$$V_j(\mathbf{x}) \equiv \max_{\eta_j(\mathbf{x})} \mathbb{E} \int_0^\infty e^{-rt} \left(\pi_j(\mathbf{z}_t) - \gamma \frac{\eta_j(\mathbf{x}_t)^2}{2} \right) dt \tag{10}$$

s.t.

$$\text{equation (8) holds for all } j.$$

There are three features of problem (10) that we want to elaborate. First, we impose that all firms follow a Markov strategy: marketing investment $\eta_j(\mathbf{x})$ is only a function of the current payoff-relevant state $\mathbf{x}_t$ instead of the full history. Secondly, the firm form ex-

pectation with respect to both its productivity process and its competitors' productivity, as described in the earlier section. Lastly, the firm also internalizes the law of motion of customer capital for all firms in the group. Using the definition of firms' problem, we now formally define the marketing equilibrium as follows:

Definition 1 *A marketing equilibrium in a given product group is marketing investment* $(\eta_j(\mathbf{x}))_{j=1}^J$, *where the* $\eta_j(\mathbf{x})$ *solves equation* (10) *for each firm* j, *given the other firms' strategies* $(\eta_\ell(\mathbf{x}))_{\ell \neq j}$.

Fully analyzing a marketing equilibrium is a challenging task for two reasons. First, due to the granular features of the firms, each individual firm's productivity shock becomes an aggregate shock to the product group. The relevant state vector thus becomes the entire J-dimensional vector. Given our empirics, a representative group has 10 firms. It becomes impossible to solve this problem directly. Second, even without the dimensionality issue, characterizing the equilibrium involves finding a fixed point of firms' strategies, which further complicates the analysis.

Approximation of Profits. To gain tractability, we consider a quadratic approximation of profit functions. After this approximation, our marketing equilibrium becomes the Markov perfect equilibrium among J firms with quadratic payoffs. This is an equilibrium we can solve easily since all firms' equilibrium strategy become linear in its customer capital. We detail this approximation and offer also the interpretations.

Our first step of this approximation involves linearizing the equation system for the market shares of firms, from equation (9). This system of J equation implicitly define the market shares as a function of $(\mathbf{a}, \mathbf{b})$. We linearized this equation around the point $(\mathbf{a}, \mathbf{0})$ using the implicit function theorem. This point is the state where firms have productivity differences but have identical customer capital. Denoting the market share at point $(\mathbf{a}, \mathbf{0})$ as $\bar{s}(\mathbf{a})_{j=1}^J$, we obtain an explicit formula for the approximated market shares:

$$\hat{s}_j(\mathbf{a}, \mathbf{b}) = \bar{s}_j(\mathbf{a}) + \zeta_j(\mathbf{a}) \left(b_j - \sum_\ell \bar{s}_\ell(\mathbf{a}) b_\ell \right), \tag{11}$$

where $\zeta_j(\mathbf{a}) = \frac{\bar{s}_j(\mathbf{a})(1-\bar{s}_j(\mathbf{a}))}{1+\bar{s}_j(\mathbf{a})(\sigma-2)}$. Equation (12) is a linear approximation of firms' market shares, where the approximated market share is written as the linear combination of the equal-demand market share $\bar{s}_j(\mathbf{a})$ and the gap between the customer capital of firm j and the weighted average of customer capitals in the group, where the weight is the

equal-demand market shares. How responsive are the market shares to customer capital is governed by $\zeta_j(\mathbf{b})$. There are two mechanisms in equation (9) that are captured by $\zeta_j(\mathbf{b})$. First, a firm with equal-demand market share is more responsive to the increase in customer capital. This is the complementarity force in the pricing equilibrium. Second, $\zeta_j(\mathbf{a}) < \bar{s}_j(\mathbf{a})$, reflecting the impacts of variable markups in determining the market shares. More precisely, if we consider a monopolisitically competitive equilibrium where firms' markups are constant, the counterpart of equation (12) is

$$s_j^*(\mathbf{a},\mathbf{b}) = \bar{s}_j^*(\mathbf{a}) + \bar{s}_j^*(\mathbf{a})\left(b_j - \sum_\ell \bar{s}_\ell(\mathbf{a}) b_\ell\right). \tag{12}$$

Due to firms' markup inventives, we can show that $\bar{s}_j(\mathbf{a}) < s_j^*(\mathbf{a})$, which implies that an increase in the relative customer capital always lead to a smaller increase in the approximated market share in an oligopolistic equilibrium than the constant markup equilibrium.

We now input the approximated market shares directly in the firms' profits as a function of market shares:

$$\hat{\pi}(\mathbf{a},\mathbf{b}) = \frac{1}{\sigma}\hat{s}(\mathbf{a},\mathbf{b}) + \frac{\sigma-1}{\sigma}\hat{s}(\mathbf{a},\mathbf{b})^2. \tag{13}$$

Since the approximated shares are linear in the customer capital and the profits are quadratic in shares, the approximated profits of firms become quadratic in customer capital. To highlight the role of variable markups, we again consider the case of monopolistic competition. In the monopolistic competition case, we instead have an approximated profit that is linear in customer capital,

$$\pi^*(\mathbf{a},\mathbf{b}) = \frac{1}{\sigma}\hat{s}^*(\mathbf{a},\mathbf{b}). \tag{14}$$

The difference between equation (13) and equation (14) highlights the second impact of oligopolistic competition. With constant markups, firms only expand their profits through the expansion of market shares, while under variable markups, firms can also expand their profits through markups. As a result, firm profits with variable markups are convex in market share. The marginal benefit of increasing firm size increases with current size. This is the sole source of nonlinearity in our approximated model. We later

show that this nonlinearity manifests into a path dependence of marketing investment: keeping productivities identical between two firms in the same product group, the firm with higher customer capital also invests more in their customer capital.

Equilibrium Marketing Investment. The approximated profits are quadratic forms in the productivity and customer capital gap, which converts our dynamic game into a quadratic game. We use the established results from the quadratic games literature to obtain analytical solutions to our equilibrium. The quadratic approximation of static profits lead to two simplifications. First, the value function takes a quadratic form in $\mathbf{x}$ and the equilibrium marketing strategy becomes linear in $\mathbf{x}$. Firms do not need to track the distribution of productivity and customer capital of other firms. This leads us to Lemma 1:

Lemma 1 *In a marketing equilibrium with approximated profits and variable markups, the equilibrium marketing investment takes the form:*

$$\eta_j(\mathbf{a}, \mathbf{b}) = \eta^0(\mathbf{a}) + \eta_j^1(\mathbf{a}) b_j + \sum_k \eta_{jk}^2(\mathbf{a}) b_k.$$

In the case of constant markup

$$\eta_j(\mathbf{x}) = \eta_j^0(\mathbf{a}).$$

This simple equilibrium admits rich features on the interaction between marketing and productivity. A central component of this simple equilibrium is the emergence of path dependency. Firms with high demand may reinforce their high demand through marketing, rather than purely demand tracking productivity with a lag, which is the classical story of customer capital (Gourio and Rudanko, 2014).

3.3 Path Dependency of Market Shares

A key insight from our model is that endogenous marketing can lead to misallocation between demand and productivity. To highlight this mechanism, we focus on the case where the firm productivity can only take value in $\{0, \lambda\}$. As shown in Lemma 1, firms' marketing decisions respond to their productivity and customer capital. For its response to productivity, there are only two strategies of marketing at each moment given two

productivity levels, we denote them as $\{\eta_n^a, \eta_n^b\}_{n=h,l}$. We now perform two pairs of firms. First, consider the case of two firms with identical customer capital, but differ in productivity. From Lemma 1, we have shown $\eta_h^a > \eta_l^a$ and $\eta_h^b > \eta_l^b$. Thus, the more productive firm invests more in their customer capital, holding their customer capital constant. We refer to this as the *productive incentive*. In the second comparison, we take two firms with identical productivity, but differ in their customer capital. We have shown that $\eta_n^b > 0$, for $n = h, l$. Thus, the firm with more customer capital invests more, holding productivity constant. This is the path dependency we emphasize. Mathematically, this path dependency comes from the quadratic terms in firm profits. Economically, this path dependency reflects the variable markups. A firm with a larger size has a larger benefit in investing in their customer capital due to a disproportionate increase in markup. We refer to this as the *markup incentive*.

The markup incentive is a novel mechanism when we bring the oligopolistic competition into a dynamic setting. When the productivity incentive is not large enough compared to the markup incentive, the unproductive firm might invest relatively more in their customer capital, leading to a persistent misallocation between customer capital and productivity. We highlight this possibility using a numerical example. In this example, we consider a case with two firms. The first firm has a high productivity, while the second firm has a low productivity. Keeping these productivities fixed, we want to trace out the evolution of the equilibrium after $t = 0$. Solving the path of customer capital, we reach the following analytical form:

$$b_{1t} = b_{10}e^{-(\rho_b - \eta_h^b)t} + \frac{\eta_h^a}{\rho_b - \eta_h^b}\left(1 - e^{-(\rho_b - \eta_h^b)t}\right)$$

$$b_{2t} = b_{20}e^{-(\rho_b - \eta_l^b)t} + \frac{\eta_l^a}{\rho_b - \eta_l^b}\left(1 - e^{-(\rho_b - \eta_l^b)t}\right)$$

To highlight the importance of path dependence, we further assume that $\rho_b = 0$. In this case, the difference in the initial customer capital between the two firms has a permanent impact. Observing from the data of firm size and firm size growth, the model predicts periods where the unproductive firm keeps growing at a faster rate and gains additional

market share. More precisely:

$$b_{1t} = b_{10}e^{\eta_h^b t} + \frac{\eta_h^a}{\eta_h^b}\left(e^{\eta_h^b t} - 1\right), \quad b_{2t} = b_{20}e^{\eta_l^b t} + \frac{\eta_l^a}{\eta_l^b}\left(e^{\eta_l^b t} - 1\right).$$

While both customer capital grow exponentially, the productive firm's customer capital grows faster asymptotically since $\eta_h^b > \eta_l^b$. However, when considering the initial transition periods, the unproductive firm can feature a higher growth rate in its customer capital. We now turn to a graphical illustration of these different mechanisms in Figure 4.

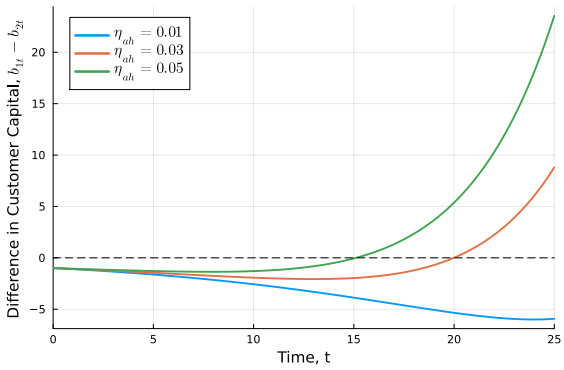
Figure 4 shows the response of customer capital, market share, aggregate productivity, and markups to a productivity shock to an increase in productivity at firm 1 in a two-firm environment. In this environment, we assume firm 1 gets a high-productivity draw when it starts with lower customer capital than firm 2. This figure presents interesting dynamics from the demand follower becoming the productivity leader. We focus on no mean reversion in customer capital and evaluate the evolution depending on three different advertising intensities, which will depend on the cost of advertising. We discuss the four subfigures of this impulse response in turn.

Each graph indicates dynamics in customer capital that will generate potentially different patterns from a model without endogenous demand and oligopolistic strategic dynamics. We see lower initial customer capital for firm 1. Even if it chooses a high marketing intensity ($\eta = 0.05$), it takes a long time to surpass firm 2, about 15 years. This comes from the strong path dependency in investment at the demand leader. One sees the same phenomena with market share, as the productivity leader takes 12 years to surpass the productivity laggard. The group-level productivity *falls* in response to the shock (in Figure 4c). This is due to aggressive marketing from the productivity laggard. Eventually, the leading firm's productivity overtakes the market, but this can take significant time, and will depend on the dynamics of demand. We also see the eventual increase in markups from this leadership.

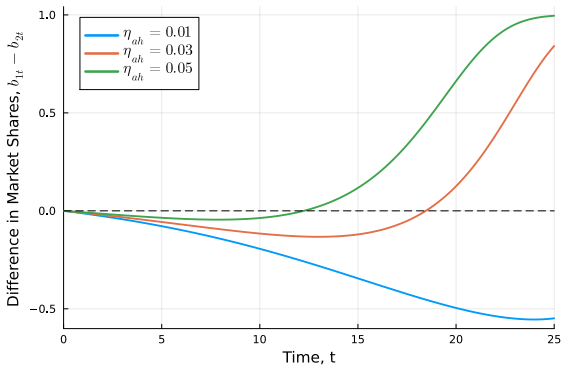
The intuition from this two-firm case extends to the more general oligopolistic setting. The slow convergence to the world where the productivity improvements transmit to the aggregate is unique to a framework where firms with significant pre-existing size have a mechanism such as investment in customer capital to protect their market share. This is a central component of this paper, both for positive analysis of the evolution of

Figure 4: The Evolution of Market Dynamics

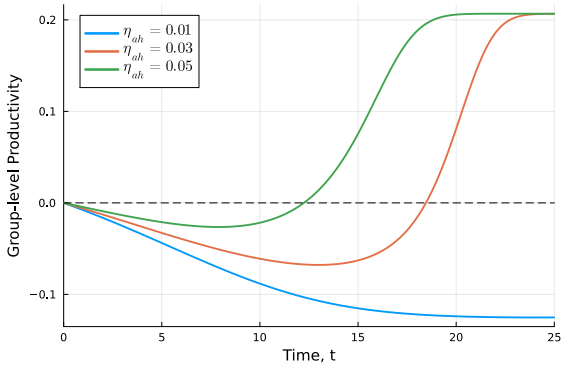
(a) Firm-level Customer Capital



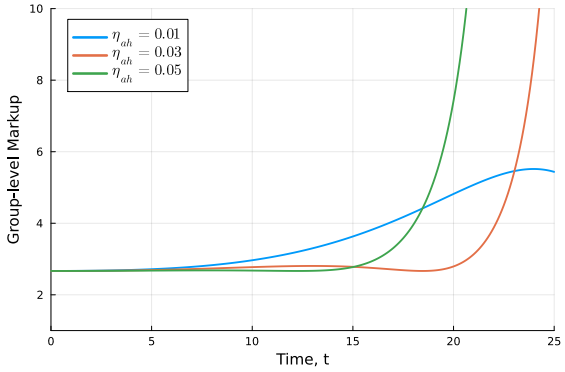
(b) Firm-level Market Shares



(c) Group-level Efficient Productivity



(d) Group-level Markup



productivity and normative implications of standard policies in the literature. We now turn to the more formal aggregation in product group k with multiple firms.

3.4 Aggregation and Welfare

The household makes its static consumption and labor decisions, as well as the dynamic savings decision. We assume the household has a flow utility of $\log \mathbf{C}_t - \mathbf{L}_t$. The household spends on consumption and holds assets $\mathbf{A}_t$ that evolve over time. The household can borrow and save in a representative portfolio of all firms, such that the aggregate profit $\mathbf{\Pi}_t$ is rebated to the household as a dividend. We define ρ to be the discount rate, r_t to be the interest rate, and normalize the wage to be 1. We write the household's problem as

$$\max_{c_{jkt}, \mathbf{L}_t} \int_0^\infty e^{-\rho t} \left(\log \mathbf{C}_t - \mathbf{L}_t \right) dt,$$

s.t.

$$\dot{\mathbf{A}}_t = r_t \mathbf{A}_t + \mathbf{L}_t + \mathbf{\Pi}_t - \int_0^1 \sum_{j=1}^{J} p_{jkt} c_{jkt} dk,$$

and

$$\mathbf{C}_t \text{ given equations (1) and (7)}$$

We now turn to the connection between firm-level choices and aggregate outcomes. In the main discussion of the model, we abstracted from group level k, and in the first step of aggregation, we start by focusing on welfare group by group. We then aggregate over groups.

We present a heuristic discussion of overall welfare which we expand on in Section 5. The household's utility depends on both the dispersion of markups and on whether customer capital is allocated toward the productive firm. We write out the consumption from a product group with productivity gap a and brand gap b.

Lemma 2 (Aggregation) *In a steady-state equilibrium, the discounted utility of the household is given by:*

$$\mathbf{W} = \int_{\mathbf{x}} W(\mathbf{x}) dG(\mathbf{x}), \quad W(\mathbf{x}) = \frac{1}{\rho}\left(\log \frac{Z(\mathbf{x})}{M(\mathbf{x})} - \frac{1}{M(\mathbf{x})} - D(\mathbf{x})\right)$$

where

1. $Z(\mathbf{x})$ *is the group-level labor productivity*

$$Z(\mathbf{x}) = \left(\frac{\sum_{j=1}^{J} e^{a_j + b_j}}{\left(\sum_{j=1}^{J} e^{b_j}\right)^{\theta}}\right)^{\frac{1}{\sigma-1}};$$

2. $M(\mathbf{x})$ *is the aggregate markup*

$$M(\mathbf{x}) = \left(\frac{\sum_{j}^{J} e^{a_j + b_j} \mu_j(\mathbf{x})^{1-\sigma}}{\sum_{j}^{J} e^{a_j + b_j}}\right)^{\frac{1}{1-\sigma}};$$

3. $D(\mathbf{x})$ *is the aggregate labor cost in marketing*

$$D(\mathbf{x}) = \frac{d_0}{2} \sum_j^J \eta_j(\mathbf{x})^2.$$

Returning to firms' reallocation decisions, there are two externalities created by firms to the representative household. First, the dispersion of markups between firms creates a misallocation of labor. Firms choose markups to maximize individual profit and not overall welfare. This misallocation reduces the productivity of labor. Second, the firms do not fully internalize the benefit of matching transferable customer capital towards more productive firms. Firms may stick with mismatch if it builds more monopoly power. The second externality is a novel insight from our paper. This equilibrium is not necessarily efficient, which creates room for policy interventions.

We wait to discuss optimal policies until we have set up the planner's solution in Section 5. However, the analysis so far hints at the main role of policy. A policy that aims to improve efficiency should induce firms to sort customer capital in the right direction and reduce markups.

4 Estimation

In this section, we unite the model and the data by estimating the central parameters of the model. We start by identifying a small set of parameters that match central moments in the data. We then proceed to discuss the fit of our model with out-of-sample moments.

4.1 Moments and Parameters

We calibrate our model to the moments in steady state, but evaluate the transitional dynamics in our policy counterfactuals. We calibrate 10 parameters. We start by discussing the external parameters, three of which we take from the literature. We then discuss seven parameters calibrated jointly with simulated method of moments (SMM).

Parameters from the Literature. We follow our earlier discussion of taking some central but well-understood parameters from the literature. As in the empirical section, we take from the literature the substitution elasticity, $\sigma = 3.9$. This value is the median

across-firm substitution elasticity from Hottman et al. (2016), estimated from a similar demand system and RMS NielsenIQ Scanner data. We calibrate the model to a yearly frequency, which requires setting the discount rate to the level of one-year risk-free rate, $\rho = 0.03$. We vary this parameter in policy counterfactuals in Section 5 and find the main results hold. These elasticities imply the minimum profit margin of firms is around 0.25, and the maximum is 1.0. The substitution elasticity σ has important implications for our counterfactuals, as the degree to which consumers substitute across brands affects the incentives of firms. We report the quantitative analysis given different values of substitution elasticity.

Calibrated Parameters. In addition to our references to the literature, we focus on six parameters from within our estimation. This consists of the number of firms in a product group J, which we count as large firms.[4] We then turn to the persistence of productivity (ρ_a) and customer capital (ρ_b), which are estimated from our AR(1) regressions. In matching the dynamic response of productivity and customer capital, we also obtain the dispersion of these permanent shocks, ν_a^P and ν_b^P respectively. Finally, we connect the advertising cost d to the responsiveness of customer capital to productivity.

In addition to our baseline parameters (σ, ρ), we have six more, $(J, \rho_a, \rho_b, \nu_a, \nu_b, d)$. We present the results in Table 6.

Table 6: Estimation Moments and Parameters

Parameter		Value	Main Identification
Discount Rate	ρ	0.03	Annual Risk-free Rate
Substitution Elasticity	σ	3.90	Hottman et al. (2016)
Number of Firms	J	10	Data
Autoregressive productivity	ρ_a	-0.84	AR(1) regression
Autoregressive demand	ρ_b	-0.71	AR(1) regression
St.Dev. of permanent productivity	ν_a	0.46	Autoregressive dynamics
Cost of advertisement	d_0	1.54	Autoregressive dynamics

Notes: Parameters estimated separately (top panel) and jointly (bottom panel). Source: RMS NielsenIQ, USPTO and author calculations.

[4] We focus on the top 10 firms, though the general result holds when we include more or fewer firms. We test robustness to this estimate by varying the number of firms from 2 to 20.

4.2 Goodness of Fit

We now discuss how well the model fits components of the data that were not directly targeted. We first focus on autocovariance functions: how firm-level variables are correlated with past measures of the firm. We then discuss how the correlation between customer capital and productivity evolve dynamically with and without correlated residuals.

Figure 5 plots the empirical autocovariance function (blue solid line) for the productivity process a alongside the theoretical autocovariance implied by our estimated AR(1) parameters (red dashed line). This is matched based on parameters σ_a and ρ_{aa}. Only the first observation ; $var(\Delta a_t)$ is directly matched. The impulse response for lag 1–5 is untargeted. Visually, we see that both curves have similar shapes, suggesting the AR(1) structure performs a reasonable job capturing the main persistence features of productivity. The zero-lag point corresponds to the process variance, while the gradual decay at higher lags reflects the estimated persistence.

In practical terms, these results imply that shocks to productivity exhibit moderate persistence: when a firm experiences a positive productivity draw, it tends to remain more efficient than average for roughly one or two periods before regressing back to its long-run mean. The close match between empirical and theoretical autocovariances reinforces our identification strategy, where lagged levels of a serve as valid instruments. Any large discrepancy at early lags would have suggested a more serious misspecification, but here the alignment helps validate our approach and justifies using AR(1) for the baseline quantitative analysis.

This consistency between empirical and theoretical moments implies that, although driven partly by firm-specific productivity (a) and partly by broader market factors, the demand side's dynamic behavior can still be well approximated by the simple AR(1) model we employ. In business terms, it means that fluctuations in demand conditions tend to linger but gradually dissipate over one or two periods, aligning with many marketing and IO theories of partial adjustment. Overall, the side-by-side comparisons provide confidence that our estimated parameters for b capture real demand persistence in the data.

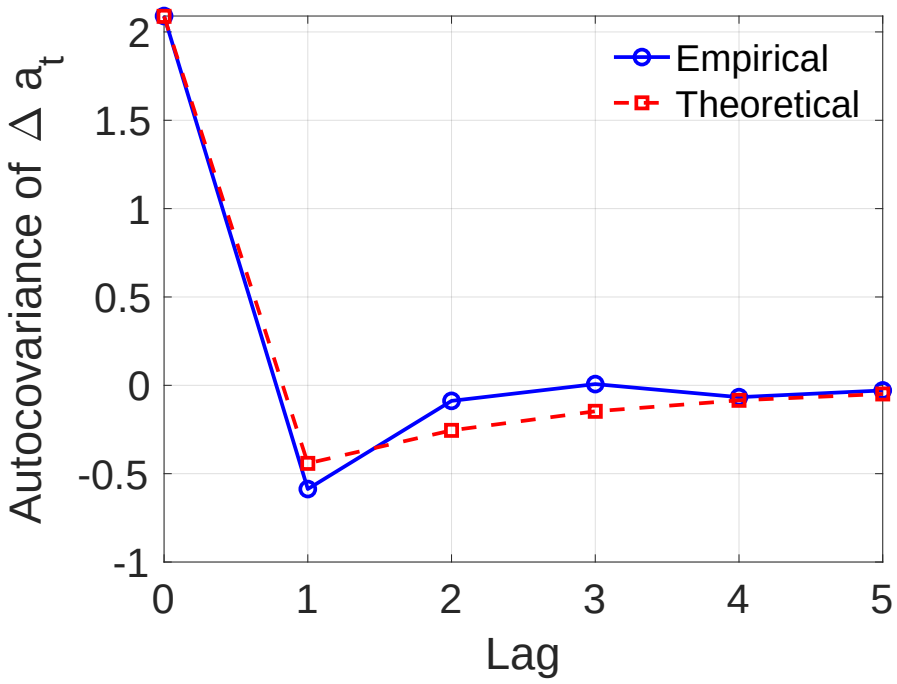


Figure 5: AR(1) for Productivity

Notes: This figure presents the productivity series a, which is derived from firm-level production records, aggregated annually, and normalized within each product-market segment. The autocovariances are computed from the differenced series to ensure stationarity, and the theoretical AR(1) curve is obtained via Arellano–Bond estimates of ρ_a and σ_a.

4.3 Impact of Demand-Productivity Matching: Reduced Form

Having established the empirical significance of customer capital in determining firm market share, we now analyze its implications for aggregate welfare through a modified Baqaee and Farhi (2019) decomposition. Our framework extends their methodology to explicitly account for customer capital and customer awareness in the welfare calculus.

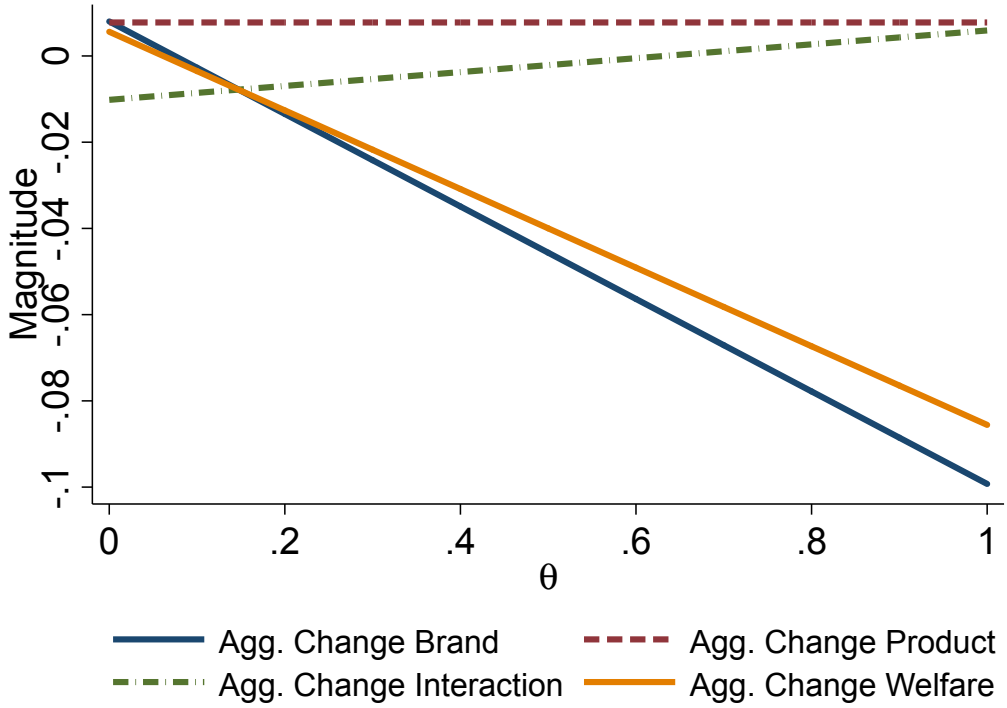
$$\left(1+\frac{\alpha(\sigma-1)}{\sigma}\right) d\log C = \underbrace{\frac{\sigma}{\sigma-1}\left(\sum_j s_j d\log b_j - \theta\sum_j m_j d\log b_j\right)}_{\text{Utility Effect of Branding}}$$
$$\underbrace{-\frac{\alpha}{\sigma}\left(\sum_j \iota_j d\log b_j - \theta\sum_j m_j d\log b_j\right)}_{\text{Production Effect of Branding}}$$
$$\underbrace{+\sum_j\left(s_j - \frac{\alpha(\sigma-1)}{\sigma}\iota_j\right) d\log z_j}_{\text{Net Effect of Productivity}} + \underbrace{\alpha\sum_j(\iota_j - s_j) d\log \mu_j}_{\text{Net Effect of Markups}} \tag{15}$$

The decomposition reveals several important economic mechanisms. First, customer capital affects welfare through both direct utility and production channels, with the magnitude of these effects moderated by the business-stealing parameter θ. When θ approaches unity, indicating pure business-stealing, the welfare gains from customer capital are substantially attenuated as production-side inefficiencies offset utility benefits. This interaction produces non-monotonic welfare effects across the parameter space. We present the decomposition given different assumptions on θ in Figure 6.

A couple of remarks are in order for the aggregation. First, the direct effect of branding is the most essential component of changes in aggregate welfare. Changes in productivity and reallocation have a positive effect on welfare, but for large values of θ this is muted.

Overall in Figure 6, we find that the distribution of customer capital reduces welfare by about 8% given our baseline assumption of limited attention $\theta = 1$. This suggests the imperfect matching between productivity and demand has important implications for welfare and market share. The key element is that *residual* demand is negatively correlated with productivity. Firms that get significantly productivity shocks do not have as much demand as would be predicted by their productivity alone. This is consistent with the fact that it takes time to build customers and may also be consistent with distorted incentives in how firms build customer capital.

Figure 6: Aggregate Changes in Welfare



Notes: This figure disaggregates the change in welfare (yellow) due to change in customer capital (blue) productivity (dashed red) and their interaction (dotted green). The decomposition is available in equation (15). Authors' calculations.

5 Quantitative Analysis

This section studies the positive and normative implications of the matching between demand and productivity. We start by asking how much of the matching pattern is driven by exogenous forces coming from the lag between changes in productivity and changes in demand. We then move to study policy counterfactuals that target firm-level decisions.

5.1 What Drives Matching between Demand and Productivity?

We have documented empirically that the demand of firms does not always match productivity. The model allows us to interpret this mismatch into endogenous and exogenous mechanisms. This decomposition is important because it allows us to quantify how much of the mismatch is due to distortions that can be undone by policies.

In the model, the exogenous mismatch is driven by the persistence of demand factors

ρ_b while the endogenous mismatch is driven by the endogenous marketing efforts η_j. We show in the following tables this split. For our baseline model, we match the AR(1) regression coefficients to replicate the dependence of demand factors on past productivity and demand itself. Using the estimated model, we shut down the exogenous component and the endogenous component and resimulate a large sample (for 1000 years) of firm growth. We then perform the same AR(1) regression on the simulated data. The table reports the coefficients on past demand.

Table 7: Decomposition of Path-dependence of Demand

	Baseline ($\sigma = 3.9$)	High Markup ($\sigma = 2$)	Low Markup ($\sigma = 20$)
Total Response	0.70	0.71	0.70
Endogenous (η)	0.13	0.22	0.06
Exogenous (ρ_b)	0.51	0.49	0.61

In Table 7, we reperform the estimation for the baseline calibration and for the two alternative values of the substitution elasticity. In both cases, our model can replicate the empirical estimates in the AR(1) regression of demand on the past realization of demand and productivity. In our baseline model, the majority of the path-dependence is due to an exogenous force with a sizable component that is endogenous. 21% of the path dependence is due to endogenous forces (0.13/0.64), which exacerbates the demand persistence that already drives a wedge between demand and productivity. This fraction increases to 30% when considering the model with high markups. This decomposition manifests in our policy implications, with potentially large room for improvement, especially for the high-markup calibration.

5.2 Policy Implications

By introducing demands and productivity as dynamically evolving joint determinants of market share, we provide new insights on the impact of standard policy counterfactuals. Here, we perform analysis of different policy counterfactuals. As a benchmark, we start by setting up the planner's problem in our economy. This will serve as the main reference point for our policy analysis. We show that the gap between the decentralized equilibrium and the planner's allocation reflects the static markup distortion and the dynamic mismatch between firms' customer capital and core productivity.

The Planner's Solution. The planner makes both static decisions and dynamic decisions to maximize the representative household's discounted utility. Statically, she chooses how much firms produce given their productivity and their customer capital. Dynamically, she chooses how intensely each firm should invest in their customer capital. The planner aims to sort customer capital to the frontier firm and solve the static inefficiency from markups.

In the static allocation of production, the social planner chooses production (and thus, consumption) given the distribution of brand gaps across product groups. The optimal static labor allocation is standard. Compared to the equilibrium, there are no markup distortions in the planner's allocation. Two implications follow. The planner sets the production labor to 1, and the aggregation consumption equals the aggregate productivity $\mathbf{C} = \mathbf{Z}$, where $\mathbf{Z}$ is defined in Lemma 2.

We now turn to the planner's dynamic decision, which is where the role of customer capital leads to novel insights. The dynamic decision can be made group by group. This comes from that (i) the real consumptions from different groups are aggregated via a Cobb-Douglas aggregator and (ii) the disutility of labor from different groups is linearly additive. More precisely we can write the optimal utility for the representative household, under the planner's solution, as

$$\mathbf{W}^* = \int_0^1 W_k dk, \tag{16}$$

where W_k is the expected utility from product group k:

$$W_k = \max_{\eta_{jt}(\mathbf{x}^t)} \mathbb{E}_0 \int_0^\infty e^{-\rho t} \left[\log Z_{kt}(\mathbf{x}^t) - 1 - \gamma \sum_j^J \frac{\eta_j^2}{2}(\mathbf{x}^t) \right] dt \tag{17}$$

Comparing the planner's problem to the problem of firms in the decentralized equilibrium, we notice two types of wedges. Firstly, the firms internalize only the profits from the static pricing equilibrium, while the planner factors in the full surplus to the representative household. Conditional on the state $\mathbf{x}$, the oligopolistic equilibrium leads to distortions due to variable markups and due to the level of markup. This is the distortion well studied in the literature, in seminal papers such as Edmond et al. (2023). The second kind of wedges is novel in our setting, when we understand firms' size in the lens of two separate dimensions. Firms become big either due to demand or productiv-

ity. The split of these two margins is irrelevant, as the profits only depend on the vector of composite productivity $\mathbf{z}$. The social surplus, however, does differ. As an illustrative example, consider the case where all firms have the same productivity and $\theta = 1$. From the planner's perspective, any configuration of demand leads to the same social surplus as the demand heterogeneity is purely reallocative. To the firms, becoming bigger in terms of demand does lead to a higher profit.

The optimal policy that corrects both types of wedges thus involves correcting markups and making firms internalize the matching between productivity and demand. We now show, with alternative policy instruments, that ignoring the matching between demand and productivity can lead to further distortions. That is, a policy that ignores the split between demand and supply can actually further exacerbate the very distortions it aims to correct.

To fully characterize the planner's solution, we use a similar linearization strategy as in the equilibrium. First, we linearized the planner's period return function $\log Z(\mathbf{x})$ around the equal-demand state. Doing so we obtain the following function:

$$w^*(\mathbf{a}, \mathbf{b}) = \log \frac{\sum_j \exp(a_j)}{J^\theta} + \sum_j \left(\bar{s}_j^*(\mathbf{a}) - \frac{\theta}{J} \right) b_j. \tag{18}$$

In the planner's payoff as in (18), the planner's equal-demand payoff is the simple average of productivity across all firms. The endogenous component of the planner's payoff reflects the difference between the weighted average of customer capital by the equal-demand market shares and the simple average of customer capital, adjusted by the level of rivalry θ.

We now compare the planner's value to the equilibrium profits of firms, as in equation (13) and equation (14). Since the dynamic costs are identical for the planner and for firms in the equilibrium, the difference in the dynamic decisions must root in the difference in static payoffs. First, the planner does not have the markup incentives. More precisely, the linearity in the planner's static payoff means that her optimal marketing strategy only depends on the productivity of firms, not the accumulated customer capital. Second, the planner internalizes the business stealing externality among firms. In firms' profits, an increase in their own customer capital always increases profits. Thus, the marketing investment of the firms in the equilibrium can never be zero; Planner considers the difference between productivity gain from demand reallocation to the busi-

ness stealing externality. Since the net gain $\bar{s}_j^*(\mathbf{a}) - \frac{\theta}{J}$ can fall below zero, the planner will set the marketing investments for the least productive firms to be zero. This is a gap between planner and equilibrium with or without variable markups; last, there is a standard lack of appropriability problem. We compare the planner's payoff to the constant-markup firm profit to see this. As $1/\sigma < 1$, the firms always internalize less gains from reallocating customer capitals.

This difference reflects the business stealing externalities calculated by the planner. A special case to consider is where all productivities are equal and the rivalry is $\theta = 1$. In this case, customer capital is purely reallocated, as no social gain is created by shifting the customer capital across firms.

Size-based Policy. We start by considering a size-based subsidy when the policy makers have access to both demand and productivity differences of firms, separately. The optimal policy that corrects the distortions due to markups removes the revenue incentives of firms and reimburses the firms with the full social surplus they created. In practice, this involves setting a gross transfer:

$$T(q) = \frac{\theta}{\theta-1}\frac{e^{b_j/\theta}q^{\frac{\theta-1}{\theta}}}{\sum_k \exp(b_k+a_k)} - \frac{e^{b_j/\theta}q_j^{\frac{\theta-1}{\theta}}}{\sum_{k\neq j}\exp(b_k+a_k) + e^{b_j/\theta}q^{\frac{\theta-1}{\theta}}}.$$

Given that the other firms all price at the marginal cost, the post-transfer optimization of the focal firm j becomes:

$$\max_q \frac{\theta}{\theta-1}\frac{e^{b_j/\theta}q^{\frac{\theta-1}{\theta}}}{\sum_k \exp(b_k+a_k)} - qe^{b_j/(1-\theta)}.$$

The optimal choice of quantity is $q_j^* = \frac{e^{a_j+b_j\frac{\theta}{\theta-1}}}{\sum_k \exp(b_k+a_k)}$ and the price is the marginal cost $p_j^* = e^{b_j/(1-\theta)}$. This implements the static optimal allocation. In the equilibrium with such a subsidy, every firm receives the share of surplus they created:

$$T_j^* = \frac{1}{\theta-1}\frac{e^{a_j+b_j}}{\sum_k \exp(b_k+a_k)}$$

Size-based Policy with Productivity Difference. The first policy we consider is the size-based subsidy that aims to correct the static distortions due to markups, with the

interpretation that the entire heterogeneity in firms' sizes comes from productivity. More precisely, suppose firms are subject to a subsidy $T(q)$ such that they price at the marginal cost. More precisely, the optimal size-based subsidy aims to implement the allocation:

$$s_j = \frac{\exp(a_j + b_j)}{\sum_{k=1}^{J} \exp(a_k + b_k)}.$$

However, the policy designer interprets all size differentials of firms as productivity, $\hat{b}_j = a_j + b_j$. Aiming to incentivize all firms to price at their (misinterpreted) marginal cost, $\exp\left(\frac{\hat{b}_j}{\theta-1}\right)$, the optimal size-based subsidy is:

$$\hat{T}(q) = \frac{\theta}{\theta-1} \frac{q^{\frac{\theta-1}{\theta}}}{(C^*)^{\frac{\theta-1}{\theta}}} - \frac{q_j^{\frac{\theta-1}{\theta}}}{(C^*)^{\frac{\theta-1}{\theta}} - (q_j^*)^{\frac{\theta-1}{\theta}} + q^{\frac{\theta-1}{\theta}}}$$

Marketing Tax/Subsidy. The second type of policy we consider is a flat tax on marketing. More precisely, starting from the calibrated equilibrium, we assume the government imposes a flat tax on the marketing costs τ_m. The total tax revenue is eventually reimbursed to the representative household as a lump-sum transfer.

Results. Table 8 reports the productivity losses in three environments: The first environment is the average case in our sample Hottman et al. (2016), which takes the average elasticity of substitution from their paper. We add a "high-markup" world ($\sigma = 2$) and a "low-markup" environment ($\sigma = 8$), and compare overall welfare under the efficient scenario to the size-dependent policy and the reallocation policy. Consumption equivalent welfare is made up of consumption, labor in production and reallocation labor.

6 Conclusion

Customer capital and productivity jointly determine the market share of firms and the aggregate economy. This process governs the link between firm-level productivity and aggregate productivity statically and dynamically. We find that customer capital is a central driver of firm market share and has significant aggregate impact. We build a dynamic model to study the formation of demand and productivity. We find that

Table 8: Welfare Changes under Alternative Policies

	Efficient	No Marketing	Size-based	Size-based (Productivity)
Baseline $\sigma = 3.9$				
Chg.Welfare,	41.42	4.62	21.47	-20.55
Productivity,	10.21	-0.28	0.003	-1.32
Markup,	26.4	0.11	26.4	-5.92
Marketing Cost,	-4.81	-4.70	4.95	13.31
High Markup $\sigma = 2.0$				
Chg.Welfare,	69.63	8.28	-122.04	-31.53
Productivity,	16.32	-4.16	15.52	-6.23
Markup,	29.33	0.70	29.33	-7.42
Marketing Cost,	-23.98	-8.55	147.21	17.98
Low Markup $\sigma = 20.0$				
Chg.Welfare,	15.15	0.50	7.01	-1.77
Productivity,	13.26	-0.01	0.00	-0.02
Markup,	0.57	0.001	0.57	-0.03
Marketing Cost,	-1.32	-0.50	-6.38	1.72

the dynamic interaction of these forces can have a significant impact on the transmission of productivity shocks from firm-level to aggregate. Further, standard policies for managing size-based distortions can backfire through encouraging the accumulation of customer capital at the firm level.

Our results highlight the importance of understanding the sources of firm performance beyond the single-dimensional productivity measure. Whether market power comes from productivity or accumulated customer capital leads to different conclusions regarding the efficiency of the market economy. We discuss some possible extensions of our framework here. First, we believe our two-dimensional case on customer capital and productivity can be extended to think about other forces driving firm size: worker prestige, political connections, or location, that interact with productivity. Second, changes in customer capital may endogenously feed back to productivity through firm-level investments in technology. This could then create long-run *growth* effects beyond the allocative efficiency dynamics discussed here. Third, the nature of entry and acquisitions in markets may significantly interact with the forces discussed here. We believe these threads are fruitful extensions for further research.

References

Acemoglu, D., U. Akcigit, H. Alp, N. Bloom, and W. Kerr: 2018, 'Innovation, Reallocation, and Growth'. *American Economic Review* **108**(11), 3450–3491.

Afrouzi, H., A. Drenik, and R. Kim: 2023, 'Concentration, Market Power, and Misallocation: The Role of Endogenous Customer Acquisition'. Technical report, National Bureau of Economic Research.

Aghion, P., C. Harris, P. Howitt, and J. Vickers: 2001, 'Competition, Imitation and Growth with Step-by-Step Innovation'. *Review of Economic Studies* **68**(3), 467–492.

Aghion, P. and P. Howitt: 1992, 'A Model of Growth through Creative Destruction'. *Econometrica* **60**(2), 323–351.

Akcigit, U. and S. T. Ates: 2021, 'Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory'. *American Economic Journal: Macroeconomics* **13**(1), 257–98.

Akcigit, U. and S. T. Ates: 2023, 'What happened to US business dynamism?'. *Journal of Political Economy* **131**(8), 2059–2124.

Akcigit, U. and W. R. Kerr: 2018, 'Growth through Heterogeneous Innovations'. *Journal of Political Economy* **126**(4), 1374–1443.

Amiti, M., O. Itskhoki, and J. Konings: 2019, 'International Shocks, Variable Markups, and Domestic Prices'. *The Review of Economic Studies* **86**(6), 2356–2402.

Arellano, M. and S. Bond: 1991, 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations'. *Review of Economic Studies* **58**(2), 277–297.

Argente, D., D. Fitzgerald, S. Moreira, and A. Priolo: 2021, 'How Do Firms Build Market Share?'. Technical report. SSRN Working Paper.

Argente, D., M. Lee, and S. Moreira: 2018, 'Innovation and Product Reallocation in the Great Recession'. *Journal of Monetary Economics* **93**, 1–20. Carnegie-Rochester-NYU Conference on Public Policy held at the Stern School of Business at New York University.

Argente, D., M. Lee, and S. Moreira: 2020, 'The Life Cycle of Products: Evidence and Implications'. Technical report.

Atkeson, A. and A. Burstein: 2008, 'Pricing-to-Market, Trade Costs, and International Relative Prices'. *American Economic Review* **98**(5), 1998–2031.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen: 2020, 'The Fall of the Labor Share and the Rise of Superstar Firms*'. *The Quarterly Journal of Economics* **135**(2), 645–709.

Bain, J. S.: 1956, *Barriers to New Competition: Their Character and Consequences in Manufacturing Industries*. Harvard University Press.

Baqaee, D. R. and E. Farhi: 2019, 'Productivity and Misallocation in General Equilibrium*'. *The Quarterly Journal of Economics* **135**(1), 105–163.

Berger, D. W., K. F. Herkenhoff, and S. Mongey: 2019, 'Labor Market Power'. Working Paper 25719, National Bureau of Economic Research.

Bhandari, A. and E. R. McGrattan: 2020, 'Sweat Equity in U.S. Private Business*'. *Quarterly Journal of Economics* **136**(2), 727–781.

Boar, C. and V. Midrigan: 2019, 'Markups and inequality'. Technical report, NBER Working Paper.

Bornstein, G. and A. Peter: 2022, 'Nonlinear Pricing and Misallocation'. Technical report, Working Paper.

Bronnenberg, B. J., S. K. Dhar, and J.-P. Dubé: 2009, 'Brand History, Geography, and the Persistence of Brand Shares'. *Journal of Political Economy* **117**(1), 87–115.

Bronnenberg, B. J., J.-P. Dubé, and M. Gentzkow: 2012, 'The Evolution of Brand Preferences: Evidence from Consumer Migration'. *American Economic Review* **102**(6), 2472–2508.

Cavenaile, L., M. A. Celik, J. Perla, and P. Roldan-Blanco: 2023, 'A Theory of Dynamic Product Awareness and Targeted Advertising'. *Working Paper*.

Cavenaile, L., M. A. Celik, P. Roldan-Blanco, and X. Tian: 2022, 'Style over substance? advertising, innovation, and endogenous market structure'. *Working Paper*.

Cavenaile, L. and P. Roldan-Blanco: 2021, 'Advertising, Innovation, and Economic Growth'. *American Economic Journal: Macroeconomics* **13**(3), 251–303.

Crouzet, N. and J. C. Eberly: 2019, 'Understanding weak capital investment: The role of market concentration and intangibles'. Technical report, National Bureau of Economic Research.

David, J. M. and V. Venkateswaran: 2019, 'The Sources of Capital Misallocation'. *American Economic Review* **109**(7), 2531–67.

De Loecker, J. and J. Eeckhout: 2018, 'Global Market Power'. National Bureau of Economic Research Working Paper 24768.

De Loecker, J., J. Eeckhout, and G. Unger: 2020, 'The Rise of Market Power and the Macroeconomic Implications*'. *Quarterly Journal of Economics* **135**(2), 561–644.

De Ridder, M.: 2024, 'Market power and innovation in the intangible economy'. *American Economic Review* **114**(1), 199–251.

Deaton, A.: 1997, *The analysis of household surveys: a microeconometric approach to development policy*. World Bank Publications.

Dinlersoz, E. M. and M. Yorukoglu: 2012, 'Information and Industry Dynamics'. *American Economic Review* **102**(2), 884–913.

Edmond, C., V. Midrigan, and D. Y. Xu: 2018, 'How Costly Are Markups?'. Working Paper 24800, National Bureau of Economic Research.

Edmond, C., V. Midrigan, and D. Y. Xu: 2023, 'How Costly Are Markups?'. *Journal of Political Economy* **131**(7), 1619–1675.

Eggertsson, G. B., J. A. Robbins, and E. G. Wold: 2018, 'Kaldor and Piketty's Facts: The Rise of Monopoly Power in the United States'. National Bureau of Economic Research Working Paper 24287.

Einav, L., P. Klenow, J. D. Levin, and M.-G. Ravid: 2021, 'Customers and Retail Growth'. Working Paper.

Eslava, M., J. Haltiwanger, and N. Urdaneta: 2024, 'The size and life-cycle growth of plants: The role of productivity, demand, and wedges'. *Review of Economic Studies* **91**(1), 259–300.

Foster, L., J. Haltiwanger, and C. Syverson: 2008, 'Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?'. *American Economic Review* **98**(1), 394–425.

Foster, L., J. Haltiwanger, and C. Syverson: 2016, 'The Slow Growth of New Plants: Learning about Demand?'. *Economica* **83**(329), 91–129.

Gourio, F., T. Messer, and M. Siemer: 2014, 'What Is the Economic Impact of the Slowdown in New Business Formation?'. *Chicago Fed Letter* (Sep).

Gourio, F. and L. Rudanko: 2014, 'Customer Capital'. *The Review of Economic Studies* **81**(3), 1102–1136.

Greenwood, J., Y. Ma, and M. Yorukoglu: 2021, "You Will:' A Macroeconomic Analysis of Digital Advertising'. *SSRN Working Paper*.

Gutiérrez, G. and T. Philippon: 2017, 'Declining Competition and Investment in the U.S.'. National Bureau of Economic Research Working Paper 23583.

Hall, R. E.: 2018, 'New Evidence on Market Power, Profit, Concentration, and the Role of Mega-Firms in the U.S. Economy'. Standofrd University, mimeo.

Haltiwanger, J., S. Scarpetta, and H. Schweiger: 2014, 'Cross Country Differences in Job Reallocation: The Role of Industry, Firm Size and Regulations'. *Labour Economics* **26**, 11–25.

Haskel, J. and S. Westlake: 2017, *Capitalism without capital: the rise of the intangible economy*. Princeton University Press.

Heathcote, J., K. Storesletten, and G. L. Violante: 2010, 'The macroeconomic implications of rising wage inequality in the United States'. *Journal of Political Economy* **118**(4), 681–722.

Hopenhayn, H. A.: 1992, 'Entry, Exit, and Firm Dynamics in Long Run Equilibrium'. *Econometrica* **60**(5), 1127–1150.

Hottman, C. J., S. J. Redding, and D. E. Weinstein: 2016, ' Quantifying the Sources of Firm Heterogeneity *'. *The Quarterly Journal of Economics* **131**(3), 1291–1364.

Hsieh, C.-T. and P. J. Klenow: 2009, 'Misallocation and Manufacturing TFP in China and India'. *Quarterly Journal of Economics* **124**(4), 1403–1448.

Ignaszak, M. and P. Sedlácek: 2022, 'Customer acquisition, business dynamism and aggregate growth'. Technical report, Working Paper.

Jaravel, X.: 2018, 'The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector*'. *The Quarterly Journal of Economics* **134**(2), 715–783.

Jovanovic, B.: 1982, 'Selection and the Evolution of Industry'. *Econometrica* **50**(3), 649–70.

Klette, T. J. and S. Kortum: 2004, 'Innovating Firms and Aggregate Innovation'. *Journal of Political Economy* **112**(5), 986–1018.

Liu, E., A. Mian, and A. Sufi: 2022, 'Low Interest Rates, Market Power, and Productivity Growth'. *Econometrica* **90**(1), 193–221.

Mongey, S.: 2021, 'Market structure and monetary non-neutrality'. Technical report, National Bureau of Economic Research.

Olmstead-Rumsey, J.: 2022, 'Market concentration and the productivity slowdown'. *Working Paper*.

Pearce, J. and L. Wu: 2022, 'Brand Reallocation, Concentration, and Growth'. *Available at SSRN 4279983*.

Peters, M.: 2020, 'Heterogeneous Markups, Growth, and Endogenous Misallocation'. *Econometrica* **88**(5), 2037–2073.

Restuccia, D. and R. Rogerson: 2017, 'The Causes and Costs of Misallocation'. *Journal of Economic Perspectives* **31**(3), 151–74.

Schmalensee, R.: 1978, 'A Model of Advertising and Product Quality'. *Journal of Political Economy* **86**(3), 485–503.

Schmalensee, R.: 1982, 'Product Differentiation Advantages of Pioneering Brands'. *The American Economic Review* **72**(3), 349–365.

Shapiro, C.: 1983, 'Premiums for High Quality Products as Returns to Reputations*'. *The Quarterly Journal of Economics* **98**(4), 659–679.

Syverson, C.: 2011, 'What Determines Productivity?'. *Journal of Economic Literature* **49**(2), 326–65.

Syverson, C.: 2019, 'Macroeconomics and Market Power: Facts, Potential Explanations, and Open Questions'. *Brookings Institution Report*. Accessed January 21, 2019. https://www.brookings.edu/research/macroeconomics-and-market-power-facts-potential-explanations-and-open-questions/.

Appendix

The Appendix contains three sections. Appendix A discusses the theoretical proofs and expands on the firm's dynamic problem. Appendix B discusses the data background and merges across datasets. Appendix C discusses the empirical analysis, connections to the literature, and empirical robustness.

A Theoretical Appendix

A.1 Details of the Approximation of Profits

This section discusses the details of the approximation of profits. We take the following steps, given the nonlinear system of equations

$$s_j = \frac{\exp(a_j + b_j)(1 - s_j)^{\theta-1}}{\sum_k \exp(a_k + b_k)(1 - s_k)^{\theta-1}} = \frac{\alpha_j \beta_j (1 - s_j)^{\theta-1}}{\sum_k \alpha_k \beta_k (1 - s_k)^{\theta-1}}$$

or using our definition that $z_j = a_j + b_j$:

$$s_j = \frac{\exp(z_j)(1 - s_j)^{\theta-1}}{\sum_k \exp(z_k)(1 - s_k)^{\theta-1}}$$

We approximate the equilibrium shares with a Taylor expansion to the second order, around the point where $a_j + b_j = 0$, $j = 1, ..., J$. When $a_j + b_j = 0$, firms split the market equally, so $s_j = \frac{1}{J}$. The equilibrium market share $s_j(\mathbf{z})$ can be approximated by

$$s_j(\mathbf{z}) \approx \frac{1}{J} + \sum_k \frac{\partial s_j}{\partial z_k} z_k + \frac{1}{2} \sum_k \sum_m \frac{\partial^2 s_j}{\partial z_k \partial z_m} z_k z_m$$

and the profit of firm j can be approximated by

$$\begin{aligned} \pi_j(s_j(\mathbf{z})) \approx & \pi_0 + \left(\frac{1}{\theta} + 2\frac{\theta - 1}{\theta} \frac{1}{J} \right) \sum_k \frac{\partial s_j}{\partial z_k} z_k \\ & + \frac{1}{2} \left(\frac{1}{\theta} + 2\frac{\theta - 1}{\theta} \frac{1}{J} \right) \sum_k \sum_m \frac{\partial^2 s_j}{\partial z_k \partial z_m} z_k z_m + \frac{\theta - 1}{\theta} \sum_k \sum_m \frac{\partial s_j}{\partial z_k} z_k \frac{\partial s_j}{\partial z_m} z_m \end{aligned}$$

To find the derivatives, we need to fully differentiate the equation system that defines the shares, totally differentiate the system on both sides with respect to z_k we get:

$$\begin{pmatrix} \frac{\partial s_1}{\partial z_k} \\ \frac{\partial s_2}{\partial z_k} \\ \cdots \\ \frac{\partial s_J}{\partial z_k} \end{pmatrix} = s_k \mathbf{e}_k - \begin{pmatrix} s_j s_k \\ s_j s_k \\ \cdots \\ s_j s_k \end{pmatrix} - (\theta - 1) \begin{pmatrix} s_1 & -\frac{s_1 s_2}{1-s_2} & \cdots & -\frac{s_1 s_J}{1-s_J} \\ -\frac{s_2 s_1}{1-s_1} & s_2 & \cdots & -\frac{s_2 s_J}{1-s_J} \\ \cdots & & & \\ -\frac{s_J s_1}{1-s_1} & -\frac{s_J s_2}{1-s_2} & \cdots & s_J \end{pmatrix} \begin{pmatrix} \frac{\partial s_1}{\partial z_k} \\ \frac{\partial s_2}{\partial z_k} \\ \cdots \\ \frac{\partial s_J}{\partial z_k} \end{pmatrix}$$

There is much symmetry involved in the system. Specifically, we denote $\Delta_1 = \frac{\partial s_k}{\partial z_k}$ and $\Delta_2 = \frac{\partial s_j}{\partial z_k}$, where $j \neq k$. This equation system simplifies into:

$$\begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix} = \begin{pmatrix} s_k \\ 0 \end{pmatrix} - \begin{pmatrix} s_k^2 \\ s_j s_k \end{pmatrix} - (\theta - 1) \begin{pmatrix} s_k & -(J-1)\frac{s_j s_k}{1-s_j} \\ -\frac{s_j s_k}{1-s_k} & s_j - (J-2)\frac{s_j s_j}{1-s_j} \end{pmatrix} \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}$$

Evaluated at the expansion point:

$$\begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{J} \\ 0 \end{pmatrix} - \begin{pmatrix} \frac{1}{J^2} \\ \frac{1}{J^2} \end{pmatrix} - (\theta - 1) \begin{pmatrix} \frac{1}{J} & -\frac{1}{J} \\ -\frac{1}{J(J-1)} & \frac{1}{J} - \frac{J-2}{J(J-1)} \end{pmatrix} \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}$$

Solving this system we get:

$$\Delta_1 = \frac{(J-1)^2}{J^2(J+\theta-2)}, \quad \Delta_2 = -\frac{J-1}{J^2(J+\theta-2)}$$

It is useful to denote:

$$A = \begin{pmatrix} \frac{1}{J} & -\frac{1}{J} \\ -\frac{1}{J(J-1)} & \frac{1}{J} - \frac{J-2}{J(J-1)} \end{pmatrix}$$

To obtain the second-order derivatives, we first consider the case with $m = k$. Again follow the symmetry logic, we denote $\delta_1 = \frac{\partial^2 s_k}{\partial z_k^2}$ and $\delta_2 = \frac{\partial^s s_j}{\partial z_k^2}$. Differentiating the system

to a second order we get

$$
\begin{aligned}
(I+(\theta-1)A)\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix} &= \begin{pmatrix}\Delta_1\\ 0\end{pmatrix} - \begin{pmatrix}2\frac{1}{J}\Delta_1\\ \frac{1}{J}(\Delta_1+\Delta_2)\end{pmatrix} \\
&\quad -(\theta-1)\begin{pmatrix}\Delta_1 & -(J-1)\frac{\frac{1}{J}\frac{J-1}{J}(\Delta_1+\Delta_2)+\Delta_2\frac{1}{J^2}}{(\frac{J-1}{J})^2}\\ -\frac{\frac{1}{J}\frac{J-1}{J}(\Delta_1+\Delta_2)+\Delta_1\frac{1}{J^2}}{(\frac{J-1}{J})^2} & \Delta_2-(J-2)\frac{2\frac{J-1}{J^2}\Delta_2+\frac{1}{J^2}\Delta_2}{(\frac{J-1}{J})^2}\end{pmatrix}\begin{pmatrix}\Delta_1\\ \Delta_2\end{pmatrix} \\
&= \begin{pmatrix}\Delta_1\\ 0\end{pmatrix} - \begin{pmatrix}2\frac{1}{J}\Delta_1\\ \frac{1}{J}(\Delta_1+\Delta_2)\end{pmatrix} \\
&\quad -(\theta-1)\begin{pmatrix}\Delta_1 & -(\Delta_1+\Delta_2)+\Delta_2\frac{1}{J-1}\\ -\frac{1}{J-1}(\Delta_1+\Delta_2)+\Delta_1\frac{1}{(J-1)^2} & \Delta_2-2\frac{J-2}{J-1}\Delta_2+\frac{J-2}{(J-1)^2}\Delta_2\end{pmatrix}\begin{pmatrix}\Delta_1\\ \Delta_2\end{pmatrix}
\end{aligned}
$$

From symmetry, we denote :

$$
\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix} = \begin{pmatrix}\Delta_1\\ 0\end{pmatrix} - \begin{pmatrix}\frac{2}{J}\Delta_1\\ \frac{1}{J}(\Delta_1+\Delta_2)\end{pmatrix} - (\theta-1)\begin{pmatrix}\Delta_1 & -\left(\Delta_1 J+\frac{J}{J-1}\Delta_2\right)\\ -() & \frac{1}{J}-\frac{J-2}{J(J-1)}\end{pmatrix}\begin{pmatrix}\Delta_1\\ \Delta_2\end{pmatrix}
$$

We expand the terms:

$$
\frac{\partial}{\partial z_m}\begin{pmatrix}s_j s_k\\ s_j s_k\\ \dots\\ s_j s_k\end{pmatrix} = \frac{1}{J}\begin{pmatrix}s_j s_k\\ s_j s_k\\ \dots\\ s_j s_k\end{pmatrix}
$$

$$
\frac{\partial}{\partial z_m}\begin{pmatrix}s_1 & -\frac{s_1 s_2}{1-s_2} & \dots & -\frac{s_1 s_J}{1-s_J}\\ -\frac{s_2 s_1}{1-s_1} & s_2 & \dots & -\frac{s_2 s_J}{1-s_J}\\ \dots & & & \\ -\frac{s_J s_1}{1-s_1} & -\frac{s_J s_2}{1-s_2} & \dots & s_J\end{pmatrix} = \frac{\partial}{\partial z_m}\begin{pmatrix}s_1 & -\frac{s_1 s_2}{1-s_2} & \dots & -\frac{s_1 s_J}{1-s_J}\\ -\frac{s_2 s_1}{1-s_1} & s_2 & \dots & -\frac{s_2 s_J}{1-s_J}\\ \dots & & & \\ -\frac{s_J s_1}{1-s_1} & -\frac{s_J s_2}{1-s_2} & \dots & s_J\end{pmatrix}
$$

where Π_0 is the vector of profits when $z_j = 0$: $\Pi_0 = \left(\frac{1}{\theta}\frac{1}{J}+\frac{\theta-1}{\theta}\frac{1}{J^2},\dots,\frac{1}{\theta}\frac{1}{J}+\frac{\theta-1}{\theta}\frac{1}{J^2}\right)^T$; $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ collects the first-order and the second-order derivatives, respectively. Using the chain rule:

$$
[\mathbf{\Pi}_1]_{jm} = \left(\frac{1}{\theta}+2\frac{\theta-1}{\theta}s_j\right)\frac{\partial s_j}{\partial z_m}
$$

$$
[\mathbf{\Pi}_2]_{jm} = \begin{cases}2\frac{\theta-1}{\theta}\frac{\partial s_j}{\partial z_j}+\left(\frac{1}{\theta}+2\frac{\theta-1}{\theta}s_j\right)\frac{\partial^2 s_j}{\partial z_j^2} & \text{if } j=m\\ \left(\frac{1}{\theta}+2\frac{\theta-1}{\theta}s_j\right)\frac{\partial s_j}{\partial z_m} & \text{if } j\neq m\end{cases}
$$

A.2 Uniqueness of Equilibrium

In this section, we analyze the uniqueness of the equilibrium. For notation simplicity, we index the vector $\mathbf{a}$ by index $\alpha = 1, ..., N^J$. We start by re-writing firm j's dynamic problem, given the competitors follow the strategy $\eta_j(\alpha, \mathbf{b}) = \gamma_j(\alpha) + \epsilon_j(\alpha)'\mathbf{b}$, for $j \neq i$.

$$\max_{\eta_i(\alpha,\mathbf{b})} \mathbb{E}_0 \int_0^\infty e^{-rt} \left(\bar{\pi}_i(\alpha) + P_i(\alpha)'\mathbf{b} + \mathbf{b}^T Q_i(\alpha)\mathbf{b} - \frac{\kappa}{2}\eta_i(\alpha,\mathbf{b})^2 \right) dt, \tag{A1}$$

s.t.

$$\dot{\mathbf{b}}(\alpha) = G_i(\alpha)' + (E(\alpha)' - \rho\mathbf{I})\mathbf{b} + \eta_i(\alpha,\mathbf{b})\mathbf{e}_i$$

where

$$G(\alpha)_j = \begin{cases} \gamma_j(\alpha) & \text{if } j \neq i \\ 0 & \text{if } j = i \end{cases}, \quad E(\alpha)_{:,j} = \begin{cases} \epsilon_j(\alpha) & \text{if } j \neq i \\ 0 & \text{if } j = i \end{cases}$$

and $\mathbf{e}_i$ is the unit vector at index j that takes value 1 a the i-th dimension and zeros in all other dimensions.

We write out the HJB equation from this problem:

$$rV_i(\alpha,\mathbf{b}) = \max_{\eta_i(\alpha,\mathbf{b})} \hat{\pi}(\alpha,\mathbf{b}) - \kappa\frac{\eta_i(\alpha,\mathbf{b})^2}{2} + \Delta V_i(\alpha,\mathbf{b})'\dot{\mathbf{b}}(\alpha) + \lambda\sum_{\alpha'}(V_i(\alpha',\mathbf{b}) - V_i(\alpha,\mathbf{b}))F_{\alpha,\alpha'}.$$

We now guess and verify the value function follows the quadratic form. More precisely, we guess that

$$V_i(\alpha,\mathbf{b}) = v_i(\alpha) + \kappa\left(\Gamma_i'(\alpha)\mathbf{b} + \frac{1}{2}\mathbf{b}'\mathcal{E}_i(\alpha)\mathbf{b}\right)$$

From this guess, the gradient can be written as:

$$\Delta V_i(\alpha,\mathbf{b}) = \kappa\left(\Gamma_i(\alpha) + \mathcal{E}_i(\alpha)\right)$$

The first-order condition with respect to $\eta_i(\alpha,\mathbf{b})$ requires that:

$$\eta_i(\alpha,\mathbf{b}) = \gamma_i(\alpha) + \epsilon_i(\alpha)'\mathbf{b},$$

where we denote the i-th element of $\Gamma_i(\alpha)$ as $\gamma_i(\alpha)$ and the i-th column of $\mathcal{E}_i(\alpha)$ as $\epsilon_i(\alpha)$. With this optimal solution, we now verify the guess. In the equation for V_i, plugging the

guess resulting solution form. We have:

$$\begin{aligned}
&(r+\lambda)\left(v_i(\alpha)+\kappa\left(\Gamma_i'\mathbf{b}+\frac{1}{2}\mathbf{b}'\mathcal{E}_i\mathbf{b}\right)\right)\\
=&\hat{\pi}(\alpha,\mathbf{b})-\frac{\kappa}{2}\left(\gamma_i(\alpha)+\epsilon_i(\alpha)'\mathbf{b}\right)^2\\
+&\kappa\left(\Gamma_i'(\alpha)+\mathbf{b}'\mathcal{E}_i(\alpha)\right)\left(G_i(\alpha)+\gamma_i(\alpha)\mathbf{e}_i+\left(E_i(\alpha)'-\rho\mathbf{I}+\mathbf{e}_i\epsilon_i(\alpha)'\right)\mathbf{b}\right)\\
+&\lambda\sum_{\alpha'}\left(v_i(\alpha')+\gamma\left(\Gamma_i(\alpha')'\mathbf{b}+\frac{1}{2}\mathbf{b}'\mathcal{E}_i(\alpha')\mathbf{b}\right)\right)F_{a,a'},
\end{aligned}$$

This verifies the quadratic guess, and we end up with three equations for the matrices. We only state the two equation systems that characterize Γ_i and $\mathcal{E}_i$

$$(r+\lambda)\Gamma_i(\alpha)=\frac{1}{\kappa}P_i(\alpha)+\gamma_i(\alpha)\epsilon_i(\alpha)+((E(\alpha)-\rho I)\Gamma_i(\alpha)+G_i(\alpha)\mathcal{E}_i(\alpha))+\lambda\sum_{\alpha'}\Gamma_i(\alpha')F_{\alpha,\alpha'}$$

$$(r+\lambda)\mathcal{E}_i(\alpha)=\frac{1}{\kappa}Q_i(\alpha)+\frac{1}{2}\mathcal{E}_i(\alpha)\mathbf{e}_i\mathbf{e}_i'\mathcal{E}_i(\alpha)+\mathcal{E}_i(\alpha)(E_i'(\alpha)-\rho I)+\lambda\sum_{\alpha'}\mathcal{E}_i(\alpha')F_{\alpha,\alpha'}$$

or

$$0=\frac{1}{\kappa}Q_i(\alpha)+\frac{1}{2}\mathcal{E}_i(\alpha)\mathbf{e}_i\mathbf{e}_i'\mathcal{E}_i(\alpha)+\mathcal{E}_i(\alpha)(E(\alpha)'-(\rho+r)I)+\lambda\sum_{\alpha'}(\mathcal{E}_i(\alpha')-\mathcal{E}_i(\alpha))F_{\alpha,\alpha'}$$

In the equilibrium, we can write:

$$E_i(\alpha)=\sum_{j\neq i}\mathbf{e}_i\mathbf{e}_i'\mathcal{E}_j(\alpha).$$

We thus end up with I equations for the following form

$$0=\frac{1}{\kappa}Q_i(\alpha)+\frac{1}{2}\mathcal{E}_i(\alpha)\mathbf{e}_i\mathbf{e}_i'\mathcal{E}_i(\alpha)+\mathcal{E}_i(\alpha)\left(\sum_{j\neq i}\mathbf{e}_j\mathbf{e}_j'\mathcal{E}_j(\alpha)-(\rho+r)I\right)+\lambda\sum_{\alpha'}(\mathcal{E}_i(\alpha')-\mathcal{E}_i(\alpha))F_{\alpha,\alpha'}$$

Stability:

$$\sum_{i=1}^{J}\mathbf{e}_i\mathbf{e}_i'\mathcal{E}_i(\alpha)-\rho I \text{ negative definite}$$

In the standard riccatti form, we write $\tilde{\mathcal{E}}_i=-\mathcal{E}_i$:

$$0=-\frac{1}{\kappa}Q_i(\alpha)-\frac{1}{2}\tilde{\mathcal{E}}_i(\alpha)\mathbf{e}_i\mathbf{e}_i'\tilde{\mathcal{E}}_i(\alpha)+\tilde{\mathcal{E}}_i(\alpha)(E(\alpha)-(\rho+r)I)+\lambda\sum_{\alpha'}(\tilde{\mathcal{E}}_i(\alpha')-\tilde{\mathcal{E}}_i(\alpha))F_{\alpha,\alpha'}$$

Identical Productivity. We first consider the case where the productivity distribution is degenerate. In this case, we look for a symmetric equilibrium. Denote $\bar{Q} = Q_i$. In this symmetric equilibrium, firms use the same strategy, $\mathcal{E}_i = \bar{\mathcal{E}}$. The equation becomes

$$0 = \frac{\bar{Q}}{\kappa} + \frac{1}{2}\bar{\mathcal{E}}\mathbf{e}_i\mathbf{e}_i'\bar{\mathcal{E}} + \sum_{j\neq i}\bar{\mathcal{E}}\mathbf{e}_j\mathbf{e}_j'\bar{\mathcal{E}} - (\rho + r)\bar{\mathcal{E}}.$$

As firms are identical $\bar{s}_i = \frac{1}{J}$ and we can write $\bar{Q} = \xi qq'$, where

$$q = \begin{cases} J-1 & \text{if } j = i \\ -1 & \text{if } j \neq i \end{cases}$$

and

$$\xi = \frac{\sigma - 1}{\sigma}\frac{(J-1)^2}{(J+\sigma-2)^2 J^4}.$$

We guess the solution takes the form $\bar{\mathcal{E}} = \epsilon qq'$. We know that

$$\sum_{j\neq i}\bar{\mathcal{E}}\mathbf{e}_j\mathbf{e}_j'\bar{\mathcal{E}} = \bar{\mathcal{E}}\bar{\mathcal{E}} - \bar{\mathcal{E}}\mathbf{e}_i\mathbf{e}_i'\bar{\mathcal{E}}$$

and

$$\bar{\mathcal{E}}\mathbf{e}_i\mathbf{e}_i'\bar{\mathcal{E}} = \epsilon^2(J-1)^2 qq'$$

Thus, our equation becomes

$$0 = \frac{\xi}{\kappa}qq' + \epsilon^2(q'q)qq' - \frac{1}{2}\epsilon^2(J-1)^2 qq' - (\rho + r)\epsilon qq' = 0$$

Since qq' is positive definite, we can invert the equation to get

$$\frac{\xi}{\kappa} + \epsilon^2(q'q) - \frac{1}{2}\epsilon^2(J-1)^2 - (\rho + r)\epsilon = 0$$

It remains to calculate $q'q = J(J-1)$. Plugging into the equation we get a quadratic equation

$$0 = \frac{\xi}{\kappa} + \frac{J^2-1}{2}\epsilon^2 - (\rho + r)\epsilon.$$

This equation has real solutions if and only if

$$(\rho + r)^2\kappa \geq 2(J^2-1)\xi.$$

and we focus on the smaller solution (due to the transversality condition)

$$\frac{(\rho+r)-\sqrt{(\rho+r)^2-\frac{2\xi}{\kappa}(J^2-1)}}{J^2-1}$$

For the solution to be stable, we need $\bar{\mathcal{E}}-\rho I$ to be negative definite. Since $\bar{\mathcal{E}}=\epsilon qq'$ has only one non-zero eigenvalue $\epsilon J(J-1)$. The stability condition is equivalent to

$$\epsilon J(J-1)\leq\rho$$

, equivalent to

$$\left((\rho+r)-\sqrt{(\rho+r)^2-\frac{2\xi}{\kappa}(J^2-1)}\right)\frac{J}{J+1}<\rho$$

As $J\to\infty$, $\xi\to 0$.

$$(\rho+r)^2\kappa\geq 0$$

and the stability condition requires that

$$0<\rho$$

Fixed Heterogeneous Productivity. Suppose $\lambda=0$. We can write $Q_i=\xi_i q_i q_i'$, where

$$q_{i,j}=\begin{cases}(1-\bar{s}_i) & \text{if } j=i\\ -\bar{s}_j & \text{otherwise}\end{cases}$$

$$0=\frac{1}{\kappa}\xi_i q_i q_i'+\frac{1}{2}\mathcal{E}_i\mathbf{e}_i\mathbf{e}_i'\mathcal{E}_i+\mathcal{E}_i\left(\sum_{j\neq i}\mathbf{e}_j\mathbf{e}_j'\mathcal{E}_j-(\rho+r)I\right)$$

With this guess

$$\mathcal{E}_i\mathbf{e}_j\mathbf{e}_j'\mathcal{E}_j=\epsilon_j\epsilon_j(-\bar{s}_j)(1-\bar{s}_j)q_iq_j'$$

A.3 Details of Two-firm Case

We discuss the details in the derivation of the results in the two-firm case. More specifically, we assume the productivities take value in $\{0,\lambda\}$. In this case, we can write the approximated market share as

$$\hat{s}_j(\mathbf{a},\mathbf{b})=\bar{s}_j(\mathbf{a})+\zeta_j(\mathbf{a})\bar{s}_{-j}(\mathbf{a})(b_j-b_{-j}),$$

where we used the fact $\bar{s}_j(\mathbf{a}) + \bar{s}_{-j}(\mathbf{a}) = 1$. Using this result, we can write the approximated profit as

$$\hat{\pi}_j(\mathbf{a}, \mathbf{b}) = \bar{\pi}_j(\mathbf{a}) + \left(\frac{1}{\sigma} + \frac{\sigma - 1}{\sigma} 2\bar{s}_j(\mathbf{a})\right) \zeta_j(\mathbf{a}) \bar{s}_{-j}(\mathbf{a})(b_j - b_{-j}) + \frac{\sigma - 1}{\sigma} \zeta_j(\mathbf{a})^2 \bar{s}_{-j}(\mathbf{a})^2 (b_j - b_{-j})^2.$$

We guess that

$$V_i(\alpha, \mathbf{b}) = \nu_i(\alpha) + \kappa \gamma_i(\alpha)(b_i - b_{-i}) + \frac{\kappa}{2} \epsilon_i(\alpha)(b_i - b_{-i})^2.$$

With this guess, we take the partial derivatives

$$\frac{\partial V_i}{\partial b_i}(\alpha, \mathbf{b}) = \kappa(\gamma_i(\alpha) + \epsilon_i(\alpha)(b_i - b_{-i})),$$

and

$$\frac{\partial V_i}{\partial b_{-i}}(\alpha, \mathbf{b}) = -\kappa(\gamma_i(\alpha) + \epsilon_i(\alpha)(b_i - b_{-i})).$$

Taking the first-order condition with respect to the marketing investment:

$$\eta_i(\alpha, \mathbf{b}) = \kappa(\gamma_i(\alpha) + \epsilon_i(\alpha)(b_i - b_{-i}).$$

$$\eta_{-i}(\alpha, \mathbf{b}) = \kappa(\gamma_{-i}(\alpha) + \epsilon_{-i}(\alpha)(b_{-i} - b_i).$$

The net value from marketing is

$$\eta_j(\mathbf{a}, \mathbf{b}) \frac{\partial V_j}{\partial b_j}(\mathbf{a}, \mathbf{b}) - \frac{\gamma}{2} \eta_j(\mathbf{a}, \mathbf{b})^2 = \frac{1}{2\gamma} \left(v_j(\mathbf{a}) + \omega_j(\mathbf{a})(b_j - b_{-j})\right)^2$$

We plug these decisions to verify the guess into the HJB equation:

$$\begin{aligned}
& r\left(\nu_i(\alpha) + \kappa\gamma_i(\alpha)(b_i - b_{-i}) + \frac{\kappa}{2}\epsilon_i(\alpha)(b_i - b_{-i})^2\right) \\
=& \bar{\pi}_i(\alpha) + \left(\frac{1}{\sigma} + \frac{\sigma - 1}{\sigma} 2\bar{s}_i(\alpha)\right) \zeta_i(\alpha) \bar{s}_{-i}(\alpha)(b_i - b_{-i}) + \frac{\sigma - 1}{\sigma} \zeta_i(\alpha)^2 \bar{s}_{-i}(\alpha)^2 (b_i - b_{-i})^2 \\
+& \frac{\kappa}{2} \left(\gamma_i(\alpha) + \epsilon_i(\alpha)(b_i - b_{-i})\right)^2 - \kappa\rho \left(\gamma_i(\alpha) + \epsilon_i(\alpha)(b_i - b_{-i})\right)(b_i - b_{-i}) \\
-& \kappa \left(\gamma_i(\alpha) + \epsilon_i(\alpha)(b_i - b_{-i})\right)\left(\gamma_{-i}(\alpha) + \epsilon_{-i}(\alpha)(b_{-i} - b_i)\right) + \mathcal{A}(V_j)
\end{aligned}$$

Matching terms, we find the guessed form solve the HJB equation and we now have the following separate HJB equations for the coefficients:

$$\rho v_j(\mathbf{a}) = \bar{\pi}_j(\mathbf{a}) + \frac{1}{2\gamma} v_j(\mathbf{a})^2 - \frac{1}{\gamma} v_j(\mathbf{a}) v_{-j}(\mathbf{a}) + \mathcal{A} v_j(\mathbf{a})$$

$$\rho v_j(\mathbf{a}) = \left(\frac{1}{\sigma} + \frac{\sigma-1}{\sigma} 2\bar{s}_j(\mathbf{a})\right) \zeta_j(\mathbf{a}) \bar{s}_{-j}(\mathbf{a}) + \frac{1}{\gamma} v_j(\mathbf{a}) \omega_j(\mathbf{a}) - \frac{1}{\gamma} \left(\omega_j(\mathbf{a}) v_{-j}(\mathbf{a}) - v_j(\mathbf{a}) \omega_{-j}(\mathbf{a})\right) + \mathcal{A} v_j(\mathbf{a})$$

$$(r + 2\rho + \lambda)\epsilon_i(\alpha) = \frac{2}{\kappa}\frac{\sigma-1}{\sigma} \zeta_i(\alpha)^2 \bar{s}_{-i}(\alpha)^2 + \epsilon_i(\alpha)^2 + 2\epsilon_i(\alpha)\epsilon_{-i}(\alpha) + \lambda \sum_{\alpha'} F_{\alpha,\alpha'} \epsilon_i(\alpha')$$

Identical Productivity. For this case we set $\lambda = 0$. Starting with the higher-order term. The common response to customer capital gap is the solution to:

$$\frac{2}{\kappa}\frac{\sigma-1}{\sigma} \zeta^2 \bar{s}^2 + 3\epsilon^2 - (r + 2\rho)\epsilon = 0$$

$$\dot{b} = (\epsilon - \rho) b$$

In this case, knowing $\bar{s} = \frac{1}{2}$, we can further simplify this equation to

$$3\epsilon^2 - (r + 2\rho)\epsilon + \frac{\sigma-1}{8\kappa\sigma^3} = 0$$

This equation has real solutions when

$$(r + 2\rho)^2 > \frac{3(\sigma-1)}{2\kappa\sigma^3}$$

When this condition is satisfied, we have the solutions in the form of

$$\frac{(r + 2\rho) + \sqrt{(r + 2\rho)^2 - \frac{3(\sigma-1)}{2\kappa\sigma^3}}}{6}$$

and

$$\frac{(r + 2\rho) - \sqrt{(r + 2\rho)^2 - \frac{3(\sigma-1)}{2\kappa\sigma^3}}}{6}$$

Only the small one satisfies the transversality condition.

Heterogeneous Productivity with $\lambda = 0$. Denoting $q_i = \frac{2}{\kappa}\frac{\sigma-1}{\sigma}\zeta_i(\alpha)^2\bar{s}_{-i}(\alpha)^2$, we have a two-equation system:

$$(r+2\rho)\epsilon_1 = q_1 + \epsilon_1^2 + 2\epsilon_1\epsilon_2$$

$$(r+2\rho)\epsilon_2 = q_2 + \epsilon_2^2 + 2\epsilon_1\epsilon_2$$

Defining $\epsilon^s = \epsilon_1 + \epsilon_2$,

A.4 Proof of Lemma 1

A.1 Bellman Equation Formulation

We consider a system of J Bellman equations, where the value function $V_j(\mathbf{a}, \mathbf{b})$ for firm j satisfies the recursive equation:

$$\rho V_j(\mathbf{a},\mathbf{b}) = \max_{\eta_j} \bar{\pi} j(\mathbf{a}) + \zeta j(\mathbf{a}) \left(b_j - \sum_k \bar{s}k(\mathbf{a}) bk \right) + \xi_j(\mathbf{a}) \left(b_j - \sum_k \bar{s}k(\mathbf{a}) bk \right)^2 \qquad -\frac{\gamma}{2}\eta_j^2 + \eta_j \frac{\partial V_j}{\partial b_j} + \sum_{k\neq j} \eta$$

where $\mathcal{A}$ represents the infinitesimal generator of a Poisson jump process governing $\mathbf{a}$.

A.2 Value Function Guess

We postulate the following quadratic form for V_j:

$$V(\mathbf{a},\mathbf{b}) = v^0(\mathbf{a}) + v_j^1(\mathbf{a}) b_j + v_j^2(\mathbf{a}) b_j^2 + b_j \sum_{k\neq j} u_{jk}(\mathbf{a}) b_k \qquad + \sum_{k\neq j} w_{jk}^1(\mathbf{a}) b_k + \sum_{k\neq j} w_{jk}^2(\mathbf{a}) b_k^2.$$

A.3 First-Order Conditions

Taking partial derivatives of V_j with respect to b_j and b_k gives:

$$\frac{\partial V_j}{\partial b_j} = v_j^1(\mathbf{a}) + 2v_j^2(\mathbf{a}) b_j + \sum_{k\neq j} u_{jk}(\mathbf{a}) b_k, \; \frac{\partial V_j}{\partial b_k} \qquad = u_{jk}(\mathbf{a}) b_j + w_{jk}^1(\mathbf{a}) + 2w_{jk}^2(\mathbf{a}) b_k.$$

The optimal marketing investment η_j is obtained from the first-order condition:

$$\eta_j = \frac{1}{\gamma}\left[v_j^1(\mathbf{a}) + 2v_j^2(\mathbf{a})b_j + \sum_{k\neq j} u_{jk}(\mathbf{a})b_k\right].$$

A.4 Substituting into the Bellman Equation

Substituting η_j into the Bellman equation and collecting terms, we obtain the system:

$$\rho v_j^0(\mathbf{a}) = \bar{\pi} j(\mathbf{a}) + \frac{1}{2\gamma} v j^1(\mathbf{a})^2 + \frac{1}{\gamma}\sum_{k\neq j} v_k^1(\mathbf{a}) w_{jk}^1(\mathbf{a}) + \mathcal{A}(v_j^0(\mathbf{a})),\ \rho v_j^1(\mathbf{a}) \quad = \zeta_j(\mathbf{a}) + \frac{1}{\gamma} v_j^1(\mathbf{a}) v_j^2(\mathbf{a}) + \frac{1}{\gamma}\sum_{k\neq j} v_k^1$$

In this section, we compute the dynamic decision of firms. To start, we guess the equilibrium strategy of all firms follows the proposed form in the lemma. Given this guess, we write out the value function of an individual firm j:

$$\rho V_j(\mathbf{a},\mathbf{b}) = \max_{\eta_j} \hat{\pi}(\mathbf{a},\mathbf{b}) - \gamma\frac{\eta_j^2}{2} + \sum_{\ell\neq j}\eta_\ell(\mathbf{a},\mathbf{b})\frac{\partial V_j}{\partial b_\ell} + \mathcal{A}(V_j),$$

where we use $\mathcal{A}$ to denote the differential generator of the productivity process. With this Bellman equation, we want to verify that the value function takes a quadratic form in customer capital:

$$V_j(\mathbf{a},\mathbf{b}) = v_j^0(\mathbf{a}) + v_j^1(\mathbf{a})'\mathbf{b} + \mathbf{b}' v_j^2(\mathbf{a})\mathbf{b}.$$

With this quadratic form, the optimal solution of marketing investment is:

$$\eta_j(\mathbf{a},\mathbf{b}) = \frac{1}{\gamma}\left(v_j^1(\mathbf{a})[j] + 2v_j^2(\mathbf{a})[j,:]\mathbf{b}\right).$$

This strategy indeed follows the proposed form in the lemma. It remains to verify this quadratic form does solve the Belman equation. Since the strategy for all firms is quadratic, every term involving $\eta_\ell(\mathbf{a},\mathbf{b})\frac{\partial V_j}{\partial b_\ell}$ remains quadratic. Since $\hat{\pi}$ is quadratic as well, all terms on the RHS of the Bellan equation is quadratic. This verifies our guess.

A.5 Derivation of the Stationary Distribution

In this section, we derive the details of the stationary distribution from the model's perspective. We take as given the two equations for the evolution of productivity and

customer capital:

$$da_{jt} = -\rho_a a_{jt} dt + \sigma_a \left(dW^a_{jt} - \frac{1}{J-1} \sum_{k \neq j} dW^a_{kt} \right),$$

and

$$db_{jt} = m_a a_{jt} dt + (m_b - \rho_b)\, b_{jt} dt + \sigma_b \left(dW^b_{jt} - \frac{1}{J-1} \sum_{k \neq j} dW^b_{kt} \right).$$

The joint distribution between $(\mathbf{a}, \mathbf{b})$ converges to a normal distribution with mean zero. We will show that:

$$\mathrm{Var}(a_j) = \frac{\sigma_a^2}{2\rho_a}\Big(1 + \tfrac{1}{J-1}\Big), \quad \mathrm{Var}(b_j) = \Big(1 + \tfrac{1}{J-1}\Big) \frac{1}{2(\rho_b - m_b)} \Big(\tfrac{m_a^2}{\rho_a + \rho_b - m_b} \tfrac{\sigma_a^2}{\rho_a} \; + \; \sigma_b^2 \Big),$$

$$\mathrm{Cov}(a_j, b_j) = \frac{m_a}{\rho_a + \rho_b - m_b} \frac{\sigma_a^2}{2\rho_a}\Big(1 + \tfrac{1}{J-1}\Big), \quad \mathrm{Cov}(a_j, a_k) = -\frac{1}{J-1} \frac{\sigma_a^2}{\rho_a},$$

$$\mathrm{Cov}(a_j, b_k) = -\frac{m_a}{\rho_a + \rho_b - m_b} \frac{1}{J-1} \frac{\sigma_a^2}{\rho_a}, \quad \mathrm{Cov}(b_j, b_k) = -\frac{1}{J-1} \frac{m_a}{\rho_b - m_b} \Big(\tfrac{m_a}{\rho_a + \rho_b - m_b} \tfrac{\sigma_a^2}{\rho_a} \; + \; \sigma_b^2 \Big).$$

We start by deriving $\mathrm{Var}(a_j) = \mathbb{E}(a_j^2)$ under stationarity. Applying Itô's lemma to a_j^2, we have

$$a_j^2 = a_j da_j + (da_j)^2, \tag{A2}$$

From the equation for da_j, we can write

$$(da_j)^2 = \sigma_a \left(1 + \frac{1}{J-1}\right) dt,$$

Plugging this second-order term to equation (A2) and dividing both sides by dt, we have

$$\frac{d}{dt}\, \mathbb{E}(a_j^2) = -2\rho_a Var(a_j) + \sigma_a^2 \left(1 + \frac{1}{J-1}\right).$$

Imposing symmetry, we have an equation for $Var(a_j)$, which gives us the formula at the begining.

$$Var(a_j) == \frac{\sigma_a^2}{2\rho_a} \left(1 + \frac{1}{J-1}\right),$$

Using similar logic, we can derive an equation for $Var(b_j)$:

$$0 = \frac{d}{dt}\mathbb{E}(b_j^2) = 2(m_a\mathbb{E}(a_j b_j) + (m_b - \rho_b)\mathbb{E}(b_j^2) + \sigma_b^2\left(1 + \frac{1}{J-1}\right).$$

Imposing symmetry, we write $Var(b_j)$ as

$$Var(b_j) = \frac{1}{2(\rho_b - m_b)}\left(2m_a Cov(a_j, b_j) + \sigma_b^2\left(1 + \frac{1}{J-1}\right)\right).$$

In order to solve for this variance, we need to also derive $Cov(a_j, b_j)$. We now move on to this term. From the law of motions, we can write

$$\frac{d}{dt}(a_j b_j) = m_a a_j^2 + (m_b - \rho_a - \rho_b)_j b_j,$$

which gives us

$$Cov(a_j, b_j) = \frac{m_a Var(a_j)}{\rho_a + \rho_b - m_b} = \frac{m_a}{\rho_a + \rho_b - m_b}\frac{\sigma_a^2}{2\rho_a}\left(1 + \frac{1}{J-1}\right),$$

where in the last equation, we plug in the expression for $Var(a_j)$. Plugging in this covariance for the expression of $Var(b_j)$, we have

$$Var(b_j) = \left(1 + \frac{1}{J-1}\right)\frac{1}{2(\rho_b - m_b)}\left(\frac{m_a^2}{\rho_a + \rho_b - m_b}\frac{\sigma_a^2}{\rho_a} + \sigma_b^2\right).$$

We now move to the other terms, the cross covariances. Starting from $Cov(a_j, a_k)$ where $j \neq k$:

$$0 = \frac{d}{dt}\mathbb{E}(a_j a_k) = -2\rho_a\mathbb{E}(a_j a_k) - 2\frac{1}{J-1}\sigma_a^2,$$

which implies

$$Cov(a_j, a_k) = -\frac{1}{J-1}\frac{\sigma_a^2}{\rho_a}.$$

For $Cov(a_j, b_k)$, similar logic implies that

$$0 = \frac{d}{dt}\mathbb{E}(a_j b_k) = (m_b - \rho_a - \rho_b)\mathbb{E}(a_j b_k) + m_a\mathbb{E}(a_j a_k),$$

and, thus,

$$Cov(a_j, b_k) = \frac{m_a Cov(a_j, a_k)}{\rho_a + \rho_b - m_b} = -\frac{m_a}{\rho_a + \rho_b - m_b}\frac{1}{J-1}\frac{\sigma_a^2}{\rho_a}.$$

Lastly, for $Cov(b_j, b_k)$:

$$0 = \frac{d}{dt}\mathbb{E}(b_j b_k) = 2m_a\mathbb{E}(a_j b_k) + 2(m_b - \rho_b)\mathbb{E}(b_j^2) - \frac{1}{J-1}\sigma_b^2$$

Thus

$$Cov(b_j, b_k) = \frac{m_a Cov(a_j, b_k) - \frac{1}{J-1}\sigma_b^2}{\rho_b - m_b} = -\frac{1}{J-1}\frac{m_a}{\rho_b - m_b}\left(\frac{m_a}{\rho_a + \rho_b - m_b}\frac{\sigma_a^2}{\rho_a} + \sigma_b^2\right)$$

B Data Appendix

This section addresses the set of data sources relevant for the analysis and the data examples that motivate our investigation. Appendix B.1 expands on the details of the merge across datasets.

B.1 Data and Definitions

Merge. Our main merge links USPTO Trademark data with RMS NielsenIQ Scanner data. We proceed by linking firms and products separately. Our merge matches over 80% of sales-weighted products. Some problems still emerge with short-names. We use "tokens" and fuzzy matches to deal with the names. Firms and products follow similar procedures and we discuss them in turn.

Datasets. We use four "parent" datasets in our study. We make use of brand-level data, which has brand ID, firm ID, product group code, sales, prices, and year. We make use of firm-level data which contains firm-level sales by group and year. We also make use of customer-level data, which has household ID, product-level detail, and year. We finally also make use of retail-level data for our local market regressions.

Product Definition. A product is defined as a UPC code (12-digit identifier) linked to the *NielsenIQ* parent firm. Products lie underneath the umbella of a brand. Brands also have brand codes which correspond to an umbrella aggregate across all brands in the main maturity specification to avoid brand $\times$ product variation. We study the life cycle of both products and brands in Appendix C.2.

Transaction Definition. Transactions are defined at both the NielsenIQ and USPTO level. The reason we define transactions using both is as follows. We note that when

we plot the results applying only USPTO transaction information we find as follows. Multiple serial numbers per brand.

Firms. For matching firms, we first standardize on a large set of firm tags, eliminating common firm words, e.g. "CORP", "INC", "ESTABLISHMENT").[1] We then take the cleaned and standardized name and match according to a tokenized bigram matching procedure.

Brands. By focusing on brands, we direct our attention to long-running products held by firms. USPTO Trademark data provides the "tm_name" or the name associated with a registered trademark. RMS NielsenIQ follows a similar format, which has a "brand_name". We join the two by employing a token name matching. For brand names, there are no further removals of tokens beyond the firm-level analysis.[2] For brand age, we focus on the "prior" brand, as in the broader brand umbrella of the production. For transacted brands, we observe the level of the transaction and focus on this.

Transactions. For this paper, we use transaction data to map to brand reallocation. This occurs when a parent company of a given brand is different in period t from $t-1$. We use transactions in both USPTO and RMS NielsenIQ Scanner data. Overall, we get approximately 20% of brand transactions from USPTO Trademark data and 80% of transactions from RMS NielsenIQ. While there are more transactions observed in trademark data, there are some within firm transactions we drop, as we generate a text similarity threshold above which we do not consider transactions. Further, some transactions in USPTO do not have companies that are both identified in the RMS NielsenIQ Scanner Data. We also append transactions manually that are listed in Refinitiv M&A data and matched to firms listed. Approximately 2% of our transactions are manually imputed.

[1]The full list is here ('AB', 'AG', 'BV', 'CENTER', 'CO', 'COMPANY', 'COMPANIES', 'CORP', 'CORPORATION', 'DIV', 'GMBH', 'GROUP', 'INC', 'INCORPORATED', 'KG', 'LC', 'LIMITED', 'LIMITED-PARTNERSHIP', 'LLC', 'LP', 'LTD', 'NV', 'PLC', 'SA', 'SARL', 'SNC', 'SPA', 'SRL', 'TRUST', 'USA', 'KABUSHIKI', 'KAISHA', 'AKTIENGESELLSCHAFT', 'AKTIEBOLAG', 'SE', 'CORPORATIN', 'GROUP', 'GRP', 'HLDGS', 'HOLDINGS', 'COMM', 'INDS', 'HLDG', 'TECH', 'GAISHA', 'AMERICA', 'AMERICAN', 'NORTH', 'OPERATIONS', 'OPERATION', 'DIVISION', 'COMPAGNIE','INTERNATIONAL', 'NORTH AMERICA', 'InBev').

[2]Standardizations include removing any relevant firm names as discussed in the firms section, but does not do any further standardizations and tracks the token grams within each brand name.

C Empirical Appendix

This section explores some additional evidence on a couple core empirical messages on brands and branding, addresses references in the main text, and discusses the robustness of our empirical analysis. We apply broader data from the USPTO to indicate the fact that large firms build large portfolios of brands and their acquired brands drive a larger share of their portfolio. We then discuss the product life cycle with reference to the literature and discuss the integration of the product life cycle with our firm-level analysis. In each case, we explore the robustness of our results to varying definitions.

We focus on the role of reallocation in the firm size distribution in Appendix C.1, returning to the study of the sources of concentration, and evaluate the robustness of the empirical results on firms. We also discuss the types of reassignment in the trademark data, which is in part a please for further investigation on the sources and implications of trademark reallocation. We expand on the brand maturity and life cycle in Appendix C.2, focusing on the interaction of age and sales, and the evidence for the importance of product maturity and sales dispersion over time. We further discuss our connection to the literature on the product life cycle and then turn to the robustness of product-level results.

C.1 Drivers of Firm and Brand

Market Share Decomposition. To implement our decomposition empirically, we build on the measurement framework of Hottman et al. (2016). While they identify a broader set of demand and supply factors, we focus specifically on three key components: customer capital (demand), productivity, and markups. Using their production function estimates as a starting point, we measure each component as follows:

$$\begin{aligned}
\Delta^k \text{Demand} &= (\sigma^F - 1)\Delta^k \log \phi_{jkt} + \left(\frac{\sigma^F - 1}{\sigma^U - 1}\right) \Delta^k \log N_{jkt} \\
\Delta^k \text{Productivity} &= -\omega \log MC + (1-\omega)[\Delta^k \log E - \Delta^k \log P] \\
\Delta^k \text{Scale} &= (1-\omega) \log MC - (1-\omega)[\Delta^k \log E - \Delta^k \log P] \\
\Delta^k \text{Markup} &= \Delta^k \log \mu_{jkt}
\end{aligned}$$

The Δ^k denotes firm difference from the average in group k, which is the locus of competition for firms. σ^F is the elasticity of substitution across firms within a group, while σ^U is the elasticity across products within the firm and ω is the coefficient that

governs diminishing returns at the firm level. We take these parameters from Hottman et al. (2016) in our empirical exercises. ϕ_{jk} is the residual demand at the product level for each firm, with N_{jk} is the number of products. $\ln MC$ is the marginal cost. $\Delta^k E$ and $\Delta^k P$ are the deviations from the group means for expenditure and price at the firm level. μ_{jk} is the markup at the firm level. We now turn to the data construction and summary statistics, which will then allow us to decompose the sources of market share dispersion.

C.2 Brand Maturity

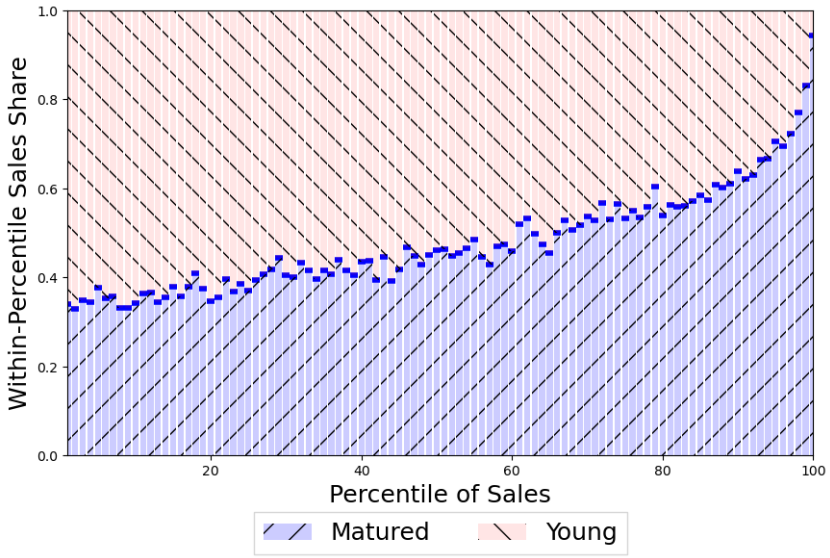
This paper treats brand capital as a transferable form of intangible that would otherwise take time to produce. When firms want to acquire customer capital, they can build it directly, which is costly in time, or acquire it. The core tenets of this point can be found in recent macroeconomic literature. Gourio and Rudanko (2014) and Foster et al. (2016), among many others, have noted that customer capital is not built in a day. This complements the importance of reallocation of existing brands for economic analysis. This section expands on this point by describing the lifecycle of brands. In particular, we note the time-to-build nature of brand capital. By looking at trademark data and NielsenIQ data, one can observe the importance of senior brands. Figure C1 takes data from 2016. We plot the brand percentile in terms of overall sales on the x-axis. On the y-axis, we plot the share of sales in this group that belongs to brands older than 10 years and brands younger than 10 years.[3]

For brands created in 2006 and earlier, they maintain large sales share into the future. By 2016, those brands are still dominant in the top 1% of brands. Within the top 1% of brands, brands created before 2006 make up 92% of sales. Overall, old brands make up over 70% of sales, but only about 1/3rd of products. For the median brand in terms of sales, older brands make up less than half (38%) of total sales.

The dominance of mature brands could come from two forces. First, if few brands achieve such large sales, there may be a selection process. Young brands have less of a chance than old brands to have high customer capital, as the brands that survive to maturity must have a high quality draw. The composition only selects for the best. Second, brands could increase their sales over the life cycle such that only mature brands have significant sales share. We aim to understand this by linking a brand to its specific age. By employing the USPTO-NielsenIQ merge, we are able to extend this life cycle beyond the current work in the literature, but we review our connection to the current

[3]We omit brands with less than $1000 in sales over an entire year, to have only brands that at least have a product line.

Figure C1: Brand Percentile and Maturity



Note: This figure shows the sales share within a percentile bin of products, split by those born before 2006 ("Matured") and after 2006 ("Young").

Source: RMS NielsenIQ Scanner Data.

literature benchmarks here.

Literature Benchmark: The Product Life Cycle. As discussed in the main text, our findings on the brand life cycle are significantly longer than the life cycle discussed in recent work (e.g. Argente et al., 2018). Here, we crosswalk our results to existing work on the product life cycle to benchmark where we diverge. Argente et al. (2018) focus on the life cycle of products applying NielsenIQ Scanner Data. This work is able to identify new products and brands and document their life cycle patterns. However, it is not able to link brands and products to their history, and is thus unable to speak to the longer time horizon of persistent brands. We perform similar life cycle regressions to this current paper in the literature, in particular focusing on defining age in two different ways, to ensure the differences in the age profile does not simply come from applying a dataset with different age measures. Equation (A3) presents the regression:

$$\log y_{it} = \alpha + \sum_{a=0}^{4} \beta_a D_a + \gamma_b + \lambda_t + \epsilon_{it} \tag{A3}$$

Where the coefficients of interest are the coefficients on age (β_a) with controls for cohort and time effects (and an adjustment on cohort from Deaton, 1997). Table C1 engages in the same specification as Argente et al. (2018) in the UPC data (panels 1 and

2) and Trademark merged data (panels 3 and 4) respectively.

Table C1: Log Sales, by NielsenIQ and Trademark Age

	(1) Log Sales	(2) Log Sales	(3) Log Sales	(4) Log Sales
Age 1	0.939***	1.095***	0.917***	0.953***
	(0.00)	(0.00)	(0.00)	(0.00)
Age 2	0.857***	1.159***	1.019***	1.060***
	(0.00)	(0.00)	(0.00)	(0.00)
Age 3	0.632***	1.016***	0.834***	0.832***
	(0.00)	(0.00)	(0.00)	(0.00)
Age 4	0.169***	0.644***	0.412*	0.488***
	(0.00)	(0.00)	(0.00)	(0.00)
N	668993	89203	3402	4136
R^2	0.138	0.179	0.256	0.050
Variation	UPC	Brand-Group	TM Brand	TM Brand-Group

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Notes: Balanced Panel Life Cycle Regressions of Log Sales on Age, utilizing different age sources and different variation. Source: USPTO Trademark and RMS NielsenIQ

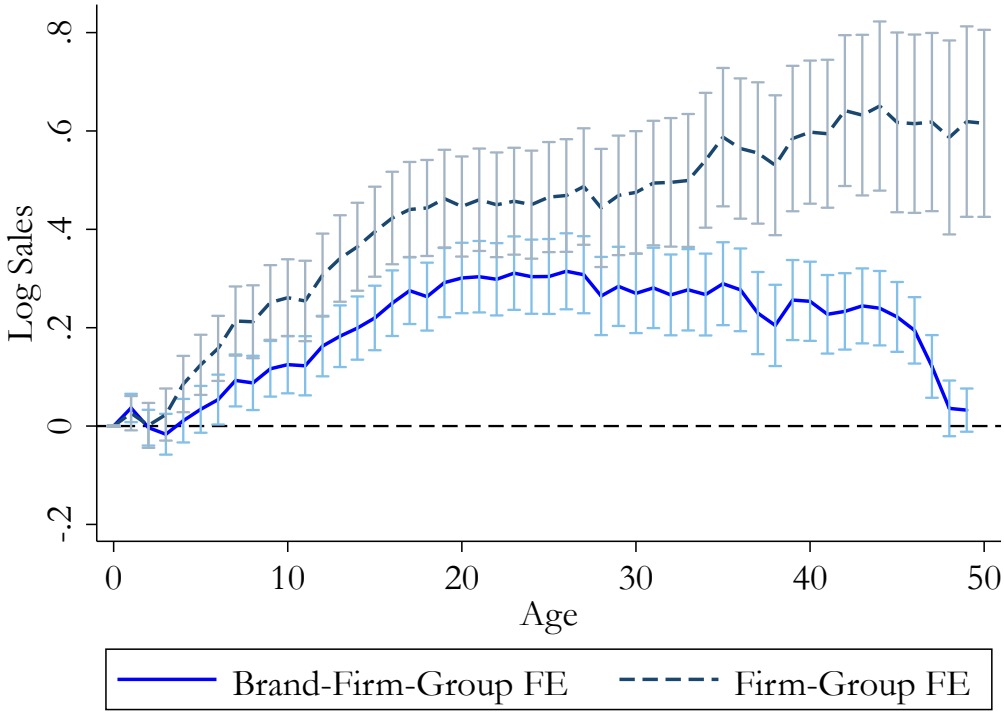
We find that sales increase in early years (often increasing from Age 1 to Age 2), but decline steadily post Age 2. This is broadly consistent with Argente et al. (2018). Note that while at the level of brands and trademarks there are significantly fewer observations, the same general pattern holds. This indicates how age is picking up something similar in our context, yet due to the broader horizon of historical data we are able to connect brands to their histories, indicating a significantly longer brand life cycle than found in Argente et al. (2018). We also show here similar general trends as in the main text when we evaluate the life cycle of products, controlling for brand-firm-group level.

Figure C3 evaluates the life cycle profile within a given product group code. We follow the regression in the main text, except in prices we weight by sales share. Equation (A4) illustrates the structure of the regression.

$$\log y_{ijkt} = \alpha + \sum_{a=1}^{50} \beta_a D_a + \gamma_b + \lambda_t + \theta_{ikj(i)} + \epsilon_{ijkt} \tag{A4}$$

The regression in Equation (A4) considers the sales and prices of brand i with firm j in group k at time t, $\log y_{ijkt}$ as a function of a constant (α), brand age indicators from 1 to

Figure C2: Life Cycle Regressions



Note: Plots of log sales on age regression coefficients, controlling for brand-group and controlling only for firm-group. 95% confidence intervals plotted alongside coefficients. Source: RMS NielsenIQ and USPTO.

50, D_a, and fixed effects for cohort (γ_b) and time (λ_t).[4] The $\theta_{ikj(i)}$ indicates a brand-group or firm-group fixed-effect. Figure C3 plots the regressions by age coefficient β_a.

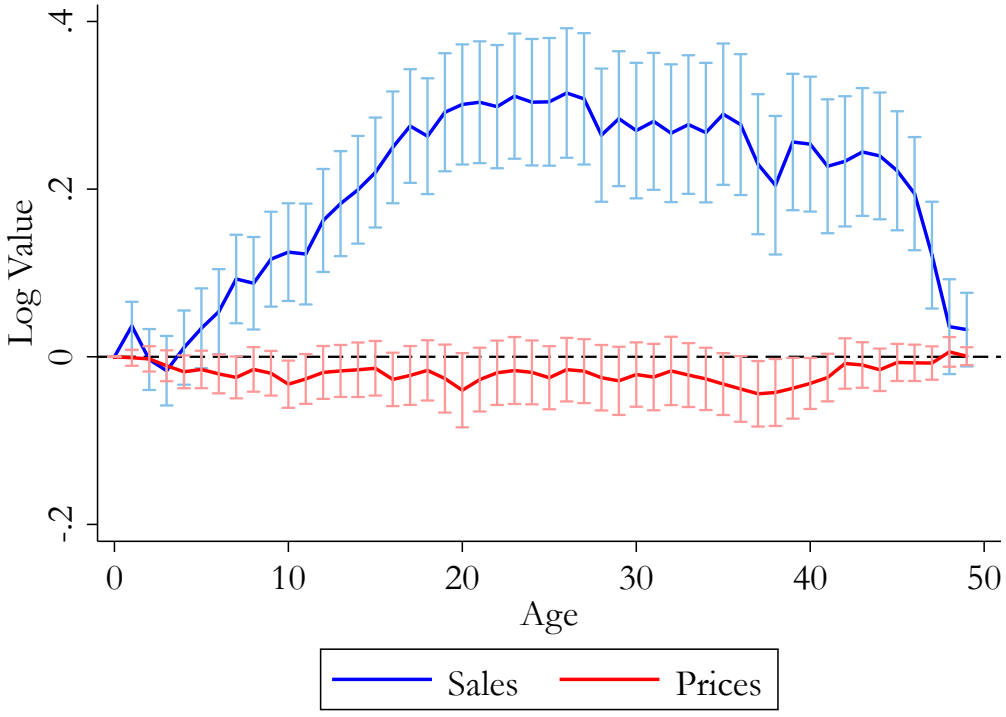
We note that the inverted-U profile is still persistent within group, though with a slightly lower peak than in the brand's overall life cycle. We also note that the life cycle of prices shows on average somewhat minimal activity for the brand across age. This means that the strategic pricing firms engage in does not appear to be correlated with age, though as we have noted from events there are shifts in prices, consistent with previous evidence in the literature.

D Estimation Appendix

This appendix details our two-step estimation procedure for recovering the structural parameters of our model. We first estimate the autoregressive parameters using Arellano-Bond dynamic panel methods, then recover variance components through GMM mo-

[4]Given the linear relationship between age, time, and cohort, we follow a method developed by Deaton (1997) to correct for this issue. The normalization orthogonalizes the cohort trends such that growth components move with age and time effects.

Figure C3: Life Cycle Regressions, Prices and Sales within Group



Note: Plots of log sales and prices on age regression coefficients, controlling for brand-group, as in Equation (A4). 95% confidence intervals plotted alongside coefficients. Source: RMS NielsenIQ and USPTO.

ment matching.

D.1 Model Structure

Our empirical model decomposes both productivity and brand capital into persistent and transitory components. For firm j at time t:

$$a_{jt} = a^P_{jt} + \zeta^a_{jt} + \alpha_j$$
$$b_{jt} = b^P_{jt} + \zeta^b_{jt} + \beta_j$$

where ζ^a_{jt} and ζ^b_{jt} are the transitory shocks, they are independent of the current and past persistent shocks; α_j and β_j are the fixed effects; ν^a_{jt} and ν^b_{jt} are the persistent shocks. In this note, we assume that we have already recovered ρ_{aa}, ρ_{ba}, ρ_{bb} from Arellano-Bond and we want to recover the variances $\left(\sigma^2_{aT}, \sigma^2_{bT}, \sigma^2_{aP}, \sigma^2_{aP}\right)$. We will recover these parameters from the auto-covariance function of the first difference. More precisely, we

denote

$$\Sigma_k^x = Cov\left(x_{jt} - x_{jt-1}, x_{jt-k} - x_{jt-k-1}\right), \quad x = a, b$$

It is useful to write out the values in terms of past realizations. More precisely, for any $k \geq 1$

$$a_{jt}^P = \rho_{aa}^k a_{jt-k}^P + error$$

where the error is independent to a_{jt-k}^P. We can do the same with b:

$$b_{jt}^P = \rho_{bb}^k b_{jt-k}^P + \left(\sum_{l=0}^{k-1} \rho_{bb}^l \rho_{ba} \rho_{aa}^{k-l-1}\right) a_{jt-k}^P + error$$

For notation simplicity, lets define

$$R_k = \sum_{l=0}^{k-1} \rho_{bb}^l \rho_{ba} \rho_{aa}^{k-l-1}$$

Thus

$$b_{jt}^P = \rho_{bb}^k b_{jt-k}^P + R_k a_{jt-k}^P + error$$

Productivity To start, we impose stationarity and denote the stationary variance of a^P as $Var(a^P)$:

$$Var\left(a^P\right) = \rho_{aa}^2 Var\left(a^P\right) + \sigma_{aP}^2 \implies Var\left(a^P\right) = \frac{\sigma_{aP}^2}{1 - \rho_{aa}^2}$$

We start with $k = 0$:

$$\begin{aligned}
\Sigma_0^a = Var\left(a_{jt} - a_{jt-1}\right) &= Var\left(a_{jt}^P - a_{jt-1}^P + \zeta_{jt}^a - \zeta_{jt-1}^a\right) \\
&= Var\left(a_{jt}^P - a_{jt-1}^P\right) + Var\left(\zeta_{jt}^a - \zeta_{jt-1}^a\right) \\
&= Var\left(\rho_{aa} a_{jt-1}^P + \nu_{jt}^a - a_{jt-1}^P\right) + 2\sigma_{aT}^2 \\
&= (1 - \rho_{aa})^2 Var\left(a^P\right) + \sigma_P^2 + 2\sigma_T^2
\end{aligned}$$

where the first equality uses the definition, the second equality uses the fact ζ^a are independent of the permanent components, the third equality uses the definition of the

variance of the transitory variance, and the last equation uses fact ν_{jt}^a is independent of past values. We denote the stationary value of the persistent variance as $Var(a^P)$. Plugging the stationary variance we have

$$\Sigma_0^a = \frac{2}{1+\rho_a}\sigma_{aP}^2 + 2\sigma_{aT}^2$$

Now we move on to $k = 1$:

$$\begin{aligned}
\Sigma_0^a =& Cov\left(a_{jt}^P - a_{jt-1}^P + \zeta_{jt}^a - \zeta_{jt-1}^a, a_{jt-1}^P - a_{jt-2}^P + \zeta_{jt-1}^a - \zeta_{jt-2}^a\right) \\
=& Cov\left(a_{jt}^P - a_{jt-1}^P, a_{jt-1}^P - a_{jt-2}^P\right) + Cov\left(\zeta_{jt}^a - \zeta_{jt-1}^a, \zeta_{jt-1}^a - \zeta_{jt-2}^a\right) \\
=& Cov\left(a_{jt}^P - a_{jt-1}^P, a_{jt-1}^P - a_{jt-2}^P\right) - \sigma_{aT}^2 \\
=& Cov\left(a_{jt}^P, a_{jt-1}^P\right) - Var\left(a^P\right) - Cov\left(a_{jt}^P, a_{jt-2}^P\right) + Cov\left(a_{jt-1}^P, a_{jt-2}^P\right) - \sigma_{aT}^2
\end{aligned}$$

where the first equality uses the fact ζ^a is independent of the persistent values, the second equality write out the variance, the third equality expand the terms. Now we inspect the terms one by one. Using the sequential form we derived:

We now unpack the terms

$$Cov\left(a_{jt}^P, a_{jt-1}^P\right) = Cov\left(a_{jt-1}^P, a_{jt-2}^P\right) = \rho_{aa} Var(a^P) = \rho_{aa} Var(a)$$

$$Cov\left(a_{jt}^P, a_{jt-2}^P\right) = \rho_{aa}^2 Var(a_{jt-2}^P) = \rho_{aa}^2 Var(a)$$

Plugging these values we have

$$\Sigma_1^a = -\frac{1-\rho_a}{1+\rho_a}\sigma_{aP}^2 - \sigma_{aT}^2$$

Now we move on the $k \geq 2$

$$\begin{aligned}\Sigma_k^a =& Cov\left(a_{jt}^P - a_{jt-1}^P + \zeta_{jt}^a - \zeta_{jt-1}^a, a_{jt-k}^P - a_{jt-k-1}^P + \zeta_{jt-k}^a - \zeta_{jt-k-1}^a\right)\\ =& Cov\left(a_{jt}^P - a_{jt-1}^P, a_{jt-k}^P - a_{jt-k-1}^P\right)\\ =& \rho_{aa}^k Var\left(a^P\right) - \rho_{aa}^{k-1} Var\left(a^P\right) - \rho_{aa}^{k+1} Var\left(a^P\right) + \rho_{aa}^k Var\left(a^P\right)\\ =& -\rho_{aa}^{k-1}\frac{1-\rho_{aa}}{1+\rho_{aa}}\sigma_{aP}^2\end{aligned}$$

This gives us the general formula

$$\Sigma_k^a = \begin{cases} \frac{2}{1+\rho_a}\sigma_{aP}^2 + 2\sigma_{aT}^2 & k=0 \\ -\frac{1-\rho_a}{1+\rho_a}\sigma_{aP}^2 - \sigma_{aT}^2 & k=1 \\ -\rho_{aa}^{k-1}\frac{1-\rho_{aa}}{1+\rho_{aa}}\sigma_{aP}^2 & k>1 \end{cases}$$

Demand Similarly, we start by imposing stationarity:

$$\left(1-\rho_{bb}^2\right) Var\left(b^P\right) = \rho_{ba}^2 Var\left(a^P\right) + \rho_{ab}\rho_{ab} Cov\left(a^P, b^P\right) + \sigma_{aP}^2$$

$$Cov(a, b^P) = \rho_{aa}\rho_{ba} Var(a^P) + \rho_{aa}\rho_{bb} Cov(a^P, b^P)$$

With $Var(a)$ known from earlier steps. These two equations pin down $Var\,(b)$ and $Cov(a,b)$. From here on, we treat the two values as known.

$k=0$

$$\begin{aligned}\Sigma_0^b &= Var\left(b_{jt}^P - b_{jt-1}^P + \zeta_{jt}^b - \zeta_{jt-1}^b\right)\\ \text{(independence of } \zeta^b\text{)} &= Var\left(b_{jt}^P - b_{jt-1}^P\right) + Var\left(\zeta_{jt}^b - \zeta_{jt-1}^b\right)\\ \text{(expand terms)} &= Var\left(\rho_{ba} a_{jt-1}^P + (\rho_{bb}-1) b_{jt-1}^P + \nu_{jt}^b\right) + 2\sigma_{bT}^2\\ \text{(expand terms)} &= \rho_{ba}^2 Var(a^P) + (1-\rho_{bb})^2 Var\left(b^P\right) + \sigma_{bP}^2 - 2\rho_{ba}(1-\rho_{bb}) Cov(a^P, b^P) + 2\sigma_{aT}^2\end{aligned}$$

For higher order terms, it is convenient to find a sequential form for b_{jt}^P. Using the definition, for any $k \geq 1$, we can write:

where the error terms are independent of the values from $t-k$.

$k=1$

$$\begin{aligned}
\Sigma_1^b =& Cov\left(b_{jt}^P - b_{jt-1}^P + \zeta_{jt}^b - \zeta_{jt-1}^b, b_{jt-1}^P - b_{jt-2}^P + \zeta_{jt-1}^b - \zeta_{jt-2}^b\right) \\
\text{(independence of } \zeta^b\text{)} =& Cov\left(b_{jt}^P - b_{jt-1}^P, b_{jt-1}^P - b_{jt-2}^P\right) - \sigma_T^2 \\
\text{(expand terms)} =& Cov\left(b_{jt}^P, b_{jt-1}^P\right) - Var\left(b_{jt-1}^P\right) - Cov\left(b_{jt}^P, b_{jt-2}^P\right) + Cov\left(b_{jt-1}^P, b_{jt-2}^P\right) - \sigma_T^2
\end{aligned}$$

Now we can use the sequential form.

$$\begin{aligned}
Cov\left(b_{jt}^P, b_{jt-1}^P\right) = Cov\left(b_{jt-1}^P, b_{jt-2}^P\right) &= Cov\left(\rho_{bb} b_{jt-1}^P + R_1 a_{jt-1}^P + error, b_{jt-1}^P\right) \\
&= R_1 Cov\left(a^P, b^P\right) + \rho_{bb} Var\left(b^P\right)
\end{aligned}$$

$$\begin{aligned}
Cov\left(b_{jt}^P, b_{jt-2}^P\right) &= Cov\left(\rho_{bb}^2 b_{jt-2}^P + R_2 a_{jt-2}^P + error, b_{jt-2}^P\right) \\
&= R_2 Cov\left(a^P, b^P\right) + \rho_{bb}^2 Var\left(b^P\right)
\end{aligned}$$

Plugging in

$$\Sigma_1^b = -\left(1 - \rho_{bb}\right)^2 Var\left(b^P\right) + \left(2R_1 - R_2\right) Cov\left(a^P, b^P\right) - \sigma_T^2$$

$k > 1$

$$\begin{aligned}
\Sigma_k^b =& Cov\left(b_{jt}^P - b_{jt-1}^P + \zeta_{jt}^b - \zeta_{jt-1}^b, b_{jt-k}^P - b_{jt-k-1}^P + \zeta_{jt-k}^b - \zeta_{jt-k-1}^b\right) \\
\text{(independence of } \zeta^b\text{)} =& Cov\left(b_{jt}^P - b_{jt-1}^P, b_{jt-k}^P - b_{jt-k-1}^P\right) \\
=& Cov\left(b_{jt}^P, b_{jt-k}^P\right) - Cov\left(b_{jt-1}^P, b_{jt-k}^P\right) - Cov\left(b_{jt}^P, b_{jt-k-1}^P\right) + Cov\left(b_{jt-1}^P, b_{jt-k-1}^P\right) \\
=& \left(2\rho_{bb}^k - \rho_{bb}^{k-1} - \rho_{bb}^{k+1}\right) Var\left(b^P\right) + \left(2R_k - R_{k-1} - R_{k+1}\right) Cov(a, b) \\
=& -\rho_{bb}^{k-1}\left(1 - \rho_{bb}\right)^2 Var(b^P) + \left(2R_k - R_{k-1} - R_{k+1}\right) Cov(a, b)
\end{aligned}$$

This gives us the general formula

$$\Sigma_k^b = \begin{cases} \rho_{ba}^2 Var(a) + (1-\rho_{bb})^2 Var(b) + \sigma_{bP}^2 - 2\rho_{ba}(1-\rho_{bb}) Cov(a,b) + 2\sigma_{aT}^2 & k=0 \\ -(1-\rho_{bb})^2 Var\left(b^P\right) + (2R_1 - R_2) Cov\left(a^P, b^P\right) - \sigma_T^2 & k=1 \\ -\rho_{bb}^{k-1}(1-\rho_{bb})^2 Var(b^P) + (2R_k - R_{k-1} - R_{k+1}) Cov(a,b) & k>1 \end{cases}$$

D.2 First Stage: Arellano-Bond Estimation

We first estimate the autoregressive parameters $(\rho_{aa}, \rho_{ba}, \rho_{bb})$ using the Arellano-Bond GMM estimator. The key insight of this approach is using lagged levels as instruments for first differences to address the correlation between fixed effects and regressors.

D.2.1 Productivity Persistence

For productivity, we estimate:

$$\Delta a_{jt} = \rho_{aa} \Delta a_{jt-1} + \Delta \epsilon_{jt}$$

where $\Delta \epsilon_{jt} = \Delta \nu_{jt}^a + \Delta \zeta_{jt}^a$. The moment conditions are:

$$E[a_{jt-s} \Delta \epsilon_{jt}] = 0 \quad \text{for } s \geq 2$$

D.2.2 Brand Capital Dynamics

For brand capital, we estimate:

$$\Delta b_{jt} = \rho_{ba} \Delta a_{jt-1} + \rho_{bb} \Delta b_{jt-1} + \Delta \eta_{jt}$$

where $\Delta \eta_{jt} = \Delta \nu_{jt}^b + \Delta \zeta_{jt}^b$. The moment conditions are:

$$E[x_{jt-s} \Delta \eta_{jt}] = 0 \quad \text{for } x \in \{a, b\}, s \geq 2$$

D.2.3 Implementation and Results

We implement the estimator using firms with at least 4 consecutive years of data. The baseline specification uses up to 4 lags as instruments and a two-step efficient GMM estimator with Windmeijer-corrected standard errors.

Our estimates are:

$$\hat{\rho}_{aa} = 0.73 \quad (0.04)$$
$$\hat{\rho}_{ba} = 0.24 \quad (0.05)$$
$$\hat{\rho}_{bb} = 1.02 \quad (0.01)$$

Standard errors in parentheses. Specification tests support the validity of our instruments:

- Hansen J-test fails to reject overidentifying restrictions (p = 0.42)
- AR(2) test finds no evidence of second-order serial correlation in residuals (p = 0.38)

The high estimate of ρ_{bb} suggests brand capital is highly persistent, while the positive ρ_{ba} indicates productivity improvements help build brand capital. These first-stage estimates are treated as known in the second stage of estimation. Before turning to the second stage, we discuss identification concerns.

Identification Concerns. Our estimation strategy faces several potential threats to identification that we consider here. We organize these challenges into four main categories and detail our approaches to addressing each concern.

A. Reverse Causality. The primary identification concern is potential feedback from brand capital to productivity. While our baseline specification assumes productivity evolves independently, firms might adjust their production processes in response to brand-related shocks. We address this by estimating an expanded system that allows for bidirectional effects:

$$\text{Baseline:} \quad a_{jt} = \rho_{aa} a_{jt-1} + \zeta_j + \nu_{ajt} \tag{A5}$$
$$\text{Alternative:} \quad a_{jt} = \rho_{aa} a_{jt-1} + \rho_{ab} b_{jt-1} + \zeta_j + \nu_{ajt} \tag{A6}$$

Through vector autoregression estimation and systematic Granger causality testing, we examine the timing and direction of relationships between productivity and brand capital innovations. The results, detailed in Section 4.2, suggest limited evidence of reverse causality affecting our main estimates.

B. Unobserved Heterogeneity. A second concern is that time-varying shocks might simultaneously affect both productivity and brand capital, violating our moment conditions.

For example, quality improvements could drive both measures. We implement an instrumental variables approach that leverages industry-level variation:

$$\text{First Stage:} \quad \Delta a_{jt} = \pi_1 Z_{jt} + \eta_{jt} \tag{A7}$$

$$\text{Second Stage:} \quad \Delta b_{jt} = \rho_{ba} \Delta \hat{a} jt - 1 + \rho bb \Delta b_{jt-1} + \epsilon_{jt} \tag{A8}$$

where Z_{jt} includes industry-level shifters of productivity that are plausibly exogenous to firm-specific brand capital. We further examine heterogeneity across market structures to validate our identification strategy.

C. Validity of Lag Instruments. The Arellano-Bond approach relies crucially on the validity of lagged levels as instruments. We conduct extensive specification testing through Hansen tests of overidentifying restrictions and serial correlation tests in first-differenced errors. Additionally, we examine sensitivity to instrument set construction by varying lag depth and implementing collapsed instrument matrices. These robustness checks support the validity of our identification strategy.

D. Non-linear Dynamics. Our linear AR(1) specification may miss important non-linearities in the evolution of productivity and brand capital. We estimate an expanded specification allowing for state-dependent parameters:

$$\Delta b_{jt} = \rho_{ba}(\phi(s_{jt-1}))\Delta a_{jt-1} + \rho_{bb}(\phi(s_{jt-1}))\Delta b_{jt-1} + \Delta \nu_{bjt} \tag{A9}$$

where $\phi(s_{jt-1})$ is a flexible function of firm market share. We complement this analysis with threshold regression models and non-parametric tests for non-linear dependence. The results suggest our baseline linear specification captures the first-order dynamics while missing limited higher-order effects.

D.3 Second Stage: GMM Moment Matching

Given these first-stage estimates, we recover the variance parameters $(\sigma^2_{aT}, \sigma^2_{bT}, \sigma^2_{aP}, \sigma^2_{bP})$ by matching theoretical and empirical autocovariance functions.

D.3.1 Autocovariance Functions

We denote the autocovariance of first differences as:

$$\Sigma^x_k = Cov(x_{jt} - x_{jt-1}, x_{jt-k} - x_{jt-k-1}), \quad x = a, b$$

The theoretical autocovariance functions are:

For productivity ($k \geq 0$):

$$\Sigma_k^a = \begin{cases} \frac{2}{1+\rho_a}\sigma_{aP}^2 + 2\sigma_{aT}^2 & k = 0 \\ -\frac{1-\rho_a}{1+\rho_a}\sigma_{aP}^2 - \sigma_{aT}^2 & k = 1 \\ -\rho_{aa}^{k-1}\frac{1-\rho_{aa}}{1+\rho_{aa}}\sigma_{aP}^2 & k > 1 \end{cases}$$

For brand capital ($k \geq 0$):

$$\Sigma_k^b = \begin{cases} \rho_{ba}^2 Var(a) + (1-\rho_{bb})^2 Var(b) + \sigma_{bP}^2 - 2\rho_{ba}(1-\rho_{bb})Cov(a,b) + 2\sigma_{bT}^2 & k = 0 \\ -(1-\rho_{bb})^2 Var(b^P) + (2R_1 - R_2)Cov(a^P, b^P) - \sigma_{bT}^2 & k = 1 \\ -\rho_{bb}^{k-1}(1-\rho_{bb})^2 Var(b^P) + (2R_k - R_{k-1} - R_{k+1})Cov(a,b) & k > 1 \end{cases}$$

where R_k captures cross-persistence effects:

$$R_k = \sum_{l=0}^{k-1} \rho_{bb}^l \rho_{ba} \rho_{aa}^{k-l-1}$$

D.3.2 Implementation Details

Key numerical considerations:

- Near unit root in brand capital ($\rho_{bb} \approx 1$) requires careful handling of variance terms
- Regularization in matrix inversions:

$$A = \begin{bmatrix} 1 - \rho_{bb}^2 + 1e-6 & -\rho_{ba} \\ -\rho_{aa}\rho_{ba} & 1 - \rho_{aa}\rho_{bb} \end{bmatrix}$$

- Treatment of initial conditions

D.3.3 Results

Our second-stage estimates are:

$\sigma_{aP} = 0.5534$ (Productivity innovation std)
$\sigma_{aT} = 0.3348$ (Transitory productivity shock std)
$\sigma_{bP} = 2.1619$ (Brand capital innovation std)
$\sigma_{bT} = 0.0034$ (Transitory brand shock std)

The estimates reveal:

1. Brand capital innovations have larger variance than productivity innovations
2. Productivity has significant transitory components while brand capital is primarily persistent
3. Model fits productivity autocovariance well but shows some deviation in brand capital autocovariance at longer lags

D.4 Identification Discussion

The separate identification of variance components comes from:

1. Transitory shocks $(\sigma^2_{aT}, \sigma^2_{bT})$ primarily identified by $k = 1$ autocovariances
2. Persistent shock variances $(\sigma^2_{aP}, \sigma^2_{bP})$ identified by decay rates at higher lags
3. Cross-persistence ρ_{ba} helps identify relative contribution of productivity versus brand capital shocks

The level difference in brand capital autocovariance suggests potential model extensions:

1. Non-stationary brand capital processes
2. Richer cross-persistence structures
3. Time-varying parameters

D.5 Different Selection Criteria

For our main results, we use all firms and do not condition on a balanced panel. Qualitatively, we find very similar results when we

MORE ON AB:

To address the incidental parameters problem arising from firm fixed effects and the dynamic structure, we employ first differences:

$$\Delta a_{jt} = \rho_{aa}\Delta a_{jt-1} + \Delta \nu_{ajt} \tag{A10}$$

$$\Delta b_{jt} = \rho_{ba}\Delta a_{jt-1} + \rho_{bb}\Delta b_{jt-1} + \Delta \nu_{bjt} \tag{A11}$$

Table D2: Persistence Parameters by Firm Size and Time Period

Size	Year	ρ_{aa}	ρ_{ba}	ρ_{bb}
$1/10^2$	0	0.641	0.186	0.841
$1/10^2$	7	0.671	0.164	0.853
$1/10^2$	13	0.569	0.194	0.853
$1/10^3$	0	0.577	0.397	1.043
$1/10^3$	7	0.616	0.384	1.025
$1/10^3$	13	0.776	0.244	0.955
$1/10^4$	0	0.735	0.249	1.056
$1/10^4$	7	0.749	0.285	1.037
$1/10^4$	13	0.813	0.161	0.966
$1/10^5$	0	0.733	0.248	1.039
$1/10^5$	7	0.755	0.292	1.034
$1/10^5$	13	0.819	0.143	0.945
$1/10^6$	0	0.732	0.241	1.021
$1/10^6$	7	0.745	0.292	1.022
$1/10^6$	13	0.803	0.157	0.932

Notes: ρ_{aa} represents productivity persistence, ρ_{ba} represents the effect of lagged productivity on demand, and ρ_{bb} represents demand persistence. Size indicates minimum average firm size, and Year indicates minimum number of years a firm must be present in the sample (13 years is a balanced panel).

The first-differenced specification eliminates the firm fixed effects but introduces correlation between Δa_{jt-1} and $\Delta \nu_{ajt}$ through the shared ν_{ajt-1} term. Following Arellano and Bond (1991), we construct moment conditions using lagged levels as instruments:

$$E[a_{jt-s}\Delta\nu_{ajt}] = 0 \quad \text{for } s \geq 2 \tag{A12}$$

$$E[b_{jt-s}\Delta\nu_{bjt}] = 0 \quad \text{for } s \geq 2 \tag{A13}$$

These moment conditions exploit the assumption that productivity and brand capital levels from $t-2$ and earlier are uncorrelated with the differenced errors. We estimate the system using two-step GMM with optimal weighting matrix and windmeijer-corrected standard errors to account for potential finite-sample bias. The validity of our estimation approach relies on two key assumptions. First, sequential exogeneity, as follows,

$$E[\nu_{ajt}|a_{j1}, ..., a_{jt-1}, \zeta_j] = 0 \tag{A14}$$

$$E[\nu_{bjt}|b_{j1}, ..., b_{jt-1}, a_{j1}, ..., a_{jt-1}, \xi_j] = 0, \tag{A15}$$

and no serial correlation in the error terms:

$$E[\nu_{ajt}\nu_{ajt-s}] = 0 \quad \text{for } s \geq 1 \tag{A16}$$

$$E[\nu_{bjt}\nu_{bjt-s}] = 0 \quad \text{for } s \geq 1. \tag{A17}$$

We test these assumptions using the Arellano-Bond test for serial correlation in the first-differenced errors and the Hansen test of overidentifying restrictions.