Investment in Demand and Dynamic Competition for Customers

Joachim Hubmer
University of Pennsylvania

Lukas Nord
University of Pennsylvania

September 15, 2025

Abstract. How demand is allocated across firms is a central determinant of aggregate productivity. Firms affect this allocation through their pricing and their spending on sales and marketing—investment in demand. We develop a framework in which investment in demand shapes allocations directly, by matching customers to suppliers, and indirectly, by altering price competition. A quantitative version generates rich industry dynamics and matches salient empirical facts on how firms compete for customers. We highlight two main results: First, firms over-invest in marketing due to a business-stealing externality. A social planner would reduce aggregate marketing spending by 22%. Second, marketing has a pro-competitive effect, mitigating the welfare cost of imperfect competition. Equilibrium interactions of marketing and pricing create complementarities between policies that target over-investment in demand and markup distortions. Quantitatively, the observed increase in marketing since the 1980s can account for the observed rise in industry concentration without raising market power, while boosting productivity and GDP.

Contact: jhubmer@sas.upenn.edu; lksnord@sas.upenn.edu. We thank Guido Menzio as well as conference and workshop participants at Einaudi, Penn, Philadelphia Fed, SED 2025, and USC Marshall for useful comments and suggestions.

1 Introduction

The allocation of economic activity across firms is a central determinant of aggregate productivity and output (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009). In markets with imperfect competition, heterogeneous markups distort not only the level of production but also the distribution of demand across firms (Edmond, Midrigan and Xu, 2022; Baqaee and Farhi, 2020). Standard theories of misallocation typically view demand as static, with imperfect competition arising from product differentiation.[1] Recent work instead emphasizes that firms accumulate demand dynamically by investing in sales and marketing to build their customer base (Argente et al., 2025; Fitzgerald, Haller and Yedid-Levi, 2024). This *investment in demand* now rivals outlays on physical capital and constitutes the largest component of intangible investment (He, Mostrom and Sufi, 2024). Firms' willingness to invest in marketing to expand their customer base aligns with theories that highlight search frictions as a source of market power in product markets (Menzio, 2024a,b).

This paper studies how firms' investment in demand affects allocative efficiency in an economy where demand is dynamic and market power stems from search frictions. Our main contribution is a novel framework in which pricing and competition among firms is endogenous to their investment in marketing. A quantitative version replicates key empirical patterns of how firms expand demand. We highlight two main results: First, on average firms over-invest in marketing to accumulate customers due to a business-stealing externality. Second, marketing has a pro-competitive effect that lowers the misallocative cost of imperfect competition. Calibrating the model to match an increase in marketing expenditures since the 1980s, we show that lower marketing cost can explain higher industry concentration without greater market power, and that this shift has contributed positively to GDP growth.

We model competition for customers in general equilibrium with households, final good producers, and intermediate good producers. Households supply labor, invest in physical capital, hold claims to all firms, and consume the final good. Final good firms combine intermediates into the final good through a continuum of divisions, each making imperfectly substitutable contributions. These divisions are the customers in the market for intermediates. Intermediate producers—the suppliers—are heterogeneous in stochastic productivity, set prices, and spend on marketing to reach potential customers. The market for intermediates is subject to search frictions: each division must match with a supplier to procure inputs, but can only purchase from suppliers they are in contact with. Upon meeting a supplier, customers observe only its price and decide whether to start purchasing. Customers and

[1] Afrouzi, Drenik and Kim (2023) is a notable exception, considering misallocation in an economy with differentiated products and dynamic demand. We relate our contribution to their seminal work in detail below.

suppliers form persistent relationships that can be unilaterally dissolved at any time, and existing customers may be poached by other suppliers. Each division contracts with at most one supplier, but final good firms substitute across divisions, allowing variation in demand along both the extensive (number of customers) and intensive (sales per customer) margin.

We solve for a Bayesian Markov equilibrium in which customers form beliefs about their suppliers' payoff-relevant state from observed prices. We focus on separating equilibria where prices are strictly decreasing in productivity. Assuming constant returns to scale in the marketing technology, firms' choices of price and marketing do not depend on their existing customer base. In equilibrium, heterogeneous markups arise from firms' marketing decisions. When setting prices, firms internalize that their customers may be contacted by competitors. Greater marketing by rivals increases the perceived elasticity of demand, inducing firms to set lower prices and markups relative to the static optimum. Markups are heterogeneous, reflecting how strongly each firm is threatened by competitors given its production costs.

The framework matches several empirical facts emphasized in recent work on how firms grow demand. First, it replicates the relative contribution of the number of customers versus sales per customer to overall sales variation across firms (Einav et al., 2021), which disciplines demand elasticities at both margins. Second, it generates a positive relation between sales per customer and markups—but no relation between markups and the number of customers—in line with observed variation in market power (Afrouzi, Drenik and Kim, 2023). Third, firms expand their customer base through marketing while prices remain largely insensitive to customer base size, consistent with evidence on market share growth (Fitzgerald, Haller and Yedid-Levi, 2024; Argente et al., 2025). Fourth, this is despite the number of customers being responsive to prices, making prices allocative at the customer margin (Paciello, Pozzi and Trachter, 2019; Roldan-Blanco and Gilbukh, 2021).

We asses efficiency by comparing the social planner's allocation to the decentralized equilibrium. The planner faces the same technological constraints and search frictions but can directly assign marketing, match formation, and quantities within each match. We show analytically that in equilibrium, standard markup distortions arise as in e.g. Edmond, Midrigan and Xu (2022): the aggregate markup drives a wedge in households' factor supply condition, while dispersion in markups distorts the intensive margin allocation of demand across divisions. By contrast, match formation at the extensive margin is efficient as in Menzio (2024a,b): customers always switch to more productive suppliers when matched, since prices fall monotonically with productivity. Marketing investment, however, is generally inefficient. Two externalities operate in opposite directions. First, firms under-invest because they bear the full cost of contacting customers but do not capture all static surplus, mirroring the failure of the Hosios (1990) condition for efficiency in search markets. Second, they over-invest because marketing diverts customers from competitors, a business-stealing externality that firms do not internalize.

Quantitatively, the business stealing externality dominates, and the social planner would reduce aggregate marketing expenditures by 22% relative to the calibrated equilibrium. The magnitude and sign of the inefficiency is heterogeneous across firms. While the planner would cut marketing spending for most suppliers, it would increase marketing for the 3% most productive firms. This variation reflects differences in both the strength of the business-stealing externality and the surplus customers obtain along the productivity distribution. Eliminating the marketing inefficiency raises welfare by about 0.5% along the transition from the decentralized equilibrium to the planner's allocation.

The framework allows us to revisit the quantitative cost of misallocation from markups. Here, markups arise due to customers' inability to access all suppliers simultaneously. Eliminating the aggregate markup raises welfare by 1.5% along the transition. The implied steady-state gain is much larger (8.5%) but does not account for the buildup of capital. Eliminating distortions from markup dispersion adds another 1.6% along the transition. This result aligns quantitatively with Edmond, Midrigan and Xu (2022), albeit for different reasons: Their finding of a small cost of markup dispersion (as compared to, e.g., Baqaee and Farhi (2020)) relies on targeting a low correlation between markups and total sales. We target the higher covariance between markups and sales per customers as in Afrouzi, Drenik and Kim (2023), generating higher dispersion in markups. We still find comparable welfare losses from markup dispersion because we separate the *source* of market power from the parameter governing its *cost*. In our model, market power stems from search frictions, while the cost of markups depends on the elasticity of demand per customer. This elasticity is identified from variation in sales per customer, which is empirically small. If instead we counterfactually imposed a higher elasticity consistent with the aggregate markup, the productivity loss from markup dispersion would rise from 1.2% to 6.4%. Finally, as markups are endogenous to marketing, counterfactually doubling marketing costs increases the welfare costs of markups by about a quarter. Taken together, eliminating aggregate markups, removing markup dispersion, and correcting marketing inefficiencies would raise welfare by 3.6% along the transition.

We solve for the welfare-maximizing product market policy and uncover a complementarity between taxing marketing and subsidizing production. We restrict attention to two constant, flat instruments: a marketing tax and a production subsidy. The joint optimum features a 41% tax on marketing and an 18% subsidy to production, raising welfare by 1.7% along the transition. Each instrument on its own is weaker: the optimal marketing tax falls to 18% when applied in isolation, while the optimal production subsidy falls to 16%. The complementarity reflects the two-way interaction between marketing and profits. A production subsidy raises profits per customer and thus stimulates additional (inefficient) marketing, limiting the optimal subsidy in isolation. Conversely, a marketing tax reduces competitive pressure in prices, which amplifies distortions from markups. Without a subsidy to offset these additional distortions, the optimal tax is substantially lower.

Finally, we use the model to quantify the macroeconomic consequences of the rise in marketing spending since the 1980s. We calibrate a change in the cost of marketing to match the increase in both marketing expenditures relative to GDP and in sales concentration within industries. The implied reduction in marketing costs raises GDP by 4.6% relative to 1980. The accompanying increase in concentration occurs without higher market power and therefore does not raise the misallocative costs of imperfect competition. Despite higher marketing outlays, the intangible value of firms' customer bases declines, as the increase in competitors' investment in demand and implied greater turnover lowers the value of each customer relationship. These results caution against using industry concentration as a measure of imperfect competition or capitalization of spending as a measure of intangible asset values.

Our analysis builds on the literature studying welfare losses from misallocation in the tradition of Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), with particular emphasis on distortions arising from markup dispersion (Baqaee and Farhi, 2020; Edmond, Midrigan and Xu, 2022). In contemporaneous work, Pearce and Wu (2025) study the role of marketing for aggregate productivity in a model with rationally inattentive consumers and granular firms. Most closely related is Afrouzi, Drenik and Kim (2023), who incorporate an extensive margin of demand into a model of imperfect competition due to differentiated varieties. They find a combined welfare loss of markup dispersion and customer misallocation of 11%, higher than our result (2%) and driven by the exentensive margin of demand. Three differences in assumptions account for the deviating results: First, we generate heterogeneous markups through search frictions rather than product differentiation, separating the source and cost of market power. Second, their matching process allows a planner to achieve any customer allocation without increasing resource costs relative to the equilibrium allocation, while our planner faces the same cost of adjusting matches as firms do in equilibrium. Third, they assume customers cannot transition between suppliers and require exogenous separation before rematching, while we allow for poaching of customers and transitions directly from supplier to supplier.[2] Supplier-to-supplier transitions directly imply that the number of a firm's customers responds to its price, making demand elastic at the extensive margin. As argued above, these features are central to understanding the allocative effects of marketing and the welfare costs of misallocation in our setting.

[2] Consistently, Afrouzi, Drenik and Kim (2023) report that under a parametric restriction on the matching process and separation of all customers each period, the allocation of customers in their framework is efficient. This does not hold under their baseline calibration, which is informed by similar targets as our calibration.

A number of papers study customer-base dynamics in models with directed search. In a seminal contribution, Gourio and Rudanko (2014) emphasize the role of customer capital for firms' physical investment dynamics. Rudanko (2025) studies implications for firms' use of sales vs. teaser pricing. Roldan-Blanco and Gilbukh (2021) introduce heterogeneous productivity and dynamic contracts to study firm growth and the cyclicality of aggregate markups. Chiavari (2024) emphasizes the growing importance of intangible assets, rationalizing macroeconomic trends in a model where intangible investment raises firms' returns to scale, as commonly assumed in the literature (Crouzet et al., 2022, 2024; De Ridder, 2024). In contrast, we focus on intangible investment in demand in shaping competition and industry dynamics, motivated by the evidence in He, Mostrom and Sufi (2024) that sales and marketing expenditures account for a significant share of intangible investment.

In a model with random search, Paciello, Pozzi and Trachter (2019) study how endogenous search effort by customers shapes the pass-through of productivity shocks to prices. Dinlersoz and Yorukoglu (2012) analyze marketing investment in an industry equilibrium with fully transitory productivity shocks. In complementary work, Shen (2025) shows that rising marketing intensity can account for declining business dynamism and increasing industry concentration. Her model of industry dynamics features permanent heterogeneity in the quality of firms' products and their efficiency to attract unmatched customers, and highlights a congestion externality in matching customers without pre-existing suppliers as a source of over-investment in marketing. We deviate from this literature by embedding marketing into a general-equilibrium framework with persistent productivity shocks, an intensive margin of demand, and customer poaching from existing supplier relationships. These features allow us to characterize and quantify both the allocative efficiency of marketing and its interaction with the broader misallocation costs of market power. Endogenous marketing investment and variation in demand per customer are also important deviations relative to Menzio (2024a,b), who establish efficiency as a benchmark in a static random-search model with unit demand and sellers unable to influence the number of customers they contact.[3]

A number of papers examine frictions in or cost of attracting customers for business cycle dynamics (Bai, Rios-Rull and Storesletten, 2025; Fernández-Villaverde et al., 2024), the effect of monetary policy (Morlacco and Zeke, 2021), international trade (Arkolakis, 2010; Drozd and Nosal, 2012; Heise, 2024) and investment in R&D (Cavenaile and Roldan-Blanco, 2021; Ignaszak and Sedláček, 2023). Cavenaile et al. (2025) study the interaction between advertising and R&D for competition and misallocation. In their framework, competition is driven by market shares, and marketing increases markups by raising concentration. In

[3]Our findings are also consistent with complimentary results on the efficiency of labor market equilibria in models with random search and job-to-job transitions in Fukui and Mukoyama (2025).

contrast, we show how marketing can foster competition in the presence of search frictions in goods markets. Greenwood, Ma and Yorukoglu (2024), Baslandze et al. (2023) and Cavenaile et al. (2023) consider theories of targeted advertising. We complement these studies by focusing on the role of general sales and marketing activities for the misallocation of demand.

The remainder of the paper is organized as follows: Section 2 presents the theoretical framework and characterizes an equilibrium. Section 3 describes the quantitative implementation. Section 4 reports the main results on marketing investment and misallocation. Section 5 analyzes the macroeconomic effects of rising marketing. Section 6 concludes.

2 A Theory of Competition for Customers

The economy is populated by three types of agents: Households, final good producers, and intermediate good producers. Households supply labor, save in physical capital, own all firms, and consume the final good. Competitive final good producers combine intermediate goods into the final good. They may either produce generic inputs in-house or source from intermediate good producers.[4] Search frictions imply that they can procure intermediates only from suppliers with whom they have established contact. Intermediate good producers differ in idiosyncratic productivity and engage in costly marketing activities to contact potential customers. This section outlines the optimization problem for each type of agent, the market structure, and the definition of equilibrium.

2.1 The Problem of Households

Time is discrete. A representative household lives forever and values consumption of the final good, C_t, and leisure. Preferences are given by

$$\sum_{t=0}^{\infty} \beta^t \big\{ u(C_t) - v(L_t) \big\}, \tag{1}$$

where L_t denotes labor supplied, $u(\cdot)$ is increasing and concave, $v(\cdot)$ is increasing and convex, and $0 < \beta < 1$ is a time discount factor. The household holds the capital stock K_t, owns all firms, and receives profits Π_t. The final good serves as the numeraire, with its price

[4] We model final good producers as the customers, rather than households directly. This choice reflects that much of our firm-level evidence concerns business-to-business marketing and sales, and our use of the final good as an intermediate input for marketing. A formulation with households as CES consumers of intermediates would deliver similar insights.

normalized to one. Each period, the budget constraint is

$$C_t + K_{t+1} = w_t L_t + (1 + r_t) K_t + \Pi_t, \quad \forall t, \tag{2}$$

where w_t is the wage rate and r_t the net return on capital. The household chooses $\{C_t, L_t, K_{t+1}\}_{t\geq 0}$ to maximize (1) subject to (2). The optimality conditions are the familiar intratemporal and intertemporal equations:

$$v'(L_t) = u'(C_t) w_t, \tag{3}$$

$$u'(C_t) = \beta(1 + r_{t+1}) u'(C_{t+1}). \tag{4}$$

2.2 The Problem of Final Good Producers

A representative final good producer consists of a continuum of divisions, indexed by $j \in [0, 1]$. Each division j contributes a component j to the final good. The final good producer aggregates all components into output of the final good Y_t using a CES technology

$$Y_t = \left(\int_0^1 Y_{jt}^{\frac{\sigma-1}{\sigma}} dj \right)^{\frac{\sigma}{\sigma-1}}, \tag{5}$$

where Y_{jt} is the quantity of component j used in period t and $\sigma > 1$ is the elasticity of substitution between components.[5] Cost minimization delivers the demand for each component,

$$Y_{jt} = \left(\frac{P_{jt}}{P_t} \right)^{-\sigma} Y_t, \tag{6}$$

where P_{jt} is the shadow cost of obtaining component j, and

$$P_t = \left(\int_0^1 P_{jt}^{1-\sigma} dj \right)^{\frac{1}{1-\sigma}} \tag{7}$$

is the unit cost of producing the final good. Since the market for the final good is perfectly competitive and we normalize its price as the numeraire, equilibrium requires $P_t = 1$, which implies $Y_{jt} = P_{jt}^{-\sigma} Y_t$.

[5] Unlike in standard models of monopolistic competition with CES preferences, our framework generally admits a well-defined solution even for $\sigma < 1$, since the extensive margin of demand precludes infinite markups from being optimal as $\sigma \downarrow 1$. For ease of comparison to the monopolistic competition benchmark, however, we restrict attention to $\sigma > 1$.

In-house production. Each division j can produce its component in-house using a Cobb–Douglas technology,

$$Y_{jt} = K_{jt}^{\alpha} L_{jt}^{1-\alpha}, \tag{8}$$

with in-house productivity normalized to one. Cost minimization yields the unit cost of in-house production,

$$\overline{mc}_t = \left(\frac{r_t + \delta_K}{\alpha}\right)^{\alpha} \left(\frac{w_t}{1-\alpha}\right)^{1-\alpha}, \tag{9}$$

where δ_K is the depreciation rate of physical capital. The associated unit factor requirements are $\bar{\ell}_t = \left(\frac{(1-\alpha)(r_t+\delta_K)}{\alpha w_t}\right)^{\alpha}$ for labor and $\bar{k}_t = \left(\frac{\alpha w_t}{(1-\alpha)(r_t+\delta_K)}\right)^{1-\alpha}$ for capital. Since all divisions face the same technology, these quantities are constant across j.

External sourcing. Instead of producing its component in-house, a division may source an intermediate input externally and transform it one-for-one into its component. To do so, the division must be matched with an intermediate good producer, referred to as its *supplier*. If matched, the division sources its entire demand for the component at the supplier's price p; if unmatched, it must produce in-house at cost $\overline{mc}_t$. The shadow price of component j is therefore

$$P_{jt} = \begin{cases} p & \text{if matched with a supplier charging } p \\ \overline{mc}_t & \text{if not matched with a supplier} \end{cases} \tag{10}$$

The static gain to the final good producer of switching a division from in-house production to sourcing from a supplier charging p is a marginal reduction in total production cost $P_t Y_t$, given by

$$\eta_t(p) = \frac{p^{1-\sigma} - \overline{mc}_t^{1-\sigma}}{\sigma - 1} P_t^{\sigma} Y_t = \frac{p^{1-\sigma} - \overline{mc}_t^{1-\sigma}}{\sigma - 1} Y_t. \tag{11}$$

We describe in Section 2.4 how divisions are contacted by potential suppliers and how these static gains translate into the dynamic value of maintaining a match.

2.3 The Problem of Intermediate Good Producers

A continuum of atomistic intermediate good producers is heterogeneous in their productivity z and in the number of divisions n they are matched with at the beginning of a period. We refer to these divisions as a producer's *customers*. Productivity z evolves stochastically according to a discrete-time Markov process with conditional CDF $F(z'|z)$ on $(0,\infty)$. We assume that the transition kernel is continuous in both arguments. Moreover, we assume that for any $z_1 < z_2$ and any z', $F(z'|z_2) \le F(z'|z_1)$ with strict inequality for some z'; i.e., $F(\cdot \mid z_2)$ first-order stochastically dominates $F(\cdot \mid z_1)$. At the end of each period, producers exit exogenously with probability δ_F. In addition, after observing their productivity at the start of a period, producers may choose to exit endogenously.

Production. Intermediate good producers produce output y with a Cobb-Douglas production technology

$$y = zk^{\alpha}\ell^{1-\alpha}, \tag{12}$$

where k and ℓ denote capital and labor inputs. The marginal cost of producing one unit of output at productivity z follows from the cost-minimization problem,

$$mc_t(z) = \frac{1}{z}\left(\frac{r_t + \delta_K}{\alpha}\right)^{\alpha}\left(\frac{w_t}{1-\alpha}\right)^{1-\alpha} = \frac{\overline{mc}_t}{z}, \tag{13}$$

where $\overline{mc}_t$ captures the common component of marginal cost. Accordingly, optimal factor demands per unit of output are $\ell_t(z) = \frac{\bar{\ell}_t}{z}$ and $k_t(z) = \frac{\bar{k}_t}{z}$. Final good producers' problem implies static demand $y_t(p) = p^{-\sigma}Y_t$ per matched customer for any supplier charging price p. Hence, the static profit per customer of an intermediate producer with productivity z and price p is

$$\pi_t(p,z) = (p - mc_t(z))\, y_t(p) = \left(p - \frac{\overline{mc}_t}{z}\right) p^{-\sigma} Y_t. \tag{14}$$

We denote as $p_t^S(z) = \frac{\sigma}{\sigma-1} mc_t(z)$ the static optimal price that maximizes profits per customer for a given productivity z and is the solution to $\frac{\partial \pi_t(p,z)}{\partial p} = 0$.

Marketing. Intermediate good producers can engage in costly marketing activities to contact additional customers. To contact $\hat{n}$ new customers, a firm must spend $\hat{m}(\hat{n},n)$ units of marketing resources. We assume that the final good is used for marketing so that its price is normalized to one. We further assume that the cost function satisfies $\hat{m}(\hat{n},n) = m\left(\frac{\hat{n}}{n}\right)n$, where $m(\cdot)$ is strictly increasing and strictly convex. We denote $i = \frac{\hat{n}}{n}$ as the *marketing intensity* with which a firm contacts new customers.[6]

[6] These properties follow from a meeting technology that is constant returns to scale (CRS) in marketing spending and existing customers. By construction, $\hat{m}(\lambda\hat{n},\lambda n) = m\left(\frac{\lambda\hat{n}}{\lambda n}\right)\lambda n = \lambda\hat{m}(\hat{n},n)$, so the cost function is CRS. Moreover, $\hat{m}(\hat{n},n)$ is increasing and convex in $\hat{n}$ (from the properties of $m(\cdot)$) and decreasing in n, since $\frac{\partial \hat{m}}{\partial n} = m\left(\frac{\hat{n}}{n}\right) - m'\left(\frac{\hat{n}}{n}\right)\frac{\hat{n}}{n} < 0$ by strict convexity of $m(\cdot)$. Equivalently, the cost function can be expressed

Intermediate good producers contact prospective new customers before production takes place. Upon contact, potential customers decide whether to accept or reject a match. Both newly contacted and existing customers observe only a supplier's price, not its productivity or customer base, so acceptance decisions depend on price alone. Let $e_t(p)$ denote the fraction of contacted customers that accept a match with a producer charging price p; we refer to $e_t(p)$ as the *conversion rate*. Existing customers may also be approached by alternative suppliers. Let $q_t(p)$ denote the fraction of a firm's existing customers (when it charges price p) that accept a match with a competitor; we refer to $q_t(p)$ as the *quit rate*. Both $e_t(p)$ and $q_t(p)$ are equilibrium objects characterized below. Given $e_t(p)$ and $q_t(p)$, the total number of customers at the time of production and sales, as a function of marketing intensity i, price p, and initial customer base n, is

$$n' = n - q_t(p)n + e_t(p)\hat{n} = (1 - q_t(p) + e_t(p)i)n \equiv \rho_t(p,i)n \tag{15}$$

where $\rho_t(p,i)$ denotes the firm's net customer growth factor.

Dynamic Problem. At the beginning of each period, intermediate good producers choose a price p and marketing intensity i, conditional on their current productivity z and customer base n.[7] Both p and i affect not only current profits but also the evolution of the customer base, and hence future profits. Firms therefore take into account these dynamic effects, discounting future payoffs at rate r_t consistent with firm ownership. The value function is

$$\tilde{V}_t(z,n) = \max_{p,i}\left\{-m(i)n + \rho_t(p,i)n\pi_t(p,z) + \frac{1-\delta_F}{1+r_{t+1}}\mathbb{E}\left[\max\left\{\tilde{V}_{t+1}\left(z',\rho_t(p,i)n\right),0\right\}|z\right]\right\}, \tag{16}$$

where the first term represents marketing costs, the second term is the static profit, and the third term is the expected discounted continuation value. The expectation is taken over next-period productivity z', and the max-operator captures the option of endogenous exit. To ensure that $\tilde{V}_t(z,n)$ is well defined, we assume $\lim_{t\to\infty}\mathbb{E}_0 Q_{0,t}\left[\tilde{V}_t(z_t,n_t^*)|z_0,n_0\right] = 0$, where $Q_{0,t} = \prod_{\tau=1}^{t}(1+r_\tau)^{-1}$ is the discount factor and n_t^* the number of customers implied by optimal choices.[8] As $\tilde{V}_t(z,n)$ is a contraction mapping for increasing functions and $\pi_t(p,z)$ is strictly

as a matching technology in existing customers n and marketing spending $\hat{m}$: $\hat{n}(n,\hat{m}) = n\,m^{-1}\left(\frac{\hat{m}}{n}\right) \equiv nf\left(\frac{\hat{m}}{n}\right)$, with $f(\cdot)$ strictly concave. The matching technology is CRS since $\hat{n}(\lambda n,\lambda\hat{m}) = \lambda nf\left(\frac{\lambda\hat{m}}{\lambda n}\right) = \lambda\hat{n}(n,\hat{m})$ and increasing in both arguments. The positive contribution of existing customers to meetings with new customers can be interpreted as the benefit of word-of-mouth advertising or peer effects in product adoption, consistent with empirical evidence in Bailey et al. (2022) and Argente, Méndez and Van Patten (2024).

[7] We assume firms set a single price for all customers, abstracting from price discrimination. This assumption makes our framework directly comparable to the literature on market power without search frictions (e.g. Edmond, Midrigan and Xu, 2022; Afrouzi, Drenik and Kim, 2023).

[8] This condition rules out infinite customer growth with positive probability. It must hold in equilibrium, since the mass of matched divisions is bounded above by one and hence no firm can be allowed to accumulate

increasing in z, it follows that $\tilde{V}_t(z,n)$ is strictly increasing in z. Moreover, the firm value scales linearly with n, so we can write $\tilde{V}_t(z,n) = nV_t(z)$, where

$$V_t(z) = \max_{p,i} \left\{ -m(i) + \rho_t(p,i) \left(\pi_t(p,z) + \frac{1-\delta_F}{1+r_{t+1}} \mathbb{E}\left[\max\left\{ V_{t+1}(z'), 0 \right\} | z \right] \right) \right\} \tag{17}$$

is the value per customer. The value of a firm is thus given by the number of its customers multiplied by the value of each customer, which depends on the firm's productivity. As $V_t(z,n)$ is strictly increasing in z, $V_t(z)$ is strictly increasing in z. Hence, existing suppliers choose to exit the market at the beginning of period t if and only if $z < \bar{z}_t$, where $V_t(\bar{z}_t) = 0$. From the property $\tilde{V}_t(z,n) = nV_t(z)$, it follows that the optimal choices of p and i are independent of the number of existing customers n. If $\rho_t(p,i)$ is differentiable, the optimal price $p_t(z)$ and marketing intensity $i_t(z)$ for a firm with productivity z satisfy the first order conditions[9]

$$\rho_t(p,i)\frac{\partial \pi_t(p,z)}{\partial p} = -\frac{\partial \rho_t(p,i)}{\partial p} W_t(p,z), \tag{18}$$

$$\frac{\partial m(i)}{\partial i} = \frac{\partial \rho_t(p,i)}{\partial i} W_t(p,z), \tag{19}$$

where we define $W_t(p,z) \equiv \pi_t(p,z) + \frac{1-\delta_F}{1+r_{t+1}} \mathbb{E}\left[\max\left\{ V_{t+1}(z'), 0 \right\} | z \right]$ as the value per customer at production, comprising current profit and continuation value. In equilibrium, firms choose $p_t(z)$; we denote the corresponding value by $W_t^*(z) \equiv W_t(p_t(z), z)$.

Firms' price setting (18) balances static and dynamic effects on profits. The left-hand side is the marginal benefit of raising the price on static profits from current customers, while the right-hand side captures the dynamic cost of slower customer growth, scaled by both current profit and continuation value. The optimal choice for marketing intensity (19) trades off the static cost of marketing against the value of expanding the customer base, as is standard in the literature on customer dynamics.[10] Crucially, the two choices interact: the returns to marketing depend on the price charged, both through its effect on customer acceptance and the profit margin per customer. In turn, marketing intensity i affects the gains from adjusting the price to attract additional customers.

infinite customers.

[9] We show in section 2.5 that $\rho_t(p,i)$ is differentiable in the class of equilibria we consider.

[10] See, e.g., Gourio and Rudanko (2014), Einav et al. (2021), Afrouzi, Drenik and Kim (2023), Fitzgerald, Haller and Yedid-Levi (2024), Rudanko (2025).

2.4 The Market for Intermediate Goods

Final good firms act as customers, and intermediate good producers as suppliers, in the market for intermediates. The market is characterized by frictions in the matching process: suppliers must engage in costly marketing activities to contact potential customers and form matches. Suppliers cannot target specific customers. Instead, they contact customers at random, and every customer is equally likely to be approached, regardless of whether they are currently unmatched or already in a relationship with another supplier. Upon being contacted, customers observe only the current price charged by the potential supplier and form expectations about the path of future prices and the expected duration of the match. Customers in an existing match may terminate it at any time, and they may re-match with a new supplier if contacted. Similarly, unmatched customers may accept or decline a match if contacted. We assume that customers cannot recall previous matches or contacts.

The market for intermediates unfolds in two stages: In the first stage, at the beginning of each period, exogenous exit shocks and the productivity of existing suppliers are revealed. Suppliers then decide whether to continue operating or exit the market. Continuing suppliers choose their marketing intensity i and price p. New suppliers enter, are matched with initial customers from the pool of unmatched final-good divisions, and also choose their price and marketing intensity. In the second stage, both continuing and new suppliers contact potential customers according to their chosen marketing intensity. Contacted customers decide whether to accept a match and begin purchasing from the supplier, or to decline. Finally, production and sales take place. We detail both stages of the matching process below.

Market Entry and Exit. In each period, a fraction χ of unmatched customers is contacted by a potential entrant.[11] After contacting a potential customer, entrants draw their initial productivity from a distribution $F_0(z)$ with support $(0, z^{max})$, where $1 < z^{max} \leq \infty$. We assume that F_0 is continuous and has strictly positive density throughout its support, so that the distribution is connected (i.e., contains no gaps). Entrants decide to operate if and only if $z \geq \bar{z}_t$ (equivalently, $V_t(z) \geq 0$). Entrants that start operating do so with one initial customer.

Simultaneously with entry, existing suppliers decide whether to exit the market whenever $z < \bar{z}_t$. Let $G_t(z)$ denote the mass of customers matched with suppliers of productivity z or lower during production in period t, and u_t is the mass of unmatched customers.[12] As an intermediate step, let $G_t^i(z)$ and u_t^i denote, respectively, the corresponding masses of matched

[11] One interpretation is that in divisions with in-house production, with probability χ an employee spins off a new company to supply their previous employer.

[12] $G_t(z)$ is a mass function and does not integrate to one. However, $\int_0^\infty 1 dG_t(z) + u_t = 1$.

and unmatched customers after the exit and entry of suppliers in period t. Their laws of motion are

$$(20)\quad G_t^i(z) = \begin{cases} \chi u_{t-1}\big(F_0(z) - F_0(\bar{z}_t)\big) + (1-\delta_F)\int_0^\infty \big(F(z|\hat{z}) - F(\bar{z}_t|\hat{z})\big)\, dG_{t-1}(\hat{z}), & \text{if } z \geq \bar{z}_t, \\ 0, & \text{if } z < \bar{z}_t, \end{cases}$$

$$(21)\quad \begin{aligned} u_t^i &= \Big(1 - (1 - F_0(\bar{z}_t))\chi\Big)u_{t-1} + (1-\delta_F)\int_0^\infty F(\bar{z}_t|z)\, dG_{t-1}(z) + \delta_F(1 - u_{t-1}) \\ &= 1 - \lim_{z\to\infty} G_t^i(z). \end{aligned}$$

Re-Matching between Customers and Suppliers. New entrants and continuing suppliers contact potential customers at random. The number of potential customers contacted by a supplier is $i \cdot n$, where i is the chosen marketing intensity and n the number of customers at production in the previous period (or the number of initial customers in the case of entrants). We assume each customer is contacted by at most one (additional) firm.[13] Upon contact, customers observe only the posted price p and not the supplier's productivity. Let $M_t^i(p)$ denote the value of entering the re-matching stage in an existing match with a supplier charging p, and let U_t^i denote the value of entering unmatched. These values are

$$(22)\quad M_t^i(p) = M_t(p) + \int_0^\infty \max\{M_t(p_t(z)) - M_t(p), 0\}\, i_t(z)\, dG_t^i(z),$$

$$(23)\quad U_t^i = U_t + \int_0^\infty \max\{M_t(p_t(z)) - U_t, 0\}\, i_t(z)\, dG_t^i(z),$$

where $M_t(p)$ and U_t are, respectively, the values of being matched with a supplier charging p or being unmatched at the start of production in period t.[14] These values are given by

$$(24)\quad M_t(p) = \max\left\{\eta_t(p) + \frac{1-\delta_F}{1+r_{t+1}}\int_0^\infty \int_{\bar{z}_{t+1}}^\infty M_{t+1}^i(p_{t+1}(z'))dF(z'|z)dH_t(z|p) + \frac{\gamma_t(p)U_{t+1}^i}{1+r_{t+1}}, U_t\right\},$$

$$(25)\quad U_t = \frac{1}{1+r_{t+1}}\left[U_{t+1}^i + \chi\int_{\bar{z}_{t+1}}^\infty \big(M_{t+1}^i(p_t(z)) - U_{t+1}^i\big)\, dF_0(z)\right],$$

where $H_t(z|p)$ is the conditional distribution of beliefs about current productivity z of a supplier charging p in period t and $\gamma_t(p) = \delta_F + (1-\delta_F)\int_0^\infty F(\bar{z}_{t+1}|z)dH_t(z|p)$ is the perceived probability that a supplier charging p in t exits at the beginning of period $t+1$.[15] Customers

[13] To sustain no duplicate matches, we ensure that $0 \leq i(z) \leq 1\ \forall z$ by imposing conditions on $m(i)$ such that $\frac{\partial m(i)}{\partial i}|_{i=0} = 0$ and $\frac{\partial m(i)}{\partial i}|_{i=1} = \infty$.

[14] Because customers can always dissolve a match after the re-matching stage, and since matched and unmatched customers are contacted at the same rate, it follows that $M_t^i(p) \geq U_t^i$ for all p. We therefore abstract from the possibility of dissolving a match before re-matching without loss of generality.

[15] Customers do not need to form expectations about the size of suppliers' customer base n, as all decisions are independent of the number of existing customers and hence n is not payoff relevant.

matched with a firm charging p_1 switch to a new supplier charging p_2 during re-matching iff $M_t(p_2) \geq M_t(p_1)$. Unmatched customers accept a match iff $U_t \leq M_t(p)$. Matched customers dissolve a match before production iff $M_t(p) < U_t$.

Denote $\lambda_t = \int_0^\infty i_t(z) dG_t^i(z)$ as the probability that a given customer is contacted by a new supplier. The updated mass $G_t(z)$ of matched customers at firms with productivity less than z is given by

$$G_t(z) = G_t^i(z) + \underbrace{u_t^i \int_0^z \mathbb{I}_{M_t(p_t(\hat{z})) \geq U_t} i_t(\hat{z}) dG_t^i(\hat{z})}_{\text{new customers, from unmatched}} + \underbrace{\int_0^z \int_z^\infty \mathbb{I}_{M_t(p_t(\hat{z})) \geq \max\{M_t(p_t(z')), U_t\}} dG_t^i(z') i_t(\hat{z}) dG_t^i(\hat{z})}_{\text{new customers, poached}} - \underbrace{\int_0^z \mathbb{I}_{M_t(p_t(\hat{z})) < U_t} dG_t^i(\hat{z})}_{\text{customers leaving to unmatched}} - \underbrace{\int_0^z \int_z^\infty \mathbb{I}_{U_t < M_t(p_t(\hat{z})) < M_t(p_t(z'))} i_t(z') dG_t^i(z') dG_t^i(\hat{z})}_{\text{contacted customers leaving to competitors}}. \tag{26}$$

The mass of unmatched customers u_t during production in period t is given by

$$\begin{aligned} u_t &= u_t^i - u_t^i \int_0^\infty \mathbb{I}_{M_t(p_t(z)) \geq U_t} i_t(z) dG_t^i(z) + (1 - \lambda_t) \int_0^\infty \mathbb{I}_{M_t(p_t(z)) < U_t} dG_t^i(z) \\ &\quad + \int_0^\infty \int_0^\infty \mathbb{I}_{\max\{M_t(p_t(z')), M_t(p_t(z))\} < U_t} i_t(z') dG_t^i(z') dG_t^i(z) \\ &= 1 - \lim_{z \to \infty} G_t(z). \end{aligned} \tag{27}$$

We can now define the customer quit rate $q_t(p)$ and conversion rate $e_t(p)$ as

$$q_t(p) = (1 - \mathbb{I}_{M_t(p) \geq U_t}) + \mathbb{I}_{M_t(p) \geq U_t} \int_0^\infty \mathbb{I}_{M_t(p) \leq M_t(p_t(z))} i_t(z) dG_t^i(z), \tag{28}$$

$$e_t(p) = \mathbb{I}_{M_t(p) \geq U_t} \left[u + \int_0^\infty \mathbb{I}_{M_t(p) \geq M_t(p_t(z))} dG_t^i(z) \right]. \tag{29}$$

2.5 Equilibrium

Since the economy is not subject to aggregate risk, we focus on steady-state equilibria and omit time subscripts for ease of notation.[16] Because customers in the market for intermediate goods observe only current prices, they must form expectations about suppliers' payoff relevant state z. We therefore adopt a Bayesian equilibrium concept, similar to Coles and Mortensen (2016), and restrict attention to Bayesian equilibria in Markov strategies.

[16] We later consider non-stochastic perturbations of the steady state and transition paths between steady-state equilibria. The equilibrium concept and properties described below apply with only minor adjustments.

Equilibrium Definition

A Bayesian steady-state equilibrium in pure Markov strategies consists of prices w and r, aggregate profits Π, gross output Y, consumption C, physical capital K, and labor supply L; intermediate good firms' policies $p(z)$, $i(z)$, and value $V(z)$; customer values $M(p)$, U and beliefs $H(z|p)$; an ergodic distribution of customers across firms $G(z)$ and mass of unmatched customers u; as well as customer loss and conversion rates $q(p)$ and $e(p)$, such that:

1. Given w, r, Π, consumption C, physical capital K and labor L solve households' problem in (1) and (2).

2. Given $w, r, q(p), e(p)$, intermediate firms' policies $p(z)$, $i(z)$, and value $V(z)$ solve (17). Intermediate firms exit endogenously (do not enter) whenever $V(z) < 0$.

3. Customers' beliefs $H(z|p)$ are consistent with Bayes' law given intermediates firms' policies.

4. Customers' values $M(p)$ and U are given by (24) and (25). Customers switch from suppliers charging p to those charging p' iff $M(p') \geq M(p)$ and from p to unmatched iff $M(p) < U$.

5. Given intermediate good firms' and customers' choices, the mass of matched customers $G(z)$ and unmatched customers u are a fixed point to (26) and (27), and the customer loss and conversion rates satisfy (28) and (29).

6. Given intermediate good firms' pricing, $G(z)$ and u, aggregate gross output Y is implied by (5), (6), (7), and (10).

7. The labor-market clears: $L = \bar{\ell}\left[u(\bar{mc})^{-\sigma}Y + \int_0^\infty \frac{1}{z}p(z)^{-\sigma}Y dG(z)\right]$.

8. The capital market clears: $K = \bar{k}\left[u(\bar{mc})^{-\sigma}Y + \int_0^\infty \frac{1}{z}p(z)^{-\sigma}Y dG(z)\right]$.

9. The final-goods market clears: $Y = C + \delta_k K + \Gamma$, where $\Gamma = \int_0^\infty m(i(z))dG^i(z)$ is aggregate spending on marketing.

10. Aggregate profits are given by $\Pi = \int_0^\infty \pi(p(z), z)dG(z)$.

In what follows, we will restrict attention to separating equilibria in which prices in the intermediate goods market are strictly decreasing in suppliers' productivity; i.e., $p(z_1) > p(z_2)$ for $z_1 < z_2$. As we show below, these equilibria feature efficient transitions of customers among suppliers conditional on matches, which is a desirable property for our purpose.[17] In addition, we restrict attention to equilibria with monotone beliefs about future prices, such that $H(z|p_1) \le H(z|p_2)$ for $p_1 < p_2$ and any $z \in (0, \infty)$.[18] Before we proceed, we highlight some useful properties of equilibria in this class.

Properties of Equilibria with Strictly Decreasing Prices

Any equilibrium in which prices are strictly decreasing in productivity is fully separating by construction. Strict monotonicity of $p(z)$ ensures that the inverse $\hat{z}(p) = p^{-1}(z)$ exists, so customers can unambiguously infer a supplier's productivity type from its price. Define the image of the pricing function as $\mathcal{P} = \{p \in \mathbb{R} : \exists z \in (\bar{z}, \infty) \text{ for which } p(z) = p\}$, where $\bar{z}$ is the exit threshold productivity. Consistency of customer beliefs with Bayes' law and the pricing policy then requires

$$H(z|p) = \begin{cases} 0 & \text{if } z < \hat{z}(p) \\ 1 & \text{if } z \ge \hat{z}(p) \end{cases} \tag{30}$$

for all $p \in \mathcal{P}$. Monotonicity of beliefs further implies that for any $p \notin \mathcal{P}$ and $p_1 > p > p_2$ with $p_1, p_2 \in \mathcal{P}$, we must have $H(z|p_1) \ge H(z|p) \ge H(z|p_2)$ $\forall z \ge \bar{z}$ and moreover $H(\bar{z}|p) = 1$ whenever $p > p(\bar{z})$. Finally, consistency of beliefs about exit with suppliers' choices requires $\gamma(p) = F(\bar{z}|\hat{z}(p))$. Given equilibrium beliefs, the value of a match to a customer simplifies to

$$M(p) = \max\left\{\eta(p) + \frac{1-\delta_F}{1+r}\left[\int_{\bar{z}}^{\infty} M^i(p(z'))dF(z'|\hat{z}(p)) + F(\bar{z}|\hat{z}(p))U^i\right] + \frac{\delta_F U^i}{1+r}, U\right\}. \tag{31}$$

As we focus on equilibria where $p(z)$, and hence $\hat{z}(p)$, is strictly decreasing, it follows from the properties of $F(z'|z)$ that $M(p)$ is strictly decreasing in p. The following Lemma is an immediate implication.

Lemma 1. *There exists a reservation price $\bar{p}$ such that $M(\bar{p}) = U$, with $M(p) > U$ if and only if $p < \bar{p}$. In particular, customers quit to being unmatched whenever $p > \bar{p}$. Moreover, if customers are simultaneously in contact with two firms of productivities $z_1 < z_2$, then the monotonicity of $p(z)$ implies $p(z_1) > p(z_2)$ and hence $M(p(z_1)) < M(p(z_2))$, so customers always move to the more productive supplier.*

[17] Like the assumption of no-price-discrimination across customers, this restriction ensures comparability with the literature studying market power in models without search frictions (e.g. Edmond, Midrigan and Xu, 2022; Afrouzi, Drenik and Kim, 2023), where prices are generally a decreasing function of cost.

[18] Under the restriction on $p(z)$ being strictly decreasing, monotone beliefs are the only beliefs consistent with Bayes Law for prices posted in equilibrium and impose additional assumptions only on off-equilibrium beliefs.

Given these properties of customer choice, the quit rate $q(p)$ and conversion rate $e(p)$ simplify to

$$q(p) = \begin{cases} \int_{\hat{z}(p)}^{\infty} i(z)\, dG^i(z), & \text{if } p \leq \bar{p}, \\ 1, & \text{if } p > \bar{p}, \end{cases} \tag{32}$$

$$e(p) = \begin{cases} u + G^i(\hat{z}(p)), & \text{if } p \leq \bar{p}, \\ 0, & \text{if } p > \bar{p}. \end{cases} \tag{33}$$

We impose no restrictions on the pricing function $p(z)$ beyond it being strictly decreasing in equilibrium. We can establish additional properties of $p(z)$, by showing that any violation leads to profitable deviations for a firm and thus cannot be sustained in equilibrium.

Proposition 1. *Any equilibrium with strictly decreasing prices and monotone beliefs satisfies the following properties:*

(i) All firms set prices below the static optimum, $p(z) \leq p^S(z)$*,*

(ii) the firm at the exit threshold sets the reservation price, $p(\bar{z}) = \bar{p}$*,*

(iii) and the pricing function $p(z)$ *is continuous.*

Proof. To prove (i), note from (32) and (33) that the customer growth rate $\rho(p, i)$ is weakly decreasing in p for any given i. Now suppose that a firm with productivity z sets a price $p > p^S(z)$. By lowering its price, the firm would strictly increase profits per customer $\pi(p, z)$ without reducing its growth rate $\rho(p, i)$. This strictly raises its total profits, implying that the original price cannot be part of an equilibrium.

To show (ii), suppose that a firm with exit cut-off productivity $\bar{z}$ charged a price $p < \bar{p}$. Since $p(z)$ is strictly decreasing, p would then be the highest price posted in equilibrium. The firm could profitably raise its price infinitesimally without reducing $\rho(p, i)$, thereby increasing profits, because $p < p^S(\bar{z})$ from property (i). Indeed, while property (i) only establishes the weak inequality $p \leq p^S(z)$ for general z, equality cannot hold at the cut-off. If $p = p^S(\bar{z})$, then a firm with productivity $\bar{z}$ would earn strictly positive static profits. Since the continuation value is an option value and hence non-negative, this would imply $V(\bar{z}) > 0$, contradicting the definition of $\bar{z}$ as the threshold type with $V(\bar{z}) = 0$. Therefore $p < p^S(\bar{z})$. Hence, the original price cannot be part of an equilibrium. Only at $p = \bar{p}$ does the firm lose customers to being unmatched if it increases its price further. It follows that the equilibrium price at the lowest active productivity level, equivalent to the highest price posted in equilibrium, must satisfy $p(\bar{z}) = \bar{p}$.

To prove (iii), suppose for a contradiction that there is a gap: there exists $\tilde{z}$ with $p_2 \equiv \lim_{z\uparrow\tilde{z}} p(z) > p_1 \equiv \lim_{z\downarrow\tilde{z}} p(z)$, so $(p_1, p_2) \notin \mathcal{P}$. These one-sided limits exist since $p(\cdot)$ is monotone. First, it must be that the price at the lower end of the gap is strictly below the static optimum, $p_1 < p^S(\tilde{z})$; otherwise $p_2 > p_1 = p^S(\tilde{z})$ would contradict (i). Second, given that $p_1 < p^S(\tilde{z})$, a firm of productivity $\tilde{z}$ posting price p_1 has a profitable deviation: raising its price to some p in between p_1 and p_2 increases its static profits but has no negative effect on the growth rate $\rho(p, i)$, since no other firm posts a price in (p_1, p_2) and beliefs are monotone. The same logic applies if firm $\tilde{z}$ posts price p_2: for any sequence $z_n \downarrow \tilde{z}$ with $p(z_n) \uparrow p_1$; the firm z_n for n large still has a profitable deviation by raising its price slightly into (p_1, p_2). Therefore, such a gap cannot exist, and the pricing function $p(\cdot)$ must be continuous. □

The flow equations for the masses of matched and unmatched customers simplify to

$$G(z) = G^i(z) + u^i \int_{\bar{z}}^{z} i(\hat{z})\, dG^i(\hat{z}) - G^i(z) \int_{z}^{\infty} i(z')\, dG^i(z'), \tag{34}$$

$$u = u^i - u^i \int_{\bar{z}}^{\infty} i(z)\, dG^i(z) = 1 - \lim_{z\to\infty} G(z). \tag{35}$$

For any differentiable $G^i(z)$, the derivatives of the customer growth rate are

$$\frac{\partial\rho(p,i)}{\partial p} = 2i(\hat{z}(p))g^i(\hat{z}(p))\frac{\partial\hat{z}(p)}{\partial p}, \tag{36}$$

$$\frac{\partial\rho(p,i)}{\partial i} = e(p) = u + G^i(\hat{z}(p)), \tag{37}$$

where $g^i(z)$ is the derivative of $G^i(z)$. We can now simplify the first-order conditions (18) and (19) using the above expressions for customer growth. They reduce to

$$\frac{\partial m(i(z))}{\partial i} = \left(u + G^i(z)\right) W_t^*(z), \tag{38}$$

$$\frac{\partial p(z)}{\partial z} = -\frac{2\, i(z)\, g^i(z)}{\rho(p(z), i(z))} \cdot \frac{W_t^*(z)}{\frac{\partial\pi(p(z),z)}{\partial p}}, \tag{39}$$

where $W_t^*(z)$, as defined earlier, denotes the equilibrium value per customer, and we impose the consistency condition $\hat{z}(p(z)) = z$. The boundary condition associated with (39) is $p(\bar{z}) = \bar{p}$.[19]

[19] Since $p(z)$ must be continuous, any continuous $G^i(z)$ implies a continuous $i(z)$. The continuous $i(z)$ then maps a continuous $G^i(z)$ into a continuous $G(z)$ via (34), and hence again a continuous $G^i(z)$. We initialize our numerical implementations of the model with a continuous $G^i(z)$, and use the first order conditions to solve for firms' policy functions. We then numerically verify continuity of $i(z)$, continuity of the distributions

Condition (38) shows that the marketing decision collapses to a simple trade-off: the marginal cost of marketing equals the marginal value of an additional contact. Relative to the earlier formulation, the conversion rate of contacts into new customers, $e(p) = u + G^i(z)$, now appears explicitly as the share of customers who are either unmatched or served by less-productive firms. Condition (39) characterizes the pricing function as a differential equation. The slope of the price schedule is proportional to the effective amount of *local competition*—the product of the density of firms and their marketing intensity at z—scaled by the value per customer and the price sensitivity of per-customer profits.

3 Calibration and Quantitative Properties

In this section, we quantify the model. We first describe the calibration strategy, and then discuss the model's quantitative properties, and its relation to additional empirical findings.

3.1 Calibration

We treat the model economy as representative of an average U.S. industry. Consequently, we calibrate its parameters to empirical targets that are averages of within-industry moments. When possible, we use publicly available micro-data on the universe of U.S. firms. For measures of markups and marketing costs, we rely on data for a selected sample of public firms in the Capital IQ database. We filter the model counterpart appropriately to mimic selection for these moments. Building on Ottonello and Winberry (2020), we filter firms in the model such that older and larger firms are over-represented, and discuss the details of this procedure in Appendix B.2.

Capital IQ. Following He, Mostrom and Sufi (2024), we use Capital IQ's sales and marketing expenses from a Capital IQ–Compustat matched sample as our measure of marketing spending. They show that this measure captures forward-looking investment in demand—i.e., it predicts future firm value—whereas the residual component of the commonly used and more comprehensive Selling, General, and Administrative Expenses (SGA) variable does not. As markup measure, we use the ratio of sales to cost of goods sold. We find a robust positive relationship between the marketing spending to revenue ratio and the markup in the cross-section of firms within industries, documented in detail in Appendix B.1. The cross-sectional elasticity between these two key objects of our theory is significantly positive and equals around 1.7, a number we target in our model calibration. Since we only use both measures in logs in cross-sectional regressions, the relevant assumption is that the measured ratios vary

$G^i(z)$ and $G(z)$, and the FOCs at the converged equilibrium solution.

across firms proportional to the true ratios. For example, the sales and marketing expenses variable may not capture all forward-looking customer investments, or may capture some operating expenses; our maintained assumption is that the missing portion is proportionate to the measured variable (controlling for observables).[20]

Internal Calibration. Table 1 summarizes the calibration of the model. We choose nine parameters to target a total of ten moments. While the parameters are jointly identified, in what follows we relate each model parameter to one particularly informative moment. The framework has no issue in accounting for all empirical moments simultaneously.

Table 1: Calibrated parameters and targeted moments

Parameter		**Value**	
Log TFP: unconditional dispersion	σ_z	0.389	
Log TFP: auto-correlation	ρ_z	0.556	
Log TFP: mean	μ_z	0.610	
Marketing cost: level	$\bar{m}$	0.423	
Marketing cost: curvature	ψ	4.193	
Intensive-margin elasticity	σ	1.953	
Exogenous firm exit rate	δ_F	0.085	
Entrants per unmatched	χ	0.161	
Capital elasticity	α	0.311	
Targeted moment	**Model**	**Data**	**Source**
Top 10% sales share	0.681	0.724	SUSB
Top 1% sales share	0.437	0.423	SUSB
Firm entry (exit) rate	0.086	0.085	BDS
Relative employment of entrants	0.541	0.540	Sterk, Sedláček and Pugsley (2021)
Aggregate labor share	0.615	0.615	BLS
Average customer quit rate	0.150	0.150	Gourio and Rudanko (2014)
OLS slope $\ln \frac{marketing}{revenue}$ on $\ln markup$	$\ast$ 1.731	1.729	Capital IQ
OLS slope $\ln markup$ on $\ln \frac{revenue}{customers}$	$\ast$ 0.187	0.187	Afrouzi, Drenik and Kim (2023)
Sales variation: intensive-margin share	0.154	0.150	Einav et al. (2021)
Aggregate markup	$\ast$ 1.250	1.250	Capital IQ

Notes: $\ast$ indicates model moments computed after applying a filter that mimics selection into Capital IQ / Compustat; see Appendix B.2 and Table 7 for details.

First, we assume that intermediate good producers' log TFP is described by a Gaussian AR1 process with parameters $(\sigma_z, \rho_z, \mu_z)$. As in most models of firm dynamics, sales con-

[20]For our purposes it would be equivalent to estimate markups using the production function approach. The difference to our measure $\mu = \frac{\text{sales}}{\text{cogs}}$ is the variable cost elasticity. This elasticity is commonly assumed to be constant within industries and therefore absorbed into industry fixed effects in our regressions. Relatedly, Grassi, Morzenti and de Ridder (2022) argue that cross-sectional variation in markups obtained from the production function approach accurately reflects true markup variation.

centration is increasing in the unconditional dispersion of $\ln z$, σ_z. We target both the top 10% and the top 1% sales share in the Statistics of U.S. Businesses (SUSB). More specific to our setup, the annual auto-correlation of $\ln z$, ρ_z, has a strong effect on the relation between firms' marketing expenditures and markups: the more positive it is, the stronger the relation between these two key objects, as today's high-markup firms expect to be highly profitable in the future as well, increasing their incentive to accumulate persistent customer relationships. Thus, we target the cross-sectional elasticity of the marketing expenses-to-revenue ratio to the markup, which equals 1.729 in Capital IQ. Moreover, the mean of $\ln z$ (relative to the normalization of in-house TFP at one), μ_z, affects how markups vary with size: the higher it is, the fewer firms are affected by the outside option of customers, and the weaker is the markup–size relation. We follow Afrouzi, Drenik and Kim (2023) and target the elasticity of the markup to sales per customer, which they estimate as 0.187.[21]

Second, we assume a functional form for the marketing investment cost akin to the search cost function in Mitman and Rabinovich (2015):

$$m(i) = \frac{\bar{m}}{1+\psi}\left((1-i)^{-1-\psi} - 1\right) - \bar{m}i \tag{40}$$

This functional form implies optimal choices of $i \in (0,1)$, which guarantees that quit rates are well-defined (below one). The marketing cost level $\bar{m}$ positively affects the aggregate markup, as with higher marketing costs firms are shielded from competition and optimally set higher markups. We target an aggregate markup of 1.25 in line with the literature (Edmond, Midrigan and Xu, 2022; Afrouzi, Drenik and Kim, 2023). While the curvature parameter ψ also affects the markup level, it has a particularly strong impact on the customer quit rate: more convex marketing costs imply less reallocation to top firms, and hence a lower quit rate. While the customer quit rate is difficult to pin down and varies substantially across industries, we follow the literature and target a firm-level average quit rate of 15% following Gourio and Rudanko (2014).

Empirically, the majority of overall sales variation across firms is accounted for by the extensive margin of customers. Using credit card data, Einav et al. (2021) estimate it to account for 85%, while only 15% of overall sales variation is accounted for by variation in sales per customer. In our model, the substitution elasticity across final good firms' divisions σ controls the intensive margin share.[22]

[21] See Table A.4, which uses variation from all products in the Nielsen Consumer Panel sample and accounts for marginal cost variation by saturating the model with fixed effects.

[22] Alternatively, Afrouzi, Drenik and Kim (2023) use Nielsen data to estimate a three-way decomposition of total sales variation into number of customers (80.7%), sales per customer (11.4%), and covariance (7.9%). Our model also compares favorably with this decomposition, the respective shares are 76.6%, 7.5%, and 15.9%. The central role of the number of customers in explaining sales variation is consistent with Bernard et al.

The dynamics of entry and exit are governed by the exogenous firm exit rate δ_F and the number of entrants per unmatched χ. δ_F is closely related to the firm exit rate, which we estimate to be 0.085 using data from the Business Dynamics Statistics (BDS). χ, which can also be interpreted as the innovation rate of unmatched divisions, positively affects the relative size of entrants: the higher χ is, the smaller is the subsequent growth of entrants. We use as target the average employment of entrants relative to incumbents of age 1-19, which Sterk, Sedláček and Pugsley (2021) estimate to be 0.540.

Lastly, we target an aggregate labor share of 0.615 (Bureau of Labor Statistics: non-farm business sector in 2017) by varying the capital elasticity in production, α.

Fixed parameters and normalizations. We fix a small number of parameters ex ante to typical values. We set the annual discount factor to $\beta = 0.95$, which results in an empirically reasonable physical capital-to-GDP ratio of 2.7. We assume log consumption utility and a dis-utility of labor supply $\nu(L) = \omega \frac{L^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}$. We normalize labor supply to one in the calibrated baseline steady state, which requires $\omega = 0.711$, and assume a Frisch labor supply elasticity of $\gamma = \frac{1}{2}$. This elasticity is irrelevant for the baseline steady state, it only matters for counterfactuals. Finally, we set the physical capital depreciation rate to $\delta_K = 0.05$.

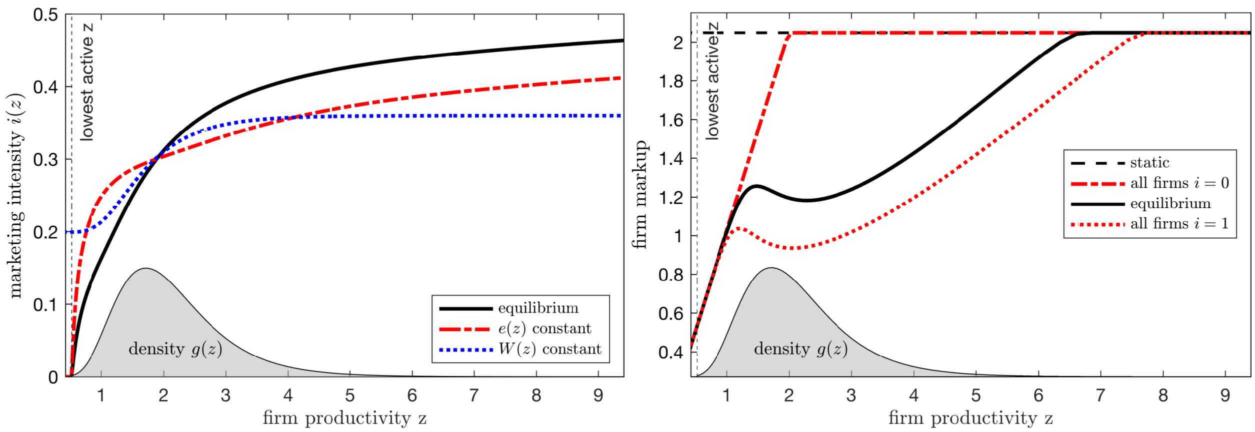
3.2 Model Properties and Relation to Empirical Literature

Firm Decision Rules. Figure 1 displays firms' decision rules. The left panel shows that the marketing intensity is increasing in firm TFP z. This is both because customer conversion rates $e(z)$ are increasing in z in equilibrium, and because values per customer at production $W(z)$ are increasing in z, as shown by the two counterfactual decision rules holding constant either margin.

The right panel displays firm markups, which are non-monotone (solid black line). To understand this, note first that under CES demand without outside options, static optimal markups are constant and equal to $\mu = \frac{\sigma}{\sigma-1}$. The possibility of in-house production creates an initially increasing and then flat markup even without any poaching and customer acquisition, as indicated by the counterfactual markup when all firms abstain from marketing. The equilibrium markup is non-monotone since firms in the middle of the distribution face the stiffest competition for their customers—the relevant metric of competition is the product of the marketing intensity and customer density $g(z)$. The higher other firms' marketing, the higher is the extensive margin price elasticity of demand. In turn, higher competition lowers markups. Moreover, marketing and lower markups are also complementary actions

(2022), who study firm-to-firm transactions in Belgium. Since they do not report a comparable decomposition, we cannot benchmark directly to their results.

Figure 1: Firm decision rules



(a) Marketing intensity

(b) Markup

Notes: The left panel displays firms' equilibrium marketing intensity as solid black line. The red dash-dotted line shows a counterfactual marketing intensity when holding the conversion rate $e(z)$ constant, while the blue dotted line shows a counterfactual holding constant the value per customer $W(z)$. The right panel shows firms' equilibrium markups as solid black line. The horizontal dashed blacked line indicates the optimal static markup under CES demand. The red dash-dotted (dotted) line is the counterfactual markup when all firms set their marketing intensities to the minimum of $i = 0$ (maximum of $i = 1$) for one period only. In both panels, $g(z)$ is the density of customers across firms of type z.

for an individual firm: if a given firm increases its marketing spending, it comes into contact with more potential new customers; hence, the incentive to lower its markup below the static markup increases. In the far right tail, even though marketing intensities are high, the density of customers vanishes and the equilibrium markup approaches the static monopoly markup.

Untargeted Moments. The model generates an aggregate marketing spending to value added ratio of 0.093. In Capital IQ, this ratio equals 0.080. The literal interpretation of this small discrepancy is that Capital IQ's sales and marketing spending variable captures most, but not all, of customer base investment.[23]

Our calibration results in an annual auto-correlation of log firm physical TFP $\rho_z = 0.556$. As explained above, we identify this parameter based on cross-sectional variation, in particular the relation between marketing spending and markups. Direct evidence for this parameter is difficult to obtain, especially outside of manufacturing. Foster, Haltiwanger and Syverson (2008) estimate a value of 0.79 in U.S. manufacturing. To validate our calibration including non-manufacturing sectors, we compare the auto-correlation of firm revenue. Log firm revenue follows close to a random walk in both data (0.991) and model (0.998). The model's

[23]We do not target the level of sales and marketing spending in Capital IQ precisely to allow for the possibility that it does not capture all customer investment. See Appendix B.1 for further details.

customer base dynamics thus generate significant additional persistence in firm revenue compared to the underlying TFP fundamental.

Since sales per customer and markups are a function of z only, the model generates zero correlation between markups and the number of customers, conditional on sales per customer. This mirrors precisely the empirical finding in Afrouzi, Drenik and Kim (2023).

Turning to the customer growth margins, Afrouzi, Drenik and Kim (2023) find that the elasticity of the number of customers with respect to marketing spending equals 0.09. In the model, the unconditional elasticity is higher at 0.615, but most of this effect is not causal (high-z firms spend more on marketing, and have accumulated more customers). Holding constant z, the model elasticity drops to 0.046. Little representative evidence exists for the response of the quit rate to price. Paciello, Pozzi and Trachter (2019) estimate a range of monthly semi-elasticities of the quit rate to the price of 0.013–0.16 for outlets of a large US retail chain. Aggregating these to annual frequency yields a range of 0.1–1.5. Our model's annual semi-elasticity equals 0.216, at the lower end of this range. Given that the literature has generally assumed exogenous separations of customers—corresponding to an elasticity of the quit rate to the price of zero—we consider the model implied elasticity a conservative deviation from existing work given the results of Paciello, Pozzi and Trachter (2019). Finally, Baker, Baugh and Sammon (2023) find that annual customer churn equals 40.8%. Our model, targeting the 15% customer loss rate reported in Gourio and Rudanko (2014), generates a comparable magnitude at 29.5%.

A key finding in the literature is that firms rely on non-price actions to acquire new customers, while they do not use lower prices to attract customers when entering new markets (Argente et al., 2025; Fitzgerald, Haller and Yedid-Levi, 2024). One way to account for this finding in models of customer dynamics is to rule out that prices impact customer acquisition and retention (e.g. Einav et al., 2021; Afrouzi, Drenik and Kim, 2023). This finding, however, does not imply that the extensive margin of customers does not react to prices (Paciello, Pozzi and Trachter, 2019; Roldan-Blanco and Gilbukh, 2021). Indeed, in our model prices are allocative—they affect both the customer acquisition and the customer loss margin, in addition to sales per customer. Under constant returns to marketing and existing customers, firms choose optimally not to vary prices with the number of existing customers.

4 Competition for Customers and Misallocation

This section assesses the efficiency of the market allocation. Inefficiencies arise on two margins. First, search frictions in the intermediate-input market prevent customers from accessing all suppliers, allowing positive markups. Conditional on a given distribution of matches, the *level* and *dispersion* of markups distort both the aggregate scale of production and the intensive-margin allocation across suppliers, nesting the sources of misallocation studied e.g. in Baqaee and Farhi (2020) and Edmond, Midrigan and Xu (2022). Second, the distribution of customers across suppliers with different productivities is endogenous to firms' marketing, entry, and exit. These choice reflect suppliers' private profits of matching with a customer, and need not coincide with social values. We characterize both margins of inefficiencies in the market equilibrium relative to a planner benchmark and quantify each component in the calibrated economy. In addition, we derive implications for product-market policy, including taxes on marketing and subsidies to production.

4.1 Efficient Allocation of Demand

We begin with the planner benchmark, subject to the same search frictions and technologies as in the market economy. We proceed in two steps: First, holding fixed the distribution of customers across suppliers, we isolate the welfare cost of an aggregate markup wedge and dispersion in markups across suppliers as standard in the literature. Second, we additionally allow the planner to choose marketing, entry, and exit of intermediate good producers to study the efficiency of the customer allocation itself.

Markup Distortions Given a Distribution of Customers

We begin with the intensive-margin allocation of demand, taking as given the distribution of matches $G(z)$ and the mass of unmatched customers u at the production stage. Appendix A.1 shows that the aggregate markup $\mathcal{M}$, aggregate productivity in the market economy Z^M, and aggregate productivity under the planner allocation Z^P absent markups are

$$\mathcal{M} = u\frac{\bar{\ell}y(\bar{mc})}{L} + \int_0^\infty \frac{\bar{\ell}y(p(z))}{zL}\mu(z)dG(z), \tag{41}$$

$$Z^M = \left(u\left(\frac{1}{\mathcal{M}}\right)^{-\sigma} + \int_0^\infty z^{\sigma-1}\left(\frac{\mu(z)}{\mathcal{M}}\right)^{-\sigma} dG(z)\right)^{\frac{1}{\sigma-1}}, \tag{42}$$

$$Z^P = \left(u + \int_0^\infty z^{\sigma-1}dG(z)\right)^{\frac{1}{\sigma-1}}. \tag{43}$$

These expressions allow us to separate the welfare cost of the aggregate markup from the cost of markup dispersion across suppliers.

Cost of Aggregate Markup. Consider a social planner who chooses $\{C_t, L_t, K_t\}_{t=0}^{\infty}$ to maximize household utility, taking aggregate productivity Z^M as given. The planner solves

$$U^P(Z^M, \Gamma^M) = \max_{C_t, L_t, K_{t+1}} \sum_{t=0}^{\infty} \beta^t \{u(C_t) - v(L_t)\} \tag{44}$$

$$\text{s.t. } Z^M K_t^{\alpha} L_t^{1-\alpha} + (1-\delta_k) K_t = C_t + K_{t+1} + \Gamma^M \tag{45}$$

for given K_0. Here $\Gamma^M = \int_0^\infty m(i(z))\, dG^i(z)$ is the cost of marketing required to sustain the distribution of matches underlying Z^M, which the planner takes as given. In steady state, the FOCs for the planner's choice of labor and capital hold as

$$v'(L) = (1-\alpha) Z^M K^{\alpha} L^{-\alpha} u'(C), \tag{46}$$

$$1 = \beta \left(\alpha Z^M K^{\alpha-1} L^{1-\alpha} + 1 - \delta_K \right). \tag{47}$$

The presence of an aggregate markup distorts the aggregate quantities of inputs into production in the market economy. For labor supply, note that $\mathcal{M} = \frac{(1-\alpha)Y}{wL} \Rightarrow w = \frac{1}{\mathcal{M}}(1-\alpha) Z^M K^{\alpha} L^{-\alpha}$ and for capital supply $\mathcal{M} = \frac{\alpha Y}{(r+\delta_K)K} \Rightarrow r + \delta_K = \frac{1}{\mathcal{M}} \alpha Z^M K^{\alpha-1} L^{1-\alpha}$. Substituting into the FOCs for labor and capital in (3) and (4) yields

$$v'(L) = w u'(C) = \frac{1}{\mathcal{M}} (1-\alpha) Z^M K^{\alpha} L^{-\alpha} u'(C), \tag{48}$$

$$1 = \beta (1 + r) = \beta \left(\frac{1}{\mathcal{M}} \alpha Z^M K^{\alpha-1} L^{1-\alpha} + 1 - \delta_K \right). \tag{49}$$

Thus, the aggregate markup acts like a uniform tax on consumption, distorting labor and capital supply to below their efficient level. Welfare in the market allocation, U^M, is obtained by solving the household problem in (1)–(2) at market prices. We define $\Delta_{\mathcal{M}} = U^P(Z^M, \Gamma^M) - U^M$ as the welfare loss from the aggregate markup.

Cost of Static Misallocation. Dispersion in markups distorts the allocation at the intensive margin of demand—sales per customer—but not at the extensive margin, which governs customers' switching decisions across suppliers. At this extensive margin, a customer deciding whether to transition from one supplier to another cares only about which firm offers the lower price, not by how much it is lower. Thus, the allocation depends on the ranking of prices, not on their relative distances. Since prices are strictly decreasing in productivity z, customer switching assigns demand efficiently across firms, as in Menzio (2024a,b). By contrast, at the intensive margin relative price differences matter for quantities purchased, so dispersion in markups misallocates sales and lowers aggregate productivity.

To isolate the cost of static misallocation, we compare aggregate productivity in the market and planner economies, (42) and (43). Relative to the efficient allocation, markup dispersion places too little weight on the productivity of firms charging above-average markups and too much on that of firms charging below-average markups. This induces a static misallocation of demand across suppliers at the intensive margin: high-markup firms produce too little and low-markup firms too much. We denote $\Delta_Z = U^P(Z^P, \Gamma^M) - U^P(Z^M, \Gamma^M)$ the welfare loss from static misallocation due to markup dispersion.

Cost of Static. To isolate the cost of static misallocation, we compare aggregate productivity in the market and planner economies, (42) and (43). Relative to the efficient allocation, markup dispersion places too little weight on high-markup (high-price) firms' productivity and too much on low-markup firms', inducing misallocation of demand across suppliers at the intensive margin—high-markup firms produce too little and low-markup firms too much. We denote $\Delta_Z = U^P(Z^P, \Gamma^M) - U^P(Z^M, \Gamma^M)$ the welfare loss from static misallocation due to markup dispersion.

Allocating Customers

So far, we have taken the distribution of customers across suppliers as given. We now expand the planner's problem to include the choice of each supplier's marketing intensity i and the entry and exit of firms. The resulting allocation, with welfare denoted by U^*, characterizes the efficient distribution of customers across suppliers. We denote $\Delta_G = U^* - U^P(Z^P, \Gamma^M)$ as the welfare loss from customer misallocation. Appendix A.2 outlines the full planner problem; here we summarize the key results to build intuition for how inefficiencies arise.

When solving for U^*, we take the market economy as an initial condition and consider the full transition to a new steady state. Here, to ease exposition, we impose that both the market and planner economies are in steady state. Combining suppliers' optimality condition (38) with the equilibrium condition $\frac{1}{1+r} = \beta$ yields the optimality condition for suppliers' marketing intensity $i(z)$ in the market economy as

$$\frac{\partial m(i(z))}{\partial i} = \left(u + G^i(z)\right)\left(\pi(p(z), z) + \beta(1-\delta_F)\int_0^\infty f(z'|z)\max\left\{V\left(z'\right), 0\right\} dz'\right) \tag{50}$$

The planner's optimal solution for $i(z)$, derived in Appendix A.2, satisfies

$$\begin{aligned}\frac{\partial m(i_t(z))}{\partial i} &= (n^u + N(z))\left[\frac{z^{\sigma-1}-1}{\sigma-1}(Z^*)^{1-\sigma}Y^* + \beta(1-\delta^F)\int_0^\infty f(z'|z)\max\left\{\hat{V}(z'), 0\right\} dz'\right] \\ &\quad - (n^u + N(z))\beta(1-\delta_F)\int_0^\infty \chi f_0(z')\max\left\{\hat{V}(z'), 0\right\} dz' \\ &\quad - \int_0^z n(\hat{z})\left[\frac{\hat{z}^{\sigma-1}-1}{\sigma-1}(Z^*)^{1-\sigma}Y^* d\hat{z} + \beta\int_0^\infty (1-\hat{\delta}(z'))(f(z'|\hat{z}) - \chi f_0(z'))\hat{V}(z') dz'\right] d\hat{z}\end{aligned} \tag{51}$$

where n^u and $n(z)$ are the mass of unmatched customers and mass of customers matched with firms of type z at the beginning of a period, $N(z) = \int_0^z n_t(\hat{z})d\hat{z}$, Z^* and Y^* are aggregate productivity and output implied by the planner's optimal solution. $\hat{V}(z)$ is the value of a match to the planner, given by[24]

$$(52)\quad \hat{V}(z) = -m(i(z)) + \hat{\rho}(z,i(z))\left[\frac{z^{\sigma-1}-1}{\sigma-1}(Z^*)^{1-\sigma}Y^* + \beta(1-\delta^F)\int_0^\infty (f(z'|z)-\chi f_0(z'))\max\left\{\hat{V}(z'),0\right\}dz'\right]$$
$$-\int_0^z n(\hat{z})(i(z)+i(\hat{z}))\left[\frac{\hat{z}^{\sigma-1}-1}{\sigma-1}(Z^*)^{1-\sigma}Y^*\beta(1-\delta^F)\int_0^\infty (f(z'|\hat{z})-\chi f_0(z'))\max\left\{\hat{V}(z'),0\right\}dz'\right]d\hat{z}$$

The planner's optimal choice for marketing differs from firms' equilibrium choice for three reasons. First, the static surplus of a match differs between the planner and the firm. To see this, note that we can write firms' static surplus (profit) per customer as

$$(53)\quad \pi(p(z),z) = \left(p(z) - \frac{\bar{m}c^M}{z}\right)p(z)^{-\sigma}Y^M = \frac{z^{\sigma-1}(Z^M)^{1-\sigma}Y^M}{\sigma-1}\left[(\sigma-1)\left(\left(\frac{\mu(z)}{\mathcal{M}}\right)^{1-\sigma} - \frac{1}{\mathcal{M}}\left(\frac{\mu(z)}{\mathcal{M}}\right)^{-\sigma}\right)\right]$$

We can compare this to the static surplus for the planner given by $\frac{z^{\sigma-1}-1}{\sigma-1}(Z^*)^{1-\sigma}Y^*$. The main difference between the two is that the planner takes into account the surplus of a match to the customer, adjusted for the outside option of being unmatched and producing with productivity 1. The absence of customers' surplus from the suppliers' decision induces *under-investment* in marketing in equilibrium. We can think of this as mirroring the failure of the condition described by Hosios (1990) for efficiency in search markets: firms bear the entire cost of marketing to generate matches, but do not internalize the entire surplus. In addition, the value of a match differs due to the distortions from markups and differences in aggregate productivity and output.

Second, firms do not internalize that by increasing their marketing to generate matches for themselves, they also limit matches with other suppliers as the overall number of customers is fixed. Part of firms' new customers are poached directly from competitiors (third line in (51)), destroying existing matches. In addition, customers that are matched in the current period will not be contacted by potential entrants at the beginning of the following period (second line in (51)), reducing the number of customers at entering firms. Finally, an additional customer that is matched with a supplier of productivity z leads to more poaching by z firms and less poaching by non-z firms in the following period, due to the effect of the number of existing customers on the cost of marketing (second line in (52)). This business stealing externality at three margins induces *over-investment* in marketing in equilibrium.

Third, the benefit of marketing depends on the conversion rate of contacts into matches, given by the fraction of customers that are unmatched or matched with firms of productivity

[24]We show in Appendix A.2 that the planner chooses a firm to exit (not enter) whenever $\hat{V}(z) < 0$.

lower than z. As differences in marketing policies between the planner solution and the market equilibrium induce different distributions of customers, the conversion rate differs between the two economies.

Decomposition of Welfare Cost. Based on the discussion above, we can decompose the aggregate welfare loss from inefficiencies in the market economy, either due to markups or inefficient investment in demand, into the three margins as follows:

$$\Delta = U^* - U^M = \underbrace{U^P(Z^M, \Gamma^M) - U^M}_{\Delta_{\mathcal{M}} \text{ (markup wedge)}} + \underbrace{U^P(Z^P, \Gamma^M) - U^P(Z^M, \Gamma^M)}_{\Delta_Z \text{ (static misallocation)}} + \underbrace{U^* - U^P(Z^P, \Gamma^M)}_{\Delta_G \text{ (customer misallocation)}} \tag{54}$$

We quantify each term in this decomposition for our calibrated economy below. We report welfare changes in terms of consumption equivalents (CEV): given two life-time utility values U_1 and U_2 (and log consumption utility), it is given as

$$CEV(U_1, U_2) \equiv \exp\left((1-\beta)(U_2 - U_1)\right) - 1. \tag{55}$$

4.2 Quantifying Sources of Misallocation

Table 2 compares the decentralized equilibrium to the social planner allocation.[25] The planner would raise labor and capital inputs significantly, increasing GDP by 19.45% eventually. Welfare rises by 10.33% in steady state, but only by 3.61% along the transition, reflecting the build-up cost of a significant increase in the capital stock. The planner chooses to reduce marketing investment by 21.81%. Aggregate TFP remains essentially unchanged. To understand these outcomes, we apply the theoretical decomposition introduced above to isolate the contribution of the aggregate markup wedge, dispersion in markups, and misallocation of customers across firms.

Cost of Aggregate Markup. The aggregate markup wedge accounts for a large part of the efficiency losses, as shown in column $\Delta_{\mathcal{M}}$ of Table 2. Eliminating it would raise welfare by 8.52% in steady state but only by 1.52% along the transition, reflecting the gradual build-up of the capital stock. The distortion reflects under-investment in capital and inefficiently low labor input. This result is standard in the literature: it arises from the economy-wide markup of $\bar{\mu} = 1.21$ and can be replicated in a growth model with exogenous aggregate productivity

[25] We report here quantities across steady states as well as welfare taking into account transitional dynamics. The full transition paths for all quantities are reported in Appendix C.1.

Table 2: Efficiency Losses of Equilibrium Allocation

		Equil.	Planner	% Δ	% $\Delta_{\mathcal{M}}$	% Δ_Z	% Δ_G
Gross output	Y	1.081	1.258	16.34	17.02	2.02	-2.70
GDP	$Y - \Gamma$	1.000	1.194	19.45	18.40	2.18	-1.14
Marketing investment	Γ	0.081	0.064	-21.81	0	0	-21.81
Tangible capital	K	2.704	3.812	40.96	41.78	2.45	-3.27
Labor	L	0.251	0.268	6.69	7.31	-0.05	-0.57
Consumption	C	0.865	1.004	16.08	14.75	2.14	-0.81
Aggregate TFP	Z	2.055	2.055	0.00	0	1.22	-1.22
Welfare (CEV) – steady state				10.33	8.52	2.07	-0.25
Welfare (CEV) – transition				3.61	1.52	1.63	0.46

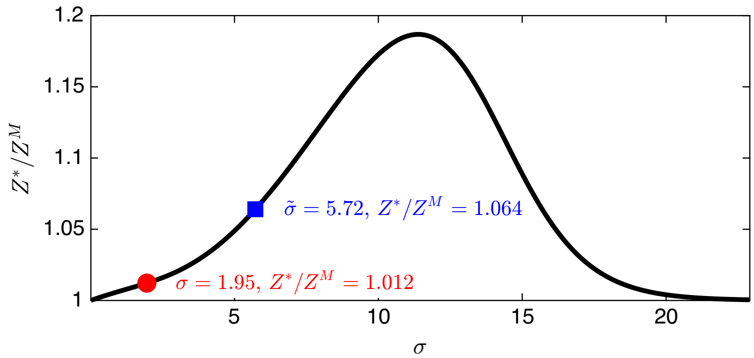
Notes: Quantities are normalized so that $GDP = 1$ in equilibrium. Column % Δ reports percent changes from the equilibrium to the social planner allocation. The last row shows welfare gains along the transition; all other rows refer to the steady state. The three rightmost columns decompose efficiency losses into (i) an aggregate markup wedge, (ii) markup dispersion (intensive-margin misallocation), and (iii) misallocation of customers across firms. See Section 4.1 for details.

set equal to that of the market economy and consistent calibration of household preferences and technology.

Cost of Markup Dispersion. Dispersion in intermediate firms' markups, and the resulting misallocation of demand at the intensive margin, lowers aggregate TFP by 1.22% and welfare by 1.63% along the transition (2.07% across steady states), as shown in column Δ_Z of Table 2. To understand this finding, note that the cost of intensive margin misallocation depends on two factors, the magnitude of markup dispersion and the cost of a given level of dispersion. In many models of imperfect competition, in which market power arises from preferences over differentiated varieties, these two factors are governed by the same parameter: Customers' elasticity of substitution at the intensive margin of demand controls both suppliers' price elasticity of demand—and thereby markups—as well as the distortions implied by a given distribution of markups to the allocation of production.

In our setting, suppliers' perceived demand elasticity—and by extension their markup—is determined primarily by the extensive margin; i.e., search frictions and marketing. We are therefore free to calibrate the intensive-margin elasticity σ separately. We set $\sigma = 1.95$ to match evidence that only about 15% of total sales variation is due to differences in sales per customer. A low σ, in turn, implies that a given level of markup dispersion generates little misallocation, since demand is less responsive to price differences. To quantify this argument, we plot the ratio of the social planner to the equilibrium aggregate TFP, $\frac{Z^P}{Z^M}$, holding constant equilibrium firm markups and the distribution of customers, for a range of

Figure 2: Intensive Margin Misallocation as Function of Intensive Margin Elasticity



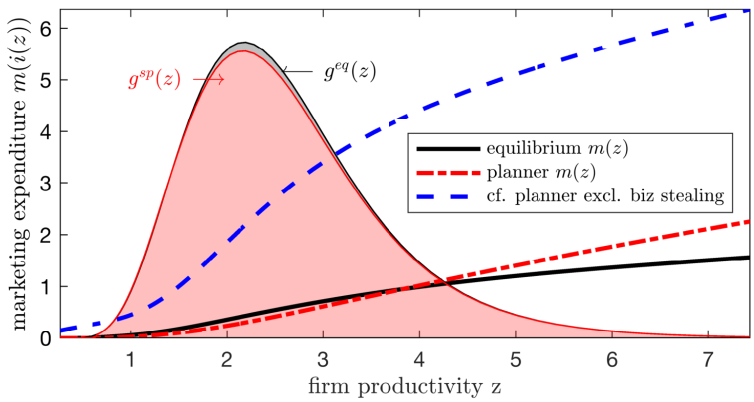
Notes: To compute $\frac{Z^*}{Z^M}$, we evaluate (42) and (43) holding constant the equilibrium distribution of customers $G(z)$ and equilibrium markups $\mu(z)$ and varying only σ.

values of σ in Figure 2. Counterfactually imposing $\tilde{\sigma} = 5.72$, the elasticity needed to generate the same aggregate markup $\bar{\mu} = 1.21$ via the relation $\bar{\mu} = \frac{\tilde{\sigma}}{\tilde{\sigma}-1}$, the efficiency cost of markup dispersion would rise more than fivefold from 1.2% to 6.4%. This result of a comparably low cost of markup dispersion is directly informed by our calibration approach.

Cost of Customer Misallocation. In equilibrium, firms over-invest in marketing by 21.81% relative to the planner. Along the transition, correcting this distortion raises welfare by 0.46%, but the steady-state comparison shows a welfare loss of 0.25% (column Δ_G of Table 2). The apparent loss arises because the planner maximizes lifetime rather than steady-state utility: equilibrium features inefficiently high marketing investment and thus an inefficiently large intangible capital stock, which boosts steady-state output but at the cost of lower lifetime welfare. As the intangible stock is run down during the transition, the associated welfare gain increases relative to a comparison of steady states.

The net adjustment of marketing relative to the decentralized equilibrium masks two opposing forces: differences in the static surplus of a match and the business stealing externality. Figure 3 illustrates their opposing effects. The black solid line shows equilibrium marketing, while the red dash-dotted line shows the planner's allocation. The planner cuts marketing for most firms but raises it in the right tail of the productivity distribution, choosing higher marketing spending than in equilibrium for 3% of firms. To isolate the different effects, we also plot a counterfactual allocation of a "naive" planner (blue dashed line) who values the full static match surplus—including consumer surplus and absent markup distortions to intensive margin demand—but ignores the business stealing externality. This counterfactual planner would choose substantially higher marketing. The difference between the blue and black lines reflects the gap between suppliers' private surplus and the full static match sur-

Figure 3: Equilibrium vs. Social Planner Marketing Spending



Notes: The black solid line shows equilibrium marketing spending across firms of type z, while the red dash-dotted line provides the planner equivalent. The blue dashed line is the counterfactual marketing spending chosen by a naive planner that ignores the business stealing externality. g^{sp} (g^{eq}) indicates the planner's (equilibrium) customer distribution.

plus, while the difference between the blue and red lines isolates the effect of the business stealing externality. A third potential inefficiency from heterogeneous conversion rates is quantitatively negligible, as the customer distributions under the planner and equilibrium remain close.

Marketing and the Cost of Markups. Table 3 illustrates how marketing affects competition for customers, markups, and misallocation through counterfactual changes in the cost parameter $\bar{m}$ of the marketing function. Halving $\bar{m}$ intensifies customer acquisition and poaching, raising suppliers' perceived demand elasticity and compressing markups. The aggregate markup and dispersion components of the welfare loss fall by 21% and 13%, respectively. Conversely, doubling $\bar{m}$ weakens competitive pressure and raises the welfare cost of both the level and the dispersion of markups. Changes in the misallocation of customers are modest, since altering the cost of marketing shifts both the equilibrium and the planner's optimal investment in demand. Overall, this exercise shows that the pro-competitive effects of marketing expenditures improve allocative efficiency by reducing distortions from markups, even though equilibrium marketing remains inefficiently high.

4.3 The Optimal Taxation of Marketing and Production Subsidies

The previous section showed that the market equilibrium yields an inefficient allocation of demand relative to the social optimum. Here we examine how much efficiency a government can restore using a restricted set of policy instruments. We consider two tools: First, a flat tax τ_M on marketing expenditures, which directly targets over-investment in marketing.

Table 3: Welfare Losses from Misallocation under Alternative Marketing Costs

	Total	Aggr. Markup	Markup Dispersion	Customer Misallocation
Baseline calibration	3.61	1.52	1.63	0.46
$\bar{m} - 50\%$	3.04	1.20	1.42	0.42
$\bar{m} + 100\%$	4.42	1.90	2.04	0.48

Notes: The table reports welfare gains along the transition to the efficient steady state (in percent of consumption), decomposed into components due to aggregate markups, markup dispersion, and customer misallocation. The first row corresponds to the baseline calibration. The second and third rows report outcomes when the cost of marketing $\bar{m}$ is cut in half or doubled, respectively.

Second, a subsidy τ_P on production inputs (capital and labor at an equal rate), designed to offset the level distortion from markups. The optimization problem of intermediate-good producers becomes

$$V_t(z) = \max_{p,i} \left\{ -(1+\tau_M)m(i) + \rho_t(p,i) \left[\pi_t \left(p, \frac{z}{1-\tau_P} \right) + \frac{1-\delta_F}{1+r_{t+1}} \mathbb{E} \left[\max \left\{ V_{t+1}\left(z'\right), 0 \right\} | z \right] \right] \right\} \tag{56}$$

We impose government budget clearing period by period, such that

$$\tau_P \bar{m} c_t \left(u_t y_t(\bar{m} c_t) + \int_0^\infty \frac{1}{z} y_t(p_t(z)) dG_t(z) \right) = T_t + \tau_M \int_0^\infty m(i_t(z)) dG_t^i(z) \tag{57}$$

where T_t is a lumpsum tax on households. The government chooses constant levels of τ_M and τ_P to maximize welfare, accounting for transitional dynamics. We consider three cases: a marketing tax only, a production subsidy only, and the joint use of both instruments. Table 4 summarizes the results.[26]

The joint optimum features a 41.10% marketing tax and an 18.16% production subsidy, raising welfare by 1.69% along the transition. This gain amounts to about half of the total welfare improvement attainable under the planner allocation, and roughly 85% of the welfare gain from eliminating the aggregate markup combined with efficient marketing. The marketing tax reduces equilibrium marketing spending by 17.32%; i.e., not by as much as under the social optimum.

Complementary policy instruments. Comparing the jointly optimal marketing tax and production subsidy to their optimal levels in isolation, we find that the two instruments are complements. The optimal production subsidy is 15.69% in isolation but rises to 18.16% when combined with a marketing tax. Likewise, the optimal marketing tax is 17.66% in

[26]The table compares steady states for all quantities and reports welfare changes along the transition. Transition paths for all variables are provided in Appendix C.2.

Table 4: Optimal Marketing Taxes and Production Subsidies

		Baseline	τ_P only		τ_m only		Both	
		Value	Value	%Δ	Value	%Δ	Value	%Δ
Marketing tax (%)	τ_M		0		17.66		41.10	
Production subsidy (%)	τ_P		15.69		0		18.16	
Lump-sum tax (% of GDP)	T		19.57		-1.21		19.85	
Gross output	Y	1.081	1.268	17.28	1.056	-2.33	1.242	14.88
GDP	$Y-\Gamma$	1.000	1.174	17.42	0.988	-1.23	1.175	17.50
Marketing investment	Γ	0.081	0.094	15.64	0.069	-15.81	0.067	-17.32
Tangible capital	K	2.704	3.780	39.79	2.626	-2.90	3.771	39.47
Labor	L	0.251	0.269	7.06	0.250	-0.65	0.269	6.93
Consumption	C	0.865	0.985	13.92	0.856	-0.97	0.986	14.06
Aggregate TFP	Z	2.055	2.072	0.83	2.035	-0.98	2.033	-1.09
Welfare (CEV) – steady state				7.95		-0.51		8.20
Welfare (CEV) – transition				1.25		0.11		1.69

Notes: Baseline refers to the calibrated economy without taxes. %Δ refers to percent changes relative to the baseline. In each of the three tax regimes, we solve for the constant levels of the marketing tax and/or production subsidy that maximize welfare taking into account the transition (from the baseline steady state to the steady state with taxes). Government budget clearing is achieved period by period by imposing a lumpsum tax on households. Quantities normalized so that $GDP = 1$ in the baseline.

isolation but more than doubles to 41.10% when paired with a production subsidy. The welfare gain of introducing both instruments jointly is 1.69%, exceeding the sum of the gains from each instrument in isolation (1.36%) by about 25%.

The complementarity stems from the two-way interaction between marketing and markups in equilibrium. Marketing rises with firms' profits per customer. A production subsidy boosts profits and thereby induces additional (inefficient) marketing, as reflected in the 15.64% increase in marketing investment when the subsidy is introduced alone. Conversely, marketing intensifies price competition and dampens markup distortions. A tax that raises the cost of marketing therefore worsens markup distortions. This is consistent with Table 3, discussed earlier, which shows that the misallocative costs of markups rise in the marketing cost parameter $\bar{m}$. With a production subsidy in place to offset aggregate markup distortions, however, the marketing tax can be set more aggressively to curb over-investment in demand.

These findings caution against evaluating different tools for product market interventions in isolation when firms use both prices and investment in demand to compete for customers. Policies that address distortions at one margin can exacerbate inefficiencies at the other. Production subsidies and marketing taxes are thus best viewed as joint tools for correcting both markup distortions and inefficient marketing.

Size-dependent policies. The two policies we analyze address average inefficiencies from marketing and markups but do not differentiate across firms and thus cannot eliminate heterogeneity in distortions across suppliers. In principle, one might consider size-dependent interventions, as often proposed in the literature, to counteract misallocation from markup dispersion and heterogeneity in over- or under-investment in marketing. Yet such policies are not only difficult to implement in practice, but also lack strong theoretical grounding in our framework. Firm size, measured by e.g. total sales, is not a sufficient statistic for productivity, markups, or marketing intensity. Much of the observed variation in sales and marketing reflects differences in customer base size, itself shaped by past productivity and past investment rather than current fundamentals. Consequently, policies conditioned only on total sales or total marketing expenditures would be poorly targeted.[27] To directly address firm productivity or markups, interventions would need to condition on the number of customers as well—an even less practical basis for policy design. For these reasons, we do not pursue size-dependent policies further.

5 Macroeconomic Effects of the Rise of Marketing

Since the 1980s, the share of marketing in aggregate revenue has increased. At the same time, industry sales concentration has increased as well. Through our theory, rising investment in demand is a natural explanation for the rise in industry concentration, as it reallocates customers towards the most productive, and larger, suppliers. In this section, we investigate technological change in marketing as a candidate driver of these developments. We treat our baseline calibration as current steady state of the U.S. economy, and compare it to a prior steady state with lower marketing spending, and lower industry concentration. We highlight implications of the rise in marketing and industry concentration for aggregate productivity, competition, and the value of intangible capital attributed to firms' customer base.

5.1 Calibration of Technological Change in Marketing

While the precise level of and trend in marketing expenditures depends on the measure applied, all indicators point towards an increase in spending as a share of output since the 1980s. Kaplan and Zoch (2020) document an increasing share and average wage of expansionary work, a broader concept including employment for sales, marketing as well as product development. Bronnenberg, Dube and Syverson (2022) find that the marketing manager share of payroll has increased 25% from 2005 to 2019. Data presented in Greenwood,

[27] Appendix Figure 12 illustrates that size distortions from markup dispersion vary systematically with productivity but show little systematic pattern across the revenue distribution.

Ma and Yorukoglu (2024) shows that the advertising-to-GDP ratio has increased 35% from 1980 to 2019. The KLEMS database provides a time series of investment in brands and customers at the level of national accounts.[28] The ratio of brand value investment to GDP has increased by 15% from 1985 to 2020. Conservatively, we target this 15% increase over time.[29] Over the same time period, industry sales concentration has also increased significantly. We target a 25.8% increase in the top 1% sales share, which we calculate as the average increase in the top 4 or top 20 firm sales share (whichever is closer to the top 1% of firms) across sectors reported in Autor et al. (2020).

We calibrate a change in the two marketing cost parameters ($\bar{m}$, ψ) to replicate the increase in marketing spending and sales concentration (top panel of Table 5). To interpret the changes, the middle panel of Table 5 reports the mean marketing cost as well as the ratio of the marketing cost at the 75th vs. 25th percentile of the marketing spending distribution, fixing marketing intensities $i(z)$ across the two columns for comparison: the mean marketing cost decreased by 28.5%, with a disproportionate decline for firms with higher marketing intensities (i.e., a decline in the convexity of $m(i)$). The change in the two parameters is identified. While both parameters affect both moments, the level parameter $\bar{m}$ primarily affects total marketing spending, while the curvature parameter ψ primarily affects sales concentration. Generating a rise in concentration through a reduction in the cost of marketing is consistent with the empirical findings of Shen (2025), who documents that sectors with a higher marketing intensity have seen a more sizable increase in concentration.

5.2 The Effects of Rising Marketing Spending

Figure 4a shows that, following the calibrated change in marketing technology, most firms increase their marketing intensity—especially those at the top of the productivity distribution, while less productive suppliers modestly reduce their efforts. This heterogeneity translates into differential customer growth rates along the productivity distribution, as illustrated in Figure 4b. The aggregate effects, reported across steady states, are shown in the bottom panel of Table 5: the rise in marketing triggers a 3.3% increase in aggregate TFP and a corresponding 4.6% increase in GDP. These gains stem from the reallocation of customers toward more productive suppliers.

The effect on the aggregate markup is theoretically ambiguous. On the one hand, more marketing increases competition for customers, reflected in higher turnover (increased quit

[28] See Bontadini et al. (2023) for a description of the KLEMS database and e.g. Corrado et al. (2022) for additional analysis on intangible investment using the data. The data can be accessed via the Luiss Lab of European Economics at `https://euklems-intanprod-llee.luiss.it`.

[29] The marketing and sales spending variable in Capital IQ, which we use to calibrate the baseline 2010s steady state, is not reliable before 2007 as explained by He, Mostrom and Sufi (2024).

and acquisition rates), lowering firms' individual markups. On the other hand, greater sales concentration reallocates economic activity toward more productive, high-markup firms, raising the aggregate markup through a composition effect. Qualitatively, these opposing forces are consistent with documented changes to firms' markups and labor shares in the U.S. economy (e.g., Autor et al., 2020; De Loecker, Eeckhout and Unger, 2020). In our model, the within-firm component is slightly larger than the between-firm component, leading to a 0.5% decline in the aggregate markup. This result highlights that with endogenous market power due to search frictions, changes in market concentration need not translate into higher markups.

Paradoxically, the intangible intensity of the capital stock—measured as the share of customer value $K^V = \int_0^\infty V(z)dG(z)$ in total capital $K^V + K$—declines even as marketing spending rises. Between the 1980s and the 2010s, intangible intensity falls by 6.5%. The reason is that while firms invest more to expand their own customer base, they also raise customer turnover at competitors. Higher churn shortens the expected duration of customer relationships, lowering the present discounted value of profits from each customer. This decline in valuations per customer dominates the increase in customer numbers, leading to lower aggregate intangible intensity. The result cautions against using the perpetual inventory method—which capitalizes expenditures to estimate intangible asset values—to measure the value of firms' customer bases.[30] Our framework suggests that marketing spending affects not only the number of customers a firm accumulates but also how each customer should be valued.

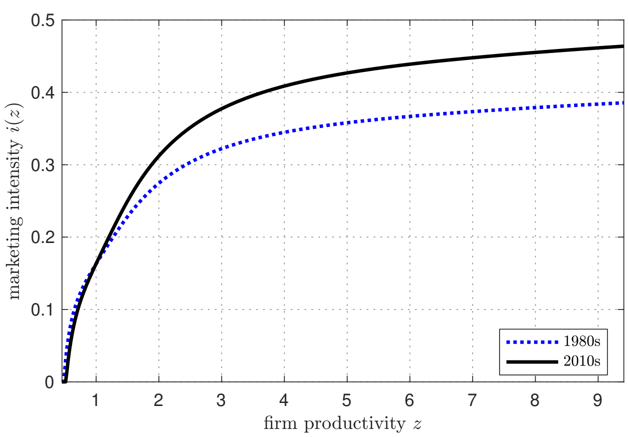
[30] For applications of the perpetual inventory method to measure general intangible capital, see Eisfeldt and Papanikolaou (2013), Eisfeldt and Papanikolaou (2014), and Peters and Taylor (2017).

Table 5: Rise in Marketing: 1980s vs. 2010s

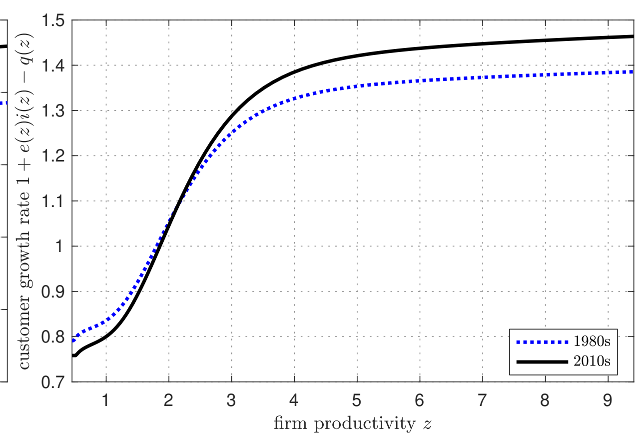
		1980s	2010s	% Change
Marketing / Gross Output (target +15%)	$\frac{\Gamma}{Y}$	0.065	0.075	15.00
Top 1% sales share (target +25.8%)		0.347	0.437	25.80
Average marketing cost (fixing $i(z)$)		0.338	0.241	-28.52
P75/P25 marketing cost (fixing $i(z)$)		2.907	2.460	-15.36
GDP	$Y - \Gamma$	1.000	1.046	4.56
Aggregate TFP	Z	1.990	2.055	3.26
Avg. customer quit rate	$\mathbb{E}_z\left[q(z)\right]$	0.126	0.150	19.31
Avg. customer acquisition rate	$\mathbb{E}_z\left[\frac{e(z)}{1+e(z)-q(z)}\right]$	0.144	0.159	10.75
Aggregate markup	$\mathcal{M}$	1.218	1.212	-0.54
Intangible capital intensity	$\frac{K^V}{K^V+K}$	0.394	0.369	-6.51

Notes: Quantities normalized so that $GDP = 1$ in the 1980s steady state. The table compares the calibrated 2010s steady state to another steady state ("1980s") with the two marketing technology parameters $(\bar{m}, \psi)$ re-calibrated to generate the targeted increase in marketing spending and sales concentration over time. Averages of quit and acquisition rates are taken across firms.

Figure 4: Technological Change in Marketing: Effects on Customer Growth



(a) Marketing intensity



(b) Customer growth

Notes: The figure compares firms' marketing intensity and customer growth in the calibrated 2010s steady state to another steady state ("1980s") with the two marketing technology parameters $(\bar{m}, \psi)$ re-calibrated to generate the targeted increase in marketing spending and sales concentration over time.

6 Conclusion

This paper develops a new quantitative framework in which investment in demand shapes how firms compete for customers. Producers are heterogeneous in productivity, set prices, and invest in marketing to reach customers in markets with search frictions. Customers can only buy from contacted suppliers and form persistent relationships, giving rise to dynamic demand accumulation.

The model generates rich industry dynamics and matches salient empirical facts on how firms compete for customers. It highlights two key effects of marketing for allocative efficiency. First, firms over-invest in marketing because they do not internalize the business-stealing externality imposed on competitors. Second, marketing exerts a pro-competitive force by increasing customer turnover and lowering markups, thereby mitigating misallocation from imperfect competition.

Quantitatively, the framework suggests that the rise in marketing expenditures between 1980 and 2010 can account for the observed increase in industry concentration without implying greater market power, while at the same time contributing positively to GDP growth. More broadly, the results underscore that policies aimed at correcting distortions from market power must account for firms' dual margins of competition—prices and demand investment.

References

Afrouzi, H., A. Drenik, and R. Kim (2023): "Concentration, Market Power, and Misallocation: The Role of Endogenous Customer Acquisition," Discussion paper, National Bureau of Economic Research.

Argente, D., D. Fitzgerald, S. Moreira, and A. Priolo (2025): "How Do Entrants Build Market Share? The Role of Demand Frictions," *American Economic Review: Insights*, 7(2), 160–76.

Argente, D., E. Méndez, and D. Van Patten (2024): "Cross-Border Product Adoption: Individual Imports, Migrant Networks, and Domestic Retailers," .

Arkolakis, C. (2010): "Market penetration costs and the new consumers margin in international trade," *Journal of political economy*, 118(6), 1151–1199.

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.

Bai, Y., J.-V. Rios-Rull, and K. Storesletten (2025): "Demand shocks as technology shocks," *Review of Economic Studies*, p. rdaf045.

Bailey, M., D. Johnston, T. Kuchler, J. Stroebel, and A. Wong (2022): "Peer effects in product adoption," *American Economic Journal: Applied Economics*, 14(3), 488–526.

Baker, S. R., B. Baugh, and M. Sammon (2023): "Customer churn and intangible capital," *Journal of Political Economy Macroeconomics*, 1(3), 447–505.

Baqaee, D. R., and E. Farhi (2020): "Productivity and misallocation in general equilibrium," *The Quarterly Journal of Economics*, 135(1), 105–163.

Baslandze, S., J. Greenwood, R. Marto, and S. Moreira (2023): "The expansion of varieties in the new age of advertising," *Review of Economic Dynamics*, 50, 171–210.

Bernard, A. B., E. Dhyne, G. Magerman, K. Manova, and A. Moxnes (2022): "The origins of firm heterogeneity: A production network approach," *Journal of Political Economy*, 130(7), 1765–1804.

Bontadini, F., C. Corrado, J. Haskel, M. Iommi, and C. Jona-Lasinio (2023): "EUKLEMS & INTANProd: industry productivity accounts with intangibles," .

Bronnenberg, B. J., J.-P. Dube, and C. Syverson (2022): "Marketing Investment and Intangible Brand Capital," *Journal of Economic Perspectives*, 36(3), 53–74.

Cavenaile, L., M. A. Celik, J. Perla, and P. Roldan-Blanco (2023): "A theory of dynamic product awareness and targeted advertising," Discussion paper, Working paper.

Cavenaile, L., M. A. Celik, P. Roldan-Blanco, and X. Tian (2025): "Style over substance? Advertising, innovation, and endogenous market structure," *Journal of Monetary Economics*, 149, 103683.

Cavenaile, L., and P. Roldan-Blanco (2021): "Advertising, innovation, and economic growth," *American Economic Journal: Macroeconomics*, 13(3), 251–303.

Chiavari, A. (2024): "Customer Accumulation, Returns to Scale, and Secular Trends," .

Coles, M. G., and D. T. Mortensen (2016): "Equilibrium Labor Turnover, Firm Growth, and Unemployment," *Econometrica*, 84(1), 347–363.

Corrado, C., J. Haskel, C. Jona-Lasinio, and M. Iommi (2022): "Intangible capital and modern economies," *Journal of Economic Perspectives*, 36(3), 3–28.

Crouzet, N., J. Eberly, A. Eisfeldt, and D. Papanikolaou (2024): "Intangible Capital, Firm Scope, and Growth," .

Crouzet, N., J. C. Eberly, A. L. Eisfeldt, and D. Papanikolaou (2022): "The economics of intangible capital," *Journal of Economic Perspectives*, 36(3), 29–52.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The rise of market power and the macroeconomic implications," *The 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.

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

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

Edmond, C., V. Midrigan, and D. Y. Xu (2022): "How Costly Are Markups?," *Journal of Political Economy*, forthcoming.

Einav, L., P. J. Klenow, J. D. Levin, and R. Murciano-Goroff (2021): "Customers and Retail Growth," Working Paper 29561, National Bureau of Economic Research.

Eisfeldt, A. L., and D. Papanikolaou (2013): "Organization capital and the cross-section of expected returns," *The Journal of Finance*, 68(4), 1365–1406.

——— (2014): "The value and ownership of intangible capital," *American Economic Review*, 104(5), 189–194.

Fernández-Villaverde, J., F. Mandelman, Y. Yu, and F. Zanetti (2024): "Search complementarities, aggregate fluctuations, and fiscal policy," *Review of Economic Studies*, p. rdae053.

Fitzgerald, D., S. Haller, and Y. Yedid-Levi (2024): "How exporters grow," *Review of Economic Studies*, 91(4), 2276–2306.

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.

Fukui, M., and T. Mukoyama (2025): "Efficiency in Job-Ladder Models," Discussion paper, National Bureau of Economic Research.

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

Grassi, B., G. Morzenti, and M. de Ridder (2022): "The Hitchhiker's guide to markup estimation," POID Working Papers 063, Centre for Economic Performance, LSE.

Greenwood, J., Y. Ma, and M. Yorukoglu (2024): ""You Will:" A Macroeconomic Analysis of Digital Advertising," *The Review of Economic Studies*, 92(3), 1837–1881.

He, B., L. I. Mostrom, and A. Sufi (2024): "Investing in Customer Capital," Working Paper 33171, National Bureau of Economic Research.

Heise, S. (2024): "Firm-to-firm relationships and the pass-through of shocks: Theory and evidence," *Review of Economics and Statistics*, pp. 1–45.

Hosios, A. J. (1990): "On the efficiency of matching and related models of search and unemployment," *The Review of Economic Studies*, 57(2), 279–298.

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

Ignaszak, M., and P. Sedláček (2023): "Customer Acquisition, Business Dynamism and Aggregate Growth," .

Kaplan, G., and P. Zoch (2020): "Markups, Labor Market Inequality and the Nature of Work," Working Paper 26800, National Bureau of Economic Research.

Menzio, G. (2024a): "Efficient Imperfect Competition with an Application to International Trade," Discussion paper, National Bureau of Economic Research.

——— (2024b): "Markups: A search-theoretic perspective," Discussion paper, National Bureau of Economic Research.

Mitman, K., and S. Rabinovich (2015): "Optimal unemployment insurance in an equilibrium business-cycle model," *Journal of Monetary Economics*, 71, 99–118.

Morlacco, M., and D. Zeke (2021): "Monetary policy, customer capital, and market power," *Journal of Monetary Economics*, 121, 116–134.

Ottonello, P., and T. Winberry (2020): "Financial Heterogeneity and the Investment Channel of Monetary Policy," *Econometrica*, 88(6), 2473–2502.

Paciello, L., A. Pozzi, and N. Trachter (2019): "PRICE DYNAMICS WITH CUSTOMER MARKETS," *International Economic Review*, 60(1), 413–446.

Pearce, J., and L. Wu (2025): "Market Concentration and Aggregate Productivity: The Role of Demand," .

Peters, R. H., and L. A. Taylor (2017): "Intangible capital and the investment-q relation," *Journal of Financial Economics*, 123(2), 251–272.

Restuccia, D., and R. Rogerson (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic dynamics*, 11(4), 707–720.

Roldan-Blanco, P., and S. Gilbukh (2021): "Firm dynamics and pricing under customer capital accumulation," *Journal of Monetary Economics*, 118, 99–119.

Rudanko, L. (2025): "Price Setting with Customer Capital: Sales, Teasers, and Rigidity," *Journal of Political Economy Macroeconomics*, forthcoming.

Shen, S. (2025): "Search frictions, customer allocation, and business dynamism," .

Sterk, V., P. Sedláček, and B. Pugsley (2021): "The Nature of Firm Growth," *American Economic Review*, 111(2), 547–79.

A Additional Theoretical Results

A.1 Aggregate Markup and Aggregate Productivity

In this appendix, we derive the aggregate markup in the market economy, as well as aggregate productivity in both the market economy and under the social planner solution. For all derivations, we take the distribution of customers across suppliers and related marketing cost as given.

We define aggregate productivity in the market economy Z^M as the Solow-residual of a hypothetical aggregate production function with identical Cobb-Douglas structure as in (8) and (12), that is

$$Z^M = \frac{Y}{K^\alpha L^{1-\alpha}} \tag{58}$$

where K and L are aggregate capital and labor input in the economy and Y is aggregate output of the final good. We can then derive Z^M as

$$\begin{aligned}
Z^M &= \frac{Y}{\left(\bar{\ell}uy(\bar{mc}) + \int_0^\infty \frac{\bar{\ell}}{z} y(p(z))dG(z)\right)^{1-\alpha}\left(\bar{k}uy(\bar{mc}) + \int_0^\infty \frac{\bar{k}}{z} y(p(z))dG(z)\right)^{\alpha}} \\
&= \left(\bar{\ell}^{1-\alpha}\bar{k}^\alpha\left(u\frac{y(\bar{mc})}{Y} + \int_0^\infty \frac{1}{z}\frac{y(p(z))}{Y}dG(z)\right)\right)^{-1} \\
&= \bar{mc}^\sigma\left(u + \int_0^\infty \frac{1}{z}\left(\frac{\mu(z)}{z}\right)^{-\sigma} dG(z)\right)^{-1} \\
&= \left(\frac{Z^M}{\mathcal{M}}\right)^\sigma\left(u + \int_0^\infty \frac{1}{z}\left(\frac{\mu(z)}{z}\right)^{-\sigma} dG(z)\right)^{-1} \\
\Rightarrow Z^M &= \left(u\left(\frac{1}{\mathcal{M}}\right)^{-\sigma} + \int_0^\infty z^{\sigma-1}\left(\frac{\mu(z)}{\mathcal{M}}\right)^{-\sigma} dG(z)\right)^{\frac{1}{\sigma-1}}
\end{aligned} \tag{59}$$

where $\mu(z) = \frac{p(z)}{\frac{\bar{mc}}{z}}$ is the markup charged by a supplier with productivity z and $\mathcal{M} = \frac{1}{\frac{\bar{mc}}{Z^M}}$ is defined as the aggregate markup in the economy. Aggregate productivity is given by a weighted average of individual productivities, where weights are given by the relative markups of firms with different productivities.

We can derive a condition for the aggregate markup $\mathcal{M}$ from the labor share of output where

$$\frac{wL}{Y} = \frac{w\frac{Y}{Z^M}\bar{\ell}}{Y} = \frac{w}{Z^M}\frac{(1-\alpha)\bar{mc}}{w} = \frac{1-\alpha}{\mathcal{M}} \tag{60}$$

The same relationship has to hold for any supplier firm, such that

$$\frac{wL}{Y}\mathcal{M} = \mu(z)\frac{w\ell(z)}{p(z)y(z)} \tag{61}$$

$$\Rightarrow \mathcal{M}\frac{p(z)y(z)}{Y} = \frac{\ell(z)}{L}\mu(z) \tag{62}$$

Integrating both sides over the distribution of matches and adding the equivalent condition for the mass of unmatched customers on both sides yields the aggregate markup as

$$\mathcal{M} = u\frac{\bar{\ell}y(\bar{mc})}{L} + \int_0^\infty \frac{\bar{\ell}y(p(z))}{zL}\mu(z)dG(z) \tag{63}$$

where we make use of the fact that we have normalized the price of the final good in (7) to one. As shown in Edmond, Midrigan and Xu (2022), the aggregate markup in an economy with heterogeneous producers is given by the input-weighted average of individual markups, evaluated at equilibrium demand for labor and the distribution of matches $G(z)$.[31]

Next, consider the optimal allocation of demand a planner would choose at the intensive margin. The planner takes the distribution of matches $G(z)$ and mass of unmatched u as given. To minimize cost (maximize aggregate productivity Z), the planner solves

$$\min_{y^u, \{y(z)\}_z} uy^u + \int_0^\infty \frac{y(z)}{z_j}dG(z) \tag{64}$$

$$\text{s.t.} \left(u(y^u)^{\frac{\sigma-1}{\sigma}} + \int_0^\infty y(z)^{\frac{\sigma-1}{\sigma}}dG(z)\right)^{\frac{\sigma}{\sigma-1}} = Y \tag{65}$$

Taking FOCs with respect to $y(z)$ yields

$$\frac{1}{z} = y(z)^{-\frac{1}{\sigma}}Y^{\frac{1}{\sigma}}\lambda \tag{66}$$

where λ is the Lagrange multiplier on the constraint.[32] Multiplying both sides by $y(z)$, integrating over z with $G(z)$ and addition the corresponding condition for y^u yields

$$uy^u + \int_0^\infty \frac{y(z)}{z}dG(z) = \lambda Y^{\frac{1}{\sigma}}\left(u(y^u)^{\frac{\sigma-1}{\sigma}} + \int_0^\infty y(z)^{\frac{\sigma-1}{\sigma}}dG(z)\right) = \lambda Y$$

$$\Rightarrow K^\alpha L^{1-\alpha} = \lambda Y$$

$$\Rightarrow \lambda = \frac{1}{Z} \tag{67}$$

[31] As the capital-labor-ratio is identical across all firms, we could have arrived at the same expression starting from the capital share of output and results are identical if we weight by capital.

[32] The derivation for y^u is analogous.

This yields the optimal relative size of firms under the planner allocation as

$$\frac{y(z)}{Y} = \left(\frac{Z}{z}\right)^{-\sigma} \tag{68}$$

and the corresponding aggregate productivity under the optimal allocation as

$$Z = \left(u\frac{y^u}{Y} + \int_0^\infty \frac{1}{z}\frac{y(z)}{Y}dG(z)\right)^{-1}$$

$$\Rightarrow Z^P = \left(u + \int_0^\infty z^{\sigma-1}dG(z)\right)^{\frac{1}{\sigma-1}} \tag{69}$$

A.2 Efficient Allocation of Customers

The complete optimization problem of a planner that solves jointly for the allocation of customers across suppliers, demand at the intensive margin, and aggregate ressources is given by

$$\max_{\{C_t, L_t, K_{t+1}, \{i_t(z), \hat{\delta}_t(z)\}_{z=0}^\infty\}_{t=0}^\infty} \sum_{t=0}^\infty \beta^t \{u(C_t) - v(L_t)\} \tag{70}$$

$$\text{s.t. } C_t + K_{t+1} + \Gamma_t \leq Z_t^* K_t^\alpha L_t^{1-\alpha} + (1-\delta_k)K_t \;\; \forall t \tag{71}$$

$$Z_t^* = \left(\hat{n}_t^u + \int_0^\infty z^{\sigma-1}\hat{n}_t(z)dz\right)^{\frac{1}{\sigma-1}} \;\; \forall t \tag{72}$$

$$\Gamma_t = \int_0^\infty m(i_t(z))n_t(z)dz \;\; \forall t \tag{73}$$

$$\hat{n}_t(z) = (1 - \hat{q}_t(z) + \hat{e}_t(z)i_t(z))n_t(z) \;\; \forall t, z \tag{74}$$

$$\hat{n}_t^u = 1 - \int_0^\infty \hat{n}_t(z)dz \;\; \forall t \tag{75}$$

$$n_t(z) = (1 - \hat{\delta}_t(z))\left[\int_0^\infty \hat{n}_{t-1}(\hat{z})f(z|\hat{z})d\hat{z} + \chi\hat{n}_{t-1}^u f_0(z)\right] \;\; \forall t, z \tag{76}$$

$$n_t^u = \hat{n}_{t-1}^u - \chi\hat{n}_{t-1}^u + \hat{\delta}_t(z)\left[\int_0^\infty \hat{n}_{t-1}(\hat{z})f(z|\hat{z})d\hat{z} + \chi\hat{n}_{t-1}^u f_0(z)\right] \;\; \forall t \tag{77}$$

$$\hat{\delta}_t(z) \geq \delta_F, \quad \hat{\delta}_t(z) \leq 1 \;\; \forall t, z \tag{78}$$

$$\hat{e}_t(z) = \int_0^z n_t(\hat{z})d\hat{z} + n_t^u \;\; \forall t, z \tag{79}$$

$$\hat{q}_t(z) = \int_z^\infty n_t(\hat{z})i_t(\hat{z})d\hat{z} \;\; \forall t, z \tag{80}$$

for given K_0, $\hat{n}_{-1}^u$ and $\{\hat{n}_{-1}(z)\}_{z=0}^\infty$. $n_t(z)$ and $\hat{n}_t(z)$ are the mass of customers matched with suppliers of productivity z at the beginning of the period and during production respectively. $\hat{\delta}_t(z)$ is the fraction of firms of type z the planner chooses to exit (not enter) the economy.

We can make some substitutions to simplify the problem

$$\max_{\{C_t,L_t,K_{t+1},\{i_t(z),\hat{\delta}_t(z)\}_{z=0}^{\infty}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \{u(C_t) - v(L_t)\} \tag{81}$$

$$\text{s.t. } C_t + K_{t+1} + \int_0^{\infty} m(i_t(z))n_t(z)dz \le Z_t^* K_t^{\alpha} L_t^{1-\alpha} + (1-\delta_k)K_t \ \forall t \tag{82}$$

$$Z_t^* = \left(1 + \int_0^{\infty} (z^{\sigma-1} - 1)\hat{\rho}_t(z, i_t(\hat{z}))n_t(z)dz\right)^{\frac{1}{\sigma-1}} \ \forall t \tag{83}$$

$$n_t(z) = (1 - \hat{\delta}_t(z))\left[\int_0^{\infty} \hat{\rho}_{t-1}(z, i_{t-1}(\hat{z}))n_{t-1}(\hat{z})(f(z|\hat{z}) - \chi f_0(z))d\hat{z} + \chi f_0(z)\right] \ \forall t, z \tag{84}$$

$$\hat{\rho}_t(z, i) = 1 + i - \int_z^{\infty} n_t(\hat{z})(i_t(\hat{z}) + i)d\hat{z} \tag{85}$$

$$\hat{\delta}_t(z) \ge \delta_F, \quad \hat{\delta}_t(z) \le 1 \ \forall t, z \tag{86}$$

For given K_0 and $\{n_{-1}(z), i_{-1}(z)\}_{z=0}^{\infty}$. Denote as λ_t^{BC} the multiplier on (82), λ_{zt}^n the multiplier on (84), and $\lambda_t^{\delta L}$ and $\lambda_t^{\delta H}$ the multipliers on (86). We use (83) and (85) only for accounting and do not attach separate multipliers; i.e., substitute them into the other equations when setting up the Langrangian.

We first determine the value of a match to the planner, which we derive by taking the FOC with respect to $n_t(z)$. This yields

$$\begin{aligned}
\beta^t \lambda_{zt}^n = &-\beta^t \lambda_t^{BC} m(i_t(z)) + \beta^t \lambda_t^{BC} \hat{\rho}_t(z, i_t(z)) \frac{z^{\sigma-1} - 1}{\sigma - 1} (Z_t^*)^{2-\sigma} K_t^{\alpha} L_t^{1-\alpha} \\
&+ \hat{\rho}_t(z, i_t(z)) \int_0^{\infty} (1 - \hat{\delta}_{t+1}(z'))(f(z'|z) - \chi f_0(z'))\beta^{t+1} \lambda_{z't+1}^n dz' \\
&- \int_0^z n_t(\hat{z})(i_t(z) + i_t(\hat{z}))\beta^t \lambda_t^{BC} \frac{\hat{z}^{\sigma-1} - 1}{\sigma - 1} (Z_t^*)^{2-\sigma} K_t^{\alpha} L_t^{1-\alpha} d\hat{z} \\
&- \int_0^{\infty} \beta^{t+1} \lambda_{z't+1}^n (1 - \hat{\delta}_{t+1}(z')) \int_0^z n_t(\hat{z})(i_t(z) + i_t(\hat{z}))(f(z'|\hat{z}) - \chi f_0(z'))d\hat{z}dz'
\end{aligned} \tag{87}$$

Denote $\hat{V}_t(z) = \frac{\lambda_{zt}^n}{\lambda_t^{BC}}$ as the value of a match to the planner at time t.[33] We get this as

$$\begin{aligned}
\hat{V}_t(z) = &-m(i_t(z)) + \hat{\rho}_t(z, i_t(z)) \frac{z^{\sigma-1} - 1}{\sigma - 1} (Z_t^*)^{2-\sigma} K_t^{\alpha} L_t^{1-\alpha} \\
&+ \hat{\rho}_t(z, i_t(z))\beta \frac{u'(C_{t+1})}{u'(C_t)} \int_0^{\infty} (1 - \hat{\delta}_{t+1}(z'))(f(z'|z) - \chi f_0(z'))\hat{V}_{t+1}(z') \\
&- \int_0^z n_t(\hat{z})(i_t(z) + i_t(\hat{z})) \frac{\hat{z}^{\sigma-1} - 1}{\sigma - 1} (Z_t^*)^{2-\sigma} K_t^{\alpha} L_t^{1-\alpha} d\hat{z} \\
&- \beta \frac{u'(C_{t+1})}{u'(C_t)} \int_0^z n_t(\hat{z})(i_t(z) + i_t(\hat{z})) \int_0^{\infty} (1 - \hat{\delta}_{t+1}(z'))(f(z'|\hat{z}) - \chi f_0(z'))\hat{V}_{t+1}(z')dz'd\hat{z}
\end{aligned}$$

[33] We transform the shadow cost of the constraint on n into units of the final good by normalizing it by the shadow cost of the budget constraint.

where we have used the fact that from the FOC with respect to C_t we have $\lambda_t^{BC} = u'(C_t)$.

To derive the optimal choice of $i_t(z)$, we take the FOC with respect to $i_t(z)$ and get

$$\begin{aligned}
m'(i_t(z))n_t(z)\beta^t\lambda_t^{BC} = &\left(1 - \int_z^\infty n_t(\hat{z})d\hat{z}\right)n_t(z)\beta^t\lambda_t^{BC}\frac{z^{\sigma-1}-1}{\sigma-1}(Z_t^*)^{2-\sigma}K_t^\alpha L_t^{1-\alpha} \\
&+ \left(1 - \int_z^\infty n_t(\hat{z})d\hat{z}\right)n_t(z)\int_0^\infty (1-\hat{\delta}_{t+1}(z'))(f(z'|z) - \chi f_0(z'))\beta^{t+1}\lambda_{z't+1}^n dz' \\
&- \int_0^z n_t(\hat{z})n_t(z)\beta^t\lambda_t^{BC}\frac{\hat{z}^{\sigma-1}-1}{\sigma-1}(Z_t^*)^{2-\sigma}K_t^\alpha L_t^{1-\alpha}d\hat{z} \\
&- \int_0^\infty \beta^{t+1}\lambda_{z't+1}^n(1-\hat{\delta}_{t+1}(z'))\int_0^z n_t(\hat{z})n_t(z)(f(z'|\hat{z}) - \chi f_0(z'))d\hat{z}dz'
\end{aligned}$$

We can rearrange this and use the defintion of $\hat{V}_t(z)$ to get

$$\begin{aligned}
m'(i_t(z)) = &\int_0^z n_t(\hat{z})\frac{z^{\sigma-1}-\hat{z}^{\sigma-1}}{\sigma-1}(Z_t^*)^{2-\sigma}K_t^\alpha L_t^{1-\alpha}d\hat{z} \\
&+ \beta\frac{u'(C_{t+1})}{u'(C_t)}\int_0^z n_t(\hat{z})\int_0^\infty (1-\hat{\delta}_{t+1}(z'))(f(z'|z) - f(z'|\hat{z}))\hat{V}_{t+1}(z')dz'd\hat{z} \\
&+ \left(1 - \int_0^\infty n_t(\hat{z})d\hat{z}\right)\frac{z^{\sigma-1}-1}{\sigma-1}(Z_t^*)^{2-\sigma}K_t^\alpha L_t^{1-\alpha} \\
&+ \left(1 - \int_0^\infty n_t(\hat{z})d\hat{z}\right)\beta\frac{u'(C_{t+1})}{u'(C_t)}\int_0^\infty (1-\hat{\delta}_{t+1}(z'))(f(z'|z) - \chi f_0(z'))\hat{V}_{t+1}(z')dz'
\end{aligned}$$

where the first two rows are the expected gains, contemporaneously and going forward, if the additional customer moves from a supplier with lower productivity and the last two rows are the gains if the customer was previously unmatched.

Finally, we can derive the optimal choice of $\hat{\delta}_t(z)$ by taking the FOC with respect to $\hat{\delta}_t(z)$ and get

$$\beta^t\lambda_{zt}^n\underbrace{\left[\int_0^\infty \hat{\rho}_t(z, i_t(\hat{z}))n_t(\hat{z})(f(z|\hat{z}) - \chi f_0(z))d\hat{z} + \chi f_0(z)\right]}_{>0} = \beta^t\lambda_t^{\delta L} - \beta^t\lambda_t^{\delta H}$$

As $\lambda_t^{BC} > 0$ due to the binding budget constraint, $\lambda_{zt}^n > 0$ iff $\hat{V}_t(z) > 0$. Hence iff $\hat{V}_t(z) > 0$ we get that $\lambda_t^{\delta L} > 0$ and so $\hat{\delta}_t(z) = \delta_F$ is optimal. If $\hat{V}_t(z) < 0$, then $\lambda_t^{\delta H} > 0$ and hence $\hat{\delta}_t(z) = 1$ is optimal.

B Data Appendix

B.1 Capital IQ

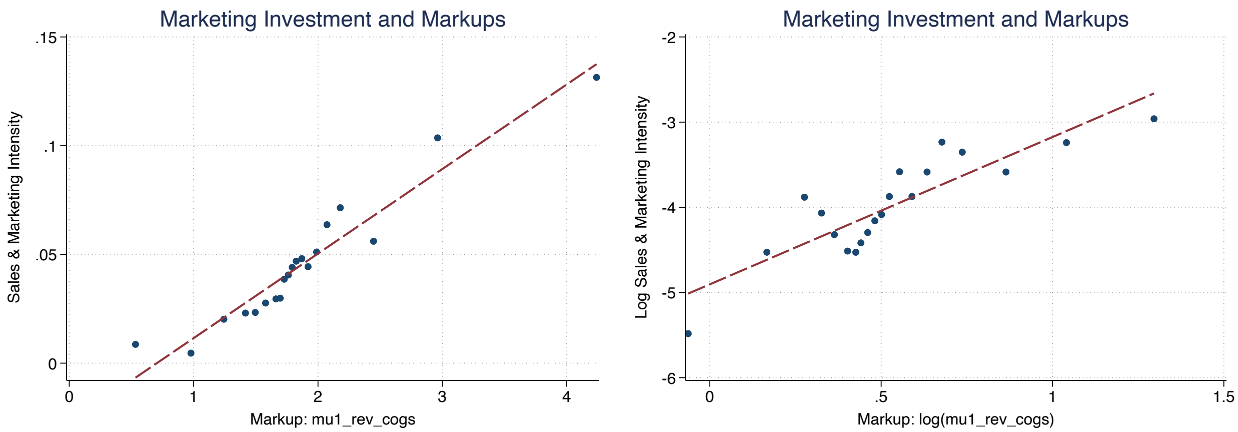
Sales and Marketing expenses. We follow He, Mostrom and Sufi (2024) in using Capital IQ's *Sales and Marketing Expenses* variable, SME_f, to proxy for firms' investment in demand. In the Capital IQ–Compustat matched sample, the aggregate (weighted average) ratio of SME_f to revenue Rev_f equals $\frac{\sum_f SME_f}{\sum_f Rev_f} = 0.045$. Because our model abstracts from other non–value-added components of revenue, we convert this into a ratio with respect to value added VA_f:

$$\frac{\sum_f SME_f}{\sum_f VA_f} = \frac{\sum_f SME_f}{\sum_f Rev_f}\frac{\sum_f Rev_f}{\sum_f VA_f} = 0.045 \times 1.80 = 0.080.$$

We compare this value to the corresponding model ratio of marketing expenses, $\int_f m(z)dG^i(z)$, to value added, $\int_f \rho\left(p(z), i(z)\right)\pi(p(z), z) - m(i(z))dG^i(z)$. For this comparison, we use the filtered model sample designed to mimic the Capital IQ–Compustat selection described in Appendix B.2.

Marketing investment and markups. Table 6 displays regressions of the log marketing–revenue ratio on firms' log markup ($= \frac{\text{Revenue}}{\text{Cost of Goods Sold}}$). Our preferred specification is column 4, which includes year and 2-digit NAICS fixed effects since the model does not feature aggregate variation or heterogeneous industries. Weighting by firm revenue reduces noise as evidenced by a higher $R^2 = 0.449$ compared to the unweighted regression in column 3 ($R^2 = 0.200$). The point estimate of 1.729 implies that a firm that has a 1% higher markup invests, on average, about 1.7% more in marketing activities, compared to the industry average in that year. Columns 1 and 2 show results without fixed effects, and columns 5 and 6 add additional controls for firm size and age (which do not explain much more of the overall variation). Overall, across specifications, a consistent picture emerges: high-markup firms invest considerably more in marketing activities.

Figure 5: Marketing Investment vs. Markups



(a) Levels (b) Logs

Source: Capital IQ. Binscatters of sales and marketing intensity, defined as $\frac{\text{Sales and marketing expenses}}{\text{Revenue}}$, vs. markups, defined as $\frac{\text{Revenue}}{\text{COGS}}$. Industry and year fixed effects; firm revenue used as weight.

Table 6: Cross-sectional relation between marketing spending and markups

	(1)	(2)	(3)	(4)	(5)	(6)
$\ln \frac{\text{Revenue}}{\text{COGS}}$	1.449***	1.970***	0.987***	1.729***	0.989***	1.615***
	(0.245)	(0.307)	(0.129)	(0.259)	(0.148)	(0.211)
R^2	0.130	0.314	0.200	0.449	0.299	0.485
Fixed Effects	—	—	Year+Ind	Year+Ind	Year+Ind	Year+Ind
Weights	—	Revenue	—	Revenue	—	Revenue
Controls	—	—	—	—	Age+Size	Age+Size

Source: Capital IQ. The dependent variable in all regressions is $\ln \frac{\text{Sales and marketing expenses}}{\text{Revenue}}$. 20,772 observations. Robust standard errors clustered at 2-digit NAICS level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure 5 plots a binscatter corresponding to the main specification, column 4 of Table 6, in the right panel. The left panel plots the relationship in levels to visualize the scale.

B.2 Model Filter to Mimic Selection into Capital IQ

Table 7 shows how model moments vary based on imposing the filter that mimics selection into Compustat/Capital IQ. We filter intermediate good firms (the suppliers) along two dimensions, age and size. For age, we simply follow Ottonello and Winberry (2020) and

Table 7: Impact of Filtering on Model Moments

Sample	All firms	Intermediate-good firms only		
Filter	none	none	age	age+size
OLS slope $\ln \frac{marketing}{revenue}$ on $\ln markup$	—	2.143	2.102	**1.731**
OLS slope $\ln markup$ on $\ln \frac{revenue}{customers}$	—	0.151	0.152	**0.187**
Aggregate markup	1.212	1.238	1.239	**1.250**
Aggregate sales & marketing share	0.075	0.090	0.091	0.093

Notes: Model moments used for calibration in bold.

restrict attention to firms above seven years old. For size, we assign an inclusion probability to each firm that increases with firm size. Specifically, we use a truncated power function of the form $\min\left(1, \left(\frac{r}{r_0}\right)^{\gamma}\right)$, where r denotes firm revenue, and the parameters r_0 and γ govern the location and steepness of the cutoff. These parameters are calibrated to match two empirical moments: (i) the average Compustat firm is 480 times larger than the average U.S. firms, and (ii) 15.2% of Compustat firms are smaller than the average U.S. firm. This filter allows the model to reproduce both the heavy right tail and the nontrivial lower tail of the empirical firm size distribution in Compustat.

In addition, divisions that are unmatched and rely on the outside option of in-house production produce competitively. This is why the economy-wide markup and marketing share of value added are lower in the aggregate model economy compared to the respective moments in the sample of all intermediate good firms.

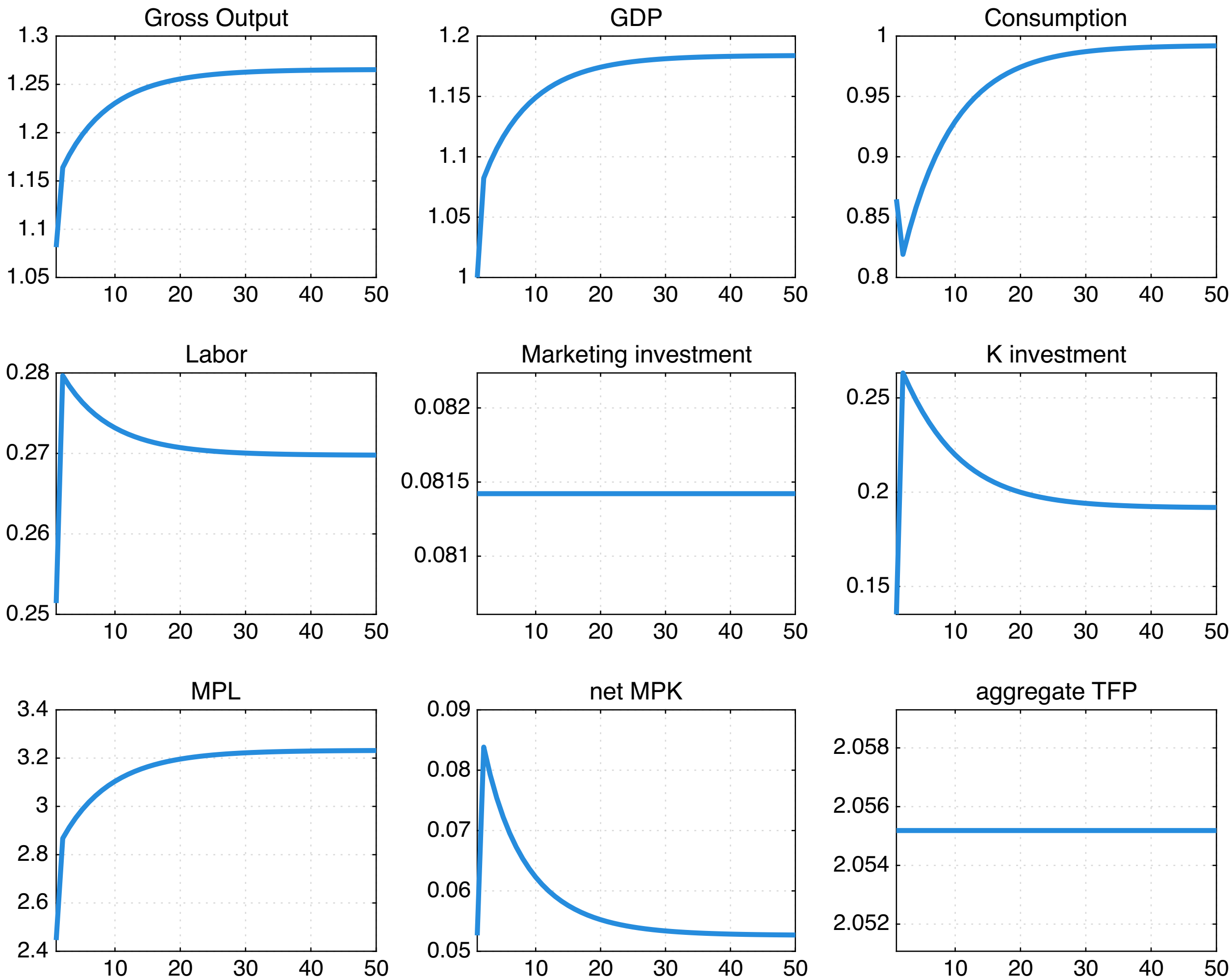
B.3 BDS and SUSB

To calibrate entry and exit rates as well as the sales concentration within industries, we rely on data from the Business Dynamics Statistics (BDS) and the Statistics of US Businesses respectively. For both datasets, we focus on the 2017 wave. For entry / exit rates, we define entrants as firms of age zero and exits as reported shutdown of firms. We compute entry and exit relative to incumbent firms within each 2-digit NAICS sector, and take the simple average across all sectors excluding agriculture, finance and insurance, as well as real estate (NAICS 11, 52, and 53). For sales concentration, we estimate pareto tails for the distribution of revenues within five digit NAICS sectors and take the simple average across sectors, again excluding observations in agriculture, finance and insurance, as well as real estate (2-digit NAICS 11, 52, and 53).

C Additional Quantitative Model Results

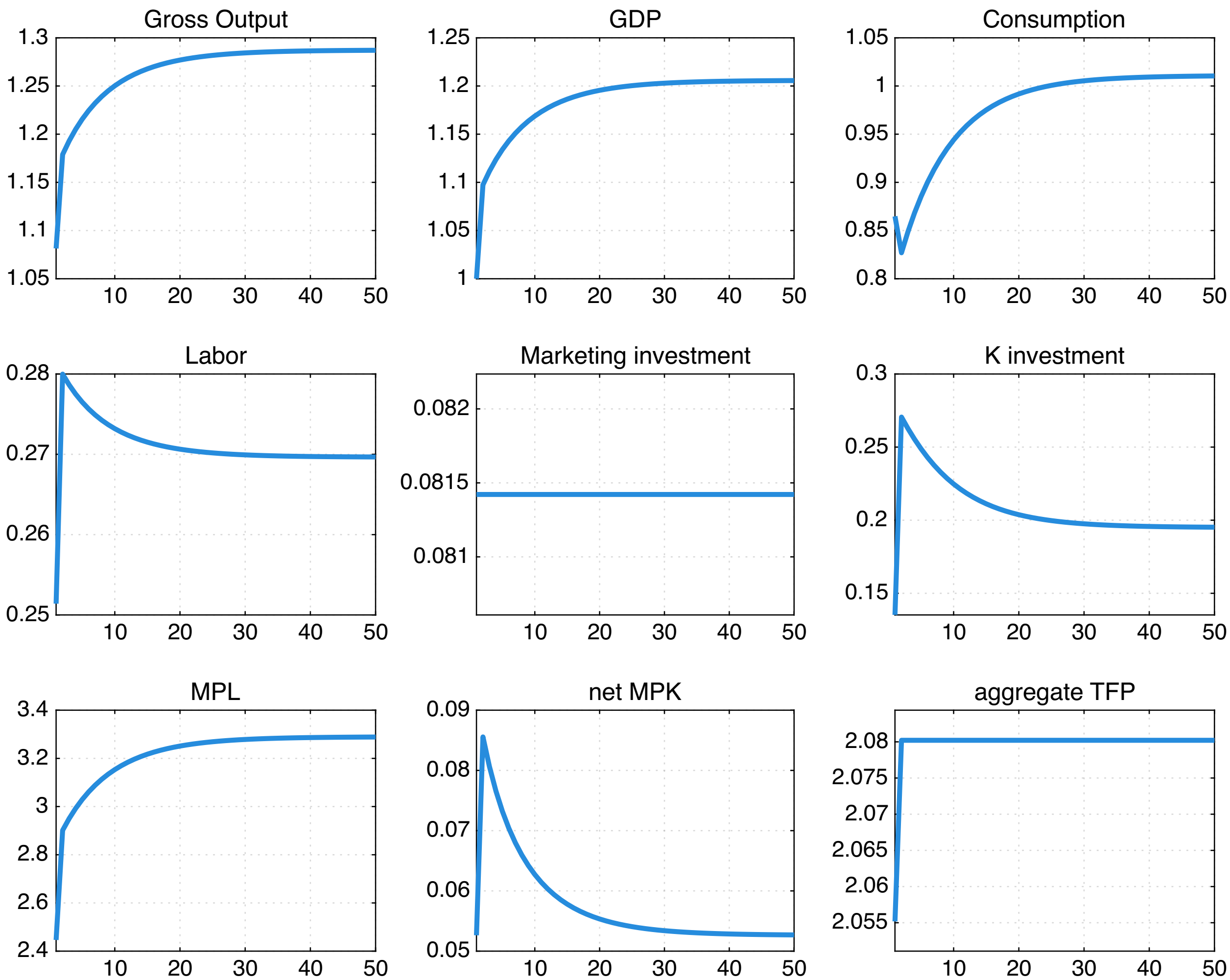
C.1 Social Planner Transition Graphs

Figure 6: Transition: planner remedies only the aggregate markup



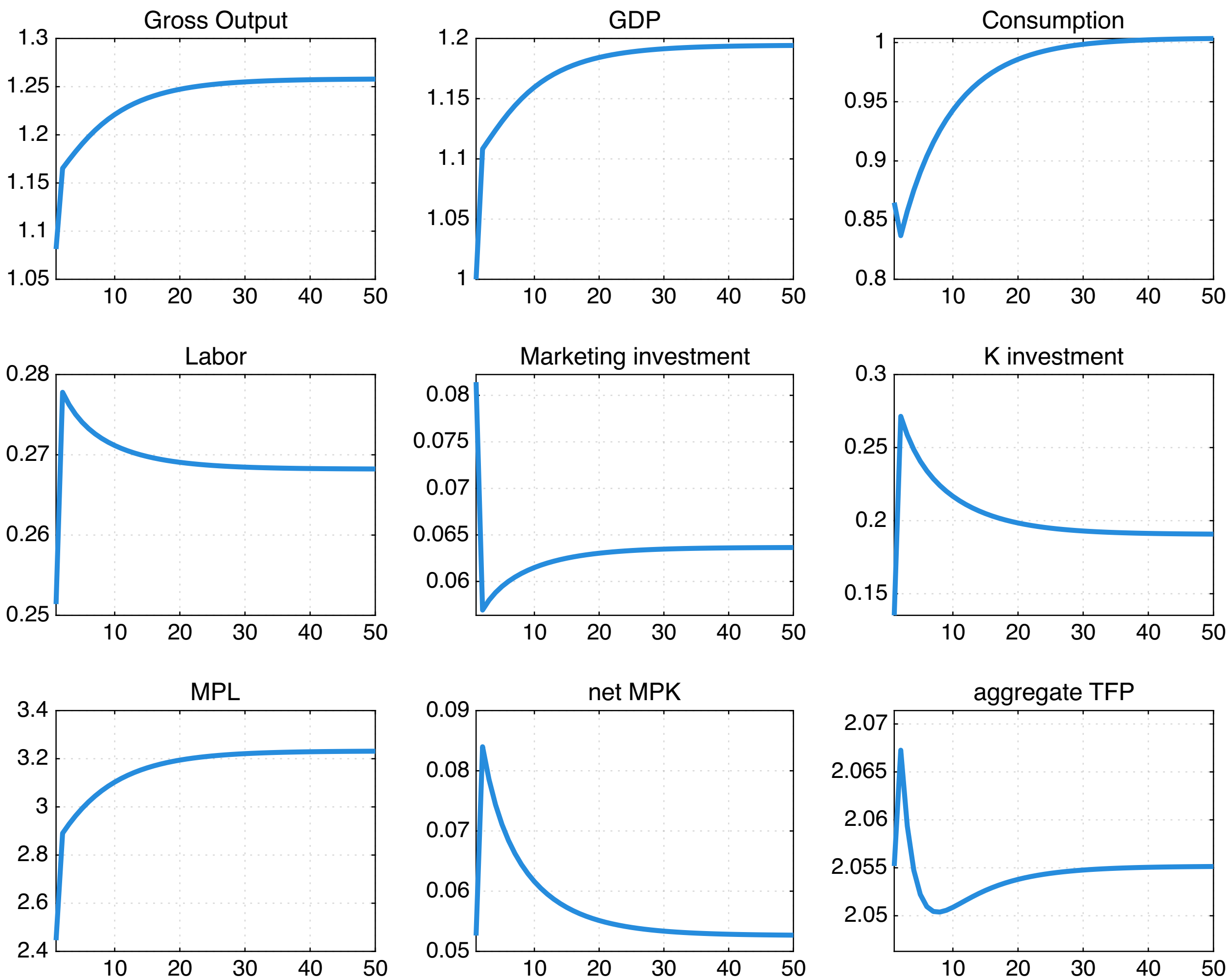
Notes: The planner removes the aggregate markup wedge while holding fixed the equilibrium objects governing demand reallocation: marketing spending Γ^M and the implied customer distribution G (extensive margin), and aggregate TFP Z^M (intensive margin). Dispersion in markups across suppliers is therefore not corrected. The planner chooses time-varying (C_t, Inv_t, L_t) subject to fixed Γ^M, G, Z^M. Period t=1 shows the decentralized equilibrium; at t=2 the planner takes over with predetermined K, while (C_t, Inv_t, L_t) are already planner-chosen. Quantities normalized so that $GDP = 1$ in the decentralized equilibrium.

Figure 7: Transition: planner remedies aggregate markup and markup dispersion



Notes: The planner removes both the aggregate markup wedge and *dispersion* in markups (intensive-margin misallocation). Marketing remains fixed at its equilibrium path, so the customer distribution G is fixed. Removing dispersion shifts aggregate TFP from the market value Z^M to the planner value Z^P. The planner chooses time-varying (C_t, Inv_t, L_t) subject to fixed Γ^M, G. Period t=1 shows the decentralized equilibrium; at t=2 the planner takes over with predetermined K, while (C_t, Inv_t, L_t) are already planner-chosen. Quantities normalized so that GDP = 1 in the decentralized equilibrium.

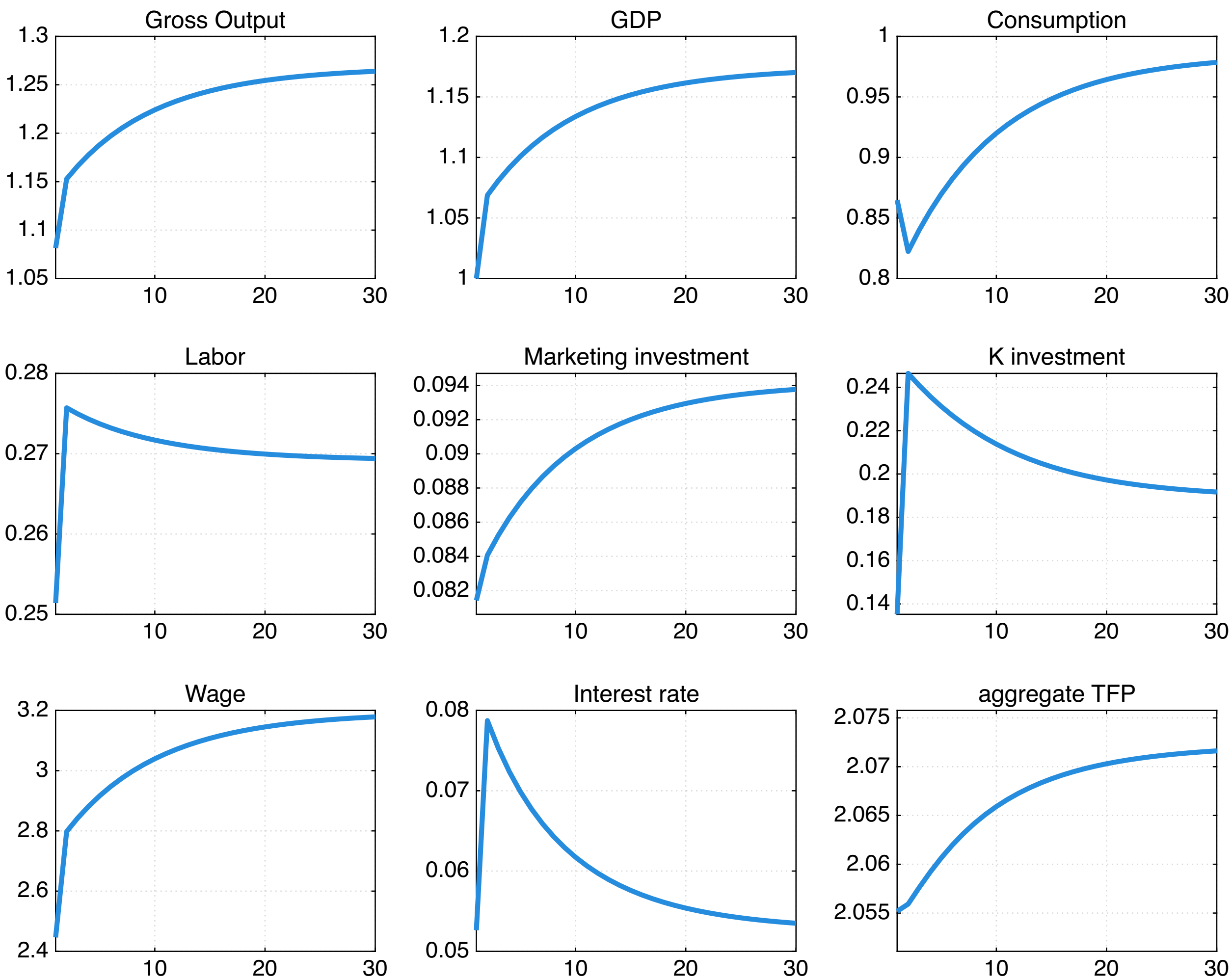
Figure 8: Transition: full social planner (markup, dispersion, and marketing)



Notes: The planner removes the aggregate markup and its dispersion *and* chooses marketing optimally. Thus, $i_t(z)$, aggregate marketing expenditures Γ_t, and the customer distribution $G_t(z)$ are endogenously time-varying, and aggregate TFP Z_t reflects both intensive-margin reallocation and evolving customer allocation. The planner chooses time-varying $(C_t, \text{Inv}_t, L_t, i_t(z))$. Period t=1 shows the decentralized equilibrium; at t=2 the planner takes over with predetermined K, while $(C_t, \text{Inv}_t, L_t, i_t(z))$ are already planner-chosen. Quantities normalized so that GDP = 1 in the decentralized equilibrium.

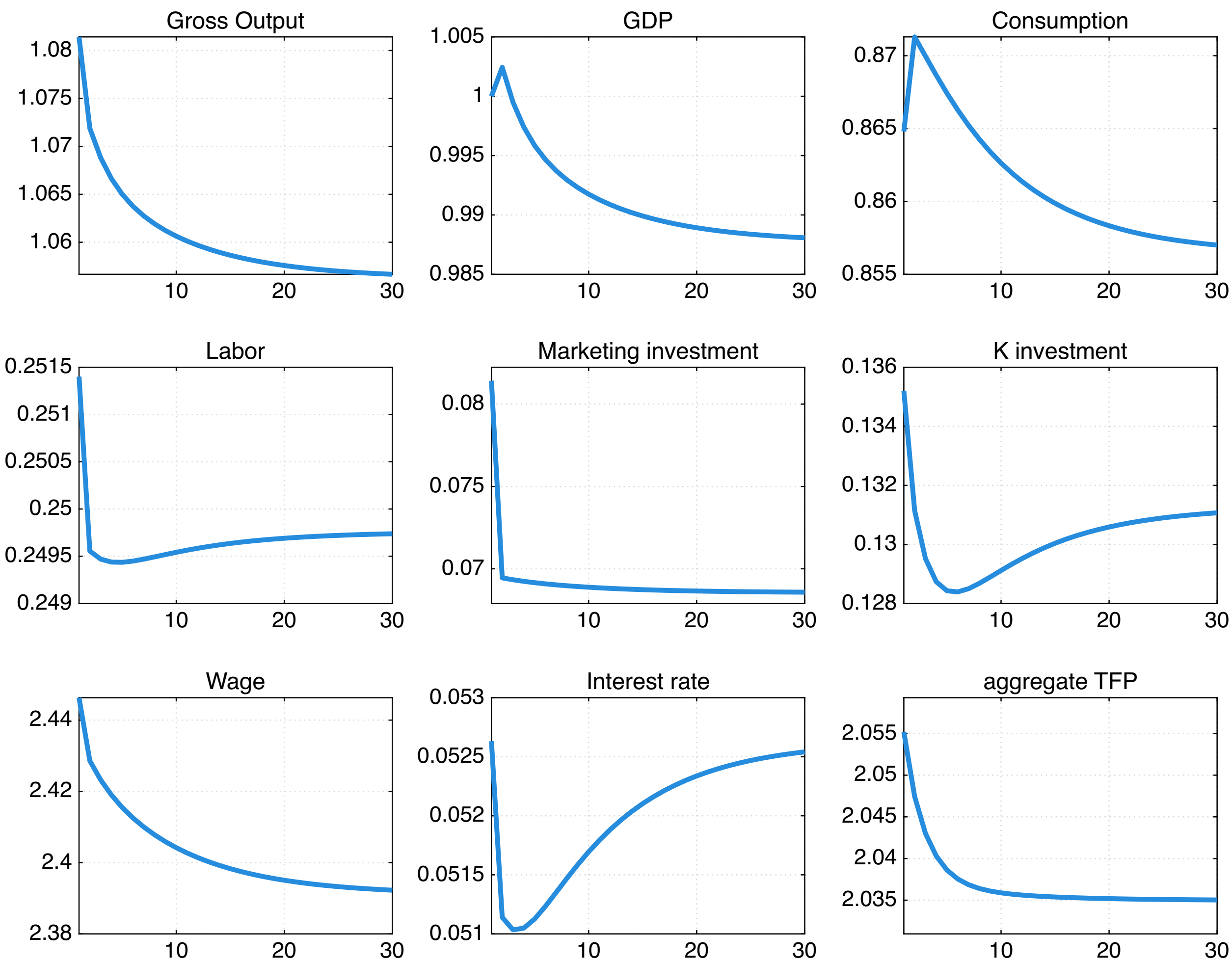
C.2 Optimal Taxation Transition Graphs

Figure 9: Transition: optimal production subsidy



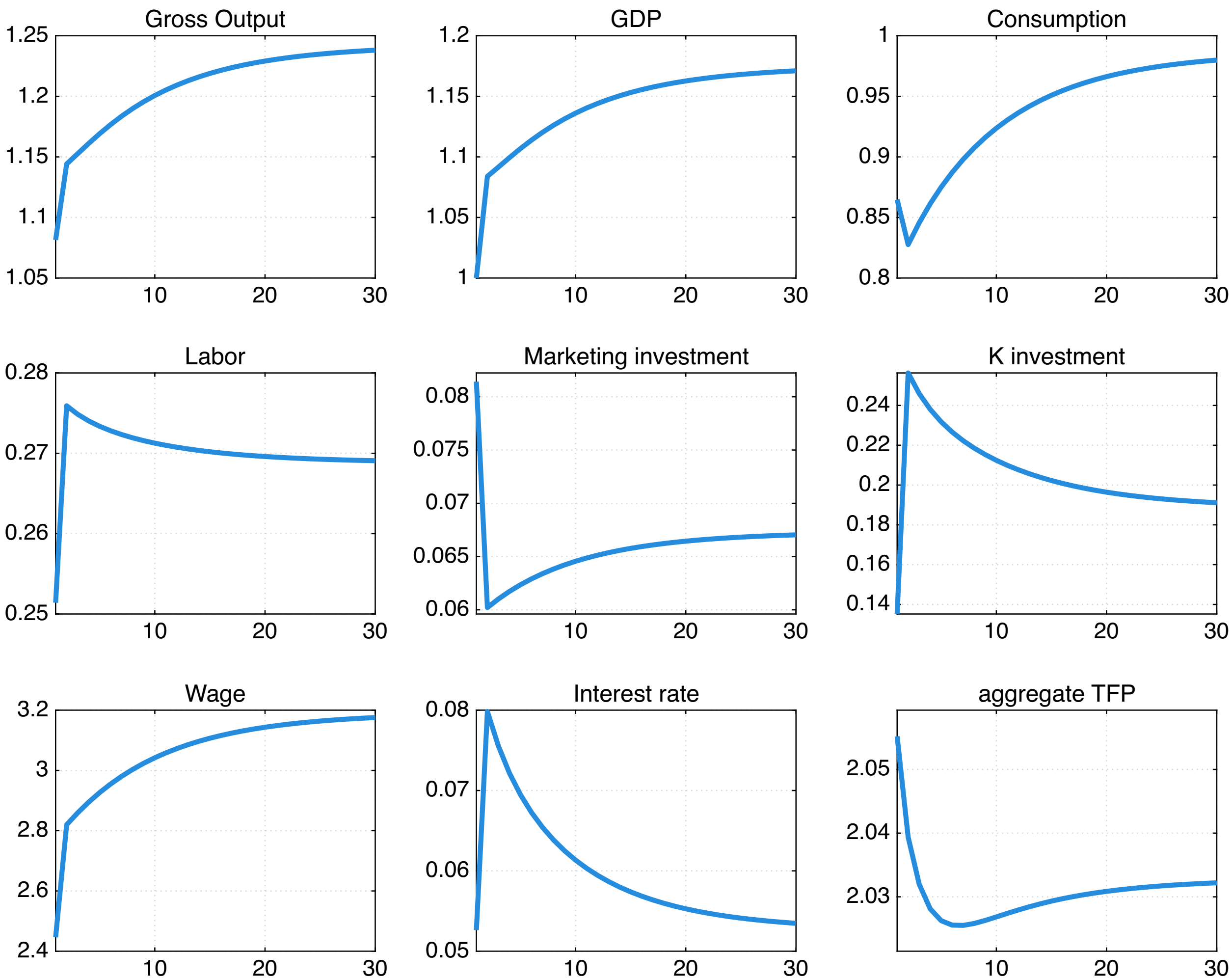
Notes: We solve for a (time-constant) production subsidy of 15.69%, which maximizes welfare along the transition from the no-tax baseline steady state. The tax is financed by a lumpsum tax (19.57% of GDP in steady state). Period t=1 shows the no-tax equilibrium; from t=2 onwards the production subsidy is imposed; agents are surprised; we solve for the resulting perfect foresight transition to the new steady state. Quantities normalized so that $GDP = 1$ in the no-tax equilibrium.

Figure 10: Transition: optimal marketing tax



Notes: We solve for a (time-constant) marketing tax of 17.66%, which maximizes welfare along the transition from the no-tax baseline steady state. The tax is rebated as lumpsum transfer (1.21% of GDP in steady state). Period t=1 shows the no-tax equilibrium; from t=2 onwards the marketing tax is imposed; agents are surprised; we solve for the resulting perfect foresight transition to the new steady state. Quantities normalized so that $GDP = 1$ in the no-tax equilibrium.

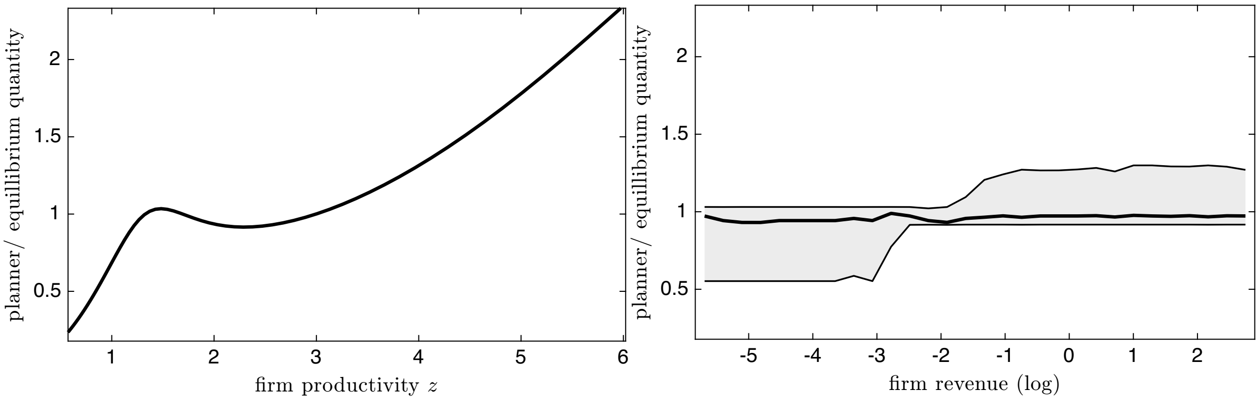
Figure 11: Transition: optimal marketing tax and production subsidy



Notes: We solve for a (time-constant) marketing tax of 41.10% and production subsidy of 18.16%, which maximize welfare along the transition from the no-tax baseline steady state. The tax is financed by a lumpsum tax (19.85% of GDP in steady state). Period t=1 shows the no-tax equilibrium; from t=2 onwards the marketing tax and production subsidy are imposed; agents are surprised; we solve for the resulting perfect foresight transition to the new steady state. Quantities normalized so that $GDP = 1$ in the no-tax equilibrium.

C.3 Additional Model Figures

Figure 12: Size distortions: planner vs. equilibrium by productivity and revenue



(a) By firm productivity z (b) By firm revenue

Each panel plots the ratio of the planner's to the equilibrium quantity demanded. In panel (a), firms are sorted by productivity z, and the ratio is a deterministic function of z. In panel (b), firms are sorted by revenue; for each revenue level, there is a distribution of ratios. The thick line shows the median and the shaded area the 10th–90th percentile range.