Investment in Demand and Dynamic Competition for Customers

Joachim Hubmer
University of Pennsylvania

Lukas Nord
University of Pennsylvania

February 12, 2026

Abstract. How demand is allocated across firms is a central determinant of aggregate productivity. Firms can affect this allocation through their prices and non-price investment. We develop a framework with search frictions and dynamic customer relationships in which non-price investment in demand shapes allocations directly by matching customers to suppliers, and indirectly by altering price competition. A quantitative version matches key facts on how firms compete for customers. In the decentralized equilibrium, distortions from markups and over-investment in demand due to business-stealing externalities cause misallocation. Equilibrium interactions of demand investment and pricing create complementarities between policies that target over-investment in demand and markup distortions. A rise in demand investment over time can account for rising industry concentration without raising market power and reduces firms' intangible value through equilibrium effects.

Contact: jhubmer@sas.upenn.edu; lksnord@sas.upenn.edu. We thank Bianca He, Lauren Mostrom, and Amir Sufi for sharing their measure of firms' sales and marketing spending with us. We thank Eric Bartelsman, Vasco Carvalho, Murat Celik, Gregor Jarosch, Ben Lester, Guido Menzio, Virgiliu Midrigan, as well as conference and seminar participants at Berkeley, DNB, Duke, Einaudi, GEA, Mannheim, NY Fed, Penn, Penn State, Philly Fed, SED 2025, USC Marshall, Vienna, and Temple 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 both the level of production and the allocation of demand across firms (Baqaee and Farhi, 2020; Edmond, Midrigan and Xu, 2022). Most theories of misallocation treat demand as static, with imperfect competition arising from product differentiation. Recent work instead emphasizes that firms accumulate demand dynamically by investing in sales and marketing to build their customer base (Fitzgerald, Haller and Yedid-Levi, 2024; Argente, Fitzgerald, Moreira and Priolo, 2025; Afrouzi, Drenik and Kim, 2025). This *investment in demand* now rivals outlays on physical capital and is the largest component of intangible investment (He, Mostrom and Sufi, 2024). Firms' willingness to invest in expanding their customer base aligns with theories in which product-market search frictions generate market power (Menzio, 2024a,b).

How does investment in demand shape allocative efficiency and market power under product-market search frictions? And what does the rise in demand investment since the 1980s imply for concentration, market power, productivity, and intangible capital? We answer these questions in a novel framework with dynamic demand in which firms' demand investment endogenously shapes pricing, competition, and market power. A quantitative version replicates key empirical patterns of how firms compete for demand. We show that search frictions generate two sources of misallocation: markup distortions, and business-stealing externalities that induce over-investment in demand. The interactions between demand investment and pricing creates complementarities between policy instruments that target over-investment in demand and markup distortions. Calibrating the model to the rise in demand investment since the 1980s, we show that declining costs of contacting customers contribute positively to GDP growth, raise concentration without raising market power, and reduce the intangible value of firms' customer base through equilibrium effects.

We model competition for customers in general equilibrium with households, final good producers, and intermediate good producers. Households choose to supply labor and physical capital. Final good firms combine imperfectly substitutable intermediates into a final good. They are the customers in the market for intermediates. Intermediate producers—the suppliers—are heterogeneous in stochastic productivity, set prices, and invest in contacts with potential customers. The market for intermediates is subject to search frictions: a customer must match with a supplier to procure inputs and can only purchase from suppliers it is in contact with. Upon meeting a supplier, customers observe only its price and decide whether to start purchasing. Matches are persistent but can be unilaterally dissolved at any time, and customers can be poached by other suppliers. For each intermediate input, a final good firm contracts with at most one supplier. Substitution across intermediate inputs allows variation in demand at the intensive (sales per customer) margin, in addition to extensive margin variation (number of customers).

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 technology for contacting customers, firms' pricing decision does not depend on their existing customer base. In equilibrium, heterogeneous markups arise from firms' demand investment decisions. When setting prices, firms internalize that their customers may be contacted by competitors. Greater demand investment by rivals increases the 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 salient empirical facts recently emphasized in work on how firms compete for demand. First, it matches the relationship between markups and firm size (Edmond et al., 2022). Second, it does so by generating a positive relation between sales per customer and markups, but no relation between markups and the number of customers (Afrouzi et al., 2025). Third, it replicates the relative contribution of the number of customers versus sales per customer to overall sales variation across firms (Bernard, Dhyne, Magerman, Manova and Moxnes, 2022; Einav, Klenow, Levin and Murciano-Goroff, 2021), which informs demand elasticities at both margins. Fourth, firms choose to expand their customer base through non-price actions while prices are insensitive to customer base size, consistent with evidence on market share growth (Fitzgerald et al., 2024; Argente et al., 2025). This is despite the number of customers being responsive to prices (Paciello, Pozzi and Trachter, 2019; Roldan-Blanco and Gilbukh, 2021), making prices allocative at the extensive margin.

We assess 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 contact rates, match formation, and quantities within each match. We show analytically that the framework nests standard markup distortions as in e.g. Edmond et al. (2022): the aggregate markup drives a wedge in households' factor supply condition, while dispersion in markups distorts the intensive margin allocation of demand across intermediate inputs. Here, markups arise due to customers' inability to access all suppliers simultaneously. In contrast to the intensive margin, markups do not distort match formation at the extensive margin as in Menzio (2024a,b): customers always switch to more productive suppliers when matched, as their value of a match increases monotonically with productivity. Demand investment, however, is generally inefficient. Two externalities operate in opposite directions. First, suppliers under-invest because they bear the full cost of contacts but do not internalize the benefit of matches to their customers. Second, they over-invest because demand investment diverts customers from competitors, generating business-stealing externalities. An increase in a supplier's demand investment raises the probability of poaching customers from competitors within the period. In addition, the induced change in the distribution of matches affects competitors' future customer growth, generating dynamic business-stealing effects only captured in a framework with persistent customer-supplier relationships and supplier-to-supplier transitions.

Quantitatively, business-stealing externalities dominate the benefit of matches to customers, and the social planner would reduce aggregate demand investment by 37% relative to the calibrated equilibrium. Dynamic externalities tied to persistent customer relationships are quantitatively more important than the within-period (static) force. The magnitude and sign of the inefficiency is heterogeneous across suppliers. The positive externality on customers continues to grow in the right tail of the productivity distribution, whereas the business-stealing externality is bounded. Accordingly, while the planner would cut demand investment for most suppliers, it would raise contacts with customers for the 2% most productive firms. Eliminating the demand investment inefficiency raises welfare by about 1.0% in consumption equivalent variation 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. Eliminating the aggregate markup raises welfare by 1.9% along the transition. Eliminating distortions from markup dispersion adds another 2.2%. We find comparably small welfare losses from markup dispersion because we separate the *source* of market power from the parameter governing its *cost*. In models of market power based on product differentiation (Kimball, 1995; Atkeson and Burstein, 2008), the elasticity of demand per customer determines both variation in markups and the cost of a given dispersion. In our model, markups arise from search frictions, while the elasticity of demand per customer is identified from variation in sales per customer, which is empirically small. Separating source and cost of markups reduces the misaollocative effects of markup dispersion. If instead we counterfactually imposed a higher elasticity consistent with the aggregate markup, the productivity loss from markup dispersion would rise from 1.6% to about 5.4%. Taken together, eliminating aggregate markups, removing markup dispersion, and correcting demand-investment inefficiencies would raise welfare by 5.1% when accounting for transitional dynamics.

We solve for the welfare-maximizing product market policy and uncover a complementarity between taxing demand investment and subsidizing production inputs. We restrict attention to two constant, flat instruments: a demand-investment tax and a production subsidy. The joint optimum features a 61.1% tax on demand investment and a 17.1% production subsidy, raising welfare by 2.5% along the transition. Each instrument on its own is weaker: the optimal demand-investment tax is 33.8% in isolation, while the optimal production subsidy is 13.3% in isolation. The complementarity reflects the two-way interaction between demand investment and profits. A production subsidy raises profits per customer and thus stimulates additional (inefficient) demand investment, limiting the optimal subsidy in isolation. Conversely, a demand-investment 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 demand investment since the 1980s. We calibrate a change in the technology for contacting customers to match the rise in demand investment and within-industry sales concentration. The implied decline in the cost of contacting customers 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. Aggregate markups change little, reflecting offsetting within-firm and composition effects: stronger competition lowers firms' markups, while reallocation toward more productive, high-markup firms raises the aggregate markup (e.g., Autor, Dorn, Katz, Patterson and Van Reenen, 2020; De Loecker, Eeckhout and Unger, 2020). Despite higher demand investment, the intangible value of firms' customer bases declines, because greater demand investment by competitors raises turnover and lowers the value of each customer relationship in equilibrium. Instead, capitalizing demand investment at a constant depreciation rate yields rising measured intangible values in both the model and the data. At the same time, firm values fan out across firms, consistent with evidence of widening dispersion in firm values and profits (Eeckhout, 2025). These results caution against using industry concentration as a measure of imperfect competition or inferring intangible values by mechanically capitalizing spending under fixed depreciation.

Related literature. Our analysis builds on the literature on welfare losses from misallocation (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009), with particular emphasis on distortions arising from markups (Baqaee and Farhi, 2020; Peters, 2020; Edmond et al., 2022; Baqaee, Farhi and Sangani, 2024a,b; Bornstein and Peter, 2025). Previous work typically generates market power from product differentiation and considers only an intensive margin of demand. A small number of papers have introduced an extensive margin and investment in customers: Cavenaile, Celik, Perla and Roldan-Blanco (2023) develop a theory of customer awareness and targeted advertising. Cavenaile, Celik, Roldan-Blanco and Tian (2025) and Pearce and Wu (2025) allow for an extensive margin under oligopolistic competition for sales per customer. Afrouzi et al. (2025) document that markups covary with the intensive rather than the extensive margin, highlighting the importance of accounting for both separately, and study implications for misallocation under monopolistic competition at the intensive margin. These papers share the assumptions that market power arises from competition among differentiated products and that customers do not decide on matches based on price. In contrast, we allow for competition at the extensive margin among perfect substitutes in a market with search frictions. These departures make the number of suppliers' matches responsive to prices, separate the source and cost of market power, and capture static and dynamic business-stealing externalities—features that are central to understanding market power, the effects of demand investment, and misallocation in our setting.

Our approach to modeling investment in demand and dynamic competition for customers relates to the literature on random search in product markets. Building on Butters (1977), Dinlersoz and Yorukoglu (2012) analyze an industry equilibrium with fully transitory productivity shocks and Greenwood, Ma and Yorukoglu (2024) study implications of digital advertising for the provision of free media goods. Paciello et al. (2019) study how customers' endogenous search effort shapes the pass-through of productivity shocks to prices. In complementary work, Shen (2025) develops an industry-dynamics model with permanent heterogeneity in firms' product quality and efficiency of attracting customers to show how rising marketing can account for declining business dynamism and increasing industry concentration, and studies congestion in matching of firms to customers without pre-existing suppliers as a cause of inefficient marketing. We deviate from previous work by embedding demand investment 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 quantify both the allocative efficiency of demand investment and its interaction with the broader misallocation costs of market power. Endogenous demand 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 in the tradition of Burdett and Judd (1983), with unit demand and sellers unable to influence the number of customers they contact.

Our result of inefficient demand investment relates to a broader literature on efficiency in random search markets (Mortensen, 1982; Hosios, 1990; Fukui and Mukoyama, 2025), where efficiency of contacts requires that actors are appropriately compensated for their contribution to matching. Work on labor markets typically captures this tradeoff by imposing a matching function and determining payoffs via bargaining or sequential auctions, emphasizing congestion of contacts as the key externality. We instead study efficiency under Poisson matching and price posting, allow existing matches to raise meeting rates, and—unlike related labor search models (Moscarini and Postel-Vinay, 2013; Coles and Mortensen, 2016; Gouin-Bonenfant, 2022; Gottfries and Jarosch, 2025)—account for an intensive margin of demand within a match. Together, these features capture important aspects of product markets, highlight the interaction between contact rates and misallocation from markups, and make business-stealing of existing matches a central driver of inefficiency.

Further related work examines frictions in, or the cost of, attracting customers in business cycle dynamics (Bai, Rios-Rull and Storesletten, 2025; Fernández-Villaverde, Mandelman, Yu and Zanetti, 2024), international trade (Arkolakis, 2010; Drozd and Nosal, 2012), R&D (Cavenaile and Roldan-Blanco, 2021; Ignaszak and Sedláček, 2023), product variety expansion (Baslandze, Greenwood, Marto and Moreira, 2023), firm growth (Roldan-Blanco and Gilbukh, 2021), and price dynamics (Rudanko, 2025). In a seminal contribution, Gourio and Rudanko (2014) study the role of intangible customer capital for firms' physical investment dynamics in a directed-search model. Chiavari (2024) emphasizes the role of intangible assets for macroeconomic trends in a customer-capital model in which intangible investment raises firms' returns to scale in production, as commonly assumed in the literature (Crouzet, Eberly, Eisfeldt and Papanikolaou, 2022, 2024; De Ridder, 2024). In contrast, we focus on the role of intangible investment in demand for competition and industry dynamics, motivated by He et al. (2024), who show that sales and marketing expenditures account for a significant share of intangible investment.

Outline. Section 2 presents the theoretical framework and characterizes an equilibrium. Section 3 describes the quantitative implementation. Section 4 reports the main results on demand investment and misallocation. Section 5 analyzes the macroeconomic effects of rising demand investment. Section 6 concludes.

2 A Theory of Competition for Customers

We study investment in demand and competition for customers in general equilibrium of an economy with households, final good producers, and intermediate good producers. Our focus is on the exchange for intermediate inputs between the two types of firms, which is subject to search frictions and requires suppliers of intermediates to invest in contacting potential customers. This section outlines the environment, the optimization problem for each type of agent, as well as the definition and properties of an equilibrium.

2.1 The Environment

Preferences. Time is discrete and indexed by t. An infinitely-lived representative household values consumption of a final good, C_t, and leisure. Preferences are given by

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

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.

Production. The final good is produced by a unit continuum of identical final good firms using intermediate goods indexed by $j \in [0,1]$.[1] Each final good firm operates technology

$$(2) \qquad Y_t = \left(\int_0^1 y_{jt}^{\frac{\sigma-1}{\sigma}} \, dj \right)^{\frac{\sigma}{\sigma-1}},$$

where y_{jt} is the quantity of intermediate j used in period t and $\sigma > 1$ is the elasticity of substitution between intermediates.[2] Each intermediate j is produced combining physical capital k_{jt} and labor ℓ_{jt} with Cobb–Douglas technology

$$(3) \qquad y_{jt} = z_{jt} k_{jt}^{\alpha} \ell_{jt}^{1-\alpha}.$$

z_{jt} is total factor productivity, $\alpha \in (0,1)$ is the capital share in production, and physical capital depreciates at rate δ_K. Two types of firms operate versions of the intermediate technology with different productivity. Final good producers operate a symmetric *in-house* technology for all intermediate goods j with productivity $z_{jt} = 1 \ \forall j$. In addition, each intermediate input j is produced by a continuum of intermediate good producers, each of which operates the technology for a single intermediate good j with idiosyncratic productivity $z_{jt} = z$. Idiosyncratic 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 and that for any $z_1 < z_2$ and any z', $F(z'|z_2) \leq F(z'|z_1)$ with strict inequality for some z'; i.e., $F(\cdot \mid z_2)$ first-order stochastically dominates $F(\cdot \mid z_1)$.

[1] A formulation with households as CES consumers of a bundle of varieties would deliver similar insights. Our choice to model the final good as a bundle of intermediates is reflected in our calibration strategy below, which targets moments for business-to-business transactions.

[2] The model generally admits a well-defined solution for $\sigma \leq 1$. This is due to an extensive margin of demand precluding infinite markups from being optimal as $\sigma \downarrow 1$ and is in contrast to models of monopolistic competition where demand is determined only by a CES-aggregator. We restrict attention to $\sigma > 1$ for comparison to monopolistic competition.

Matching. The economy is subject to search frictions between final good and intermediate good producers. To be able to trade an intermediate input, final good producers (*customers*) must match with an intermediate good producer (*supplier*). Customers can have either no supplier (*unmatched*) or one supplier (*matched*) per intermediate input. We assume that suppliers are always in contact with their customers from the previous period and may contact new customers before production takes place in any period. Contacting new customers is costly for suppliers. To contact $\hat{n}$ new customers, a supplier with n existing customers incurs cost $\hat{s}(\hat{n}, n)$, paid in the final good. We further assume that the cost function satisfies $\hat{s}(\hat{n}, n) = s\left(\frac{\hat{n}}{n}\right) n$, where $s(\cdot)$ is strictly increasing and strictly convex.[3] Suppliers contact new customers at random and with equal probability, regardless of whether the customer is currently unmatched or already matched. We assume that customers cannot observe the productivity z and existing customers n of a supplier even after being contacted and both customers and suppliers cannot recall previous matches or contacts.

Entry and exit. Intermediate good producers enter and exit the economy. Each period, a fraction $\chi \in (0, 1]$ of previously unmatched customers is contacted by a potential entrant.[4] After contacting a potential customer, entrants draw their initial productivity from a distribution $F_0(z)$, and choose whether or not to enter the economy with one initial customer. We assume that $F_0(\cdot)$ is continuous and has strictly positive density throughout its support, so that the distribution is connected (i.e., contains no gaps). At the end of each period, intermediate good producers exit exogenously with probability $\delta_F \in [0, 1]$. In addition, intermediate producers may choose to exit endogenously. The customers of exiting suppliers become unmatched, but cannot immediately be contacted by entrants.

2.2 The Problem of Households

Households consume the final good, choose how much labor L_t to supply, save in physical capital K_t, own all firms and receive profits Π_t. The final good is the numeraire, with its

[3] These properties follow from a meeting technology that is constant returns to scale (CRS) in cost paid and existing customers. By construction, $\hat{s}(\lambda\hat{n}, \lambda n) = s\left(\frac{\lambda\hat{n}}{\lambda n}\right)\lambda n = \lambda\hat{s}(\hat{n}, n)$, so the cost function is CRS. Moreover, $\hat{s}(\hat{n}, n)$ is increasing and convex in $\hat{n}$ (from the properties of $s(\cdot)$) and decreasing in n, since $\frac{\partial\hat{s}}{\partial n} = s\left(\frac{\hat{n}}{n}\right) - s'\left(\frac{\hat{n}}{n}\right)\frac{\hat{n}}{n} < 0$ by strict convexity of $s(\cdot)$. Equivalently, the cost function can be expressed as a matching technology in existing customers n and cost paid $\hat{s}$: $\hat{n}(n, \hat{s}) = n\, s^{-1}\left(\frac{\hat{s}}{n}\right) \equiv nf\left(\frac{\hat{s}}{n}\right)$, with $f(\cdot)$ strictly concave. The matching technology is CRS since $\hat{n}(\lambda n, \lambda\hat{s}) = \lambda nf\left(\frac{\lambda\hat{s}}{\lambda n}\right) = \lambda\hat{n}(n, \hat{s})$ and increasing in both arguments. The positive contribution of existing customers to meetings with new customers can be interpreted as the benefit of broader visibility, word-of-mouth advertising, or peer effects in product adoption, consistent e.g. with empirical evidence in Bailey, Johnston, Kuchler, Stroebel and Wong (2022) and Argente, Méndez and Van Patten (2024).

[4] One interpretation is that when a final good firm is unmatched and produces an intermediate input in-house, with probability χ an employee spins off a new company to supply their previous employer.

price normalized to one. Households' period t budget constraint is

$$C_t + K_{t+1} = w_t L_t + (1 + r_t) K_t + \Pi_t, \quad \forall t, \tag{4}$$

where w_t is the wage rate and r_t the net return on capital. Households choose $\{C_t, L_t, K_{t+1}\}_{t\geq 0}$ to maximize (1) subject to (4). The optimal solution satisfies

$$v'(L_t) = u'(C_t) w_t, \tag{5}$$

$$u'(C_t) = \beta(1 + r_{t+1}) u'(C_{t+1}). \tag{6}$$

2.3 The Problem of Final Good Producers

Final good producers choose a bundle of intermediates to produce final output with technology (2). Cost minimization delivers the demand for each intermediate input as

$$y_{jt} = \left(\frac{p_{jt}}{P_t}\right)^{-\sigma} Y_t, \tag{7}$$

where p_{jt} is the unit cost of input j, and

$$P_t = \left(\int_0^1 p_{jt}^{1-\sigma} dj\right)^{\frac{1}{1-\sigma}} \tag{8}$$

is the unit cost of the final good. The market for the final good is perfectly competitive and hence $P_t = 1$ and $y_{jt} = p_{jt}^{-\sigma} Y_t$ in any equilibrium.

Cost of intermediates. Final good producers either produce intermediate input j in-house or source it from a supplier they match with. If matched, we assume that the final good firm sources its entire demand for intermediate j at the supplier's price p. The cost of using intermediate j in production of the final good is therefore

$$p_{jt} = \begin{cases} p & \text{if matched with a supplier charging } p \\ \overline{mc}_t & \text{if not matched with a supplier} \end{cases} \tag{9}$$

where from cost-minimization

$$\overline{mc}_t = \left(\frac{r_t + \delta_K}{\alpha}\right)^{\alpha} \left(\frac{w_t}{1-\alpha}\right)^{1-\alpha} \tag{10}$$

is the marginal cost of producing one unit of intermediate j in-house. The associated optimal per-unit factor usage is given by $\bar{\ell}_t = \left(\frac{(1-\alpha)(r_t+\delta_K)}{\alpha w_t}\right)^{\alpha}$ for labor and $\bar{k}_t = \left(\frac{\alpha w_t}{(1-\alpha)(r_t+\delta_K)}\right)^{1-\alpha}$ for capital. As the in-house technology is symmetric across intermediates, factor use and the cost of producing in-house is constant across j.

The static gain to the final good producer per intermediate procured at price p instead of produced in-house is a marginal reduction in total production cost $P_t Y_t$, given by

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

Value of a match. Final good producers are in contact with their previous suppliers for each intermediate and may be contacted by prospective new suppliers before production takes place. Final good firms in an existing match may terminate it and return to in-house production at any time, and they may rematch with a new supplier if contacted, i.e. they cannot commit to a long-term contract. Similarly, unmatched firms may accept or decline a match if contacted. As final good firms cannot observe a supplier's idiosyncratic state, they decide about acceptance and rejection of matches solely based on the price of a supplier, forming expectations about the path of future prices and the value of a match.

Define the value of being in a match with a supplier charging price p at the start of period t before suppliers contact additional customers as $M_t^i(p)$, and the corresponding value of being unmatched as U_t^i, such that

$$M_t^i(p) = M_t(p) + \int_{M_t(p)}^{\infty} (M - M_t(p))\, d\hat{O}_t(M), \tag{12}$$

$$U_t^i = U_t + \int_{U_t}^{\infty} (M - U_t)\, d\hat{O}_t(M), \tag{13}$$

where $\hat{O}_t(M)$ is the cumulative density of the highest value of a match offered among contacting suppliers, which includes the chance of not being contacted at all. $M_t(p)$ and U_t are the values of being matched with a supplier charging p or being unmatched at the start of production in period t.[5] These values are given by

$$M_t(p) = \max\left\{\eta_t(p) + \frac{U_{t+1}^i}{1 + r_{t+1}} + \frac{(1 - \delta_F)}{1 + r_{t+1}} \int_0^{\infty} \left(M_{t+1}^i(p') - U_{t+1}^i\right) dH_{t+1}(p'|p), U_t\right\}, \tag{14}$$

$$U_t = \frac{1}{1 + r_{t+1}} \left[U_{t+1}^i + \chi \int_0^{\infty} \left(M_{t+1}^i(p) - U_{t+1}^i\right) dH_{t+1}^0(p)\right], \tag{15}$$

where $H_{t+1}(p'|p)$ is the belief that a supplier charging p in period t will offer a match at or below price p' in $t + 1$, which includes the likelihood of endogenous exit. $H_{t+1}^0(p)$ is

[5] Because customers can always dissolve a match after the rematching 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 rematching without loss of generality.

the corresponding belief about entry decisions and prices conditional on entry.[6] Customers matched with a firm charging p_1 switch to a new supplier charging p_2 during rematching 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 $M_t(p) < U_t$.

2.4 The Problem of Intermediate Good Producers

We assume all intermediate goods are symmetric and omit the index j from the intermediate good producers' problem. Intermediate good producers are heterogeneous in their productivity z and in the number of customers n purchasing from them in the previous period.

Profit per customer. Intermediate good firms produce output y with technology (3) and their idiosyncratic productivity z. The marginal cost of producing one unit of output at productivity z follows from the cost-minimization problem,

$$mc_t(z) = \frac{1}{z}\left(\frac{r_t + \delta_K}{\alpha}\right)^{\alpha}\left(\frac{w_t}{1-\alpha}\right)^{1-\alpha} = \frac{\overline{mc}_t}{z}, \tag{16}$$

where $\overline{mc}_t$ captures the common component of marginal cost and optimal per-unit factor demands 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

$$\pi_t(p, z) = (p - mc_t(z))\, y_t(p) = \left(p - \frac{\overline{mc}_t}{z}\right) p^{-\sigma} Y_t. \tag{17}$$

Customer growth. We define $i = \frac{\hat{n}}{n}$ as the *contact intensity* at which a firm approaches new customers $\hat{n}$ relative to existing customers n. 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 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 and depend only on p as customers cannot observe z or n. Contacts with new customers take place before production in each period. Given $e_t(p)$ and $q_t(p)$, i, p, and n, the total

[6] For notation, we make the implicit assumption that customers do not recall past prices. This is without loss of generality in the equilibria we consider, as current prices perfectly reveal suppliers' current idiosyncratic state. Because price and entry/exit choices are sufficient to compute customer values, forming beliefs about choices is equivalent to beliefs of states. As $H_{t+1}(p'|p)$ and $H^0_{t+1}(p)$ include the belief about supplier exit and non-entry, they do not integrate to one but to the unconditional probability of non-exit or entry for given p.

number of customers at the time of production and sales is

$$n' = n - q_t(p)n + e_t(p)\hat{n} = (1 - q_t(p) + e_t(p)i)n \equiv \rho_t(p,i)n \tag{18}$$

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

Dynamic problem. At the beginning of each period after productivity z is realized, intermediate good producers that decide to operate choose a price p and contact intensity i. We assume intermediate producers cannot commit to long term contracts and cannot price discriminate among customers, and hence reset a single price every period.[7] Both p and i affect not only current profits but also future profits through the evolution of the customer base, making both decisions dynamic. With future payoffs discounted at rate r_{t+1}, the intermediate good producers solve the recursive problem [8]

$$\tilde{V}_t(z,n) = \max_{p,i}\left\{-s(i)n + \rho_t(p,i)n\pi_t(p,z) + \frac{1-\delta_F}{1+r_{t+1}}\mathbb{E}\left[\max\left\{\tilde{V}_{t+1}\left(z', \rho_t(p,i)n\right), 0\right\}|z\right]\right\}. \tag{19}$$

We refer to the total cost of contacting customers $s(i)n$ implied by the optimal i as firms' *demand investment.* $\tilde{V}_t(z,n)$ is a contraction mapping for increasing functions and $\pi_t(p,z)$ is strictly increasing in z, such that $\tilde{V}_t(z,n)$ is strictly increasing in z iff $n > 0$. The firm value scales linearly with n such that $\tilde{V}_t(z,n) = nV_t(z)$. The value per customer $V_t(z)$ depends only on the firm's productivity z and is given by

$$V_t(z) = \max_{p,i}\left\{-s(i) + \rho_t(p,i)W_t(p,z)\right\} \tag{20}$$

where we define

$$W_t(p,z) \equiv \left(\pi_t(p,z) + \frac{1-\delta_F}{1+r_{t+1}}\mathbb{E}\left[\max\left\{V_{t+1}\left(z'\right), 0\right\}|z\right]\right) \tag{21}$$

as the value per customer at production, comprising current profit and continuation value. As $\tilde{V}_t(z,n)$ is strictly increasing in z, $V_t(z)$ is strictly increasing in z. Hence, suppliers choose to exit or not enter the market at the beginning of period t iff $z < \bar{z}_t$, 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.

[7] These assumptions make pricing in our framework directly comparable to the literature on market power without search frictions (e.g. Edmond et al., 2022).

[8] Discounting profits at rate r_{t+1} is consistent with firm ownership by households. To ensure that $\tilde{V}_t(z,n)$ is well defined, we assume $\lim_{t\to\infty}\mathbb{E}_0\left[Q_{0,t}\tilde{V}_t(z_t, n_t^{opt})|z_0,n_0\right] = 0$, where $Q_{0,t} = \prod_{\tau=1}^{t}(1+r_\tau)^{-1}$ and n_t^{opt} is the number of customers under optimal choices. This rules out infinite customer growth with positive probability and must hold in equilibrium, since the total mass of matched customers is bounded above by one.

2.5 The Allocation of Customers and Aggregation

For each intermediate good, the distribution of matches between customers and suppliers of different productivity evolves endogenously in two stages: First, productivity shocks realize for existing matches, suppliers enter and exit. Second, suppliers contact additional customers and customers decide about matches. We detail both stages below and again omit subscript j as the distribution of matches evolves symmetrically across intermediates.

Market entry and exit. Recall from above that existing suppliers stay in the market and entrants decide to operate iff $z \geq \bar{z}_t$ and that entrants start operating with one initial customer. 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 the mass of unmatched customers. As an intermediate step, let $G_t^i(z)$ and u_t^i denote, respectively, the corresponding masses of matched and unmatched customers after the exit and entry of suppliers in period t, defined by

$$(22) \qquad G_t^i(z) = \begin{cases} \chi u_{t-1}\big(F_0(z) - F_0(\bar{z}_t)\big) + (1-\delta_F)\int_0^\infty \big(F(z|\hat{z}) - F(\bar{z}_t|\hat{z})\big)\, dG_{t-1}(\hat{z}), & \text{if } z \geq \bar{z}_t, \\ 0, & \text{if } z < \bar{z}_t, \end{cases}$$

$$(23) \qquad u_t^i = \Big(1 - (1 - F_0(\bar{z}_t))\chi\Big)u_{t-1} + (1-\delta_F)\int_0^\infty F(\bar{z}_t|z)\, dG_{t-1}(z) + \delta_F(1-u_{t-1}) = 1 - \lim_{z\to\infty} G_t^i(z).$$

Distribution of contacts. Suppliers generate a total of $I_t = \int_0^\infty i_t(z) dG_t^i(z)$ contacts with new customers, distributed across different productivity levels with cumulative density $D_t(z) = \int_0^z \frac{i_t(s)}{I_t} dG_t^i(s)$. The distribution of the value of new contacts is then given as $\hat{D}_t(M) = \int_0^\infty \mathbb{I}_{M_t(p_t(z))\leq M} dD_t(z)$. As suppliers contact new customers at random, the number of contacts per customer is distributed according to a Poisson distribution with mean I_t. This means that the probability of a customer being contacted by k additional suppliers is given by $\frac{I_t^k e^{-I_t}}{k!}$ and the probability that the best outside offer for a customer with k contacts has value M or less by $\hat{D}_t(M)^k$. From here we can derive the probability that a customer's best contact is not better than M as

$$\hat{O}_t(M) = \sum_{k=0}^{\infty} \hat{D}_t(M)^k \frac{I_t^k e^{-I_t}}{k!} = e^{-I_t(1-\hat{D}_t(M))}.$$

and that the highest productivity among contacting suppliers is not higher than z accordingly as $O_t(z) = e^{-I_t(1-D_t(z))}$.

Rematching between customers and suppliers. Given the distribution of contacts and customers optimal decision to accept the match with the highest value $M_t(p)$ among contacts, the updated mass $G_t(z)$ of matched customers at firms with productivity less than z and mass of unmatched customers u_t during production in period t are given by

$$G_t(z) = G_t^i(z) + u_t^i\left(O_t(z) - e^{-I_t}\right) - G_t^i(z)\sum_{k=0}^{\infty}\frac{I_t^k e^{-I_t}}{k!}\left(1-\left(1-\int_z^{\infty}\mathbb{I}_{M_t(p_t(z'))>M_t(p_t(z))}dD(z')\right)^k\right), \tag{24}$$

$$u_t = e^{-I_t}u_t^i = 1 - \lim_{z\to\infty} G_t(z), \tag{25}$$

where e^{-I_t} is the probability of not being contacted by any supplier.[9]

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

$$q_t(p) = 1 - \hat{O}_t(M_t(p)) \tag{26}$$

$$e_t(p) = \left[u_t^i + \int_0^{\infty}\mathbb{I}_{M_t(p)\geq M_t(p_t(z))}dG_t^i(z)\right]\hat{O}_t(M_t(p)). \tag{27}$$

Aggregation. As intermediate goods are symmetric, the distribution of matched customers $G_t(z)$ and unmatched customers u_t at production for a given intermediate good j corresponds to the distribution across intermediate goods for a given customer. In Appendix A.1 we show that aggregate output of the final good is given by

$$Y_t = \left(u_t\,\overline{mc}_t^{-\sigma} + \int_0^{\infty}\frac{1}{z}(p_t(z))^{-\sigma}dG_t(z)\right)^{-1}K_t^{\alpha}L_t^{1-\alpha}. \tag{28}$$

2.6 Equilibrium

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.

Equilibrium Definition

A Bayesian equilibrium in pure Markov strategies consists of a path for prices $\{w_t, r_t\}_{t=0}^{\infty}$, aggregate quantities $\{\Pi_t, Y_t, C_t, L_t, K_{t+1}\}_{t=0}^{\infty}$; mass of customers across suppliers and mass of unmatched customers $\{G_t^i(z), u_t^i, G_t(z), u_t\}_{t=0}^{\infty}$; loss and conversion rates $\{q_t(p), e_t(p)\}_{t=0}^{\infty}$; suppliers' policies, value function, and exit threshold $\{p_t(z), i_t(z), V_t(z), \bar{z}_t\}_{t=0}^{\infty}$; customer values $\{M_t^i(p), U_t^i, M_t(p), U_t\}_{t=0}^{\infty}$ and beliefs $\{H_{t+1}(p'|p), H_{t+1}^0(p)\}_{t=0}^{\infty}$, such that given initial conditions $G_{-1}(z), u_{-1}, K_0$ for all $t \geq 0$:

[9] We simplify notation by implicitly assuming that $G_t^i(z)$ has no mass points and that no operating supplier offers a match value below U_t. The first assumption holds in all equilibira we consider. The second assumption holds generally: Any supplier posting a price p such that $M_t(p) < U_t$ must anticipate that all previous customers will quit and no new customers will accept such an offer. Hence such a supplier will choose to exit in the beginning of a period and no contact with value below U_t will occur.

1. Given $\{\Pi_t, w_t, r_t\}_{t=0}^{\infty}$, consumption $\{C_t\}_{t=0}^{\infty}$, physical capital $\{K_{t+1}\}_{t=0}^{\infty}$ and labor $\{L_t\}_{t=0}^{\infty}$ maximize households' utility (1) subject to (4).

2. Given $\{w_t, r_t, q_t(p), e_t(p)\}_{t=0}^{\infty}$, suppliers' policies $\{p_t(z), i_t(z)\}_{t=0}^{\infty}$, and value $\{V_t(z)\}_{t=0}^{\infty}$ solve (20) and $V_t(\bar{z}_t) = 0$.

3. Given $\{G_t^i(z), G_t(z), p_t(z), \bar{z}_t\}_{t=0}^{\infty}$, customers' beliefs $\{H_{t+1}(p'|p), H_{t+1}^0(p)\}_{t=0}^{\infty}$ are consistent with Bayes' law. Customers' values $\{M_t^i(p), U_t^i, M_t(p), U_t\}_{t=0}^{\infty}$ are given by (12), (13), (14) and (15). Customers switch from suppliers charging p to those charging p' iff $M_t(p') \geq M_t(p)$ and from p to unmatched iff $M_t(p) < U_t$.

4. Given suppliers' and customers' choices, the masses of matched and unmatched customers $\{G_t^i(z), u_t^i, G_t(z), u_t\}_{t=0}^{\infty}$ evolve according to (22), (23), (24) and (25), and the customer loss and conversion rates $\{q_t(p), e_t(p)\}_{t=0}^{\infty}$ satisfy (26) and (27).

5. Given w_t, r_t, $p_t(z)$, $G_t(z)$, u_t, L_t, and K_t, aggregate gross output Y_t is implied by (28).

6. The labor-market clears: $L_t = \bar{\ell}_t \left[u_t \overline{mc}_t^{-\sigma} Y_t + \int_0^{\infty} \frac{1}{z} p_t(z)^{-\sigma} Y_t dG_t(z)\right] \ \forall t \geq 0$.

7. The capital market clears: $K_t = \bar{k}_t \left[u_t \overline{mc}_t^{-\sigma} Y_t + \int_0^{\infty} \frac{1}{z} p_t(z)^{-\sigma} Y_t dG_t(z)\right] \ \forall t \geq 0$.

8. The final-goods market clears: $Y_t = C_t + (K_{t+1} - (1-\delta_k)K_t) + \Gamma_t \ \forall t \geq 0$, where $\Gamma_t = \int_0^{\infty} s(i_t(z)) dG_t^i(z)$ is aggregate demand investment.

9. Aggregate profits are given by $\Pi_t = \int_0^{\infty} \pi_t(p_t(z), z) dG_t(z) \ \forall t \geq 0$.

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., $p_t(z_1) > p_t(z_2)$ for $z_1 < z_2$. In addition, we restrict attention to equilibria with monotone beliefs, such that for $p_1 < p_2$ $H_t(p'|p_1) \geq H_t(p'|p_2)$ for any p'.[10] Before we proceed, we highlight some useful properties of equilibria in this class.

Properties of Equilibria with Strictly Decreasing Prices and Monotone Beliefs

Any equilibrium in which prices are strictly decreasing in productivity is fully separating by construction. Strict monotonicity of $p_t(z)$ ensures that the inverse $\hat{z}_t(p) = p_t^{-1}(z)$ exists,

[10] As we show below, these equilibria feature efficient transitions of customers among suppliers conditional on contacts. The restriction on decreasing prices ensures comparability with the literature studying market power in models without search frictions (e.g. Edmond et al., 2022). Under our assumptions for $F(z'|z)$ and with decreasing $p_t(z)$, only monotone beliefs are consistent with Bayes Law for prices posted in equilibrium. Our assumption binds only for off-equilibrium beliefs and rules out that prices above the equilibrium range today can be associated with beliefs of lower future prices than those posted in equilibrium.

so customers can unambiguously infer a supplier's productivity type from its price. Define the image of the pricing function as $\mathcal{P}_t = \{p \in \mathbb{R} : \exists z \in [\bar{z}_t, \infty) \text{ for which } p_t(z) = p\}$, where $\bar{z}_t$ is the exit threshold productivity. Consistency of customer beliefs with Bayes' law and the pricing policy requires $H_t(p'|p) = 1 - F(\hat{z}_{t+1}(p')|\hat{z}_t(p))$ for all $p \in \mathcal{P}_t$, $p' \in \mathcal{P}_{t+1}$, and $H_t(p' \mid p) = \sup_{\hat{p} \in \mathcal{P}_{t+1}:\ \hat{p} < p'} H_t(\hat{p} \mid p)$ for $p' \notin \mathcal{P}_{t+1}$. Monotonicity of beliefs further implies that for any $p \notin \mathcal{P}_t$ and $p_1 < p < p_2$ with $p_1, p_2 \in \mathcal{P}_t$, we must have $H_t(p'|p_1) \geq H_t(p'|p) \geq H_t(p'|p_2)$ for any p'. Finally, consistency of beliefs about entrants requires that $H_t^0(p) = 1 - F_0(\hat{z}_t(p))$. Given equilibrium beliefs, customers' value of a match and of being unmatched simplify to

$$(29)\quad M_t(p) = \max\left\{\eta_t(p) + \frac{U_{t+1}^i}{1+r_{t+1}} + \frac{1-\delta_F}{1+r_{t+1}} \int_{\bar{z}_t}^{\infty} (M_{t+1}^i(p_{t+1}(z')) - U_{t+1}^i) dF(z'|\hat{z}_t(p)), U_t\right\},$$

$$(30)\quad U_t = \frac{1}{1+r_{t+1}}\left[U_{t+1}^i + \chi \int_{\bar{z}_{t+1}}^{\infty} \left(M_{t+1}^i(p_{t+1}(z)) - U_{t+1}^i\right) dF_0(z)\right]$$

for all $p \in \mathcal{P}_t$. As we focus on equilibria where $p_t(z)$, and hence $\hat{z}_t(p)$, is strictly decreasing, it follows from the properties of $\eta_t(p)$ and $F(z'|z)$ that $M_t(p)$ is strictly decreasing in p for $p \in \mathcal{P}_t$. Further, from monotonicity of beliefs, $M_t(p)$ is strictly decreasing for any p. The following Lemma is an immediate implication.

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

Given Lemma 1, the best outside offer for any customer comes from the supplier with the highest productivity, reducing the customer value functions before rematching to

$$(31)\quad M_t^i(p) = M_t(p) + \int_{\hat{z}_t(p)}^{\infty} (M_t(p_t(z')) - M_t(p))\, dO_t(z'),$$

$$(32)\quad U_t^i = U_t + \int_{\hat{z}_t(\bar{p}_t)}^{\infty} (M_t(p_t(z')) - U_t)\, dO_t(z'),$$

Accordingly the equilibrium customer growth rate and flow of customers satisfy

$$(33)\quad \rho_t(p,i) = 1 - q_t(p) + e_t(p)i = \begin{cases} (1 + (u_t^i + G_t^i(\hat{z}_t(p)))\, i)\, O_t(\hat{z}_t(p)), & \text{if } p \leq \bar{p}_t, \\ 0, & \text{if } p > \bar{p}_t \end{cases}$$

$$(34)\quad G_t(z) = G_t^i(z) + u_t^i(O_t(z) - e^{-I_t}) - G_t^i(z)(1 - O_t(z)).$$

and $\rho_t(p,i)$ is weakly decreasing in p for any given i. The distribution of customers across suppliers is continuous in z if the initial distribution $G_{-1}(z)$ is continuous and $O_t(z)$ is continuous.[11]

We impose no restrictions on the pricing function $p_t(z)$ beyond it being strictly decreasing in equilibrium. We can establish additional properties of $p_t(z)$, 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) The price of each firm is bounded above by $p_t(z) \leq p_t^S(z)$, where $p_t^S(z) = \frac{\sigma}{\sigma-1} mc_t(z)$ is the price that maximizes static profits per customer such that $\frac{\partial \pi_t(p,z)}{\partial p} = 0$,

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

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

Proof. To prove (i), suppose that a firm with productivity z sets a price $p > p_t^S(z)$. By lowering its price, the firm would strictly increase profits per customer $\pi_t(p, z)$ without reducing its growth rate $\rho_t(p, i)$. 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 $\bar{z}_t$ charged a price $p < \bar{p}_t$. Since $p_t(z)$ is strictly decreasing, p would then be the highest price posted in equilibrium. The firm could profitably raise its price infinitesimally without reducing $\rho_t(p, i)$, thereby increasing profits, because $p < p_t^S(\bar{z}_t)$ from property (i). Indeed, while property (i) only establishes the weak inequality $p \leq p_t^S(z)$ for general z, equality cannot hold at the cut-off. If $p = p_t^S(\bar{z}_t)$, then a firm with productivity $\bar{z}_t$ would earn strictly positive static profits. Since the continuation value is an option value and hence non-negative, this would imply $V_t(\bar{z}_t) > 0$, contradicting the definition of $\bar{z}_t$ as the threshold type with $V_t(\bar{z}_t) = 0$. Therefore $p < p_t^S(\bar{z}_t)$. Hence, the original price cannot be part of an equilibrium. Only at $p = \bar{p}_t$ 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 $p_t(\bar{z}_t) = \bar{p}_t$.

To prove (iii), suppose for a contradiction that there is a gap: there exists $\tilde{z}$ with $p_2 \equiv \lim_{z \uparrow \tilde{z}} p_t(z) > p_1 \equiv \lim_{z \downarrow \tilde{z}} p_t(z)$, so $(p_1, p_2) \notin \mathcal{P}_t$. These one-sided limits exist since $p_t(\cdot)$ is monotone. First, it must be that the price at the lower end of the gap is strictly below the static optimum, $p_1 < p_t^S(\tilde{z})$; otherwise $p_2 > p_1 = p_t^S(\tilde{z})$ would contradict (i). Second, given that $p_1 < p_t^S(\tilde{z})$, a firm of productivity $\tilde{z}$ posting price p_1 has a profitable deviation: raising its

[11] The latter requires continuity of $i_t(z)$ for $z \geq \bar{z}_t$, which follows from continuity of payoffs and interior optimality.

price to some p in between p_1 and p_2 increases its static profits but has no negative effect on the growth rate $\rho_t(p,i)$, 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 $p_t(z_n) \uparrow p_1$; 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 $p_t(\cdot)$ must be continuous. □

Equilibrium Prices and Contact Intensities

Given the properties established above, we can show how competition in the market for intermediates determines suppliers' choices. The optimal price $p_t(z)$ and contact intensity $i_t(z)$ for a supplier with productivity z satisfy the first order conditions

$$\rho_t(p,i)\frac{\partial \pi_t(p,z)}{\partial p} = -\frac{\partial \rho_t(p,i)}{\partial p} W_t(p,z), \tag{35}$$

$$\frac{\partial s(i)}{\partial i} = \frac{\partial \rho_t(p,i)}{\partial i} W_t(p,z), \tag{36}$$

Suppliers' price setting (35) balances between the static benefit of raising the price on profits from current customers and the dynamic cost of slower customer growth. The optimal contact intensity (36) trades off the static cost of contacting customers against the value of expanding the customer base. Prices and contact intensities interact through their effect on customer growth $\rho_t(p,i)$ and the value of a customer $W_t(p,z)$. Importantly, suppliers' choices depend on the actions of their competitors through the customer growth rate $\rho_t(p,i)$.

To see how competitors drive suppliers decisions, we impose the equilibrium conditions derived above. First, for any differentiable $G_t^i(z)$ with derivative $g_t^i(z)$

$$\frac{\partial \rho_t(p,i_t(z))}{\partial p} = \left[2 + (u_t^i + G_t^i(\hat{z}_t(p)))i_t(z)\right] i_t(\hat{z}_t(p)) g_t^i(\hat{z}_t(p)) O_t(\hat{z}_t(p)) \frac{\partial \hat{z}_t(p)}{\partial p} < 0, \tag{37}$$

$$\frac{\partial \rho_t(p,i)}{\partial i} = \left(u_t^i + G_t^i(\hat{z}_t(p))\right) O_t(\hat{z}_t(p)). \tag{38}$$

Substituting these together with the consistency condition $\hat{z}_t(p_t(z)) = z$ into (35) and (36), we can characterize the optimal choices of suppliers in equilibrium as a system of two differential equations in $i_t(z)$ and $p_t(z)$

$$\frac{\partial s(i)}{\partial i} = \left(u_t^i + G_t^i(z)\right) O_t(z) W_t(p_t(z),z), \tag{39}$$

$$\frac{\partial p_t(z)}{\partial z} = -\frac{2 + (u_t^i + G_t^i(z))i_t(z)}{1 + (u_t^i + G_t^i(z))i_t(z)} i_t(z) g_t^i(z) \cdot \frac{W_t(p_t(z),z)}{\frac{\partial \pi_t(p_t(z),z)}{\partial p}}, \tag{40}$$

The boundary condition associated with (40) is $p_t(\bar{z}_t) = \bar{p}_t$.

Condition (39) shows how the optimal contact intensity depends on the distribution of matches and contacts in equilibrium. The benefit of contacting more customers increases in the likelihood of converting contacts into matches. In equilibrium, this conversion rate is given by $(u_t^i + G_t^i(z))O_t(z)$ for a supplier of productivity z. It is the share of customers who are previously unmatched or matched with less-productive firms and are at the same time not contacted by more productive competitors. A supplier of productivity z will find it more beneficial to contact additional customers when more customers are in contact with worse firms, and agree to a match when contacted. On the other hand, if a larger share of contacts is with more productive firms a supplier will reduce its own investment in demand due to lower effectiveness. This gives more productive firms a stronger incentive to invest demand. At the level of the individual supplier, the optimal contact intensity increases in the profit made from an additional customer, $W_t(p_t(z), z)$. This suggests that firms charging higher markups per customer should invest more in increasing their demand. We will use this firm-level relationship to discipline the calibration of the model.

Condition (40) determines the optimal pricing policy and shows how idiosyncratic marginal cost pass through into prices. This pass-through is determined by competitive forces. In equilibrium, increasing a supplier's price signals to potential customers a match of lower value and increases the chances of being outbid by a competitor. How strong this concern is differs along the distribution of productivity. From (40), the slope of the price schedule is proportional to the product of the density $g_t^i(z)$ and contact intensity $i_t(z)$ of local competitors at z, scaled by the value per customer relative to the price sensitivity of per-customer profits. The latter governs the relative incentive of to increase current profits vs. investing in future customer value. The product of $g_t^i(z)$ and $i_t(z)$ determines the effective amount of *local competition*: how many customers a supplier risks losing when raising prices slightly. When suppliers face many competitors with similar productivity who invest in contacting customers, the risk of being outbid increases substantially after a marginal price increase. This means that local competition is strong and incentivises suppliers to pass through productivity gains and set prices closer to marginal cost. Conversely, when local competition is weak, either because there are few competitors with similar productivity or because competitors do not invest in contacting customers, suppliers can afford to not pass through productivity gains without losing many potential customers, leading to higher markups.

The pricing mechanism is closely related to the limit-pricing logic common in models of endogenous growth (Grossman and Helpman, 1991; Peters, 2020). There, customers can access all suppliers simultaneously, so the most productive firm sets a price equal to the marginal cost of the second most productive firm and captures the entire market, implying a single active supplier per intermediate. Here, suppliers engage in *limit-pricing in expec-*

tation: they anticipate the most productive competitor in a customer's choice set and set prices accordingly. Because customers are only in contact with a random subset of suppliers, this expectation-based limit-pricing generates a non-degenerate equilibrium distribution of prices, markups, and productivities. Moreover, the intensity of competition is endogenous, as it depends not only on the distribution of productivity but also on firms' investment in contacting customers.

3 Calibration and Quantitative Properties

For the remainder of the paper, we study a quantitative version of the framework outlined above. This section describes our calibration strategy and summarizes key quantitative properties of the model, including firms' decision rules and additional moments for validation.

3.1 Calibration

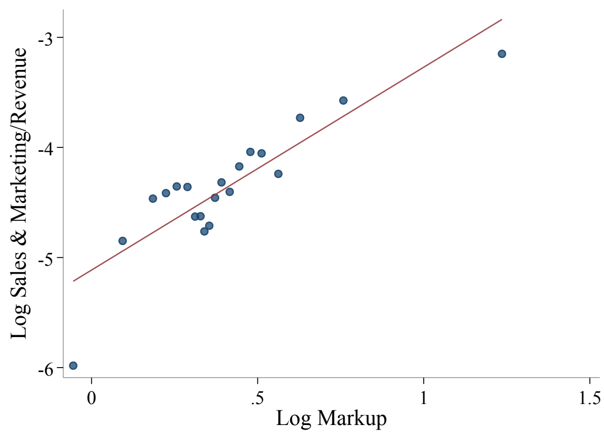
We calibrate the steady state of the model economy at annual frequency and interpret suppliers and the market for intermediate goods as a representative U.S. industry. The calibration uses moments for the period 2010-2020. We first outline how we construct a key target from the data, discuss a number of parameters that we fix externally, and then describe the internal calibration of the remaining parameters.

Demand investment and markups. A key implication of the model is that the incentive to invest in demand increases with profits per customer—i.e., with suppliers' markups (see (39)). Taking this implication to the data requires empirical measures of demand investment and markups. For demand investment, we use the measure in He et al. (2024), built from the Sales and Marketing Expenditure (S&M) variable in Capital IQ and complemented with information on sales-and-marketing employment and textual analysis of SEC filings. Unlike the residual component of Selling, General, and Administrative expenses (SG&A), this measure predicts future firm value and captures forward-looking investment in demand. In addition, He et al. (2024) document that high S&M-to-sales ratios are prevalent in business-facing industries and are correlated with the importance of a sales force as well as the need to communicate information to customers, consistent with interpreting demand investment as the cost of contacting customers in the intermediate-goods market. We measure firm-level markups as sales divided by cost of goods sold.[12]

[12]For cross-sectional regressions, this approach is equivalent to estimating markups using the production-function approach. The variable-cost elasticity included in production function estimates is commonly assumed constant within industries and therefore absorbed by industry fixed effects. For cross-sectional results we require only that, conditional on observables, measured and true ratios be proportional across firms,

In the data, we document a robust within-industry relationship between the demand-investment-to-revenue ratio and markups (Figure 1). The estimated elasticity is significantly positive and equals 1.84, which we target below.[13] Because we compute markups and demand investment from a selected sample of public firms in Capital IQ, we mirror this selection when computing the corresponding model moments.[14]

Figure 1: Demand Investment vs. Markup



Notes: Data from He et al. (2024) and Capital IQ, own calculations. Binscatter of log demand investment to revenue ratio vs. log markup. Industry and year fixed effects; cost-weighted.

External parameters and functional forms. We fix a small number of parameters ex ante at standard values. We set the physical capital depreciation rate to $\delta_K = 0.06$ and the annual discount factor to $\beta = 0.96$, implying an empirically reasonable physical capital-to-GDP ratio of 2.9. Preferences are log utility over consumption with labor disutility $\nu(L) = \omega \frac{L^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}$. We normalize steady-state labor supply to one, which pins down $\omega = 0.746$. We set the Frisch elasticity to $\gamma = 1$; it does not affect the baseline steady state but matters for counterfactuals.[15] In addition, we assume that the contact cost function takes the form $s(i) = \frac{\bar{s}}{\psi} i^{\psi}$.

so that any multiplicative measurement error is absorbed by fixed effects. Relatedly, Grassi, Morzenti and de Ridder (2022) argue that cross-sectional variation in production-function markups accurately reflects true markup variation.

[13]See Table A1 in Appendix B.1. The documented correlation is consistent with evidence from the universe of U.S. manufacturing establishments in Kehrig and Vincent (2021), who show that low-labor-share establishments spend more on advertising. This aligns with evidence on the importance of customer-related intangibles for private-firm value in Bhandari and McGrattan (2021). Afrouzi et al. (2025) document a similar relationship between markups and total SGA spending in Compustat, instead of the more restrictive measure of demand investment employed here.

[14]Building on Ottonello and Winberry (2020), we overweight older and larger firms; Appendix B.2 describes the procedure.

[15]We follow Edmond et al. (2022) for comparability.

Internal calibration. We choose nine parameters to match ten moments. Table 1 summarizes the calibration. The model fits all targeted moments closely. Although parameters are jointly determined, we relate each model parameter to one particularly informative moment.

Table 1: Calibrated Parameters and Targeted Moments

Parameter		**Value**	
Exogenous firm exit rate	δ_F	0.081	
Entrants per unmatched customer	χ	0.078	
Capital elasticity	α	0.327	
Log TFP: unconditional dispersion	σ_z	0.387	
Log TFP: auto-correlation	ρ_z	0.361	
Log TFP: mean	μ_z	0.409	
Intensive-margin elasticity	σ	2.381	
Contact cost: level	$\bar{s}$	39.792	
Contact cost: curvature	ψ	2.895	
Targeted Moment	**Model**	**Data**	**Source**
Firm entry/exit rate	0.081	0.085	BDS
Relative employment of entrants	0.526	0.540	Sterk, Sedláček and Pugsley (2021)
Aggregate labor share	0.614	0.615	BLS
Aggregate markup	$*$ 1.250	1.250	Capital IQ
Elasticity of $\frac{\text{demand investment}}{\text{revenue}}$ to markup	$*$ 1.849	1.843	Capital IQ
Elasticity of markup to revenue	0.031	0.031	Edmond et al. (2022)
Sales variation: intensive-margin share	0.262	0.260	Bernard et al. (2022)
Customer quit rate (% sales)	0.152	0.150	Dhyne, Duprez and Komatsu (2023)
Top 10% sales share	0.676	0.724	SUSB
Top 1% sales share	0.431	0.423	SUSB

Notes: $*$ indicates model moments computed after applying a filter that mimics selection into Capital IQ; see Appendix B.2 and Table A2 for details.

We discipline the exogenous firm exit rate δ_F and entrants per unmatched customer χ with the firm entry/exit rate in the Business Dynamics Statistics (BDS) and the average employment of entrants relative to incumbents (Sterk et al., 2021). We target the aggregate labor share by choosing the capital elasticity in production α.

Sales-variation moments discipline the contact technology $s(\cdot)$ and the intensive-margin elasticity of demand σ. The level $\bar{s}$ and curvature ψ of the contact cost jointly determine how easily customers reallocate across firms, shaping customer turnover and sales concentration. We target sales concentration in the U.S. from the Statistics of U.S. Businesses (SUSB).[16] For customer turnover, we target the fraction of business-to-business sales in relationships that are terminated at annual frequency. Business-to-business transactions account for more than

[16]Further details on moments from the BDS and SUSB are provided in Appendix B.3.

half of economic activity in the U.S. economy, and our interpretation of demand investment as the cost of contacts to form matches is consistent with the formation of customer-supplier relationships. As turnover in business-to-business transactions is not available at scale for the U.S. we rely on evidence from Belgium in Dhyne et al. (2023). The within-final-good substitution elasticity σ governs substitution across intermediate inputs and thus how much cross-sectional sales variation arises from the intensive margin (sales per customer) versus the extensive margin (number of customers). Consistent with the turnover in customer relationships, we discipline σ using the intensive-margin share of sales variation measured in the same Belgian data as reported in (Bernard et al., 2022).[17]

Intermediate-good producers' log TFP follows a Gaussian AR(1) process with parameters $(\sigma_z, \rho_z, \mu_z)$. From firms' optimality conditions in (39) and (40), these parameters jointly determine the distribution of markups and their relationship with firm size and demand investment. First, the lower the average productivity μ_z, the more constrained low-productivity suppliers are in their pricing by customers' outside option of in-house technology. A lower μ_z therefore reduces markups for low productive firms, strengthening the relationship between markups and firm size; accordingly, μ_z is disciplined by the elasticity of markups with respect to revenue (Edmond et al., 2022). Second, away from the constraint imposed by the outside option, higher dispersion σ_z means that suppliers are further apart in terms of productivity and thus reduces local competition as implied by (40). Lower competition raises the aggregate markup, so σ_z is disciplined by the aggregate markup in Capital IQ. Third, as variation in markups is driven by variation in productivity z, a higher persistence ρ_z makes markups more persistent. This raises the elasticity of customer value to current z, increasing the elasticity of the demand-investment-to-revenue ratio with respect to markups: high-z (high-markup) firms expect to remain high-z and therefore have a larger forward-looking payoff to investing in demand. Therefore, ρ_z is disciplined by the cross-sectional elasticity of demand investment to markups in Capital IQ as computed above.

Robustness. Some calibration moments are less commonly used or harder to measure. Appendix D therefore reports recalibrations that target alternative values for (i) the customer quit rate, (ii) the intensive-margin share of sales variation, including consumer-good-based measures for the U.S. in Einav et al. (2021), and (iii) the level of the aggregate markup. We assess the robustness of our main results to these alternative calibrations.

[17] Bernard et al. (2022) decomposes sales variation into the number of customers, customer and suppier fixed effects. As our model does not feature heterogeneity among customers, we omit customer fixed effects and compute the intensive margin share as the contribution of supplier fixed effects relative to the number of customers. To be robust to customer heterogeneity in the data, we also target sales-weighted customer quit rates instead of the fraction of relationships that are separated.

Additional moments. Table 2 shows that, beyond the targets of our calibration, the model aligns well with an additional set of empirical moments. First, although we do not target the level of demand investment, the model generates aggregate investment of 0.113 of value added, comparable to 0.080 in Capital IQ. The higher model-implied value is consistent with the interpretation that Capital IQ's sales and marketing spending captures most, but not all, demand investment.[18]

Second, despite a comparably low persistence of idiosyncratic productivity, the model generates substantial persistence in firm-level revenue consistent with the data. The high persistence of revenue is driven by the slow turnover of customer relationships, constraining the reallocation of revenue across firms. The difference in persistence of productivity and revenue suggests that customer-base dynamics are an empirically relevant amplification mechanism for firm-level revenue dynamics.

Table 2: Model Validation: Untargeted Moments

	Model	**Data**	**Source**
Demand investment over value added	0.113	0.080	Capital IQ
Log revenue auto-correlation	0.990	0.991	Capital IQ
Intermediate input elasticity σ	2.38	0.7–2.5	Peter-Ruane (2025); Carvalho et al. (2021)
Elasticity markup to # customers	0.000	−0.004	Afrouzi et al. (2025)
Elasticity markup to revenue per customer	0.152	0.187	Afrouzi et al. (2025)
Cost pass-through (annual)	0.799	0.59–0.76	Amiti, Itskhoki and Konings (2019)

Third, we calibrate σ to match the intensive-margin share of total sales variation. Alternatively, one could discipline σ using direct estimates of substitution elasticities across intermediate inputs. Our calibrated value lies at the upper end of the range reported in Carvalho, Nirei, Saito and Tahbaz-Salehi (2021) and Peter and Ruane (2025).

Finally, the model reproduces the central empirical result in Afrouzi et al. (2025): conditional on sales variation at both margins, markups are positively correlated with sales per customer but uncorrelated with the number of customers. This is driven by sales per customer responding contemporaneously to price changes while the number of customers adjusts more slowly. The model also generates average annual marginal-cost-to-price pass-through of 0.799, broadly consistent with estimates in Amiti et al. (2019), who report pass-through coefficients of up to 0.76 using Belgian firm-level data.[19]

[18]E.g. compensation for some employees building customer relationships might be reported elsewhere, such as in cost of goods sold. We do not target the level of sales and marketing spending in Capital IQ to allow for underreporting of customer investment. As long as underreporting is proportional to true investment, it does not affect the cross-sectional moment used in calibration. See Appendix B.1.

[19]The cited estimate from Afrouzi et al. (2025) is from Table A.7, which estimates the relationship under a rich set of fixed effects to control for marginal cost. Alternative estimates in the same paper show qualitatively similar results, with smaller coefficients for the elasticity of markups to revenue per customer. Our preferred

3.2 Pricing and Demand Investment

The competition for customers described in this paper speaks to two key findings of the recent literature on sales dynamics: First, variation in sales is largely driven by the extensive margin, the number of customers (Bernard et al., 2022; Einav et al., 2021; Afrouzi et al., 2025; Argente et al., 2025). Second, firms rely heavily on non-price actions to acquire new customers (He et al., 2024), but—conditional on marginal cost—do not use lower prices to attract customers when entering new markets (Argente et al., 2025; Fitzgerald et al., 2024).

While the first fact naturally arises in models with customer dynamics, the second seems at odds with pricing incentives in dynamic models of demand. One potential resolution is that demand does not respond to prices at the extensive margin.[20] An alternative explanation is that the economic forces that drive customer acquisition and retention are such that firms optimally choose not to vary prices with the number of customers. Our framework highlights this second explanation. The optimal price of suppliers depends only on productivity z, despite prices being allocative for the number of customers as both the quit and conversion rate vary with a supplier's price. Prices being independent of the number of customers follows directly from our restriction on the cost of contacts to be constant returns to scale in existing customers, which ensures that suppliers' problem scales linearly in the number of customers. The fact that firms do not vary their price with the number of customers provides a rationale for this restriction and identifies the role of existing customers for the matching technology in our framework. The allocative effect of prices at the extensive margin ensures that firms compete for customers via price. This effect is central to the results discussed below and we highlight its implications in the following sections.

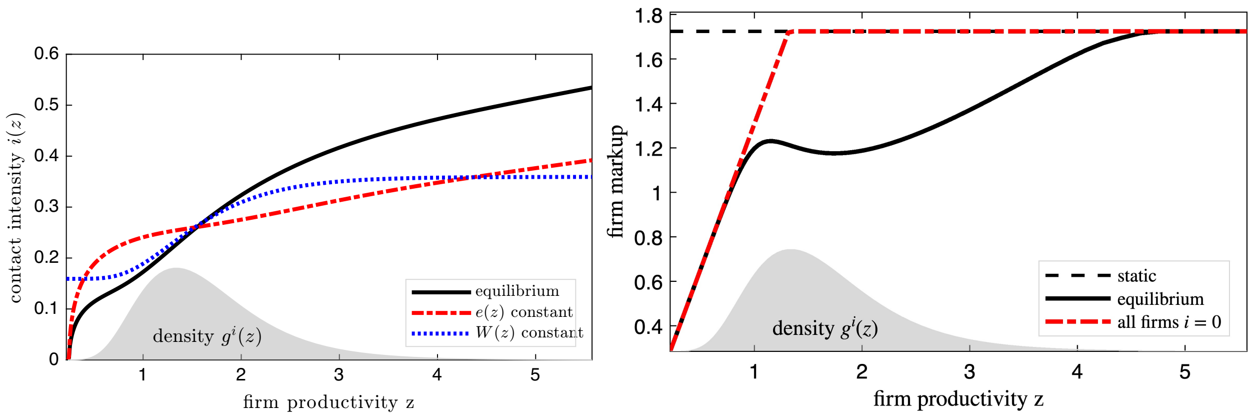
Before turning to the consequences of competition for customers, we build intuition for firms' behavior in the quantitative model. Figure 2 displays suppliers' decision rules. The left panel shows contact intensity $i(z)$ increasing in productivity z. This reflects the two forces in (39): Customer conversion rates $e(z)$ rise with z in equilibrium, as customers choose the most productive firm they are in contact with. In addition, the value of an additional customer at production, $W(p(z), z)$, is also increasing in z. The figure illustrates these channels using counterfactual decision rules that hold fixed, in turn, each of the two margins. The effect of conversion rates is strongest where the customer distribution is dense, i.e. where small productivity gains translate into large increases in customers poached from competitors.

The right panel displays suppliers' equilibrium markups as the solid black line. To under-

estimate of 0.76 in Amiti et al. (2019) is based on a specification that instruments firm-level cost with exchange rate movements based on exposure to imported intermediates.

[20] See e.g. Einav et al. (2021); Fitzgerald et al. (2024); Afrouzi et al. (2025); Cavenaile et al. (2025). Paciello et al. (2019) and Roldan-Blanco and Gilbukh (2021) provide evidence from consumer goods against this view and show that the extensive margin responds to prices.

Figure 2: Firm Decision Rules



(a) Contact intensity

(b) Markup

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

stand the drivers of equilibrium markups, we consider two counterfactuals. Under CES demand without an outside option, suppliers would charge the constant, static optimal markup at $\mu = \frac{\sigma}{\sigma-1}$. Constraining suppliers' pricing through the option of in-house production generates markups that initially rise with z and then flatten, even absent poaching or customer acquisition. This is shown by the red dashed line, which displays a counterfactual with $i(z) = 0$ for all firms. The outside option generates markups below one. Suppliers are willing to temporarily accept losses per customer to maintain the continuation value of a match, a feature directly due to the dynamic nature of competition. The equilibrium markup differs from the counterfactual with $i(z) = 0$ because of the threat of customer poaching. When suppliers invest in contacting customers, they intensify local competition, which puts downward pressure on markups. From (40), the intensity of local competition is determined by the product $i(z) \cdot g^i(z)$. The equilibrium markup is non-monotone because suppliers face the most intense local competition where the mass of competitors is dense—in the middle of the productivity distribution. In the far right tail, contact intensities remain high but potential competitors are vanishingly scarce, so the equilibrium markup converges to the static monopoly markup.

4 Competition for Customers and Misallocation

We describe how inefficiencies arise in the decentralized equilibrium and quantify the resulting sources of misallocation under our calibration. We then derive implications for optimal product-market policy, focusing on taxes on demand investment and subsidies to production.

4.1 Misallocation of Demand

We assess the efficiency of the decentralized allocation relative to a social planner facing the same technological constraints and matching frictions. Inefficiencies arise on two margins. First, search frictions in the intermediate-input market prevent customers from accessing all suppliers. These frictions generate positive markups and give rise to the sources of misallocation studied in e.g. Baqaee and Farhi (2020) and Edmond et al. (2022). Second, the distribution of customers across suppliers with different productivities is endogenous and driven by suppliers' contact choices, which reflect suppliers' private benefit and need not coincide with the social value of a contact. We discuss both margins in turn and relegate the detailed derivation of the planner solution to Appendix A.2.

Markups. Aggregate markups act like a wedge between factor payments and marginal products, distorting labor and capital supply and lowering output and welfare—the standard inefficiency from an aggregate markup with endogenous factor supply. Markup dispersion generates an additional distortion by shifting demand across intermediate inputs at the intensive margin (sales per customer). Quantities respond to relative prices at this margin and dispersion in markups drives a wedge between relative prices and relative marginal products, leading to misallocation of demand across intermediates and lowering aggregate productivity. The strength of this misallocation is governed by the elasticity of substitution across intermediates, σ, which determines customers' response to relative prices. Appendix A.3 derives these two sources of inefficiency from markups formally in our setting.

In this setting, markups and their associated distortions arise from the same friction that makes demand dynamic at the extensive margin: Search frictions make it costly to contact customers, reducing their choice set and generating market power and a distribution of markups. However, conditional on a customer's set of contacts, markups do not distort the transition of customers across suppliers. At the extensive margin, a customer cares only about which contacted supplier offers the lower price, not by how much it is lower. Since equilibrium prices are strictly decreasing in productivity z, customers switch efficiently toward more productive suppliers, as in Menzio (2024a,b). Thus, markups distort quantities within relationships (the intensive margin) but not the direction of customer reallocation across suppliers (the extensive margin).

Allocation of customers. While the presence of markups does not distort customers' choice of supplier, inefficiencies still arise at the extensive margin. This is because the distribution of contacts—and ultimately the distribution of customers across suppliers—is endogenously determined by suppliers' choice of contact intensities $i(z)$ to maximize their private profits. The planner does not generally choose the same contact intensities due to four externalities: First, suppliers do not internalize the full bilateral *private benefit* of a match as they do not account for the benefit to the customer. Second, because existing matches change the cost of contacting customers, a given match imposes a *dynamic business-stealing* externality on the rematching process in future periods. Third, any match reduces the mass of unmatched customers, *crowding out entry* in future periods. Fourth, investment in demand imposes a *static business-stealing* externality by reducing competitors' probability of forming a match. While the last effect impacts the condition for optimal demand investment directly, the other three externalities enter through the planner's valuation of a match.

We can show the first three externalities in the planner's value of a match between a customer and a supplier with productivity z at the beginning of period t, $V_t^*(z)$, measured in units of resources, and the corresponding value at production, $W_t^*(z)$. We derive both values in Appendix A.2 and obtain

$$W_t^*(z) = \frac{\partial Y_t}{\partial g_t^*(z)} + \beta \frac{u'(C_{t+1})}{u'(C_t)} \int_0^\infty V_{t+1}^*(\hat{z}) \left[\hat{f}_{t+1}(\hat{z}|z) - \chi \hat{f}_{t+1}^0(\hat{z})\right] d\hat{z}, \tag{41}$$

$$\begin{aligned} V_t^*(z) = & -s(i_t(z)) + (1 - q_t^*(z) + e_t^*(z) i_t(z)) W_t^*(z) \\ & + \int_0^\infty W_t^*(\hat{z}) g_t^{i*}(\hat{z}) \left[-\frac{\partial q_t(\hat{z})}{\partial g_t^{i*}(z)} + \frac{\partial e_t(\hat{z})}{\partial g_t^{i*}(z)} i_t(\hat{z}) \right]. \end{aligned} \tag{42}$$

Here $\hat{f}_{t+1}(z'|z)$ and $\hat{f}_{t+1}^0(z)$ denote, respectively, the transition density and entrant density of suppliers adjusted for entry and exit decisions in period $t+1$, $e_t^*(z)$ and $q_t^*(z)$ are the quit and conversion rates, and $g_t^{i*}(z)$ is the mass of customers matched with suppliers of productivity z after entry and exit in the planner's solution. Comparing these expressions to their equilibrium counterparts in (20) and (21) highlights the externalities.

First, the static private benefit of a match to a supplier is the profit per customer, $\pi(p(z), z)$. We compare this to the static benefit for the planner, $\frac{\partial Y_t}{\partial g_t^*(z)}$. The key 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 demand in equilibrium. The static benefit of a match differs further due to the distortions from markups and differences in aggregate productivity and output.

Second, the dynamic business-stealing externality enters as the integral term in (42), which adjusts the value of a match for its effect on the quit and conversion rates of other firms. Because a larger number of matches reduces the marginal cost of contacting additional customers, firms with higher matches will poach more from competitors. This effect on competitors' quit rates is thus a feature directly derived from the role of existing customers in the matching technology, and its presence is identified by evidence on the joint dynamics of prices and customers. Similarly, increasing matches at productivity z reduces the conversion rate of contacts at firms with lower productivity.

Third, as entry in our economy is directly linked to the mass of unmatched customers, increasing matches reduces entry in the following periods. The planner's value in (41) internalizes this effect by adjusting the continuation value of a match for the opportunity cost of foregone entry (the $-\chi \hat{f}^0_{t+1}(\hat{z})$ term).

Fourth, the static business-stealing externality directly enters the planner's optimality condition for the contact intensity $i^*(z)$, derived in Appendix A.2 as

$$\frac{\partial s(i^*_t(z))}{\partial i} = e^*_t(z) W^*_t(z) + \int_0^\infty W^*_t(\hat{z}) \frac{g^{i*}_t(\hat{z})}{g^{i*}_t(z)} \left[-\frac{\partial q^*_t(\hat{z})}{\partial i^*_t(z)} + \frac{\partial e^*_t(\hat{z})}{\partial i_t(z)} i^*_t(\hat{z}) \right] d\hat{z} \tag{43}$$

Comparing this condition to its equivalent in the decentralized equilibrium in (36) highlights the static business-stealing externality as the integral on the right. Suppliers do not internalize that by increasing their contact intensity they contact some customers in existing matches, increasing the quit rate at competing firms and destroying the value of the original match. In addition, suppliers do not internalize that by increasing their contact intensity, they also increase the probability that customers with multiple contacts match with them, reducing the conversion rate and hence the value of other contacts that do not translate into matches.

The crowding out of potential entry as well as the static and dynamic business-stealing externalities imply a lower value of a match or contact to the planner relative to suppliers' private value. These three externalities hence generate *over-investment* in demand in equilibrium. Capturing the dynamic externalities of matches requires a framework that can generate the persistence of customer relationships observed in the data. As the economy features both negative externalities from the effect on matches of other firms and positive externalities on customer surplus, and hence exhibits forces that can lead to either over- or under-investment in building demand by contacting customers, it is not ex ante clear whether the economy will feature too high or too low investment in demand in equilibrium. We therefore use the calibrated economy to quantify the net effect of these forces below.

In addition to the four externalities outlined above, demand investment also depends on the endogenous distribution of customers. The benefit of contacting customers depends on the conversion rate of contacts into matches $e_t^*(z)$, i.e., the fraction of customers that are unmatched or matched with firms of productivity lower than z and not contacted by firms with higher productivity. Because the planner and the decentralized equilibrium imply different contact intensities and customer distributions, (i) conversion rates $e_t(z)$ generally differ across the two economies, and (ii) aggregate investment in demand differs because it aggregates firm-level choices using different weights $g^{i*}(z)$. Finally, misalignment in the private and social value of a match can distort exit and entry decisions: we show in Appendix A.2 that the planner chooses a firm to exit (not enter) whenever $V_t^*(z) < 0$.

4.2 Quantifying Sources of Misallocation

We quantify the sources of inefficiency in the calibrated economy by comparing welfare across four allocations. Let U denote welfare in the decentralized equilibrium and U^* welfare under the full planner solution. Let U^M denote welfare in an economy that holds fixed aggregate TFP and total demand investment at their equilibrium levels, but where the planner chooses other aggregate quantities efficiently (removing the aggregate markup wedge). Finally, let U^Z denote welfare when the planner also allocates quantities within matches efficiently at the intensive margin given the equilibrium distribution of matches and demand investment (removing the loss from markup dispersion). These definitions imply the decomposition

$$\Delta \equiv U^* - U = \underbrace{U^M - U}_{\substack{\Delta_{\mathcal{M}} \\ \text{(agg. markup)}}} + \underbrace{U^Z - U^M}_{\substack{\Delta_Z \\ \text{(markup dispersion)}}} + \underbrace{U^* - U^Z}_{\substack{\Delta_G \\ \text{(customer allocation)}}} . \tag{44}$$

We quantify each term in steady state and including transitions. All transitions start from the steady state of the decentralized equilibrium as initial condition. We report welfare changes between any U_1 and U_2 in terms of consumption equivalents (CEV) as

$$CEV(U_1, U_2) \equiv \exp\left((1-\beta)(U_2 - U_1)\right) - 1. \tag{45}$$

Table 3 compares the decentralized equilibrium to the social planner allocation. Relative to the equilibrium, the planner substantially expands production by increasing factor inputs, raising GDP by 22.8% in the steady state. Welfare increases by 11.0% in steady state terms, but by only 5.1% when accounting for transition dynamics, reflecting the gradual and costly accumulation of the capital stock. At the same time, the planner sharply reduces demand investment by 36.9%. Aggregate TFP changes little, declining slightly by 0.6%. The final

Table 3: Efficiency Losses of Equilibrium Allocation

		Equil.	Planner	% Δ	% $\Delta_{\mathcal{M}}$	% Δ_Z	% Δ_G
Gross output	Y	1.092	1.286	17.7	20.8	2.7	-5.7
GDP	$Y - \Gamma$	1.000	1.228	22.8	22.7	2.9	-2.8
Demand investment	Γ	0.092	0.058	-36.9	0.0	0.0	-36.9
Tangible capital	K	2.932	4.132	40.9	44.5	3.2	-6.8
Labor	L	0.325	0.353	8.9	10.7	-0.1	-1.7
Consumption	C	0.824	0.980	18.9	18.0	2.9	-2.0
Aggregate TFP	Z	1.640	1.630	-0.6	0.0	1.6	-2.2
Welfare (CEV) – steady state				11.0	8.5	2.8	-0.3
Welfare (CEV) – transition				5.1	1.9	2.2	1.0

Notes: the top panel shows steady-state quantities; the bottom panel reports welfare gains with and without transitional dynamics. Full transition paths are reported in Appendix C.1. Quantities are normalized so that $GDP = 1$ in the decentralized equilibrium. Column % Δ reports percent changes from the equilibrium to the full social planner allocation. The three rightmost columns decompose efficiency losses into (i) an aggregate markup wedge, (ii) markup dispersion, and (iii) misallocation of customers across firms.

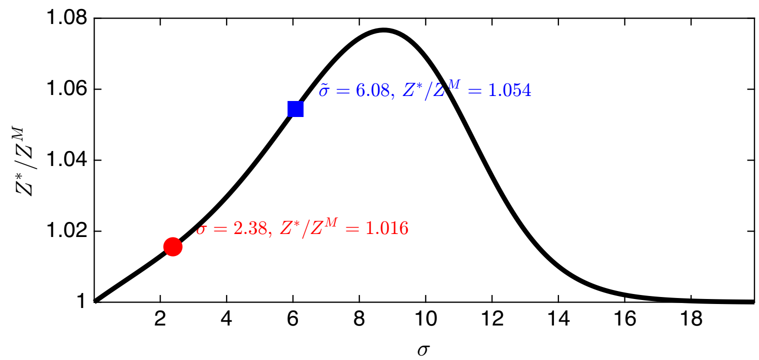
three columns of Table 3 decompose the overall effect into the aggregate markup wedge $\Delta_{\mathcal{M}}$, markup dispersion Δ_Z, and misallocation of customers across firms Δ_G.

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 3. Eliminating it would raise welfare by 8.5% in steady state but only by 1.9% along the transition, reflecting a gradual build-up of the physical capital stock. The distortion reflects under-investment in capital and inefficiently low labor input in the decentralized equilibrium. This mechanism is standard in the literature and can be replicated in a growth model with aggregate productivity fixed at the market-economy level and consistent calibration of household preferences and technology.

Cost of markup dispersion. Dispersion in intermediate producers' markups, and the resulting misallocation of demand at the intensive margin, lowers aggregate TFP by 1.6% and welfare by 2.2% along the transition (2.8% across steady states). The efficiency cost of intensive-margin misallocation depends on two objects: the degree of markup dispersion and the allocative cost of a given level of dispersion. In many models of imperfect competition, in which heterogeneous markups arise from preferences over differentiated varieties (Kimball, 1995; Atkeson and Burstein, 2008), these two objects are governed by the same parameters. In particular, the elasticity of substitution across varieties is equivalent to suppliers' price elasticity of demand and calibrated to match markups, but also determines the allocative distortions implied by a given distribution of markups.

In our setting, suppliers' relevant demand elasticity—and thus their markup—is deter-

Figure 3: Intensive Margin Misallocation as Function of Intensive Margin Elasticity



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

mined primarily by the extensive margin through search frictions and demand investment. This allows us to calibrate the intensive-margin elasticity σ separately. We set $\sigma = 2.38$ to match evidence that only about 26% of total sales variation is attributable to differences in sales per customer. The elasticity required to generate the same aggregate markup via $\mathcal{M} = \frac{\tilde{\sigma}}{\tilde{\sigma}-1}$ is instead $\tilde{\sigma} = 6.08$. The low σ implies that a given dispersion in markups generates relatively little misallocation, since demand responds weakly to price differences at the intensive margin. To quantify this mechanism, Figure 3 plots the ratio of social planner to equilibrium aggregate productivity, $\frac{Z^*}{Z^M}$, holding fixed equilibrium markups and the customer distribution while varying σ. Under $\tilde{\sigma} = 6.08$, the efficiency loss of markup dispersion increases more than threefold, from 1.6% to about 5.4%. The relatively modest role of markup dispersion in our baseline is therefore a direct implication of the calibration that separates extensive- and intensive-margin elasticities.[21]

Cost of customer misallocation. Relative to the decentralized equilibrium, the social planner substantially reduces aggregate investment in demand by 36.9%. Correcting this distortion raises welfare by 1.0% along the transition, while the steady-state comparison implies a small welfare loss of 0.3%. The difference reflects how the distribution of matches evolves over time. In equilibrium, firms sustain an inefficient distribution of customer relationships through excessive demand investment, which inflates steady-state output but lowers welfare once the full cost of building these relationships is accounted for. During the transition, the gradual unwinding of excess demand investment generates positive welfare gains.

Figure 4 illustrates both the aggregate sources and the cross-sectional heterogeneity of this

[21] Our counterfactual analysis uses the CES aggregator nested in our model. Allowing for Kimball demand yields similar conclusions: When the super-elasticity of demand is fixed to 0.16 as in Edmond et al. (2022), matching the aggregate markup implies misallocation of 4.9%, comparable to the CES case.

distortion. Panel (a) decomposes the planner-equilibrium difference in aggregate demand investment into the forces described in Section 4.1. We employ a Shapley-Owen decomposition to quantify the marginal effect of each distortion; the details of the decomposition are outlined in Appendix A.4. The planner internalizing the full private benefit of a match, including to the customer involved, would by itself lead to substantially higher demand investment than in equilibrium. This force is more than offset by static and dynamic business-stealing externalities as well as the crowding out of entry. Firms do not internalize that additional investment diverts customers from competitors, both directly through poaching and indirectly by crowding out competitors' matches in the contact process. Importantly, the dynamic externalities of existing matches on entry and customer reallocation is quantitatively more important than the static business-stealing externality generated by demand investment in the current period, highlighting the importance of accounting for dynamic effects on competition. Differences in the distribution of matches play essentially no role as in combination of all changes of demand investment across suppliers the distribution of matches before rematching moves little. As the planner chooses an exit threshold close to the equilibrium value, differences in exit and entry decisions also play a negligible role for welfare losses from customer misallocation.

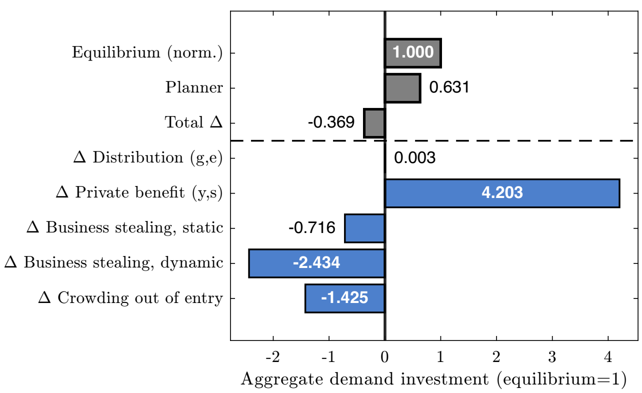
Panel (b) shows that the planner reduces demand investment for most suppliers but increases it in the right tail, choosing higher contact intensities than in equilibrium for roughly the top 2% of firms. Although business-stealing externalities are present also at high productivity levels, they flatten out, while surplus of customers grows further. While suppliers asymptotically capture only a fraction of the total private benefit of a match, the planner internalizes the increasingly large consumer surplus, leading to higher optimal investment at the top of the distribution. Aggregate demand investment nonetheless falls sharply, reflecting the dominance of the negative externalities on matches of competitors overall.

4.3 Optimal Demand Investment Taxes and Production Subsidies

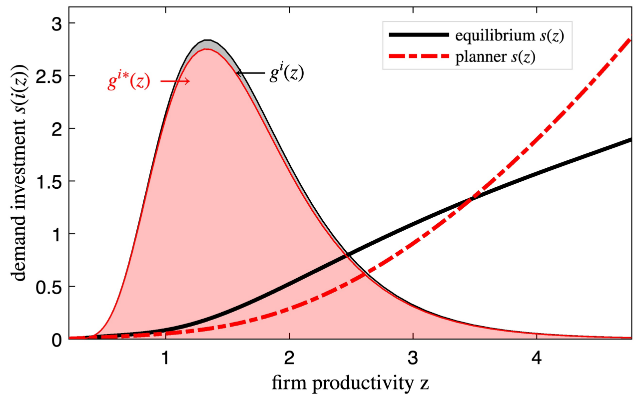
The previous section showed that the decentralized equilibrium yields an inefficient allocation of demand relative to the social optimum. Here we examine how much efficiency a government can restore using a restricted set of policy instruments. We consider two tools: First, a flat tax τ_S on demand investment $s(i(z))$, which targets over-investment. Second, a subsidy τ_P on both capital and labor inputs, designed to offset the level distortion from markups. The optimization problem of intermediate-good producers becomes

$$V_t(z) = \max_{p,i} \left\{ -(1+\tau_S)s(i) + \rho_t(p,i) \left[\pi_t \left(p, \frac{z}{1-\tau_P} \right) + \frac{1-\delta_F}{1+r_{t+1}} \mathbb{E} \left[\max \left\{ V_{t+1}(z'), 0 \right\} | z \right] \right] \right\}. \tag{46}$$

Figure 4: Equilibrium vs. Social Planner Demand Investment



(a) Aggregate decomposition



(b) By firm productivity

Notes: Panel (a) decomposes the aggregate difference in demand investment between the planner and equilibrium into differences due to distribution terms, private benefits, business-stealing (both static and dynamic), and crowding out of entry. Positive bars indicate forces that, in isolation, would lead the planner to choose higher aggregate demand investment than in equilibrium, while negative bars indicate forces that lead the planner to reduce demand investment relative to the equilibrium. Panel (b) plots demand investment per customer $s(i(z))$ by firm productivity in the equilibrium (black solid) and social planner allocation (red dashed), together with the customer distribution.

We impose government budget clearing period by period, such that

$$\tau_P \overline{mc}_t \left(u_t y_t \left((1-\tau_P)\overline{mc}_t \right) + \int_0^\infty \frac{1}{z} y_t(p_t(z)) dG_t(z) \right) = T_t + \tau_S \int_0^\infty s(i_t(z)) dG_t^i(z), \tag{47}$$

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

The joint optimum features a 61.1% demand-investment tax and a 17.1% production subsidy, raising welfare by 2.5% along the transition. This gain amounts to about half of the total welfare improvement attainable under the planner allocation. Under the joint policy, aggregate demand investment falls by 31.6% relative to the baseline.

Complementary policy instruments. Comparing the jointly optimal demand-investment tax and production subsidy to their optimal levels in isolation, we find that the two instruments are complements. The optimal production subsidy is 13.3% in isolation but rises to 17.1% when combined with a demand-investment tax. Likewise, the optimal demand-investment tax is 33.8% in isolation but rises to 61.1% when paired with a production subsidy. The welfare gain of introducing both instruments jointly is 2.5%, exceeding the sum of

Table 4: Optimal Demand Investment Taxes and Production Subsidies

		Baseline	τ_P only		τ_S only		Both	
		Value	Value	%Δ	Value	%Δ	Value	%Δ
Demand investment tax (%)	τ_S				33.8		61.1	
Production subsidy (%)	τ_P		13.3				17.1	
Lump-sum tax (% of GDP)	T		16.7		-2.1		17.8	
Gross output	Y	1.092	1.306	19.5	1.035	-5.3	1.265	15.8
GDP	$Y - \Gamma$	1.000	1.193	19.3	0.973	-2.7	1.202	20.2
Demand investment	Γ	0.092	0.113	22.4	0.062	-32.8	0.063	-31.6
Tangible capital	K	2.932	4.052	38.2	2.763	-5.8	4.072	38.9
Labor	L	0.325	0.356	9.5	0.318	-1.9	0.355	9.3
Consumption	C	0.824	0.949	15.2	0.807	-2.1	0.957	16.2
Aggregate TFP	Z	1.640	1.658	1.2	1.605	-2.1	1.606	-2.0
Aggregate Markup	$\mathcal{M}$	1.197	1.194	-0.3	1.204	0.6	1.203	0.5
Welfare (CEV) – steady state				7.0		-0.7		8.0
Welfare (CEV) – transition				1.3		0.5		2.5

Notes: We compare steady states for all quantities and report welfare changes with and without transitional dynamics. Transition paths for all variables are provided in Appendix C.2. 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 demand-investment 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.

the gains from each instrument in isolation (1.7%) by about 40%.

The complementarity stems from the two-way interaction between demand investment and markups in equilibrium. Demand investment rises with firms' profits per customer. A production subsidy boosts profits and thereby induces additional (inefficient) demand investment, as reflected in the 22.4% increase when the subsidy is introduced alone. Conversely, demand investment intensifies price competition and dampens markup distortions. A tax that raises the cost of demand investment raises markups and worsens related distortions. With a production subsidy in place to offset aggregate markup distortions, the demand-investment tax can be set more aggressively.

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

Size-dependent policies. The two policies we analyze address average inefficiencies from demand investment and markups but do not differentiate across firms and thus cannot eliminate heterogeneity in distortions. One might consider size-dependent interventions, as often proposed in the literature, to address this heterogeneity. Such policies are difficult to implement in practice, but also lack theoretical grounding in our framework. Here, firm size, measured by e.g. total sales, is not a sufficient statistic for productivity, markups, or contact intensity. Much of the observed variation in sales reflects differences in customer base size, shaped by past productivity and past investment rather than current fundamentals. Consequently, policies conditioned on total sales would be poorly targeted.[22] 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.

Robustness. To better understand our findings on efficiency and optimal policy, and to assess the robustness of the quantitative results, we conduct two sets of exercises.

First, we implement a transparent comparative static that moves the contact-cost shifter $\bar{s}$ holding all other parameters fixed. Lower costs of contacting customers intensify competition and compress markups, reducing the welfare loss from both the level and dispersion of markups; higher cost of contacts have the opposite effect (Appendix D.2.1, Table A4).

Second, we compute the welfare-loss decomposition and the optimal policy under alternative calibrations. All mechanisms discussed in this section, including the complementarity between production subsidies and demand-investment taxes, hold under alternative calibrations (Appendix D.2.2, Table A5).

5 Macroeconomic Effects of Rising Demand Investment

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

[22] Appendix Figure A7 illustrates that size distortions from markup dispersion vary systematically with productivity but show little systematic pattern across the revenue distribution.

5.1 Calibration of Technological Change in Contact Cost

While the precise level and trend in demand investment depends on the measure applied, all indicators point toward 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 in sales and 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 et al. (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.[23] The ratio of brand value investment to GDP has increased by 15% from 1985 to 2020. Conservatively, we target a 15% increase in aggregate demand investment (measured relative to aggregate output) over time.[24] 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).

Table 5 compares outcomes across the two calibrated steady states. We infer a change in the two parameters ($\bar{s}$, ψ) of the contact cost function to replicate the rise in aggregate demand investment and sales concentration. The middle panel makes the implied shift in contact cost transparent by reporting demand investment holding fixed contact intensities $i(z)$ at the 1980s allocation: mean investment per customer falls by 31.0%, and the decline is more pronounced for high contact intensity, consistent with a reduction in convexity (the P75/P25 demand-investment ratio falls by 8.3%). The change in the two parameters is identified: $\bar{s}$ primarily affects aggregate demand investment, while ψ primarily affects sales concentration.

5.2 The Effects of Rising Demand Investment

At the firm level, the calibrated change in the contact technology induces most suppliers to increase their contact intensity, with the strongest response among the most productive firms. This heterogeneity translates into differential customer growth rate responses across the productivity distribution and a reallocation of customers toward high-productivity suppliers.[25]

[23]See Bontadini, Corrado, Haskel, Iommi and Jona-Lasinio (2023) for a description of the KLEMS database and e.g. Corrado, Haskel, Jona-Lasinio and Iommi (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`.

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

[25]Figure A8 in the appendix reports how outcomes and choices change along the productivity distribution.

Table 5: Rise in Demand Investment: 1980s vs. 2010s

		1980s	2010s	%Δ
Demand investment / gross output (target +15%)	$\frac{\Gamma}{Y}$	0.073	0.085	15.0
Top 1% sales share (target +25.8%)		0.342	0.431	25.8
Average demand investment (fixing $i(z)$)		0.326	0.225	-31.0
P75/P25 demand investment (fixing $i(z)$)		3.430	3.146	-8.3
GDP	$Y - \Gamma$	1.000	1.046	4.6
Aggregate TFP	Z	1.590	1.640	3.1
Average customer quit rate	$\mathbb{E}_z\left[q(z)\right]$	0.140	0.152	8.8
Aggregate markup	$\mathcal{M}$	1.201	1.197	-0.4
Intangible capital intensity	$\frac{K^V}{K}$	0.557	0.487	-12.5

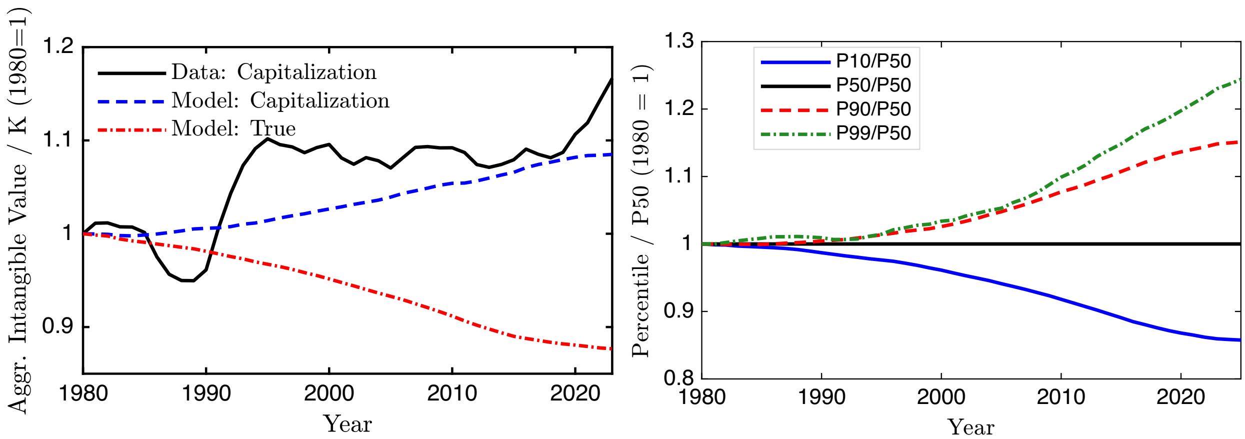
Notes: Quantities normalized so that $GDP = 1$ in the 1980s steady state. The table compares the calibrated 2010s steady state to another steady state ("1980s") with the contact cost function parameters $(\bar{s}, \psi)$ re-calibrated to generate the targeted increase in demand investment and sales concentration over time. The mean and P75/P25 demand investment measures in the middle panel are computed fixing contact intensities $i(z)$ at the 1980s level. %Δ reports percent changes relative to the 1980s steady state. The customer quit rate is computed as sales-weighted average across firms.

The aggregate effects, reported across steady states in Table 5, are sizeable. Aggregate TFP increases by 3.1%, implying a 4.6% increase in GDP. These gains reflect the reallocation of customers toward more productive suppliers.

The effect on the aggregate markup is theoretically ambiguous. On the one hand, higher demand investment intensifies competition for customers, raising turnover and 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 et al., 2020). Quantitatively, the within-firm component slightly dominates the reallocation effect, inducing a 0.4% decline in the aggregate markup. This result highlights that with endogenous market power due to search frictions, increases in market concentration need not translate into higher markups. When concentration arises from changes in the technology with which customers are matchted to suppliers, and increase in concentration can coincide with a market that becomes more competitive.

The rise in demand investment also has sharp implications for the value of customer relationships. The intangible intensity of the capital stock—measured as the customer value $K^V = \int_0^\infty V(z)dG(z)$ relative to physical capital K—declines even as demand investment rises: the intangible intensity falls by 12.5% across steady states. This result is not inconsis-

Figure 5: Technological Change in Contact Cost: Effects on Intangible Capital Intensity



(a) Aggregate Intangible Capital Intensity (b) Intangible Firm Values: Percentiles

Notes: Panel (a) shows the time series of intangible capital intensity $\frac{K^V}{K}$ following the technological change in $s(\cdot)$, using both the capitalization method (model and data) and the true model values of customer relationships, $n(z)V(z)$, where $n(z)$ denotes firms' customer base. Panel (b) shows the ratio of intangible firm values $n(z)V(z)$ at different percentiles of the productivity distribution over time. All series normalized to 1 in 1980.

tent with previous work arguing in favor of an increase in intangible capital. This literature usually constructs the stock of intangible capital from the perpetual-inventory approach as capitalized spending under a constant depreciation rate.[26] Panel (a) of Figure 5 reports capitalized spending under the perpetual inventory method in both the data and model economy, alongside the true capital intensity in the model. We apply the average customer quit rate in the 1980s economy as depreciation rate.[27] If demand investment is capitalized using a fixed depreciation rate, both the model and the data display a rise in the intangible intensity of capital of roughly 10%.

The difference to the true model value arises because demand investment endogenously shifts the customer quit rate—the depreciation rate of intangible customer capital. Greater demand investment raises customer turnover, shortening expected relationship duration and reducing the present discounted value of profits from each customer. This result cautions against mechanically applying the capitalization approach with a fixed depreciation rate to infer the value of firms' customer bases. Our framework suggests that demand investment affects not only the number of customers a firm accumulates but also how each customer should be valued in equilibrium.

[26]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).

[27]To be able to go back to the 1980s' we capitalize total SGA spending in the data. As we normalize the 1980s ratio to 1 for all time series, this approximation is valid as long as demand investment has grown in parallel with total SGA spending.

Finally, the effects on intangible values are highly heterogeneous across firms. Panel (b) of Figure 5 shows a pronounced fanning out of firm values: the intangible value of the most productive firms rises relative to the median firm as the economy transitions from the 1980s to the 2010s. This fanning out reflects opposite-signed changes in the value of customer relationships across the productivity distribution: the value of a customer $V(z)$ increases for high-productivity firms but declines for low-productivity firms. This model prediction is consistent with evidence of a widening dispersion in firm values and profits over time (Eeckhout, 2025).

Robustness. Appendix D.3 shows that the results in this section are qualitatively and quantitatively similar under alternative calibrations.

6 Conclusion

We have outlined a new quantitative framework to study how investment in demand shapes dynamic competition for customers. We show that the model matches salient empirical facts about how firms compete for customers and highlight two channels through which demand investment affects allocative efficiency. First, demand investment exerts a pro-competitive force by increasing customer turnover and lowering markups, thereby mitigating misallocation from imperfect competition. Second, firms over-invest in contacting customers because they do not internalize business-stealing externalities imposed on competitors. Our results have direct implications for the conduct of competition policy and suggest that instruments aimed at correcting distortions from market power must jointly account for firms' dual margins of competition—prices and demand investment. The framework also provides a novel perspective on macroeconomic trends, accounting for rising industry concentration without implying greater market power, and highlighting equilibrium forces that may reduce firms' intangible value despite higher investment.

Our findings relate to a broader agenda for understanding the role of customer-base dynamics in industry competition and macroeconomic outcomes. We provide a framework in which competitive forces are not driven by product differentiation, and we have focused our analysis on implications for misallocation, productivity, and the intangible value of firms. Future work may explore the role of competition for customers in the context of other macroeconomic questions where firms' margins and profitability are important drivers of outcomes, such as growth or the transmission of monetary policy to prices. Our results also call for richer measurement of customer dynamics and margins of competition beyond prices, both at the micro and macro level. We leave these questions for future research.

References

Afrouzi, H., A. Drenik, and R. Kim (2025): “Concentration, Market Power, and Misallocation: The Role of Endogenous Customer Acquisition,” .

Amiti, M., O. Itskhoki, and J. Konings (2019): “International shocks, variable markups, and domestic prices,” *The Review of Economic Studies*, 86(6), 2356–2402.

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.

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

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): “The fall of the labor share and the rise of superstar firms,” *The Quarterly Journal of Economics*, 135(2), 645–709.

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.

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

Baqaee, D. R., E. Farhi, and K. Sangani (2024a): “The darwinian returns to scale,” *Review of Economic Studies*, 91(3), 1373–1405.

Baqaee, D. R., E. Farhi, and K. Sangani (2024b): “The supply-side effects of monetary policy,” *Journal of Political Economy*, 132(4), 1065–1112.

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.

Bhandari, A., and E. R. McGrattan (2021): "Sweat equity in US private business," *The Quarterly Journal of Economics*, 136(2), 727–781.

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

Bornstein, G., and A. Peter (2025): "Nonlinear pricing and misallocation," *American Economic Review*, 115(11), 3853–3908.

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

Burdett, K., and K. L. Judd (1983): "Equilibrium price dispersion," *Econometrica: Journal of the Econometric Society*, pp. 955–969.

Butters, G. R. (1977): "Equilibrium Distributions of Sales and Advertising Prices.," *Review of Economic Studies*, 44(3).

Carvalho, V. M., M. Nirei, Y. U. Saito, and A. Tahbaz-Salehi (2021): "Supply chain disruptions: Evidence from the great east japan earthquake," *The Quarterly Journal of Economics*, 136(2), 1255–1321.

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.

Dhyne, E., C. Duprez, and T. Komatsu (2023): "The Belgian Business-to-Business Transactions Dataset 2002–2021," NBB Working Paper 444, National Bank of Belgium, Brussels.

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.

Eeckhout, J. (2025): "Presidential Address 2024: The Value and Profits of Firms," *Journal of the European Economic Association*, 23(1), 52–118.

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.

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

Gottfries, A., and G. Jarosch (2025): "Dynamic monopsony with large firms and an application to non-competes," .

Gouin-Bonenfant, É. (2022): "Productivity dispersion, between-firm competition, and the labor share," *Econometrica*, 90(6), 2755–2793.

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.

Grossman, G. M., and E. Helpman (1991): "Quality ladders in the theory of growth," *The review of economic studies*, 58(1), 43–61.

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

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.

Kehrig, M., and N. Vincent (2021): "The Micro-Level Anatomy of the Labor Share Decline," *The Quarterly Journal of Economics*, 136(2), 1031–1087.

Kimball, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 27(4), 1241–1277.

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.

Mortensen, D. T. (1982): "Property rights and efficiency in mating, racing, and related games," *The American Economic Review*, 72(5), 968–979.

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

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," .

Peter, A., and C. Ruane (2025): "Intermediate Input Substitutability," .

Peters, M. (2020): "Heterogeneous markups, growth, and endogenous misallocation," *Econometrica*, 88(5), 2037–2073.

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 Aggregation

This appendix derives the aggregate markup and aggregate productivity in the decentralized equilibrium, taking as given suppliers' price setting and the distribution of customers across suppliers. As the analysis in this section is static, we omit time subscripts for ease of notation.

We define aggregate productivity in the decentralized equilibrium Z^M as the Solow-residual of a hypothetical aggregate production function with identical Cobb-Douglas structure as in (3), that is

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

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

$$\begin{aligned} Z^M &= \frac{Y}{\left(\bar{\ell} u y(\overline{mc}) + \int_0^\infty \frac{\bar{\ell}}{z} y(p(z)) dG(z)\right)^{1-\alpha} \left(\bar{k} u y(\overline{mc}) + \int_0^\infty \frac{\bar{k}}{z} y(p(z)) dG(z)\right)^{\alpha}} \\ &= \left(\bar{\ell}^{1-\alpha}\bar{k}^\alpha \left(u \frac{y(\overline{mc})}{Y} + \int_0^\infty \frac{1}{z}\frac{y(p(z))}{Y} dG(z)\right)\right)^{-1} \\ &= \left(u\,\overline{mc}^{-\sigma} + \int_0^\infty \frac{1}{z}(p(z))^{-\sigma} dG(z)\right)^{-1} \end{aligned} \tag{A.2}$$

We can further simplify this expression to write Z^M as a function of markups

$$\begin{aligned} Z^M &= \overline{mc}^\sigma \left(u + \int_0^\infty \frac{1}{z}\left(\frac{\mu(z)}{z}\right)^{-\sigma} dG(z)\right)^{-1} \\ &= \left(\frac{Z^M}{\mathcal{M}}\right)^\sigma \left(u + \int_0^\infty \frac{1}{z}\left(\frac{\mu(z)}{z}\right)^{-\sigma} dG(z)\right)^{-1} \\ \Rightarrow Z^M &= \left(u\left(\frac{1}{\mathcal{M}}\right)^{-\sigma} + \int_0^\infty z^{\sigma-1}\left(\frac{\mu(z)}{\mathcal{M}}\right)^{-\sigma} dG(z)\right)^{\frac{1}{\sigma-1}} \end{aligned} \tag{A.3}$$

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

$$\text{(A.4)} \qquad \frac{wL}{Y} = \frac{w\frac{Y}{Z^M}\bar{\ell}}{Y} = \frac{w}{Z^M}\frac{(1-\alpha)\overline{mc}}{w} = \frac{1-\alpha}{\mathcal{M}}$$

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

$$\text{(A.5)} \qquad \frac{wL}{Y}\mathcal{M} = \mu(z)\frac{w\ell(z)}{p(z)y(z)}$$

$$\text{(A.6)} \qquad \Rightarrow \mathcal{M}\frac{p(z)y(z)}{Y} = \frac{\ell(z)}{L}\mu(z)$$

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

$$\text{(A.7)} \qquad \mathcal{M} = u\frac{\bar{\ell}y(\overline{mc})}{L} + \int_0^\infty \frac{\bar{\ell}y(p(z))}{zL}\mu(z)dG(z)$$

where we make use of the fact that we have normalized the price of the final good in (8) to one. As in Edmond et al. (2022), the aggregate markup is given by the input-weighted average of individual markups, evaluated at equilibrium demand for labor and the distribution of matches $G(z)$.[28]

A.2 Social Planner Problem

This appendix outlines and solves the problem of a social planner that maximizes welfare in the economy subject to the same technologies and search frictions as the decentralized equilibrium. In particular, the planner makes choices for households, final good firms and intermediate good firms, taking into account their full effect on aggregate outcomes. We separate the problem of the planner in two stages: First, we solve for the optimal allocation of demand per match across intermediates, taking as given the distribution of matches. As this is a static problem we can separate it from the dynamic considerations of the planner for ease of exposition. Second, we solve for the dynamic problem of the planner that chooses the allocation of customers across suppliers as well as aggregate quantities, taking the optimal static allocation of demand at the intensive margin for a given distribution of customers as given.

Intensive margin allocation. Consider the optimal allocation of demand a planner would choose at the intensive margin. The problem is static and we omit time subscripts. The

[28] 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.

planner takes the distribution of matches $G(z)$ and mass of unmatched u as given. As in-house production and intermediate firms operate the same Cobb-Douglas technology with varying productivity, to minimize cost the planner solves

$$\min_{y^u,\{y(z)\}_z} uy^u + \int_0^\infty \frac{y(z)}{z_j} dG(z) \tag{A.8}$$

$$\text{s.t.} \left(u(y^u)^{\frac{\sigma-1}{\sigma}} + \int_0^\infty y(z)^{\frac{\sigma-1}{\sigma}} dG(z)\right)^{\frac{\sigma}{\sigma-1}} = Y \tag{A.9}$$

This is equivalent to maximizing aggregate productivity $Z = \frac{Y}{K^\alpha L^{1-\alpha}}$. Taking FOCs with respect to $y(z)$ yields

$$\frac{1}{z} = y(z)^{-\frac{1}{\sigma}} Y^{\frac{1}{\sigma}} \lambda \tag{A.10}$$

where λ is the Lagrange multiplier on the constraint. The derivation for y^u is analogous. Multiplying both sides by $y(z)$, integrating over z with $G(z)$ and adding the corresponding condition for y^u yields

$$\begin{aligned} uy^u + \int_0^\infty \frac{y(z)}{z} dG(z) &= \lambda Y^{\frac{1}{\sigma}} \left(u(y^u)^{\frac{\sigma-1}{\sigma}} + \int_0^\infty y(z)^{\frac{\sigma-1}{\sigma}} dG(z)\right) = \lambda Y \\ \Rightarrow K^\alpha L^{1-\alpha} &= \lambda Y \\ \Rightarrow \lambda &= \frac{1}{Z} \end{aligned} \tag{A.11}$$

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

$$\frac{y(z)}{Y} = \left(\frac{Z}{z}\right)^{-\sigma} \tag{A.12}$$

and the corresponding aggregate productivity under the optimal allocation as

$$\begin{aligned} Z &= \left(u\frac{y^u}{Y} + \int_0^\infty \frac{1}{z}\frac{y(z)}{Y} dG(z)\right)^{-1} \\ \Rightarrow Z &= \left(u + \int_0^\infty z^{\sigma-1} dG(z)\right)^{\frac{1}{\sigma-1}} \end{aligned} \tag{A.13}$$

Efficient allocation of customers We solve for the dynamic problem of a social planner that chooses the allocation of customers across suppliers as well as aggregate quantites, taking the optimal static allocation of demand at the intensive margin given a distribution of customers as derived above as given. From the effect of a match on aggregate productivity and because of the properties of $F(z'|z)$, it is straightforward to see that given two contacts the planner will always prefer a match with higher productivity. We incorporate this rule of how to transition matches among contacts when setting up the planner problem. The full problem is defined as

$$\text{(A.14)} \quad \max_{\{C_t^*, L_t^*, K_{t+1}^*, \{i_t^*(z), g_t^{i*}(z), g_t^*(z), \delta_t^*(z), \delta_t^{E*}(z)\}_{z=0}^{\infty}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \{u(C_t^*) - v(L_t^*)\}$$

$$\text{(A.15)} \quad \text{s.t. } C_t^* + K_{t+1}^* + \int_0^\infty s(i_t^*(z)) g_t^{i*}(z) dz \le Z_t^* (K_t^*)^\alpha (L_t^*)^{1-\alpha} + (1-\delta_k) K_t^* \;\; \forall t$$

$$\text{(A.16)} \quad Z_t^* = \left(u_t^* + \int_0^\infty z^{\sigma-1} g_t^*(z) dz \right)^{\frac{1}{\sigma-1}} \;\; \forall t$$

$$\text{(A.17)} \quad g_t^*(z) = (1 - q_t^*(z) + e_t^*(z) i_t^*(z)) g_t^{i*}(z) \;\; \forall t, z$$

$$\text{(A.18)} \quad u_t^* = 1 - \int_0^\infty g_t^*(z) dz \;\; \forall t$$

$$\text{(A.19)} \quad g_t^{i*}(z) = (1 - \delta_t^*(z)) \int_0^\infty g_{t-1}^*(\hat{z}) f(z|\hat{z}) d\hat{z} + (1 - \delta_t^{E*}(z)) \chi u_{t-1}^* f_0(z) \;\; \forall t, z$$

$$\text{(A.20)} \quad u_t^{i*} = u_{t-1}^* + \int_0^\infty \delta_t^*(z) \int_0^\infty g_{t-1}^*(\hat{z}) f(z|\hat{z}) d\hat{z} dz - \int_0^\infty (1 - \delta_t^{E*}(z)) \chi u_{t-1}^* f_0(z) dz \;\; \forall t$$

$$\text{(A.21)} \quad \delta_t^*(z) \ge \delta_F, \quad \delta_t^*(z) \le 1 \;\; \forall t, z$$

$$\text{(A.22)} \quad \delta_t^{E*}(z) \ge 0, \quad \delta_t^{E*}(z) \le 1 \;\; \forall t, z$$

$$\text{(A.23)} \quad e_t^*(z) = \left[u_t^i + \int_0^z g_t^{i*}(\hat{z}) d\hat{z} \right] e^{-I_t^*(1 - D_t^*(z))} \;\; \forall t, z$$

$$\text{(A.24)} \quad q_t^*(z) = 1 - e^{-I_t^*(1 - D_t^*(z))} \;\; \forall t, z$$

$$\text{(A.25)} \quad I_t^* = \int_0^\infty i_t^*(z) g_t^{i*}(z) dz \;\; \forall t$$

$$\text{(A.26)} \quad D_t^*(z) = \int_0^z \frac{i_t^*(\hat{z})}{I_t^*} g_t^{i*}(\hat{z}) d\hat{z} \;\; \forall t, z$$

for given K_0^*, u_{-1}^* and $\{g_{-1}^*(z)\}_{z=0}^{\infty}$. ${g^{i*}}_t(z)$ and $g_t^*(z)$ are the mass of customers matched with suppliers of productivity z at the beginning of the period and during production respectively. $\delta_t^*(z)$ is the fraction of firms of type z the planner chooses to exit and $\delta_t^{E*}(z)$ the fraction not to enter the economy.

We can make some substitutions to simplify the problem

$$\max_{\{C_t^*, L_t^*, K_{t+1}^*, \{i_t^*(z), g_t^{i*}(z), g_t^*(z), \delta_t^*(z)\}_{z=0}^\infty\}_{t=0}^\infty} \sum_{t=0}^\infty \beta^t \{u(C_t^*) - v(L_t^*)\} \tag{A.27}$$

$$\text{s.t. } C_t^* + K_{t+1}^* + \int_0^\infty s(i_t^*(z)) g_t^{i*}(z) dz \le Y(K_t^*, L_t^*, \{g_t^*(z)\}) + (1-\delta_k) K_t^* \ \forall t \tag{A.28}$$

$$Y(K_t^*, L_t^*, \{g_t^*(z)\}) = \left(1 + \int_0^\infty \left(z^{\sigma-1} - 1\right) g_t^*(z) dz\right)^{\frac{1}{\sigma-1}} (K_t^*)^\alpha (L_t^*)^{1-\alpha} \tag{A.29}$$

$$g_t^*(z) \le (1 - q_t^*(z) + e_t^*(z) i_t^*(z)) g_t^{i*}(z) \ \forall t, z \tag{A.30}$$

$$\begin{aligned} g_t^{i*}(z) \le & \int_0^\infty g_{t-1}^*(\hat{z}) \left[(1 - \delta_t^*(z)) f(z|\hat{z}) - \chi(1 - \delta_t^{E*}(z)) f_0(z)\right] d\hat{z} \\ & + \chi(1 - \delta_t^{E*}(z)) f_0(z) \ \forall t, z \end{aligned} \tag{A.31}$$

$$\delta_t^*(z) \ge \delta_F, \quad \delta_t^*(z) \le 1 \ \forall t, z \tag{A.32}$$

$$\delta_t^{E*}(z) \ge 0, \quad \delta_t^{E*}(z) \le 1 \ \forall t, z \tag{A.33}$$

$$e_t^*(z) = \left[1 - \int_z^\infty g_t^{i*}(\hat{z}) d\hat{z}\right] e^{-\int_z^\infty i_t^*(\hat{z}) g_t^{i*}(\hat{z}) d\hat{z}} \ \forall t, z \tag{A.34}$$

$$q_t^*(z) = 1 - e^{-\int_z^\infty i_t^*(\hat{z}) g_t^{i*}(\hat{z}) d\hat{z}} \ \forall t, z \tag{A.35}$$

for given K_0^* and $\{g_{-1}^*(z)\}_{z=0}^\infty$. Denote as λ_t^{BC} the multiplier on the budget constraint in (A.28), λ_{zt}^i the multiplier on the law of motion for $g^{i*}(z)$ in (A.31), λ_{zt}^ρ the multiplier on the growth rate for $g^*(z)$ in (A.30), and $\lambda_t^{\delta L}$, $\lambda_t^{\delta H}$,λ_t^{EL}, λ_t^{EH} the multipliers on the boundaries for $\delta_t^*(z)$ and $\delta_t^{E*}(z)$ in (A.32) and (A.33). To ease notation, we write the transition probabilities of z including entry and exit as $\hat{f}_{t+1}(z'|z) = (1 - \delta_{t+1}^*(z')) f(z'|z)$ and $\hat{f}_{t+1}^0(z) = (1 - \delta_{t+1}^{E*}(z')) f_0(z')$.

It is useful to first take some relevant derivatives of the auxiliary equations in (A.29), (A.34), and (A.35) that we will use later. These are given as

$$\begin{aligned} \frac{\partial Y_t}{\partial g_t^*(z)} &= \frac{z^{\sigma-1} - 1}{\sigma - 1} \left(1 + \int_0^\infty \left(z^{\sigma-1} - 1\right) g_t^*(z) dz\right)^{\frac{2-\sigma}{\sigma-1}} (K_t^*)^\alpha (L_t^*)^{1-\alpha} \\ &= \frac{z^{\sigma-1} - 1}{\sigma - 1} (Z_t^*)^{2-\sigma} (K_t^*)^\alpha (L_t^*)^{1-\alpha} \\ \frac{\partial e_t^*(z)}{\partial g_t^{i*}(z')} &= -e^{-\int_z^\infty i_t^*(\hat{z}) g_t^{i*}(\hat{z}) d\hat{z}} - i_t^*(z') \left[1 - \int_z^\infty g_t^{i*}(\hat{z}) d\hat{z}\right] e^{-\int_z^\infty i_t^*(\hat{z}) g_t^{i*}(\hat{z}) d\hat{z}} \qquad \text{iff } z' \ge z, \text{ else } 0 \\ \frac{\partial q_t^*(z)}{\partial g_t^{i*}(z')} &= i_t^*(z') e^{-\int_z^\infty i_t^*(\hat{z}) g_t^{i*}(\hat{z}) d\hat{z}} \qquad \text{iff } z' \ge z, \text{ else } 0 \\ \frac{\partial e_t^*(z)}{\partial i_t^*(z')} &= -g_t^{i*}(z') \left[1 - \int_z^\infty g_t^{i*}(\hat{z}) d\hat{z}\right] e^{-\int_z^\infty i_t^*(\hat{z}) g_t^{i*}(\hat{z}) d\hat{z}} \qquad \text{iff } z' \ge z, \text{ else } 0 \\ \frac{\partial q_t^*(z)}{\partial i_t^*(z')} &= g_t^{i*}(z') e^{-\int_z^\infty i_t^*(\hat{z}) g_t^{i*}(\hat{z}) d\hat{z}} \qquad \text{iff } z' \ge z, \text{ else } 0 \end{aligned}$$

The first order condition with respect to consumption, capital, and labor yield the familiar optimality conditions

$$u'(C_t^*) = \beta u'(C_{t+1}^*)\left[(1-\delta_k) + \frac{\partial Y_{t+1}}{\partial K_{t+1}^*}\right] \tag{A.36}$$

$$u'(C_t^*)\frac{\partial Y_t}{\partial L_t^*} = v'(L_t^*) \tag{A.37}$$

The first order conditions with respect to $g_t^{i*}(z)$, $g_t^*(z)$, and $i_t^*(z)$ are given by

$$\begin{aligned} 0 =& -\lambda_t^{BC} s(i_t^*(z)) + \lambda_{z,t}^{\rho}(1 - q_t^*(z) + e_t^*(z)i_t^*(z)) - \lambda_{z,t}^i \\ &+ \int_0^\infty \lambda_{\hat{z},t}^{\rho} g_t^{i*}(\hat{z})\left[-\frac{\partial q_t^*(\hat{z})}{\partial g_t^{i*}(z)} + \frac{\partial e_t^*(\hat{z})}{\partial g_t^{i*}(z)} i_t^*(\hat{z})\right] d\hat{z} \end{aligned} \tag{A.38}$$

$$0 = \lambda_t^{BC}\frac{\partial Y_t}{\partial g_t^*(z)} - \lambda_{z,t}^{\rho} + \int_0^\infty \lambda_{\hat{z},t+1}^i \left[\hat{f}_{t+1}(\hat{z}|z) - \chi \hat{f}_{t+1}^0(\hat{z})\right] d\hat{z} \tag{A.39}$$

$$\begin{aligned} 0 =& -\lambda_t^{BC} s'(i_t^*(z)) g_t^{i*}(z) + \lambda_{z,t}^{\rho} e_t^*(z) g_t^{i*}(z) \\ &+ \int_0^\infty \lambda_{\hat{z},t}^{\rho} g_t^{i*}(\hat{z})\left[-\frac{\partial q_t^*(\hat{z})}{\partial i_t(z)} + \frac{\partial e_t^*(\hat{z})}{\partial i_t(z)} i_t^*(\hat{z})\right] d\hat{z} \end{aligned} \tag{A.40}$$

Note that while the integrals are over all possible productivities of competitors $\hat{z}$, the derivatives inside the integrals are only non-zero for $\hat{z} \leq z$ as derived above. The notation here is equaivalent to having the integral in the first and third line go only over $\hat{z} \leq z$.

Define the value of a match between a customer and a supplier with productivity z at the beginning of period, measured in units of resources, as $V_t^*(z) \equiv \frac{\lambda_{z,t}^i}{\lambda_t^{BC}}$ and the value of such a match at the production stage accordingly as $W_t^*(z) = \frac{\lambda_{z,t}^{\rho}}{\lambda_t^{BC}}$. From the first order conditions, the two values are given as

$$\begin{aligned} V_t^*(z) =& -s(i_t^*(z)) + (1 - q_t^*(z) + e_t^*(z)i_t^*(z))W_t^*(z) \\ &+ \int_0^\infty W_t^*(\hat{z}) g_t^{i*}(\hat{z})\left[-\frac{\partial q_t^*(\hat{z})}{\partial g_t^{i*}(z)} + \frac{\partial e_t^*(\hat{z})}{\partial g_t^{i*}(z)} i_t^*(\hat{z})\right] \end{aligned} \tag{A.41}$$

$$W_t^*(z) = \frac{\partial Y_t}{\partial g_t^*(z)} + \beta\frac{u'(C_{t+1}^*)}{u'(C_t^*)}\int_0^\infty V_{t+1}^*(\hat{z})\left[\hat{f}_{t+1}(\hat{z}|z) - \chi \hat{f}_{t+1}^0(\hat{z})\right] d\hat{z} \tag{A.42}$$

where we have made use of the FOCs for consumption and capital.

The optimal choice of investment for the planner solves

$$s'(i_t^*(z)) = e_t^*(z)W_t^*(z) + \int_0^\infty W_t^*(\hat{z})\frac{g_t^{i*}(\hat{z})}{g_t^{i*}(z)}\left[-\frac{\partial q_t^*(\hat{z})}{\partial i_t(z)} + \frac{\partial e_t^*(\hat{z})}{\partial i_t(z)} i_t^*(\hat{z})\right] d\hat{z} \tag{A.43}$$

Finally, we can derive the optimal choice of $\delta_t^*(z)$ and $\delta_t^{E*}(z)$ from the respective first order conditions and get

$$\text{(A.44)} \qquad \lambda_{zt}^i \underbrace{\left[\int_0^\infty g_t^*(\hat{z}) f(z|\hat{z}) d\hat{z}\right]}_{>0} = \lambda_t^{\delta L} - \lambda_t^{\delta H}$$

$$\text{(A.45)} \qquad \lambda_{zt}^i \underbrace{\chi f_0(z)\left[1 - \int_0^\infty g_t^*(\hat{z}) d\hat{z}\right]}_{>0} = \lambda_t^{EL} - \lambda_t^{EH}$$

As $\lambda_t^{BC} > 0$ due to the binding budget constraint, $\lambda_{zt}^i > 0$ iff $V_t^*(z) > 0$. Hence iff $V_t^*(z) > 0$ we get that $\lambda_t^{\delta L} > 0$ and $\lambda_t^{EL} > 0$ such that $\delta_t^*(z) = \delta_F$ and $\delta_t^{E*}(z) = 0$ is optimal. If $V_t^*(z) < 0$, then $\lambda_t^{\delta H} > 0$ and $\lambda_t^{EH} > 0$ and hence $\delta_t^*(z) = \delta_t^{E*}(z) = 1$ and the planner chooses to dissolve the match.

A.3 Markup Distortions Given a Distribution of Customers

This appendix highlights the distortions from the aggregate markup and dispersion in markups in the decentralized equilibrium. Appendix A.2 shows that the FOCs for the planner's choice of labor and capital hold as

$$v'(L_t) = (1-\alpha) Z_t^M K_t^\alpha L_t^{-\alpha} u'(C_t),$$
$$u'(C_{t-1}) = \beta\left(\alpha Z_t^M K_t^{\alpha-1} L_t^{1-\alpha} + 1 - \delta_K\right) u'(C_t).$$

The aggregate markup for the decentralized equilibrium $\mathcal{M}_t$ is derived in Appendix A.1. Note that $\mathcal{M}_t = \frac{(1-\alpha)Y_t}{w_t L_t} \Rightarrow w_t = \frac{1}{\mathcal{M}_t}(1-\alpha) Z_t^M K_t^\alpha L_t^{-\alpha}$ and $\mathcal{M}_t = \frac{\alpha Y_t}{(r_t+\delta_K)K_t} \Rightarrow r_t + \delta_K = \frac{1}{\mathcal{M}_t}\alpha Z_t^M K_t^{\alpha-1} L_t^{1-\alpha}$. Substituting into the FOCs for labor and capital in (5) and (6) yields

$$v'(L_t) = w_t u'(C_t) = \frac{1}{\mathcal{M}_t}(1-\alpha) Z_t^M K_t^\alpha L_t^{-\alpha} u'(C_t),$$
$$u'(C_{t-1}) = \beta\left(1 + r_t\right) = \beta\left(\frac{1}{\mathcal{M}_t}\alpha Z_t^M K_t^{\alpha-1} L_t^{1-\alpha} + 1 - \delta_K\right) u'(C_t).$$

Thus, the presence of an aggregate markup distorts the aggregate quantities of inputs into production in the decentralized equilibrium. The aggregate markup acts like a uniform tax on production, distorting labor and capital supply to below their efficient level.

Aggregate productivity in the decentralized equilibrium Z_t^M for a given distribution of customers $G_t(z)$ is derived in Appendix A.1 as

$$Z_t^M = \left(u_t \left(\frac{1}{\mathcal{M}_t} \right)^{-\sigma} + \int_0^\infty z^{\sigma-1} \left(\frac{\mu_t(z)}{\mathcal{M}_t} \right)^{-\sigma} dG_t(z) \right)^{\frac{1}{\sigma-1}},$$

Similarly we derive aggregate productivity under the planner allocation Z_t^P for the same distribution of customers $G_t(z)$ in Appendix A.2 as

$$Z_t^P = \left(u_t + \int_0^\infty z^{\sigma-1} dG_t(z) \right)^{\frac{1}{\sigma-1}}.$$

Comparing the two expressions, dispersion in markups distorts the allocation of demand at the intensive margin of demand—sales per customer. At the intensive margin relative price differences matter for quantities purchased, so dispersion in markups misallocates sales and lowers aggregate productivity.

A.4 Decomposing Inefficient Demand Investment

We decompose the planner's value of a match to isolate the forces that drive a wedge into the optimality condition for optimal investment of the planner vs. suppliers in equilibrium. We are applying this decomposition in steady state and will hence omit time subscripts. We first introduce some useful notation. Define the following:

$$\begin{aligned}
\rho^*(z) &\equiv 1 - q^*(z) + e^*(z) i^*(z) \\
b^*(z) &\equiv \int_0^\infty W^*(\hat{z}) \frac{g^{i*}(\hat{z})}{g^{i*}(z)} \left[-\frac{\partial q^*(\hat{z})}{\partial i(z)} + \frac{\partial e^*(\hat{z})}{\partial i(z)} i^*(\hat{z}) \right] \\
B^*(z) &\equiv \int_0^\infty W^*(\hat{z}) g^{i*}(\hat{z}) \left[-\frac{\partial q(\hat{z})}{\partial g^{i*}(z)} + \frac{\partial e(\hat{z})}{\partial g^{i*}(z)} i(\hat{z}) \right] \\
X^*(z) &\equiv -\chi \int_0^\infty \hat{f}^0(z') V^*(z') dz'
\end{aligned}$$

With this notation, define the following recursions

$$\begin{aligned}
V_P^*(z) &= -s(i^*(z)) + \rho^*(z) \left[\frac{\partial Y}{\partial g^*(z)} + \beta \int_0^\infty \hat{f}(z'|z) V_P^*(z') dz' \right] \\
V_B^*(z) &= B^*(z) + \rho^*(z) \beta \int_0^\infty \hat{f}(z'|z) V_B^*(z') dz' \\
V_X^*(z) &= \rho^*(z) \beta \left[X^*(z) + \int_0^\infty \hat{f}(z'|z) V_X^*(z') dz' \right]
\end{aligned}$$

These three terms represent, for any customer matched with a firm of productivity z at the beginning of a period before rematching, the expected present discounted value of the private joint surplus of the match to the supplier and the customer directly involved and in all future matches related to this match ($V_P^*(z)$), business stealing from active suppliers ($V_B^*(z)$) and crowding of entry ($V_X^*(z)$). We can then write the value of a customer that is matched to a firm with productivity z before rematching as

$$V^*(z) = V_P^*(z) + V_B^*(z) + V_X^*(z)$$

To see this, note that

$$\begin{aligned}
V^*(z) =& V_P^*(z) + V_B^*(z) + V_X^*(z) \\
=& -s(i^*(z)) + \rho^*(z)\left[\frac{\partial Y}{\partial g^*(z)} + \beta\int_0^\infty \hat{f}(z'|z)V_P^*(z')dz'\right] \\
&+ B^*(z) + \rho^*(z)\beta\int_0^\infty \hat{f}(z'|z)V_B^*(z')dz' \\
&+ \rho^*(z)\beta\left[X^*(z) + \int_0^\infty \hat{f}(z'|z)V_X^*(z')dz'\right] \\
=& -s(i^*(z)) + \rho^*(z)\frac{\partial Y}{\partial g^*(z)} + B^*(z) + \rho^*(z)\beta X^*(z) \\
&+ \rho^*(z)\beta\int_0^\infty \hat{f}(z'|z)\left[V_P^*(z') + V_B^*(z') + V_X^*(z')\right]dz' \\
=& -s(i^*(z)) - \int_0^\infty W^*(\hat{z})g^{i*}(\hat{z})\left[\frac{\partial q(\hat{z})}{\partial g^{i*}(z)} + \frac{\partial e(\hat{z})}{\partial g^{i*}(z)}i(\hat{z})\right]d\hat{z} + \rho^*(z)\frac{\partial Y}{\partial g^*(z)} \\
&+ \rho^*(z)\beta\left[\int_0^\infty \hat{f}(z'|z)V^*(z')dz' - \chi\int_0^\infty \hat{f}^0(z')V^*(z')dz'\right] \\
=& -s(i^*(z)) + \int_0^\infty W^*(\hat{z})g^{i*}(\hat{z})\left[-\frac{\partial q(\hat{z})}{\partial g^{i*}(z)} + \frac{\partial e(\hat{z})}{\partial g^{i*}(z)}i(\hat{z})\right] \\
&+ \rho^*(z)\left[\frac{\partial Y}{\partial g^*(z)} + \frac{1}{1+r^*}\int_0^\infty(\hat{f}(z'|z) - \chi\hat{f}^0(z'))V^*(z')dz'\right] \\
=& -s(i^*(z)) + \rho^*(z)W^*(z) + \int_0^\infty W^*(\hat{z})g^{i*}(\hat{z})\left[-\frac{\partial q(\hat{z})}{\partial g^{i*}(z)} + \frac{\partial e(\hat{z})}{\partial g^{i*}(z)}i(\hat{z})\right]
\end{aligned}$$

We can similarly decompose the value of the a match at the production stage as

$$W^*(z) = W_P^*(z) + W_B^*(z)W_X^*(z)$$

where

$$
\begin{aligned}
W_P^*(z) &= \frac{\partial Y}{\partial g^*(z)} + \beta \int_0^\infty \hat{f}(z'|z) V_P^*(z') dz' \\
W_B^*(z) &= \beta \int_0^\infty \hat{f}(z'|z) V_B^*(z') dz' \\
W_X^*(z) &= \beta \left[X^*(z) + \int_0^\infty \hat{f}(z'|z) V_X^*(z') dz' \right]
\end{aligned}
$$

To see that this decompostion holds note that

$$
\begin{aligned}
W^*(z) =& W_P^*(z) + W_B^*(z) W_X^*(z) \\
=& \frac{\partial Y}{\partial g^*(z)} + \beta X^*(z) + \beta \int_0^\infty \hat{f}(z'|z) \left[V_P^*(z') + V_B^*(z') + V_X^*(z') \right] dz' \\
=& \frac{\partial Y}{\partial g^*(z)} + \beta \int_0^\infty \left(\hat{f}(z'|z) - \chi \hat{f}_0(z') \right) V^*(z') dz'
\end{aligned}
$$

Applying the decomposition of $W^*(z)$ into its three components, we can rewrite the optimal investment condition as

$$
s'(i^*(z)) = e^*(z) \left[W_P^*(z) + W_B^*(z) + W_X^*(z) \right] + b^*(z)
$$

We can then quantify the contribution of five margins to aggregate differences between the planner's optimal demand investment and equilibrium investment:

(i) *Private benefit*: the difference between the private benefit of the match to both the customer and supplier directly involved and in all future matches related to this match ($W_P^*(z)$) relative to the private benefit to the supplier only in equilibrium ($W(p(z), z)$).

(ii) *Dynamic business-stealing*: business stealing from active suppliers due to the influence of existing matches on the rematching process ($W_B^*(z)$).

(iii) *Crowding out of entry*: crowding out of potential entry due to a reduction in the number of unmatched customers ($W_X^*(z)$).

(iv) *Static business-stealing*: business stealing through the effect of investment on the matching efficiency for other suppliers ($b^*(z)$).

(v) *Distribution effects*: differences in the distribution of customers $g^{i*}(z)$ that lead to differences in the conversion rate $e^*(z)$ and different weights when aggregating total investment in the economy.

We compute the contribution of each margin through a Shapley–Owen decomposition between the planner's terms and their equilibrium counterparts (where (ii), (iii), and (iv) are zero in equilibrium). Let $\mathcal{N} \equiv \{1, \ldots, 5\}$ index the five margins. For any subset $S \subseteq \mathcal{N}$, let $\Gamma(S)$ denote aggregate demand investment under the counterfactual in which margins in S are set to their planner values and margins in $\mathcal{N} \setminus S$ are set to their equilibrium values. For each such counterfactual, contact intensity $i_S(z)$ is recovered by inverting the demand investment FOC,

$$s'\big(i_S(z)\big) = e_S(z)\left[W_{P,S}(z) + W_{B,S}(z) + W_{X,S}(z)\right] + b_S(z),$$

and aggregate demand investment is obtained by integrating implied spending with the corresponding distribution of matched customers,

$$\Gamma(S) \equiv \int_0^\infty g_S(z)\, s\big(i_S(z)\big)\, dz.$$

This notation nests the equilibrium and planner outcomes: $\Gamma(\varnothing) = \Gamma^{eq}$ and $\Gamma(\mathcal{N}) = \Gamma^*$.

For $k \in \mathcal{N}$, define its Shapley–Owen contribution as the average incremental effect of switching margin k from equilibrium to planner across all orderings,

$$\phi_k \equiv \frac{1}{|\mathcal{N}|!} \sum_\pi \left[\Gamma\big(S_{\pi,k} \cup \{k\}\big) - \Gamma\big(S_{\pi,k}\big)\right], \qquad S_{\pi,k} \equiv \{j \in \mathcal{N} : \pi(j) < \pi(k)\},$$

where the sum is over all $|\mathcal{N}|! = 5! = 120$ permutations π of the margins. By construction, the contributions add up:

$$\Gamma(\mathcal{N}) - \Gamma(\varnothing) = \sum_{k \in \mathcal{N}} \phi_k.$$

B Data Appendix

B.1 Capital IQ

Measure of demand investment. Following He et al. (2024), we measure firms' demand investment using their augmented sales-and-marketing (S&M) measure constructed from Capital IQ line items. As in He et al. (2024), this augmented measure is constructed for a baseline sample of publicly listed firms from Compustat that is matched to Capital IQ identifiers; Capital IQ contributes the underlying line items, and missing values are supplemented using information extracted from firms' SEC filings. The resulting moments pertain to the Compustat-based public-firm universe (rather than the full set of Capital IQ-covered firms).

Aggregate demand investment. We denote the firm-level measure from He et al. (2024) by SM_f and interpret it as firm f's demand investment. The aggregate ratio of SM_f to revenue Rev_f equals $\frac{\sum_f SM_f}{\sum_f Rev_f} = 0.045$. 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 for comparison to the model analogue.We use an aggregate revenue-to-value-added ratio of $\frac{\sum_f Rev_f}{\sum_f VA_f} = 1.80$. Combining these two ratios, we obtain the aggregate demand-investment share of value added in the data as:

$$\frac{\sum_f SM_f}{\sum_f VA_f} = \frac{\sum_f SM_f}{\sum_f Rev_f} \frac{\sum_f Rev_f}{\sum_f VA_f} = 0.045 \times 1.80 \approx 0.08.$$

We compare this value to the model analogue: aggregate demand investment over value added in the intermediate-good sector, computed in the filtered model sample to mimic selection into empirical sample (Appendix B.2).

Demand investment and markups. Table A1 displays regressions of the log demand-investment intensity, $\ln\left(\frac{SM_f}{Rev_f}\right)$, on firms' log markup, $\ln\left(\frac{\text{Rev}_f}{\text{COGS}_f}\right)$. Our preferred specification is column 4, which includes year and 2-digit NAICS fixed effects (consistent with the model's lack of aggregate shocks and industry heterogeneity) and weights observations by COGS. Weighting by COGS reduces noise as evidenced by a higher $R^2 = 0.309$ compared to the unweighted regression with the same fixed effects in column 3 ($R^2 = 0.162$). The point estimate of 1.843 implies that a firm with a 1% higher markup has, on average, about a 1.84% higher demand-investment-to-revenue ratio, relative to the industry average in that year. Columns 1 and 2 report results without fixed effects, while columns 5 and 6 add controls for firm size and age. Across specifications high-markup firms consistently invest more in their demand.

Table A1: Cross-sectional Relation of Demand Investment and Markups

	(1)	(2)	(3)	(4)	(5)	(6)
$\ln \frac{\text{Rev}_f}{\text{COGS}_f}$	1.295***	2.195***	0.880***	1.843***	1.011***	1.720***
	(0.256)	(0.260)	(0.193)	(0.379)	(0.143)	(0.256)
R^2	0.095	0.176	0.162	0.309	0.300	0.339
Fixed Effects	—	—	Year+Ind	Year+Ind	Year+Ind	Year+Ind
Weights	—	Cost	—	Cost	—	Cost
Controls	—	—	—	—	Age+Size	Age+Size

Notes:: Data from He et al. (2024), Capital IQ (Compustat-matched public firms). The dependent variable in all regressions is the log of the augmented S&M measure relative to revenue, $\ln \frac{SM_f}{Rev_f}$. 30,674 observations. Robust standard errors clustered at 2-digit NAICS level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

B.2 Model Filter to Mimic Selection into Capital IQ

Table A2 shows how model moments vary based on imposing the filter that mimics selection into the publicly-listed firm sample that underlies our Capital IQ-based empirical moments. We filter intermediate good firms (suppliers) along two dimensions, age and size. For age, we follow Ottonello and Winberry (2020) and restrict the sample 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.

Given our estimated turnover in the productivity process, the important restriction on the model-based sample is firm size, not firm age. In addition, those intermediate inputs that are produced in-house are produced competitively. This is why the economy-wide markup and demand-investment share of value added are lower in the aggregate model economy compared to the respective moments in the sample of all intermediate good firms.

Table A2: Impact of Filtering on Model Moments

Sample	All firms	Intermediate-good firms only		
Filter	none	none	age	age+size
Elasticity of $\frac{\text{demand investment}}{\text{revenue}}$ to markup	—	2.084	2.079	**1.849**
Aggregate markup	1.197	1.236	1.236	**1.250**
Aggregate demand investment over value added	0.092	0.109	0.109	**0.113**

Notes: Model moments used for calibration or validation in bold.

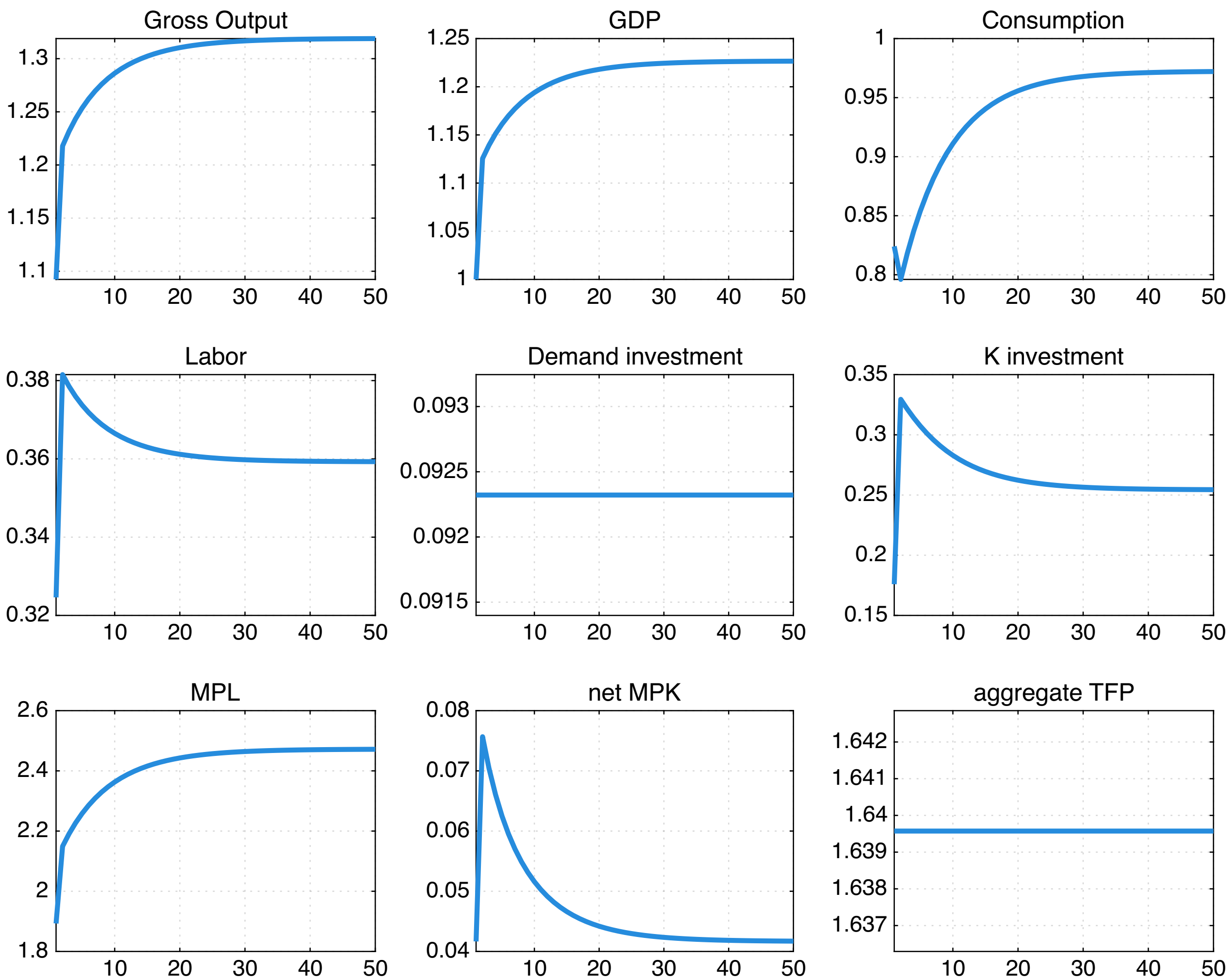
B.3 BDS and SUSB

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

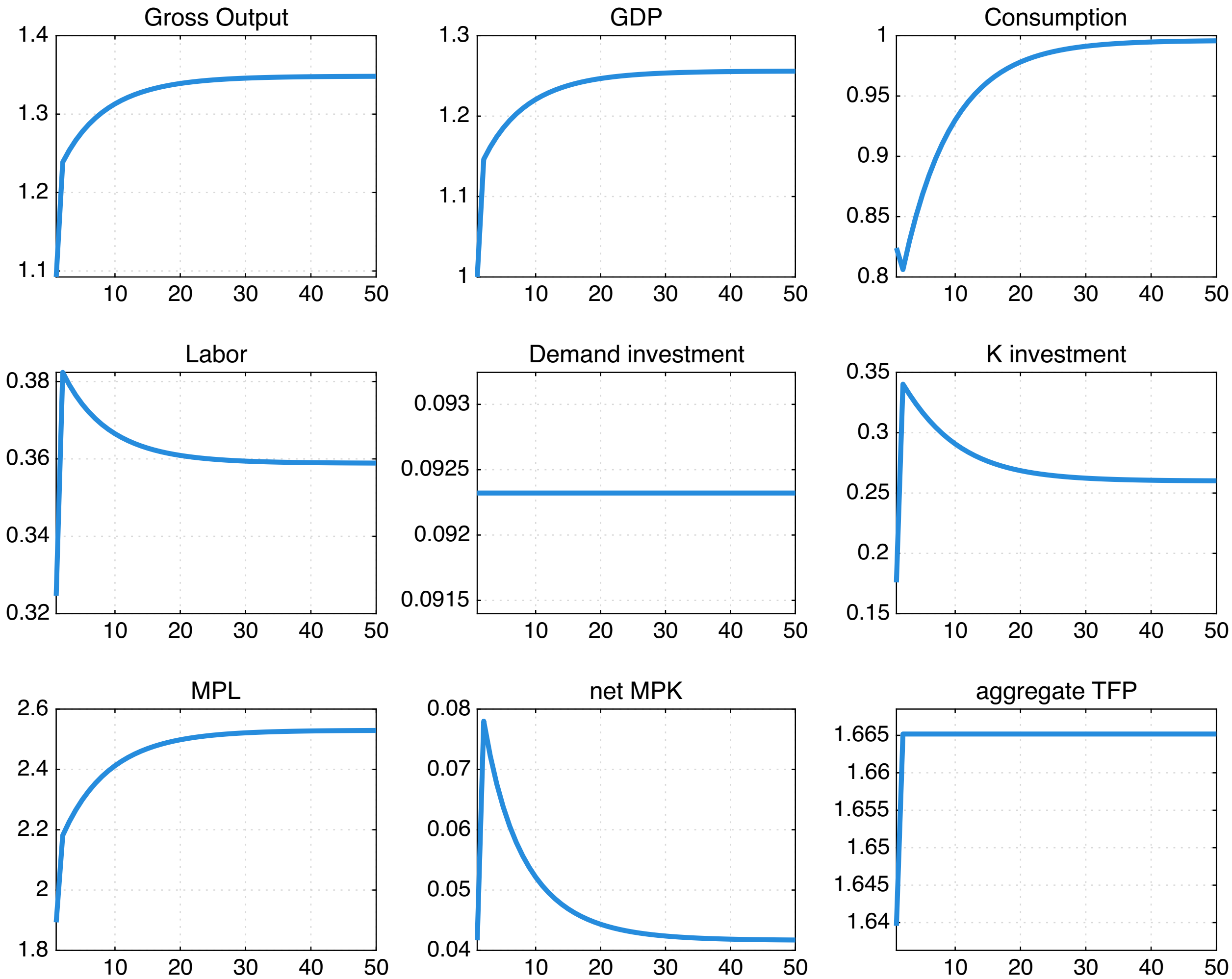
C.1 Social Planner Transition Graphs

Figure A1: Transition: Planner Remedies Only Aggregate Markup



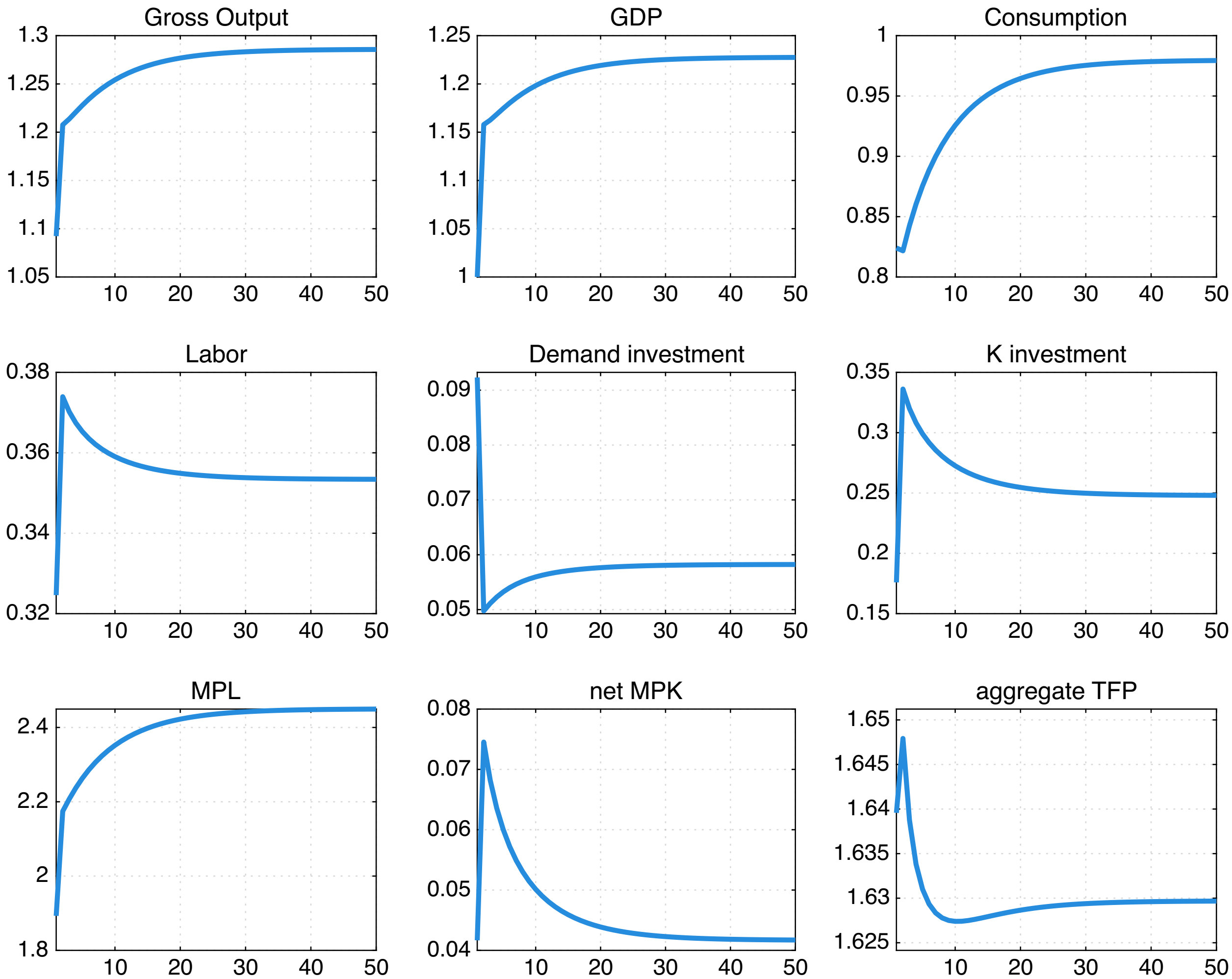
Notes: The planner removes the aggregate markup wedge while holding fixed the equilibrium objects governing demand reallocation: demand investment Γ^M and the implied customer distribution G (extensive margin), and aggregate TFP Z^M (intensive margin). Dispersion in markups across suppliers is therefore not corrected. The planner chooses time-varying (C_t, K_{t+1}, 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, K_{t+1}, L_t) are already planner-chosen. Quantities normalized so that GDP = 1 in the decentralized equilibrium.

Figure A2: Transition: Planner Remedies Aggregate Markup and Markup Dispersion



Notes: The planner removes both the aggregate markup wedge and *dispersion* in markups (intensive-margin misallocation). Selling effort 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, K_{t+1}, 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, K_{t+1}, L_t) are already planner-chosen. Quantities normalized so that GDP = 1 in the decentralized equilibrium.

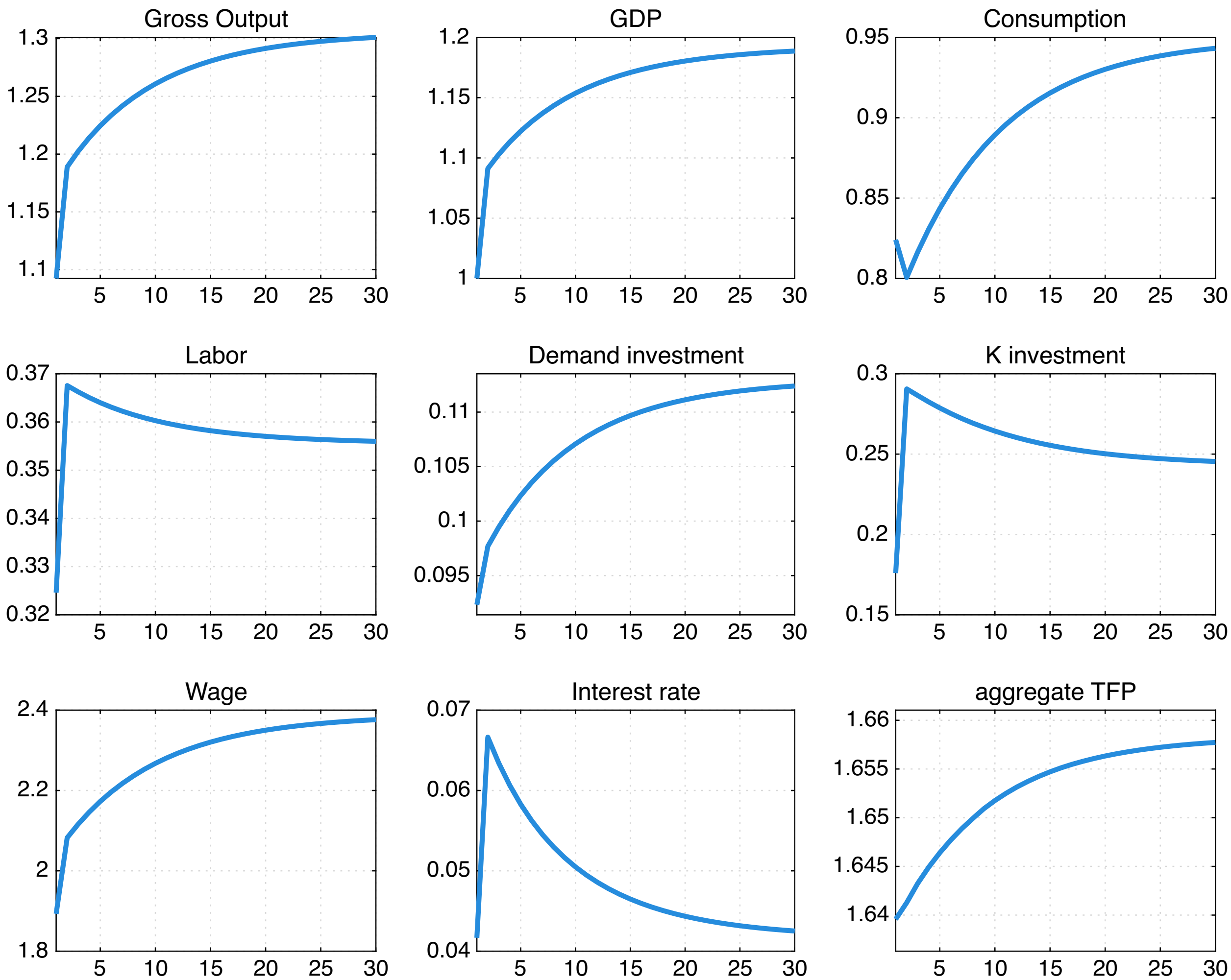
Figure A3: Transition: Full Social Planner (Markup, Dispersion, and Demand Investment)



Notes: The planner removes the aggregate markup and its dispersion *and* chooses contact intensities optimally. Thus, $i_t(z)$, aggregate demand investment 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 $(C_t, K_{t+1}, L_t, i_t(z))$. Period t=1 shows the decentralized equilibrium; at t=2 the planner takes over with predetermined K, while $(C_t, K_{t+1}, L_t, i_t(z))$ are already planner-chosen. Quantities normalized so that $GDP = 1$ in the decentralized equilibrium.

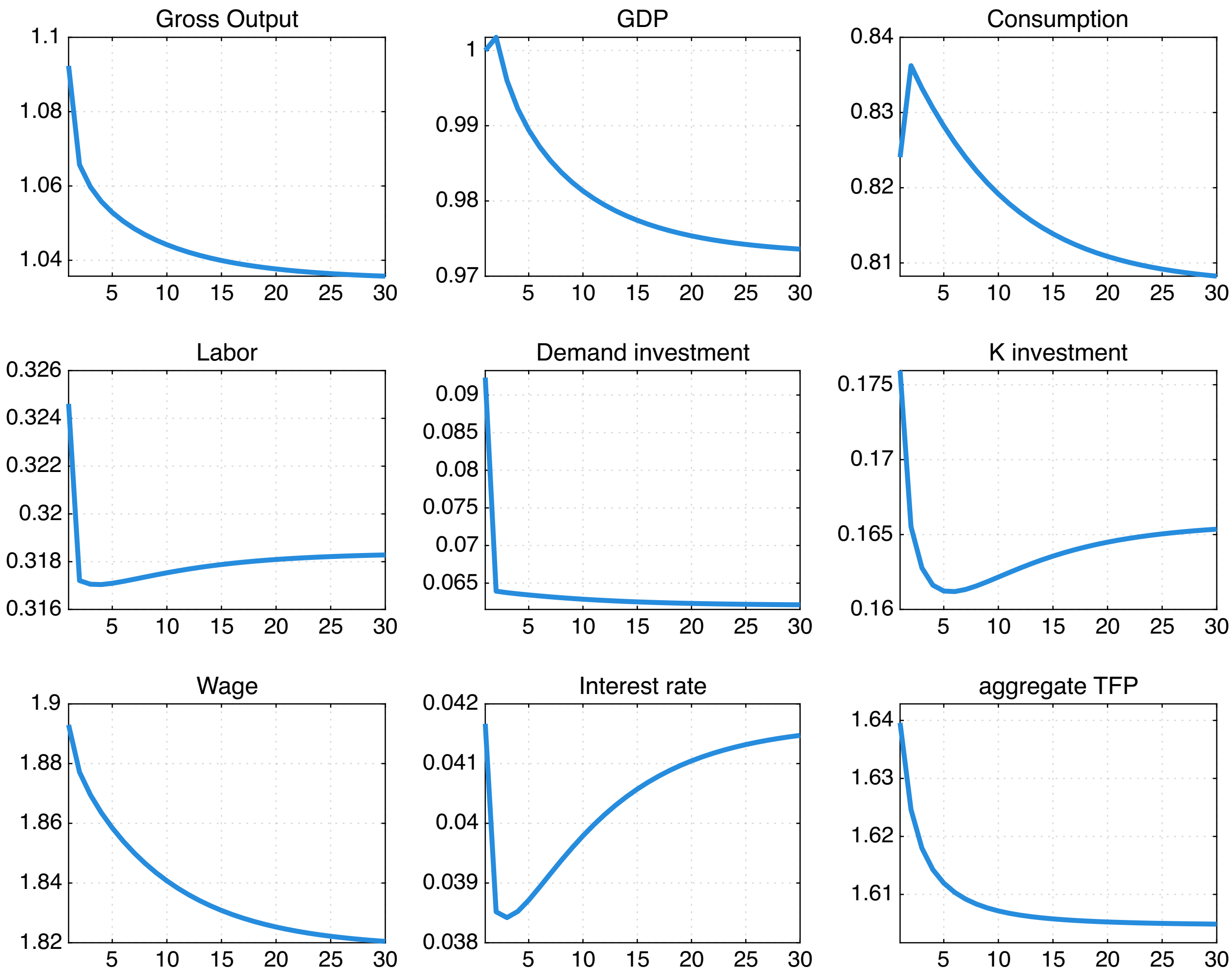
C.2 Optimal Taxation Transition Graphs

Figure A4: Transition: Optimal Production Subsidy Only



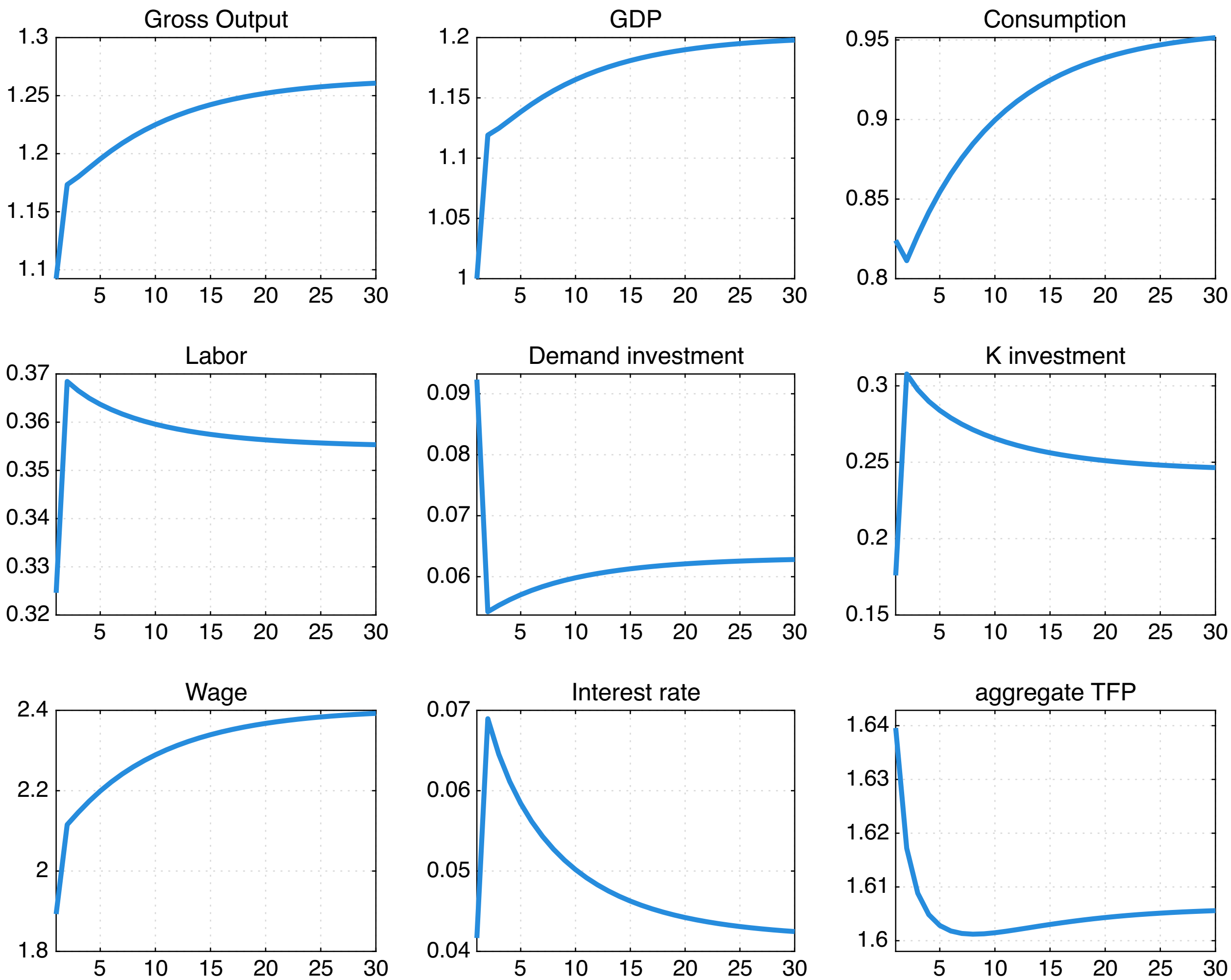
Notes: We solve for a (time-constant) production input subsidy of 13.3%, which maximizes welfare along the transition from the no-tax baseline steady state. The policy is financed by a lumpsum tax (16.7% of GDP in steady state). Period t=1 shows the no-tax equilibrium; from t=2 onwards the production 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 A5: Transition: Optimal Demand Investment Tax Only



Notes: We solve for a (time-constant) demand-investment tax of 33.8%, which maximizes welfare along the transition from the no-tax baseline steady state. The tax is rebated as a lumpsum transfer (2.1% of GDP in steady state). Period t=1 shows the no-tax equilibrium; from t=2 onwards the demand-investment 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.

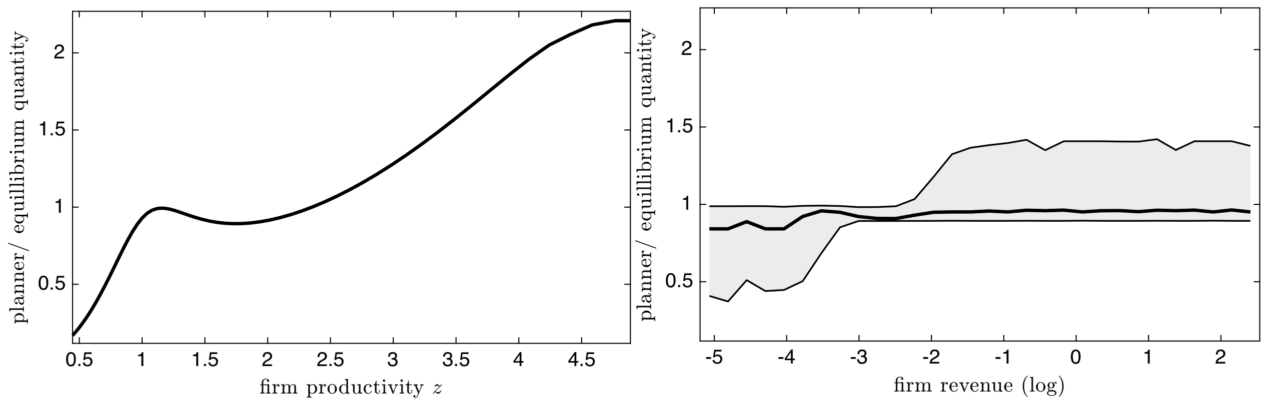
Figure A6: Transition: Optimal Demand Investment Tax and Production Subsidy



Notes: We solve for a (time-constant) demand-investment tax of 61.1% and production subsidy of 17.1%, which maximize welfare along the transition from the no-tax baseline steady state. The policy is financed by a lumpsum tax (17.8% of GDP in steady state). Period t=1 shows the no-tax equilibrium; from t=2 onwards the demand-investment 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 A7: Size Distortions: Planner vs. Equilibrium by Productivity and Revenue

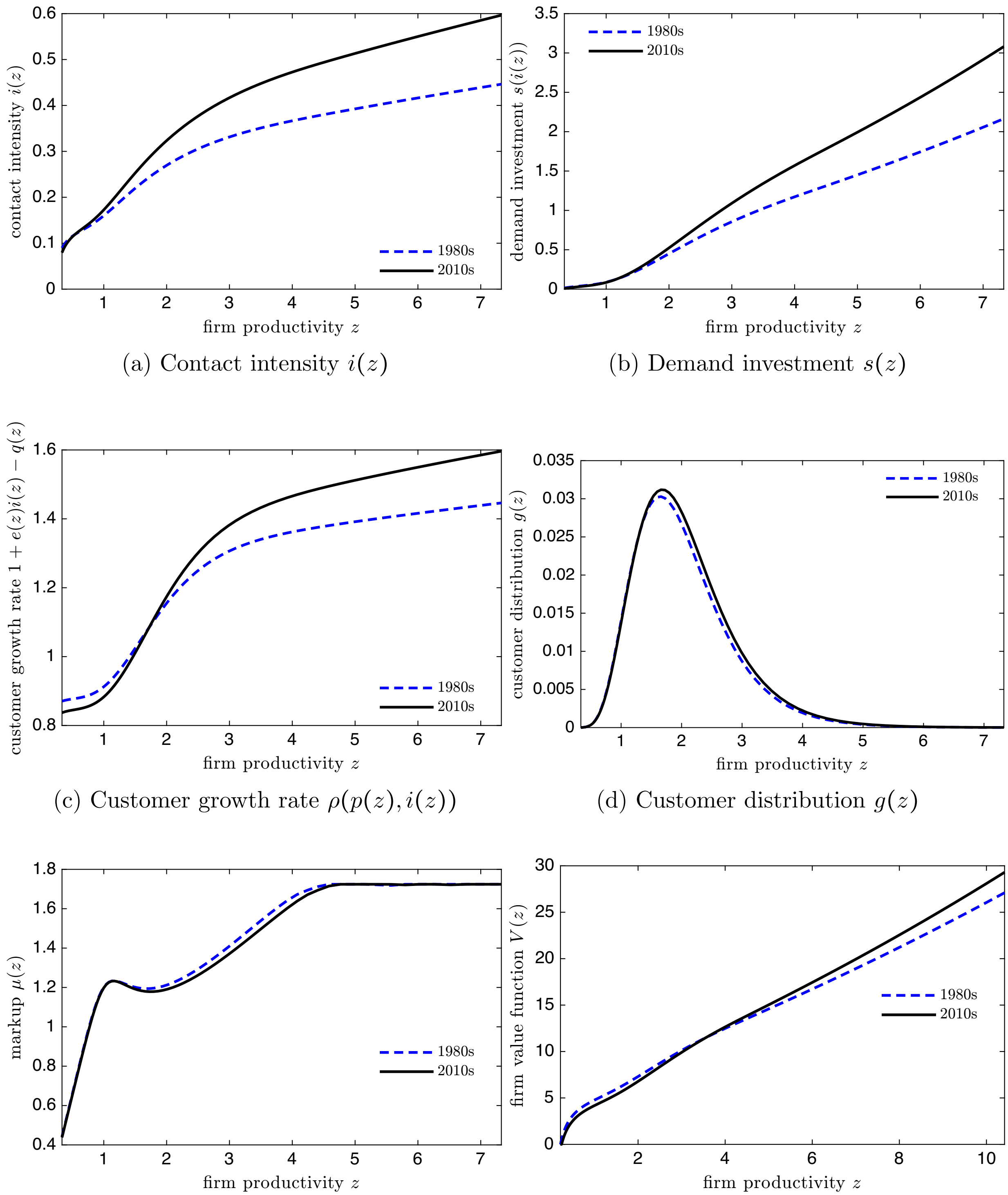


(a) By firm productivity z

(b) By firm revenue

Notes: 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.

Figure A8: Effects of Contact-Cost Technology Change Across the Firm Distribution



(a) Contact intensity $i(z)$

(b) Demand investment $s(z)$

(c) Customer growth rate $\rho(p(z), i(z))$

(d) Customer distribution $g(z)$

(e) Markup $\mu(z)$

(f) Firm value $V(z)$

Notes: Each panel compares the 1980s and 2010s steady states implied by the calibrated change in the contact-cost technology $s(i)$.

D Model Robustness

This appendix combines robustness exercises that assess the sensitivity of our quantitative findings to alternative calibration targets. We first outline the alternative calibrations and the report how our main results change across quantifications.

D.1 Alternative Calibration Targets

We re-calibrate the baseline steady state while varying one of three data targets at a time. First, we vary the sales-weighted customer quit rate. Moving from static sales variation to persistent customer relationships is a key deviation of our framework from the previous literature on markups and misallocation. To gauge how quickly results change when we move towards static demand, we consider an alternative calibration with a higher quit rate which corresponds to less persistent customer relationships and higher churn.

Second, we vary the intensive-margin share of sales variation. Again, allowing for variation at both the intensive and extensive margin of demand is a key feature of the model economy. Our baseline target (26%) is based on business-to-business transaction data from Belgium (Bernard et al., 2022). Evidence from settings in which U.S. consumers are the customers points to a lower intensive-margin share of around 15% (Einav et al., 2021; Afrouzi et al., 2025). We consider these estimates as an alternative target. Additionally, we consider a higher intensive-margin share of 35%, moving towards an economy with more flexible demand at the intensive margin.

Third, we vary the level of the aggregate markup in both directions. As considerable uncertainty remains about the level of aggregate markups, we follow Edmond et al. (2022) and consider alternative targets of 1.15 and 1.35 in addition to our baseline of 1.25.

Table A3 reports the implied parameters and shows that each specification matches all targeted moments closely. The implied parameter adjustments follow intuitive comparative statics: a higher quit rate implies lower inferred contact-cost shifter $\bar{s}$; a higher intensive-margin share implies a higher intensive-margin elasticity of substitution σ; and a higher aggregate markup implies higher productivity dispersion σ_z, in line with our calibration strategy.

Table A3: Robustness: Calibration Parameters and Targeted Moments

	Baseline	Quitrate	Int. Margin		Agg. Markup	
Modified target:		↑ 20%	↓ 15%	↑ 35%	↓ 15%	↑ 35%
Parameters (9)						
Exogenous firm exit rate (δ_F)	0.081	0.085	0.080	0.082	0.080	0.082
Entrants per unmatched customer (χ)	0.078	0.236	0.058	0.113	0.080	0.099
Capital elasticity (α)	0.327	0.326	0.324	0.332	0.344	0.295
Log TFP: unconditional dispersion (σ_z)	0.387	0.464	0.383	0.398	0.279	0.519
Log TFP: auto-correlation (ρ_z)	0.361	0.350	0.222	0.415	0.354	0.509
Log TFP: mean (μ_z)	0.409	0.603	0.360	0.448	0.264	0.688
Intensive-margin elasticity (σ)	2.381	2.380	1.939	2.814	3.042	2.003
Contact cost: level ($\bar{s}$)	39.792	25.056	59.757	30.032	89.442	74.259
Contact cost: curvature (ψ)	2.895	3.074	3.223	2.671	4.000	2.936
Targeted moments (10)						
Firm entry/exit rate	0.081	0.086	0.080	0.082	0.080	0.082
Relative employment of entrants	0.526	0.552	0.521	0.533	0.524	0.534
Aggregate labor share	0.614	0.615	0.616	0.611	0.615	0.618
Aggregate markup	0.250	0.249	0.251	0.250	0.150	0.351
Elasticity of $\frac{\text{demand investment}}{\text{revenue}}$ to markup	1.849	1.838	1.848	1.851	1.848	1.852
Elasticity of markup to revenue	0.031	0.032	0.031	0.031	0.031	0.031
Sales variation: intensive-margin share	0.262	0.266	0.150	0.355	0.263	0.263
Customer quit rate (% sales)	0.152	0.189	0.152	0.153	0.153	0.154
Top 10% sales share	0.676	0.702	0.667	0.691	0.674	0.682
Top 1% sales share	0.431	0.441	0.431	0.430	0.430	0.432

D.2 Robustness: Planner Benchmark and Optimal Policy

D.2.1 Simple Shift in Contact Cost

This section reports a simple robustness check that shifts the contact-cost shifter $\bar{s}$ holding all other parameters fixed at their baseline values, to illustrate how changes in contact cost reshape competition for customers and the welfare cost of misallocation.

Table A4 reports the resulting welfare-loss decomposition (computed along the transition to the efficient allocation). Cutting $\bar{s}$ in half lowers the aggregate-markup component by 10.8% and the markup-dispersion component by 13.0% (relative to the baseline). Doubling $\bar{s}$ raises these components by 10.8% and 22.9%, respectively.

Table A4: Welfare Losses from Misallocation under Alternative Contact Cost

	Total	Aggr. Markup	Markup Dispersion	Customer Misallocation
Baseline calibration	5.14	1.94	2.23	0.97
$\bar{s} - 50\%$	4.66	1.73	1.94	0.99
$\bar{s} + 100\%$	5.72	2.15	2.74	0.82

Notes: The table reports welfare losses 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 shifter $\bar{s}$ of the contact-cost function is cut in half or doubled, respectively.

D.2.2 Alternative Calibrations

We study the results in Section 4 under the alternative calibrations outlined in Appendix D.1. Two patterns are worth highlighting. First, targeting different values for the customer quit rate and the intensive-margin share of sales variation has relatively little effect on misallocation. Second, targeting a higher aggregate markup has a larger impact on the inefficiency results. In that case, the calibration implies a higher dispersion of productivity (higher σ_z) to reconcile higher markups with observed cross-sectional patterns, and the model features less over-investment in demand: demand investment is more likely to meaningfully improve productivity when competitors are further away. To see this, note that in the limit case where productivity dispersion approaches zero and suppliers become identical, customer poaching is a pure zero-sum game from a social perspective as old and new matches exhibit similar productivity.

As a result, the welfare losses from the level and dispersion of markups increase, while the inefficiency from demand investment declines. This is also reflected in lower optimal demand-investment taxes in the reported policy results. Despite these quantitative shifts, the qualitative insights persist across calibrations, including the complementarity between the production subsidy τ_P and the demand-investment tax τ_S.

Table A5: Robustness: Planner Benchmark and Optimal Policy

	Baseline	Quitrate	Int. Margin Sh.		Agg. Markup	
Modified target:		↓ 20%	↓ 15%	↑ 35%	↓ 15%	↑ 35%
Social Planner: % Δ relative to equilibrium (steady state)						
Gross Output	17.72	20.70	15.83	19.87	11.44	28.10
GDP	22.77	25.84	21.03	24.73	14.24	33.16
Demand investment	-36.92	-27.49	-47.46	-28.14	-46.46	-12.51
Tangible Capital	40.91	46.34	37.52	44.45	24.71	64.41
Labor	8.86	9.76	8.16	9.51	5.62	13.54
Consumption	18.90	21.47	17.55	20.44	11.79	27.54
Aggregate TFP	-0.60	0.12	-0.92	-0.14	-0.35	1.15
Social Planner: % Δ welfare relative to equilibrium (incl. transition)						
Total	5.14	5.75	5.13	5.29	3.10	6.92
Decomposition:						
Aggregate markup	1.94	2.23	1.75	2.12	0.79	3.44
Markup dispersion	2.23	2.37	2.16	2.16	1.52	2.91
Demand investment	0.97	1.15	1.22	1.02	0.79	0.58
Optimal Taxation						
Demand investment tax only						
Optimal demand investment tax	33.82	28.96	47.27	26.55	59.57	7.82
Welfare gain (% Δ)	0.46	0.36	0.69	0.34	0.51	0.05
Production subsidy only						
Optimal production subsidy	13.26	14.50	12.38	14.00	8.66	19.28
Welfare gain (% Δ)	1.28	1.55	1.07	1.47	0.52	2.76
Demand investment tax + production subsidy						
Optimal demand investment tax	61.10	58.10	75.30	53.65	84.58	34.66
Optimal production subsidy	17.05	18.31	16.30	17.66	11.25	22.52
Welfare gain (% Δ)	2.45	2.68	2.47	2.52	1.36	3.48

D.3 Robustness: Macro Effects of Rising Demand Investment

Table A6 shows that the insights from the exercise in Section 5 are similar under the alternative calibrations in Appendix D.1. Across specifications, we infer a decline in average contact costs: holding contact intensities $i(z)$ fixed, average demand investment $s(i(z))$ falls by 25–46%. We also infer a decline in the convexity of the contact cost technology: the P75/P25 ratio of $s(i(z))$, holding fixed $i(z)$, falls by 6–13%. The aggregate implications are similar as well. GDP rises by 3.6–6.0% and aggregate TFP by 2.1–4.4%; the effects are strongest under the higher-markup calibration because it implies greater dispersion in firm productivity (higher σ_z), so a given amount of reallocation toward high-productivity firms translates into larger TFP and GDP gains. Finally, results for the aggregate markup and intangible capital intensity are similar across specifications: aggregate markups fall modestly, and true intangible intensity declines by roughly 10–14% across transitions, while capitalized intangible capital measures (with constant deprecation rates) increase by 6–13%.

Table A6: Robustness: Aggregate Quantitative Exercise

	Baseline	Quitrate	Int. Margin Sh.		Agg. Markup	
Modified target:		↓ 20%	↓ 15%	↑ 35%	↓ 15%	↑ 35%
Targeted moments, % Δ across steady states (1980s vs. 2010s)						
Agg. demand investment / Gross output	15.00	15.00	15.00	15.00	15.00	15.00
Top 1% sales share	25.80	25.80	25.80	25.80	25.80	25.80
Calibrated % in contact technology, for fixed $i(z)$						
Average demand investment	-31.03	-27.99	-40.33	-24.96	-45.83	-25.89
P75/25 demand investment	-8.28	-12.57	-5.77	-10.13	-8.27	-9.09
Untargeted moments, % Δ across steady states (1980s vs. 2010s)						
GDP	4.57	4.85	4.91	4.33	3.64	6.02
Aggregate TFP	3.11	3.52	3.15	3.12	2.08	4.44
Avg. customer quit rate	8.77	7.53	10.51	7.44	9.08	7.47
Aggregate markup	-0.36	-0.14	-0.58	-0.14	-0.59	-0.34
Untargeted moments, % Δ simulated firm panel across transition path						
Intangible capital intensity (true)	-12.45	-10.33	-14.19	-10.95	-12.89	-10.32
Intangible capital intensity (cap.)	8.47	12.84	5.98	9.93	9.19	10.00

E Numerical Model Solution

To ease exposition, we first describe how we solve for a steady state equilibrium, and then describe how we solve for (perfect foresight) transitions.

E.1 Steady State Equilibrium

We discretize the productivity grid into $N_z = 100$ points $(z_j)_{j=1}^{N_z}$. The steps to solve for a steady state are as follows:

1. Compute r from the Euler equation (6): $r = \beta^{-1} - 1$.
2. Initialize guesses for aggregates $(Y, w, \bar{p})$, value functions $V(z), M(z), U$, and the customer distribution $G(z)$.
3. Jointly solve the FOCs for $p(z)$ and $i(z)$ by integrating the pricing ODE forward.
   (a) At the lowest grid point, set $p(z_1) \equiv p_1 = \bar{p}$ and $i(z_1) \equiv i_1 = 0$.[29]
   (b) For all $j \geq 2$:
      i. Given p_{j-1}, i_{j-1}, compute p_j via an explicit Euler step applied to (40). [Details below.]
      ii. Given p_j and the current guesses for $G(\cdot)$ and V_j (hence W_j^*), compute i_j from (39) using the closed-form inverse of $s(i)$. [Set $i_j = 0$ if the marginal value of an additional contact, the right-hand side of (39), is non-positive.]
4. Given decisions $(p(z), i(z))$, update the value functions $V(z), M(z), U$ and the distribution $G(z)$.
   (a) Find $\bar{z}$ such that $V(\bar{z}) = 0$.
   (b) Given $\bar{z}$, update $\bar{p}$ so that $M(\bar{z}) = U$ when firm $\bar{z}$ sets price $\bar{p}$.
5. Update (Y, w):
   (a) Using last iteration's Y and the updated firm decisions and distribution, compute labor demand L^{demand} and the capital stock K; physical investment is $\delta_K K$, aggregate demand investment is $\Gamma = \int s(i(z))\, dG(z)$, so consumption is $C = Y - \delta_K K - \Gamma$.
   (b) Given C, apply the household labor FOC (5) to obtain labor supply L^{supply}.
   (c) Update Y from labor-market clearing $L^{\text{demand}} = L^{\text{supply}}$.

[29] This assumes the grid satisfies $z_1 \leq \bar{z}$.

(d) Update w using the numeraire condition $P = 1$ and (8).

6. Repeat steps 3–5 until convergence.

Solving the pricing FOC forward. Given the boundary condition $p_1 = \bar{p}$, we integrate the ODE in (40) forward on the grid using an explicit (forward) Euler step (a left-endpoint Riemann approximation). Let $\Delta z_j \equiv z_j - z_{j-1}$. For each $j \geq 2$, given (p_{j-1}, i_{j-1}), evaluate the slope of the price schedule using (40):

$$p'(z_{j-1}) \;=\; -\frac{2 + (u^i + G^i(z_{j-1}))\, i_{j-1}}{1 + (u^i + G^i(z_{j-1}))\, i_{j-1}} \cdot i_{j-1}\, g^i_{j-1} \cdot \frac{W^*_{j-1}}{\partial \pi(p_{j-1}, z_{j-1})/\partial p}\,, \tag{E.1}$$

and take a forward Euler step to evaluate a candidate price

$$p_j^{cand} \;=\; p_{j-1} + \Delta z_j\, p'(z_{j-1}).$$

In practice, it is helpful to use domain knowledge and enforce monotonicity and the static monopoly bound at each iteration:

$$p_j \;=\; \min\{\, p_j^{cand},\; p^S(z_j),\; p_{j-1}\}\,,$$

where $p^S(z) = \frac{\sigma}{\sigma-1}\frac{\overline{mc}}{z}$.

Note that with this procedure, $p(z) = \bar{p}$ is implied for all $z \leq \bar{z}$: since $i(z) = 0$ for $z < \bar{z}$ from step 3(b)ii above, the pricing slope in (E.1) evaluates to zero in that region.

Computing model moments. For equilibrium, it suffices to compute the cross-sectional customer distribution $G(z)$. To evaluate firm-level moments, we construct the joint distribution $H(z, n, age)$ of firms by productivity, number of customers, and age. We obtain H by non-stochastic simulation on grids with 200 points for n and 200 points for age.

E.2 Transitional Dynamics

Conceptually, the transition algorithm mirrors the steady-state procedure but replaces the single joint loop with two nested loops: an outer loop for the path of aggregate objects and an inner loop that solves value functions backward in time and the distribution forward. Formally, we solve for a perfect-foresight transition from an *old* steady state at $t = 1$ to a *new* steady state at $t = T$ with a time-varying parameter vector $(\theta_t)_{t=1}^T$ as follows:

1. Initialize paths for all objects by linear interpolation between the old and new steady states, in particular aggregate objects $(r_t, w_t, Y_t, \bar{p}_t)_{t=2}^{T-1}$.

2. For each $t = T-1, T-2, \ldots, 2$, iterate backward on the value functions $V_t(z)$, $M_t(z)$, U_t, starting from the new steady state values at $t = T$. Also compute $\bar{z}_t$ such that $V_t(\bar{z}_t) = 0$, and update $\bar{p}_t$ so that $M_t(\bar{z}_t) = U_t$ when firm $\bar{z}_t$ sets price $\bar{p}_t$. Within each t, solve firms' decision rules exactly as in the steady state: integrate the pricing ODE forward along the z-grid via an explicit Euler step and obtain $i_t(z)$ pointwise from (39).

3. For each $t = 2, 3, \ldots, T-1$, iterate forward on the distribution $G_t(z)$, starting from $G_1(z) = G^{old}(z)$, and compute the implied aggregates L_t^{demand} and K_t^{demand}. Set $K_t = K_t^{demand}$ if $t \geq 3$. [In $t = 2$, $K_t = K^{old}$.]

4. For each $t = 2, 3, \ldots, T-1$, use the household problem to update (r_t, w_t, Y_t):
   (a) Updated firm decisions and the distribution imply aggregate demand investment Γ_t, physical investment $Inv_t = K_{t+1} - (1-\delta_K)K_t$, and consumption $C_t = Y_t - Inv_t - \Gamma_t$.
   (b) The labor supply condition implies L_t^{supply} given C_t and w_t.
   (c) Update Y_t from labor market clearing using L_t^{supply} and L_t^{demand}.
   (d) Update w_t using the numeraire condition $P_t = 1$ and (8).
   (e) If $t = 2$, update r_t from capital market clearing using $K_t^{supply} = K^{old}$ and K_t^{demand}.
   (f) If $t \geq 3$, update r_t from the Euler equation (6) given C_{t-1} and C_t.

5. Repeat steps 2–4 until convergence.

As usual, choose T sufficiently large so that the economy has converged to the new steady state before the terminal date; in practice, set θ_t constant for $t \geq \tau$ with $\tau \ll T$.

E.3 Social Planner

In steady state, we solve for the planner's match value $V^*(z)$ and optimal contact intensity $i^*(z)$ via value function iteration inside a joint fixed-point loop. At each iteration, $i^*(z)$ follows from the planner's optimality condition (43), where the marginal benefit of an additional contact for firm z includes not only the direct expected gain $e(z)W^*(z)$ but also externality corrections that capture how firm z's investment affects other firms' quit rates and conversion rates (business stealing). The planner's match values satisfy (42) and (41). Given $(V^*(z), i^*(z))$, we update the distribution $g^*(z)$ as in the decentralized equilibrium (entry, separation, productivity transitions, and rematching), which yields aggregate productivity Z^* and aggregate demand investment Γ^*. Given (Z^*, Γ^*), we update labor from the intratemporal FOC and capital from the steady-state Euler equation and resource constraint. We iterate on $(V^*(z), i^*(z), g^*(z), Z^*, \Gamma^*, K^*, L^*)$ until convergence.

For transition paths, given (Z_t^*, Γ_t^*) the remainder is a standard growth problem. We take $t = 1$ as the market steady state and let the planner take over at $t = 2$ with predetermined capital $K_2 = K$, so consumption can jump at $t = 2$. We compute transition paths with a one-dimensional shooting method on the initial consumption jump C_2, choosing C_2 so that terminal capital satisfies $K_T = K^*$. For any candidate C_2, we initialize paths for (K_t, C_t, L_t, Z_t) between the market steady state and the planner steady state, then iterate a backward-forward loop: backward in time for $(V_t^*(z), i_t^*(z))$ using the planner Bellman system (with time-varying continuation values), and forward in time for the customer distribution and aggregates. Given the updated distribution, we recompute Z_t and then update (K_t, C_t, L_t) using the Euler equation, the resource constraint, and the intratemporal labor FOC, and repeat until the inner loop converges. We then adjust C_2 via bracketing/bisection until the terminal condition holds.