Supplemental Materials for "Concentration, Market Power, and Misallocation: The Role of Endogenous Customer Acquisition"[*]

Hassan Afrouzi[†]
Columbia University and NBER

Andres Drenik[‡]
The University of Texas at Austin and NBER

Ryan Kim[§]
Seoul National University

September 2025

[*]We thank David Argente, Ariel Burstein, Doireann Fitzgerald, and seminar participants at various institutions and conferences for valuable comments and suggestions. Luigi Caloi provided superb research assistance. Previous versions of this manuscript were circulated under the title "Growing by the Masses: Revisiting the Link between Firm Size and Market Power." Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Researcher(s)' own analyses calculated (or derived) based on data from National Promotion Reports, LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business.

[†]Department of Economics. Email: hassan.afrouzi@columbia.edu.

[‡]Department of Economics. Email: andres.drenik@austin.utexas.edu.

[§]Department of Economics. Email: rkim59@snu.ac.kr.

1 Data Description 3

1.1 NielsenIQ, Promodata, and GS1: Variables and Data Cleaning . . . 3
1.2 Compustat: Variables and Data Cleaning . . . 4
1.3 Coverage of the NielsenIQ-Compustat Sample . . . 5
1.4 Summary Statistics . . . 5
1.5 Firm Sales Growth Decomposition . . . 6
1.6 Sales and SGA Expenses: Decomposition by Durability of Products . . . 8
1.7 The Semi-variable Nature of SGA . . . 9

2 Model with Heterogeneous Tastes and Sorting 12

2.1 Taste for Quantity . . . 12
2.2 Taste for Quality . . . 13

3 Computational Appendix 15

3.1 Recursive Formulation . . . 15
3.2 Solution Algorithm . . . 19
3.3 Estimation routine . . . 20

4 Additional Model Analysis 22

4.1 Identification of Model Parameters . . . 22
4.2 Calibration Results: Goodness of Fit . . . 24
4.3 Model Mechanisms . . . 29
4.4 Additional Analysis of Model in Steady State . . . 33

5 Calibration of the Homogeneous Customers Model 38

6 Model Robustness 40

6.1 Higher Elasticity of Production to Labor . . . 40
6.2 Correlated Production and Advertising Productivities at the Firm Level . . . 43
6.3 Hopenhayn Model of Entry . . . 46

1 Data Description

This section further describes the process to clean the NielsenIQ and Compustat data used in this paper. For this, we follow the previous studies that use these two datasets (Hottman, Redding, and Weinstein 2016 for NielsenIQ and De Loecker, Eeckhout, and Unger 2020, Traina 2019 for Compustat).

1.1. NielsenIQ, Promodata, and GS1: Variables and Data Cleaning. We construct the following variables from the NielsenIQ Homescan Panel, Promodata, and GS1 company data:

- *UPC*: scanner-level product identifier available in NielsenIQ Homescan Panel and Promo data.
- *GS1 Company Identifier*: GS1 company ID. GS1 provides both UPC and firm identifiers along with company name and headquarters location. This information allows us to identify firm boundaries in the NielsenIQ data and further merge the data with Compustat.
- *Sales*: We define sales as the sum of the total expenditures of households at different levels of aggregation: UPC-year, firm-year, and group-firm-year. We use sample weights (projection factor) to render the NielsenIQ household sample representative at the national level.
- *Number of Customers*: For each firm, product (UPC), or group-firm, we aggregate the number of households with the sample weight adjustment.
- *Sales per Customer*: At each level of aggregation, we define average sales per customer as total sales divided by the number of customers.

Sample Selection of Homescan Panel data. The Homescan Panel data we use covers 2004-2016.

1. Following Neiman and Vavra (2023), we balance the product modules across years to exclude the effect of entry and exit of modules, which mainly arises from name changes and potentially adds measurement error.
2. To render the sample representative, we drop "magnet products" in NielsenIQ data, which are fresh produce and other items without barcodes.
3. We drop products that do not have a product group identifier to render the analysis consistent across different specifications.
4. There is a small number of observations for which the sampling year of the household is different from the year their purchases took place. These reflect the fact that households were sampled in late December. While the corresponding purchases are recorded as the current year, their household panel years are recorded as the following year. We drop these observations to use coherent sample weights across households and years.

We restrict the sample in the Promodata to the years 2006-2011 since the data are incomplete for the other years. Following the manual for these data, we use both active and inactive files and drop duplicate observations in inactive files. We adjust for multi-package and unit size when using the UPC-level information; we do the same when we merge with price information from Homescan

Panel data. We exclude the small number of markets that are not common across the Homescan Panel and Promodata.

1.2. Compustat: Variables and Data Cleaning. We download and construct the following variables from Compustat:

- *Global company key* (mnemonic `gvkey`): Compustat's firm ID.
- *Year* (mnemonic `fyear`): the fiscal year.
- *Selling, general and administrative expense* (mnemonic `XSGA`): the SG&A sums "all commercial expenses of operation (such as expenses not directly related to product production) incurred in the regular course of business pertaining to the securing of operating income." They include expenses such as marketing and advertising expenses, research and development, accounting expenses, delivery expenses, etc.
- *Costs of goods sold* (mnemonic `COGS`): the COGS sums all "expenses that are directly related to the cost of merchandise purchased or the cost of goods manufactured that are withdrawn from finished goods inventory and sold to customers." They include expenses such as labor and related expenses (including salary, pension, retirement, profit sharing, provision for bonus and stock options, and other employee benefits), operating expense, lease, rent, and loyalty expense, write-downs of oil and gas properties, and distributional and editorial expenses.
- *Operating expenses, total* (mnemonic `XOPR`): OPEX represents the sum of COGS, SG&A, and other operating expenses.
- *Sales (net)* (mnemonic `SALE`): this variable represents gross sales, for which "cash discounts, trade discounts, and returned sales and allowances for which credit is given to customer" are discounted from the final value.
- *Capital*: we calculate capital in two ways. First, we simply set capital to be equal to the gross property, plant, and equipment value (mnemonic `PPEGT`) deflated by the investment goods deflator from NIPA's nonresidential fixed investment good deflator (line 9). For our second measurement of capital, we use the perpetual inventory method—we set the first observation of each firm to be equal to the gross property, plant, and equipment value and for susbsequent years we add the difference from $netPPE_t$ (mnemonic `PPENT`) and $netPPE_{t-1}$. We also deflate the difference in `PPENT` by the investment goods deflator from NIPA's nonresidential fixed investment good deflator.

We used NIPA Table 1.1.9. GDP deflator (line 1) to generate the real value for the variables `sale`, `COGS`, `XOPR`, and `XSGA`.

Sample Selection. We downloaded the dataset "Compustat Annual Updates: Fundamentals Annual," from Wharton Research Data Services, from Jan 1950 to Dec 2016. The following options were chosen:

- Consolidated level: C (consolidated)

- Industry format: INDL (industrial)
- Data format: STD (standardized)
- Population source: D (domestic)
- Currency: USD
- Company status: active and inactive

We took the following steps in the cleaning process:

1. To select American companies, we filtered the dataset for companies with Foreign Incorporation Code (FIC) equal to "USA."
2. We replace industry variables (`sic` and `naics`) with their historical values whenever the historical value is not missing.
3. We drop utilities (`sic` value in the range [4900, 4999]) because their prices are very regulated and financials (`sic` value in the range [6000,6999]) because their balance sheets are notably different than the other firms in the analysis.
4. To ensure the quality of the data, we drop missing or nonpositive observations for sales, COGS, OPEX, sic 2-digit code, gross PPE, net PPE, and assets. We also exclude observations in which acquisitions are more than 5% of the total assets of a firm.
5. A portion of the data missing for sales, COGS, OPEX, and capital between years for firms. We input these values using a linear interpolation, but we do not interpolate for gaps longer than 2 years. This exercise inputs data for 4.6% of our sample.

1.3. Coverage of the NielsenIQ-Compustat Sample. Approximately 300 firms identified in Compustat can be matched with the NielsenIQ data for 2004-2016. Although the number of firms we matched is small, they account for a significant fraction of total sales, number of UPCs, and observations in the NielsenIQ Homescan Panel data, as shown in Table SM.1.1.

Table SM.1.1: Coverage of the NielsenIQ-Compustat Sample

	Sales (b)	**# of UPCs (k)**	**# of Obs. (m)**
NielsenIQ-Compustat Sample	94.5	114.9	12.1
NielsenIQ Sample	421.2	698.9	51.6
Share (%)	22.4	16.4	23.5

Note: Sales is the projection-factor-weighted sales in NielsenIQ data and is denoted in billions US dollars. # of UPCs is in thousand UPCs, and # of Obs. is in millions of observations. All variables are annual averages.

1.4. Summary Statistics. Below we present summary statistic tables as discussed in the main text.

Table SM.1.2: Summary Statistics

	Panel A: NielsenIQ-GS1, Firm-Product Group-Year					
	count	mean	sd	p10	p50	p90
S_{igt} (in thousands USD)	557818	6708.2	64961.2	3.968	126.1	5170.5
$p_i q_{igt}$ (in USD)	557818	10.03	20.14	1.963	5.876	19.50
m_{igt} (in thousands)	557818	500.8	2789.8	0.823	19.83	639.9
m_{igt}^{New} (in thousands)	557818	250.3	989.0	0.476	16.03	424.9
m_{igt}^{Old} (in thousands)	557818	250.5	1963.1	0	1.601	194.2
	Panel B: NielsenIQ-Compustat, Firm-Year					
	count	mean	sd	p10	p50	p90
SGA_{it} (in millions USD)	2376	1957.7	4894.3	7.626	311.6	4653.6
$COGS_{it}$ (in millions USD)	2399	6950.3	18056.3	15.54	1058.6	16618.3
$OPEX_{it}$ (in millions USD)	2399	8889.3	20916.2	23.95	1549.0	23382.6
SGA-to-$OPEX_{it}$	2376	0.298	0.195	0.0767	0.265	0.574
$Sales$-to-$COGS_{it}$	2399	1.773	1.045	1.143	1.487	2.581

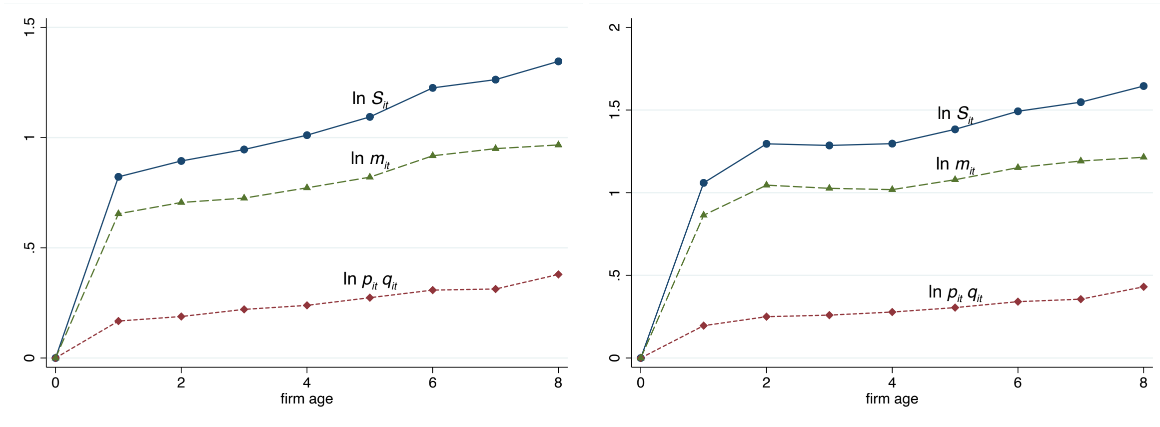
Notes: The NielsenIQ-GS1 data in Panel A has 40,418 firms and 109 product groups in the period of 2004-2016. S_{igt} denotes sales of firm i in product group g and time t, $p_{igt}q_{igt}$ average sales per customer, m_{igt} number of customers, m_{igt}^{New} new customers in year t who did not purchase products in year $t-1$, and m_{igt}^{Old} customers who purchase the products consecutively in year $t-1$ and t ($m_{igt} = m_{igt}^{\text{New}} + m_{igt}^{\text{Old}}$). S_{igt} is measured in thousands of US dollars, and m_{igt}, m_{igt}^{New}, and m_{igt}^{Old} are in thousands of customers. All NielsenIQ variables are projection-factor adjusted. The NielsenIQ-Compustat matched data in Panel B have 312 firms in the period 2004-2016. All cost-side variables are in millions of US dollars and are deflated by the GDP deflator.

1.5. Firm Sales Growth Decomposition. One concern in Figure 1 is that some firms might only appear temporarily in our data—not because of their actual behavior but due to sampling error. For example, it could be that households in our sample do not appear to purchase a firm's product even though the product was actually purchased and not recorded. In this case, the average value of sales, number of customers, and sales per customer of young firms in our analyses might be confounded with those of old firms.

To address concerns regarding the sampling error, Figure SM.1.1 uses only those firms that appear for at least 3 or 5 consecutive years. The results still show that the number of customers is a primary factor that generates an increase in firms' sales over time. There is a steeper increase in sales in the firm's early stage than our baseline results in Figure 1. The results are intuitive, since firms that survive for several years are likely to generate more sales at the beginning relative to firms that could not survive. Overall, the robustness analysis suggests that the sampling errors are not first-order concerns in our analyses.

Another concern is that firms might sell their products in a different number of months over the following years. For example, some firms might enter the market in late November or December

Figure SM.1.1: Decomposition of Firm Sales Growth by Firm Age: Survivors

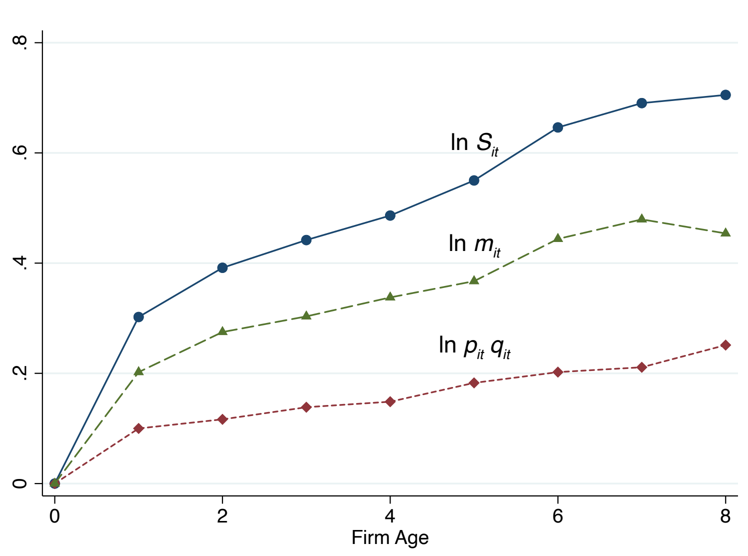


(a) Survive for 3+ Consecutive Years (b) Survive for 5+ Consecutive Years

Notes: Figures SM.1.1a and Figure SM.1.1b replicate Figure 1 by using firms that appear in the sample for at least 3 and 5 consecutive years, respectively. There are 32,242 number of observations and 6,400 firms used in Figures SM.1.1a and 19,603 number of observations and 2,997 firm used in Figure SM.1.1b.

but sell their products over many months in subsequent years. To adjust for these differences, we calculate the average monthly sales over a year per firm and redo the decomposition exercise in Figure SM.1.2. There is a smaller increase in firms' sales at age 1, which suggests that some firms enter during the late months of the initial year. However, the relative importance of the number of customers in explaining sales remains the same, and accounts for approximately 70% of sales on average.

Figure SM.1.2: Decomposition of Firm Sales Growth by Firm Age: Monthly Sales



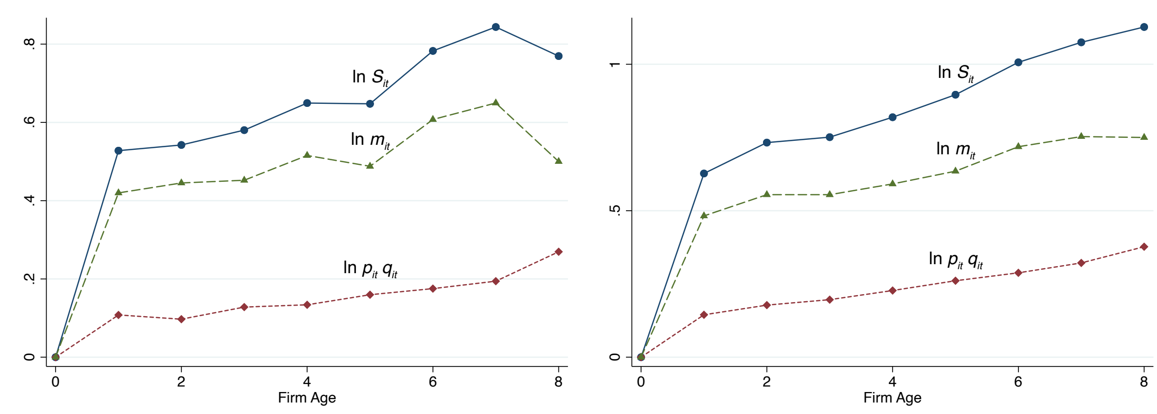
Notes: Figure SM.1.2 replicates Figure 1 by using average monthly sales per firm and year instead of yearly sales.

Also, one might be worried that the empirical pattern we observe might not apply to products outside of our sample, which is restricted to products with a barcode. For example, it could be that for more durable products that customers purchase occasionally, firms might not be able to grow as much through the extensive margin of demand because they may not face the same customers every year.

Given that there is no other consumer-producer matched dataset (to the best of our knowledge), we use our data—which cover a substantial fraction of consumer goods with a wide variety of products—to understand the underlying differences between durable and non-durable products. We closely follow Argente, Lee, and Moreira (2024) and define product group-level durability by using information on the number of shopping trips. We count the average yearly number of trips customers made to purchase products in each product group and divide product groups into durable and non-durable products based on the median value of average trips. The set of durable products include, for example, "LIGHT BULBS, ELECTRIC GOODS," "HARDWARE, TOOLS," and "AUTOMOTIVE," and the non-durable products include "MILK," "SNACKS," and "BEER."

Figure SM.1.3 presents the results. Regardless of whether we are analyzing durable or non-durable products, firms mainly grow by expanding their customer bases. The relevance of the number of customers for firm growth remains when redefining durable goods based on the 75th percentile of the trips distribution or analyzing the variance decomposition of total sales of durable or non-durable products.

Figure SM.1.3: Decomposition of Firm Sales Growth by Firm Age: By Durability



(a) > p50 durability

(b) ≤ p50 durability

Notes: Figures SM.1.3a and Figure SM.1.3b replicate Figure 1 by dividing firms based on the durability of the products they sell. There are 19,988 observations and 5,050 firms used in Figure SM.1.3a and 29,816 observations and 7,323 firms used in Figure SM.1.3b.

1.6. Sales and SGA Expenses: Decomposition by Durability of Products. Using the same durability measure we constructed and used in Figure SM.1.3, we analyze the potential heterogeneity in the relationship between SGA expenses and sales based on a product's durability. Table SM.1.3 reports the results. We find no statistically significant differences in this relationship by the durability of products.

Table SM.1.3: Sales and SGA Expenses: Decomposition by Durability of Products

	Decomposition of $\ln S_{\text{igt}}$			$\ln m_{\text{igt}}$: New vs. Old	
	(1)	(2)	(3)	(4)	(5)
	$\ln S$	$\ln pq$	$\ln m$	$\ln m^{\text{New}}$	$\ln m^{\text{Old}}$
$\ln \text{SGA}_{\text{it}}$	0.106***	0.004	0.101***	0.104***	0.028
	(0.039)	(0.017)	(0.033)	(0.033)	(0.039)
$\ln \text{SGA}_{\text{it}}$ x Durability	0.015	0.014	0.001	0.022	-0.023
	(0.057)	(0.026)	(0.039)	(0.049)	(0.050)
Observations	12514	12514	12514	12514	12514
R^2	0.963	0.910	0.965	0.944	0.961
Firm-year Controls	✓	✓	✓	✓	✓
Group-year FE	✓	✓	✓	✓	✓
SIC-year FE	✓	✓	✓	✓	✓
Group-firm FE	✓	✓	✓	✓	✓

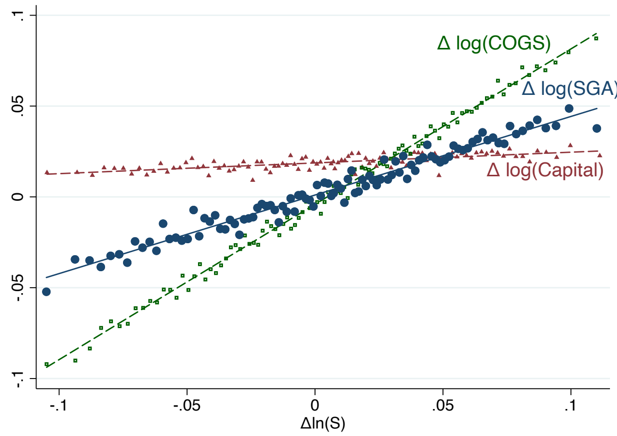
Notes: The regression specification is the same as the one estimated in Table A.8, with the exception that we include as additional regressors the durability measure and its interaction with the log of SGA expenses.

1.7. The Semi-variable Nature of SGA. This section establishes that the SGA expenses have a semi-variable nature and correlate with firms' short-run sales. Previous studies that examined the non-production cost of firms (SGA) made polar opposite assumptions on the variable nature of this cost. For instance, in measuring price-cost markups, Traina (2019) includes non-production costs as variable costs, whereas De Loecker, Eeckhout, and Unger (2020) interpret non-production costs as fixed costs in their baseline approach. We empirically assess the validity of such assumptions by comparing the comovement of sales and SGA expenses with the comovement of sales and other costs that are commonly assumed to be variable and fixed in the short run in the literature: COGS expenses and investment.

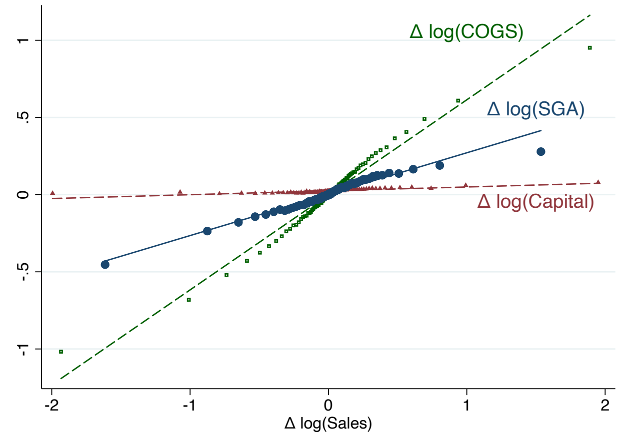
Our results suggest that SGA expenses have both variable and fixed components; They are more variable than the capital expenditure but less than COGS expenses. Figure SM.1.4a reports the binned scatter plot of changes in sales against changes in firm's costs for a range of $\Delta \ln S_{i,t}$ between -10% and 10%, which are approximately the 25th and 75th percentiles of the $\Delta \ln S_{it}$ distribution. Consistent with the view in the literature (e.g., De Loecker, Eeckhout, and Unger 2020), sales exhibit the largest comovement with production costs ($\beta = 0.894$; SE 0.008) and the lowest comovement with investment ($\beta = 0.081$; SE 0.008). Figure SM.1.4b shows that similar relationships hold in the full sample.

We consider other specifications to confirm the semi-variable nature of SGA expenses. Table SM.1.4 presents the regression results that correspond to Figure SM.1.4. The semi-variable nature

Figure SM.1.4: The Semi-variable Nature of SGA



(a) Restricted Sample



(b) Full Sample

Notes: The figure shows the binned scatter plots of the correlation between the quarterly change in log sales and the quarterly change in (i) log SGA expenses, (ii) log COGS expenses, and (iii) log stock of capital for firms in the quarterly Compustat dataset. We also plot the best linear fit for each variable. The correlations control for quarter and firm fixed effects. In Panel (a), we restrict the sample to observations with a change in the log of sales between -0.1 and +0.1. Panel (b) plots the binned scatter plot using the full sample. We adopt the perpetual inventory method following Traina (2019) and use Gross and Net Capital (PPEGT and PPENT) and deflate investment with NIPA's non-residential fixed investment good deflator to measure the capital stock. There are 16,301 firms in the 1970-2016 period used in this analysis.

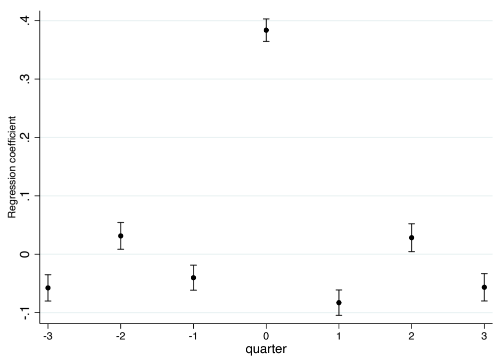
of SGA expenses is clear in this table, both with and without fixed effects. We also show that R&D expenses (a subcomponent of SGA costs) are more variable than the stock of capital but less variable than total SGA expenses. Finally, Figure SM.1.5 reports the coefficients of a regression of change in SGA expenses on leads and lags of change in total sales, which further supports the short-run variability of SGA expenses: Although there is a strong correlation between change in SGA expenses and contemporaneous change in sales, we find no large correlation with future or past change in sales.

Table SM.1.4: The Semi-variable Nature of SGA

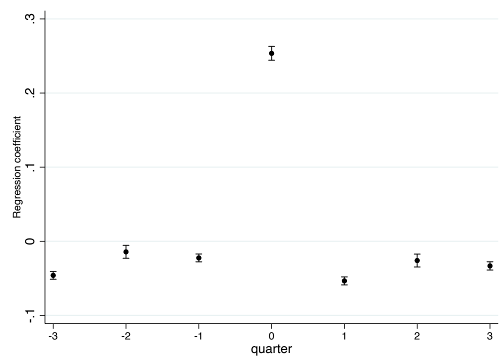
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta ln(\text{COGS}_{it})$	$\Delta ln(\text{COGS}_{it})$	$\Delta ln(\text{SGA}_{it})$	$\Delta ln(\text{SGA}_{it})$	$\Delta ln(\text{Capital}_{it})$	$\Delta ln(\text{Capital}_{it})$	$\Delta ln(\text{R\&D}_{it})$	$\Delta ln(\text{R\&D}_{it})$
ΔlnS_{it}	0.855***	0.857***	0.451***	0.438***	0.114***	0.066***	0.263***	0.226***
	(0.008)	(0.008)	(0.009)	(0.010)	(0.006)	(0.006)	(0.028)	(0.029)
R^2	0.047	0.128	0.011	0.074	0.002	0.176	0.002	0.089
Quarter FE	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	No	Yes	No	Yes	No	Yes	No	Yes
N	311298	309985	276656	275428	184323	182454	67501	66938

Notes: Dependent variables are the quarterly change in log COGS expenses, SGA expenses, and the stock of capital. The estimation method used in all columns is OLS. Standard errors (in parentheses) are clustered at the firm level.

Figure SM.1.5: Correlation between change in SGA expenses and leads and lags of change in sales



(a) Trimmed Sample



(b) Full Sample

Notes: Figure SM.1.5a uses the trimmed sample presented in Figure SM.1.4b and Figure SM.1.5b uses the full sample. 95% confidence intervals are presented for every estimate. Figure SM.1.5a replicates column (4) in Table SM.1.4 at quarter = 0.

2 Model with Heterogeneous Tastes and Sorting

In this section, we provide two model extensions to our households' preferences. In the first part, we allow for idiosyncratic preference shocks that affect the optimal quantities consumed by each consumer. In the second, we allow for shifters that affect the perceived "quality" of each variety. We discuss how each extension affects the total demand of a firm as well as the relationship between the number of customers and average sales per customer. We also provide empirical evidence for why we abstract away from these extensions in the main analysis.

2.1. Taste for Quantity. Consider the following extension of the Kimball aggregator in Equation (4):

$$\int_0^{N_t}\int_0^1 \mathbf{1}_{\{j\in m_{i,t}\}}\xi_{i,j,t}^{-1}\Upsilon\left(\frac{\xi_{i,j,t}c_{i,j,t}}{C_t}\right)dj\,di = 1, \tag{2.1}$$

where $\xi_{i,j,t}$ is now household j's "quantity" taste for variety i at time t. The implied relative demand for firm i at t is then given by

$$\frac{c_{i,t}}{C_t} = \underbrace{m_{i,t}}_{\text{number of customers}} \times \underbrace{\mathbb{E}[\xi_{i,j,t}|j\in m_{i,t}]\Upsilon'^{-1}\left(\frac{p_{i,t}}{D_t}\right)}_{q_{i,t}\equiv\text{ demand per customer}}, \tag{2.2}$$

where $\mathbb{E}[\xi_{i,j,t}|j\in m_{i,t}]$ is the average quantity taste of firm i's customers. The introduction of such shocks will affect firm i's demand depending on the nature of sorting between customers and firms. If there is no sorting or selection regarding who is matched to a firm, then $\mathbb{E}[\xi_{i,j,t}|j\in m_{i,t}] = \mathbb{E}[\xi_{i,j,t}]$ and this extension replicates the demand function in the main text. However, sorting generates a correlation between the number of customers and average sales per customer. For example, with positive sorting, firms with a larger customer base should sell less per customer on average (because the marginal customer always buys less than the average customer). Instead, with negative sorting (e.g., when marginal consumers are the ones who experiment with the product and buy more than the average consumer), firms with a larger customer base should sell more per customer on average. More generally, this model implies the following relationship between average sales per customer and a firm's number of customers:

$$\ln(p_{i,t}q_{i,t}) = \beta_0\ln(m_{i,t}) + \ln\left(p_{i,t}\Upsilon'^{-1}\left(\frac{p_{i,t}}{D_t}\right)\right), \tag{2.3}$$

where the sign and magnitude of β_0 determine the type and strength of sorting, respectively. To test this, we aggregate the NielsenIQ Homescan Panel data at the UPC-year-level. For this analysis, we use the whole sample available in the Homescan Panel data. We compute the price of each product as sales divided by the unit-adjusted quantity. Table SM.2.1 shows the results of OLS regressions of the log average sales per customer of UPC u at time t on the log number of customers and polynomials of log price that approximate the nonlinear function $\Upsilon'^{-1}(\cdot)$. Across all specifications, we consistently find a small and positive relationship between the size of the customer base and average sales per customer. Firms with 1% more customers sell on average 0.03% more per customer. Given the economic insignificance of this estimate, we do not consider this kind of preference heterogeneity in our model, which, as discussed below, is a conservative choice.

Table SM.2.1: Average Sales per Customer and the Size of the Customer Base

	(1)	(2)	(3)	(4)	(5)
$\ln m_{ut}$	0.026***	0.025***	0.023**	0.029***	0.029***
	(0.009)	(0.009)	(0.009)	(0.002)	(0.002)
$\ln p_{ut}$	0.162***	0.180***	0.132***	0.680***	0.688***
	(0.028)	(0.037)	(0.027)	(0.017)	(0.017)
$\ln p^2{}_{ut}$		0.021*	0.025	-0.003	-0.003
		(0.012)	(0.015)	(0.005)	(0.005)
$\ln p^3{}_{ut}$			0.005	-0.001	-0.001
			(0.003)	(0.001)	(0.001)
Observations	9097044	9097044	9097044	8452679	8452679
R^2	0.086	0.096	0.102	0.851	0.852
UPC FE				✓	✓
Year FE				✓	
Group-Year FE					✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Standard errors are clustered by product group. The variable $p_{ut}q_{ut}$ denotes sales per customer of UPC u at time t, m_{ut} is the number of customers, and p_{ut} is the price.

Implications of Sorting for Misallocation Results. A potential concern in our analysis of efficiency losses from the misallocation of demand is that the sorting of customers might exacerbate or reduce welfare losses from misallocation. For example, since with positive sorting the marginal customer values a variety less than the average customer, the marginal value of allocating customers to more productive firms might not be as high as in our baseline model. However, regression results in Table SM.2.1 show that, if anything, there is negative—albeit small—sorting (i.e., the marginal customer buys more than the average customer). Viewed through the lens of this model extension, these results indicate that the quantified welfare losses from the misallocation of demand in Section 5 provide a *lower* bound on the actual welfare losses once this small negative sorting is taken into account.

2.2. Taste for Quality. Another source of preference heterogeneity can be the different perception of the "quality" of a variety across customers. For this, consider the following extension of the Kimball aggregator in Equation (4):

$$\int_0^{N_t} \int_0^1 \mathbf{1}_{\{j \in m_{i,t}\}} \zeta_{i,j,t} \Upsilon \left(\frac{c_{i,j,t}}{C_t} \right) dj di, = 1 \tag{2.4}$$

where $\zeta_{i,j,t}$ is now household j's "quality" taste for variety i at time t. We assume for any i and t that $\zeta_{i,j,t}$ is i.i.d. and its distribution is scaled so that its unconditional mean is 1 ($\mathbb{E}[\zeta_{i,j,t}] = 1$). The implied relative demand of firm i is then given by

$$q_{i,t} \equiv \frac{c_{i,t}}{C_t} = \int_0^1 \mathbf{1}_{\{j \in m_{i,t}\}} \Upsilon'^{-1} \left(\frac{p_{i,t}}{\zeta_{i,j,t} D_t} \right) dj. \tag{2.5}$$

The curvature of $\Upsilon'^{-1}(\cdot)$, and thus the elasticity of demand, now interacts with the distribution of $\zeta_{i,j,t}$. However, we can show that up to a first-order approximation, sorting would imply a correlation between average demand per customer and the number of customers, similar to the one derived for quantity shocks. To see this, note that up to a first-order approximation around the point $\frac{p_{i,t}}{\mathbb{E}[\zeta_{i,j,t}]D_t}$:

$$\Upsilon'^{-1}\left(\frac{p_{i,t}}{\zeta_{i,j,t}D_t}\right) \approx \Upsilon'^{-1}\left(\frac{p_{i,t}}{D_t}\right)\left(1+\sigma_{i,t}(\zeta_{i,j,t}-1)\right), \tag{2.6}$$

where $\sigma_{i,t}$ is the elasticity of demand evaluated at the average quality taste, $p_{i,t}/D_t$, and hence depends only on the price. Now, using this approximation, we can write demand as

$$\frac{c_{i,t}}{C_t} \approx \underbrace{m_{i,t}}_{\text{number of customers}} \times \underbrace{\Upsilon'^{-1}\left(\frac{p_{i,t}}{D_t}\right)(1+\sigma_{i,t}(\mathbb{E}[\zeta_{i,j,t}|j \in m_{i,t}]-1))}_{q_{i,t}\equiv\text{demand per customer}}, \tag{2.7}$$

where $\mathbb{E}[\zeta_{i,j,t}|j \in m_{i,t}]$ is the average quality taste of firm i's customers. Similar to quantity shocks, if there is positive (negative) sorting, then $\mathbb{E}[\zeta_{i,j,t}|j \in m_{i,t}]$ is a decreasing (increasing) function of $m_{i,t}$ and we should see a negative (positive) relationship between $q_{i,t}$ and $m_{i,t}$, which brings us to the same argument in the previous section for quantity shocks.

3 Computational Appendix

In this section, we present the details of the computation algorithm that solves and calibrates the model. We first describe the recursive representation of the firm's problem. Then, we describe the law of motion of firms and characterize the stationary distribution. Next, we describe the algorithm that solves the model and the algorithm used in the calibration.

3.1. Recursive Formulation. Here we present the recursive formulation of the firms' problem, the law of motion for the number and distribution of firms over time, the steady state of the model and its general equilibrium.

3.1.1. Firm-Level Problem: Existence and Uniqueness of the Value Function.

Firm's Recursive Problem. In period t, a firm that decided to operate with a customer base of m_{-1} and productivity z solves the following dynamic programming problem (which corresponds to the sequential representation in Equation (14)):

$$
\begin{aligned}
&V_t(m_{-1},z) \equiv \max_{l_s,l_p,p}\left\{py - W_t l_p - W_t(l_s+\chi) + \beta\nu\frac{U_{c,t+1}}{U_{c,t}}\mathbb{E}\left[\max\{0, v_t(m_{-1},z)\}|z\right]\right\}\\
s.t.\quad &q = \left[1-\eta\ln\left(\frac{p}{D_t(1-\sigma^{-1})}\right)\right]^{\frac{\sigma}{\eta}}\\
&y = mqC_t = zl_p^{\alpha}\\
&m = (1-\delta)m_{-1} + \frac{l_s^{\phi}}{P_{m,t}},
\end{aligned}
$$

Law of Motion for the Measure and Distribuition of Firms. Let $\mathcal{N}_t : M\times Z \to [0,1]$—where M and Z denote the domains of m and z, respectively—denote the cumulative measure function (CMF) of incumbent firms measured after the realization of idiosyncratic productivity shocks, but before exit decisions are made—i.e., measure of firms with customers less than m and productivity less than z'. The law of motion of the CMF of firms is given by

$$
\mathcal{N}_t(m,z') = \int_{M\times Z} F(z'|z)\mathbf{1}_{\{m_t^*(m_{-1},z)\le m\}}\nu\mathbf{1}_{\{v_t(m_{-1},z)\ge 0\}}d\mathcal{N}_{t-1}(m_{-1},z) \tag{3.1}
$$

$$
+\lambda\int_{M\times Z} F(z'|z)\mathbf{1}_{\{m_t^*(0,z)\le m\}}\mathbf{1}_{\{v_t(0,z)\ge 0\}}dF^e(z), \tag{3.2}
$$

where $F(z'|z)$ is the Markov chain given by the AR(1) productivity process, $F^e(z)$ is the productivity distribution of potential entrants, and $m^*(m_{-1},z)$ denotes the optimal policy for customer acquisition. Note that with the CMF $\mathcal{N}_t$ at hand, $N_t \equiv \sup_{(m,z')\in M\times Z}\mathcal{N}_t(m,z')$ is the total number of operating firms in the economy at t and $\mathcal{L}_t(m,z') \equiv \mathcal{N}_t(m,z')/N_t$ is the CDF of the operating firms in the economy.

Steady State. Define a stationary equilibrium as one where the vector of aggregate variables $X \equiv (W, P_m, C, D)$, and the measure $\mathcal{N}$ (which encodes both the number of firms N and their distribution $\mathcal{L}$) are constant over time. The firm's dynamic problem for a given X becomes independent of t

and can be written as:

$$
\begin{aligned}
&V(m_{-1},z)\\
&=\max_{q,m}\left\{CD\Upsilon'(q)mq-W\Big(\frac{Cmq}{z}\Big)^{\alpha^{-1}}-WP_m^{\phi^{-1}}\big(m-(1-\delta)m_{-1}\big)^{\phi^{-1}}-W\chi+\beta\nu\mathbb{E}^{z'}[\max\{0,V(m,z')\}\mid z]\right\}
\end{aligned}
$$

Dividing through by DY and letting $\bar{W}=W/(CD)$ gives the normalized value function $v(m_{-1},z)$ that satisfies:

$$
v(m_{-1},z)=\max_{m}\left\{F(m,z;\bar{W}C^{\alpha^{-1}})-\bar{W}P_m^{\phi^{-1}}(m-(1-\delta)m_{-1})^{\phi^{-1}}-\bar{W}\chi+\beta\nu\mathbb{E}^{z'}[\max\{0,v(m,z')\}\mid z]\right\},
$$

where

$$
F(m,z;\bar{W}C^{\alpha^{-1}})\equiv\max_{q}\left\{\Upsilon'(q)mq-\bar{W}\Big(\frac{Cmq}{z}\Big)^{\alpha^{-1}}\right\}.
$$

For $z\in Z\equiv[0,\bar{Z}]$ and $m\in M\equiv[0,\bar{m}]$, F is continuous and bounded. Thus, the fixed point of the operator

$$
(Tv)(m_{-1},z)=\max_{m}\left\{F(m,z;\bar{W}C^{\alpha^{-1}})-\bar{W}P_m^{\phi^{-1}}(m-(1-\delta)m_{-1})^{\phi^{-1}}-\bar{W}\chi+\beta\nu\mathbb{E}^{z'}[\max\{0,v(m,z')\}\mid z]\right\}
$$

characterizes the value function of firms. The existence and uniqueness of this fixed point $Tv^*=v^*$ follows from the standard arguments in Chapter 9 of Stokey and Lucas Jr (1989) for given aggregate vector $X\geq 0$.

3.1.2. Steady-State Aggregates. Given firm policies $\{m^*(\cdot,X),q^*(\cdot,X)\}$ implied by v^* under some vector of aggregates X, one can immediately derive the stationary measure $\mathcal{N}_X$ from Equation (3.1). One can then define the aggregation operator $f(\cdot)$ that maps a conjectured vector of aggregate variables X into implied aggregates based on optimal firm decisions:

$$
X'=f(X).
$$

Letting $\omega\equiv(m_{-1},z)\in M\times Z$ denote the binary type of a firm, the function f is defined by the following equations:

$$
X\to C':\quad C'=\left(\int m^*(\omega,X)\Upsilon(q^*(\omega,X))\mathcal{N}_X(d\omega)\right)C \tag{3.3}
$$

$$
X\to P_m':\quad 1=\int m^*(\omega,X)\mathcal{N}_X(d\omega)=(1-\delta)\int m_{-1}(\omega)\mathcal{N}_X(d\omega)+\frac{1}{P_m'}\int l_s^*(\omega,X)\mathcal{N}_X(d\omega) \tag{3.4}
$$

$$
X\to W':\quad W'=\xi C\left[\int\left(l_p^*(\omega,X)+l_s^*(\omega,X)+\chi\mathbf{1}_{\{v^*(\omega,X)\geq 0\}}\right)\mathcal{N}_X(d\omega)\right]^{\psi} \tag{3.5}
$$

$$
X\to D':\quad D'=\left[\int m^*(\omega,X)q^*(\omega,X)\Upsilon'(q^*(\omega,X))\mathcal{N}_X(d\omega)\right]^{-1} \tag{3.6}
$$

where $l_s^*(.,X)=P_m^{\phi^{-1}}(m^*(\omega,X)-(1-\delta)m_{-1}(\omega))^{\phi^{-1}}$ and $l_p^*(.,X)=\left(\frac{Cm^*(\omega,X)q^*(\omega,X)}{z}\right)^{\alpha^{-1}}$ are the implied levels of advertising and production labor implied by the policies $q^*(.,X)$ and $m^*(.,X)$ under the aggregates in X.

3.1.3. A Simple Case. Assume full depreciation of customers ($\delta = 1$) so that the only relevant state for the firm is their productivity. Suppose also there are no over-head costs so that $\chi = 0$. Suppose a stationary equilibrium exists where $(W, C, D, P_m) \in (0, \infty)$. Suppose that in this stationary equilibrium the marginal distribution of $\mathscr{L}$ is $\mathscr{L}_z$ and the measure of firms is N. Furthermore, let us pick the parameters λ and ξ such that $\bar{W} = C = 1$ without loss of generality (alternatively one can see that fixing the values of these variables is equivalent to a scaling of the productivity distribution as well as a scaling of P_m). It follows that the value function of a firm (normalized by CD) is given by

$$v(z) = \max_{m,q} \left\{ \left(mq\Upsilon'(q) - \left(\frac{mq}{z}\right)^{\alpha^{-1}} \right) - P_m^{\phi^{-1}} m^{\phi^{-1}} + \beta\nu \int \max\{0, v(z')\} \mathscr{L}_z(dz') \right\}$$

We note that the choice for m and q are static and given by the program:

$$\max_{m,q} \left\{ mq\Upsilon'(q) - \left(\frac{mq}{z}\right)^{\alpha^{-1}} - P_m^{\phi^{-1}} m^{\phi^{-1}} \right\} \geq 0$$

where the max is positive because $m = q = 0$ is always feasible and yield a zero payoff. Thus, $v(.)$ is the fix point of an operator that maps positive functions to positive functions implying that $v(z) \geq 0, \forall z$. Thus, there are no endogenous exits in the model, all potential entrants enter and exit exogenously at rate $1 - \nu$ and the stationary measure of firms N satisfies:

$$N = \nu N + \lambda \implies N = \frac{\lambda}{1 - \nu}$$

Note that since we have already used λ to set $C = 1$ and ν is an exogenous parameter, this implies that instead of picking λ directly, we can set N such that $C = 1$ and then use the equation above to find the corresponding λ. Moreover, we now observe that the distribution of productivity among incumbents evolves according to

$$\mathscr{L}_z(z') = \nu \int F(z'|z) \mathscr{L}_z(dz) + (1 - \nu) \int F(z'|z) F^e(dz)$$

showing that $\mathscr{L}_z$ is completely exogenous and determined by the distribution of initial productivity, its evolution over time and the parameter ν.

Going back to the problem of the firm, we observe that the FOCs for m and q are respectively given by:

$$mq\Upsilon'(q) = \alpha^{-1} \left(\frac{mq}{z}\right)^{\alpha^{-1}} + \phi^{-1} P_m^{\phi^{-1}} m^{\phi^{-1}}$$

$$mq\Upsilon'(q) \left(1 + \frac{q\Upsilon''(q)}{\Upsilon'(q)}\right) = \alpha^{-1} \left(\frac{mq}{z}\right)^{\alpha^{-1}}$$

which jointly determine m^* and q^* as a function of z and P_m. Let us denote these as $m^*(z, P_m)$ and $q^*(z, P_m)$. In particular, noting that $\varepsilon(q) \equiv -\frac{\Upsilon'(q)}{\Upsilon''(q)q}$ is the demand elasticity of firms, we can rearrange the equations above to get

$$\phi^{-1} P_m^{\phi^{-1}} m^{*\phi^{-1}} = \frac{m^* q^*}{\varepsilon(q^*)} \implies m^*(z, P_m) = P_m^{-\frac{1}{1-\phi}} \left(\phi \frac{q^*(z, P_m)}{\varepsilon(q^*(z, P_m))}\right)^{\frac{1}{\phi^{-1}-1}}$$

$$q^*\Upsilon'(q^*)\left(1-\varepsilon(q^*)^{-1}\right)=\alpha^{-1}\left(\frac{q^*}{z}\right)^{\alpha^{-1}} m^{*\alpha^{-1}-1}=\alpha^{-1}\left(\frac{q^*}{z}\right)^{\alpha^{-1}}\left(P_m^{-\frac{1}{1-\phi}}\left(\phi\frac{q^*}{\varepsilon(q^*)}\right)^{\frac{1}{\phi^{-1}-1}}\right)^{\alpha^{-1}-1}$$

Noting that we have already reserved the parameter ξ in the labor supply curve to set $\bar{W}=1$ and N (alternatively λ) to set $C=1$, we are left with three more equilibrium conditions that need to be satisfied. Of those three, one gives us equilibrium D as a function of m^* and q^* and the other two are:

$$1=N\int m^*\Upsilon(q^*)\mathscr{L}_z(dz)\implies 1=NP_m^{-\frac{1}{1-\phi}}\int\left(\phi\frac{q^*(z,P_m)}{\varepsilon(q^*(z,P_m))}\right)^{\frac{1}{\phi^{-1}-1}}\Upsilon(q^*(z,P_m))\mathscr{L}_z(dz)$$

$$1=N\int m^*\mathscr{L}_z(dz)\implies 1=NP_m^{-\frac{1}{1-\phi}}\int\left(\phi\frac{q^*(z,P_m)}{\varepsilon(q^*(z,P_m))}\right)^{\frac{1}{\phi^{-1}-1}}\mathscr{L}_z(dz)$$

We note that the two equations above imply that

$$\int\left(\phi\frac{q^*(z,P_m)}{\varepsilon(q^*(z,P_m))}\right)^{\frac{1}{\phi^{-1}-1}}\Upsilon(q^*(z,P_m))\mathscr{L}_z(dz)=\int\left(\phi\frac{q^*(z,P_m)}{\varepsilon(q^*(z,P_m))}\right)^{\frac{1}{\phi^{-1}-1}}\mathscr{L}_z(dz)$$

which is an implicit equation in terms of P_m and pins it down. Once P_m is pinned down, we can use either of the equations; e.g.,

$$1=NP_m^{-\frac{1}{1-\phi}}\int\left(\phi\frac{q^*(z,P_m)}{\varepsilon(q^*(z,P_m))}\right)^{\frac{1}{\phi^{-1}-1}}\mathscr{L}_z(dz)$$

to recover the N (or alternatively $\lambda=(1-\nu)N$) that satisfies $C=1$.

The case of the CES aggregator. In the case above, suppose $\Upsilon(q)=q^{1-\sigma^{-1}}$ which also implies that $\varepsilon(q)=\sigma$. It follows that the equilibrium P_m is given by

$$\int q^*(z,P_m)^{\frac{1}{\phi^{-1}-1}+1-\sigma^{-1}}\mathscr{L}_z(dz)=\int q^*(z,P_m)^{\frac{1}{\phi^{-1}-1}}\mathscr{L}_z(dz)$$

where $q^*(z,P_m)$ is given by

$$q^{*1-\sigma^{-1}}\left(1-\sigma^{-1}\right)^2=\alpha^{-1}\left(\frac{q^*}{z}\right)^{\alpha^{-1}}\left(P_m^{-\frac{1}{1-\phi}}\left(\sigma^{-1}q^*\right)^{\frac{1}{\phi^{-1}-1}}\right)^{\alpha^{-1}-1}\tag{3.7}$$

$$\implies q^*(z,P_m)=\left[\frac{\alpha^{-1}\left(\frac{\phi}{\sigma}\right)^{\frac{\alpha^{-1}-1}{\phi^{-1}-1}}}{\left(1-\sigma^{-1}\right)^2}\right]^{-\frac{\alpha(1-\phi)}{1-\alpha+\alpha\sigma^{-1}(1-\phi)}} z^{\frac{1-\phi}{1-\alpha+\alpha\sigma^{-1}(1-\phi)}}P_m^{\frac{1-\alpha}{1-\alpha+\alpha\sigma^{-1}(1-\phi)}}\tag{3.8}$$

Plugging this in the first equation we obtain the equilibrium P_m as

$$P_m=\left[\frac{\int z^{\frac{\phi}{1-\alpha+\alpha\sigma^{-1}(1-\phi)}}\mathscr{L}_z(dz)}{\int z^{\frac{\phi+(1-\sigma^{-1})(1-\phi)}{1-\alpha+\alpha\sigma^{-1}(1-\phi)}}\mathscr{L}_z(dz)}\right]^{\frac{1-\alpha+\alpha\sigma^{-1}(1-\phi)}{(1-\alpha)(1-\sigma^{-1})}}\left[\frac{1}{\alpha\left(1-\sigma^{-1}\right)^2}\right]^{\frac{\alpha(1-\phi)}{1-\alpha}}\left(\frac{\phi}{\sigma}\right)^{\phi}$$

A note on non-existence with $\alpha=1$. Note that if $\alpha=1$, then the equation pinning down $q^*(z,P_m)$ becomes:

$$\Upsilon'(q^*)\left(1-\varepsilon(q^*)^{-1}\right)=\frac{1}{z}$$

which implies $q^* = q^*(z)$ and is independent of P_m. Thus, the condition that pins down P_m becomes an equality that should hold regardless of any value of P_m:

$$\int \left(\phi \frac{q^*(z)}{\varepsilon(q^*(z))}\right)^{\frac{1}{\phi^{-1}-1}} \Upsilon(q^*(z)) \mathscr{L}_z(dz) = \int \left(\phi \frac{q^*(z)}{\varepsilon(q^*(z))}\right)^{\frac{1}{\phi^{-1}-1}} \mathscr{L}_z(dz)$$

But note that all the objects in this equation are exogenous functions of the z including the distribution $\mathscr{L}_z$, so the equation does not hold for arbitrary choices of $\mathscr{L}_z$, $\Upsilon(.)$, and ϕ; contradicting the existence of a stationary equilibrium.

3.2. Solution Algorithm. The algorithm for the numerical solution of the steady state of the model is as follows:

Step 0: Set up a grid for firm's state $S = M \times Z$. We choose 15 collocation points in each dimension. For Z, we use the 0.0001 and 0.9999 percentiles of the ergodic distribution of the AR(1) productivity process as the grid bounds. For M, the lower bound of the grid is 0, and the upper bound is chosen so that the largest customer base in the solution of any version of the model is smaller than the bound. Given these bounds, we construct power grids to concentrate grid points at lower values for m_{-1} and z.

Step 1: Guess values for C, $\tilde{W} \equiv W/(CD)$ and P_m.

Step 2: Solve firm's problem given $(C, \tilde{W}, P_m)$ by scaling the value function by $1/(CD)$ (this reduces the number of aggregate variables we need to solve for by one). We solve this problem by using projection methods to approximate both the value function $v_t(m_{-1}, z)$ and its expected value $\mathbb{E}\left[V_{t+1}(m, z')|z\right]$. We approximate these functions with the tensor product of a linear spline in the z dimension and a cubic spline in the m_{-1} dimension. We follow a two-step procedure to compute optimal policies. First, for a given candidate m, we compute q, l_s, and l_p by solving the nonlinear FOC for q and using the production function and the law of motion of matches. Second, to optimize the value function with respect to m, we use the golden search method. Having approximated these values and guessed a vector of the spline's coefficients, we combine an iteration procedure and a Newton solver to find the coefficient of the basis function. To compute the expectation in $\mathbb{E}\left[V_{t+1}(m, z')|z\right]$, we rely on the following approximation:

$$\mathbb{E}\left[V_t(m, z')|z\right] = \sum_{i=1}^{50} \omega_i V_t\left(m, \exp(\rho \ln z + \varepsilon_i)\right). \tag{3.9}$$

To construct the nodes ε_i, we generate an equidistant grid of 50 points from 0.0001 to 0.9999 and invert the CDF of the $\mathcal{N}(0, \sigma_z^2)$ distribution. To construct the weights ω_i, we discretize the normal distribution with a histogram centered around the nodes.

Step 3: To approximate the ergodic distribution of firms, we construct a finer grid with 100 and 500 points in the m_{-1} and z direction, respectively. Then, we solve the firm's problem once on the new grid using the approximation to the value functions from the previous step.

To find the ergodic distribution, we rely on the nonstochastic simulation approach of Young (2010). This method approximates the distribution of firms on a histogram based on the finer grid. Since both optimal policies and productivity shocks are allowed to vary continuously, we assign values of m and z that do not fall on points in the grid in the following way. Let $s \equiv (m_{-1}, z)$ denote a firm's state. Then, the transition matrix for a firm's customer base can be constructed as

$$Q_M(s, m'(s)) = \left[\mathbf{1}_{m'(s)\in[m_{j-1},m_j]} \frac{m'(s) - m_j}{m_j - m_{j-1}} + \mathbf{1}_{m'(s)\in[m_j,m_{j+1}]} \frac{m_{j+1} - y'(s)}{m_{j+1} - m_j} \right] \tag{3.10}$$

for all states s in the grid. That is, the transition matrix allocates firms in the histogram based on the proximity of the optimal policy to each point in the finer grid. The transition matrix for productivity shocks is approximated as $Q_Z = \sum_{i=1}^{200} \omega_i Q_{z,i}$, where $Q_{z,i}$ is similarly constructed as in Equation (3.10) for $z'(s) = \exp(\rho \ln z + \varepsilon_i)$. The overall transition matrix is then given by $Q = Q_Z \otimes Q_M$. Finally, the distribution of firms is obtained by iterating until convergence the approximation to the law of motion

$$\mathcal{N} = Q' \left(\nu \mathbf{1}_{v(\boldsymbol{s})\geq 0} \mathcal{N} + \lambda \mathbf{1}_{v(\boldsymbol{s})\geq 0} F^e \right),$$

where F^e is an approximation of the distribution of entrants on the finer grid.

Step 4: Compute aggregate variable X from firms' vectorized policies $x(\boldsymbol{s})$ as $X = (\nu \mathbf{1}_{v(\boldsymbol{s})\geq 0} \mathcal{N} + \lambda \mathbf{1}_{v(\boldsymbol{s})\geq 0} F^e)' x(\boldsymbol{s})$. Compute the residual vector

$$1 = \int_0^N m_i di, \quad 1 = \int_0^N m_i \Upsilon(q_i) di, \text{ and } \quad 1 = \frac{W}{\xi C^\gamma L^\psi}.$$

If the distance is small, stop. Otherwise, update $(C, \tilde{W}, P_m)$ with a Newton method and go to **Step 2**.

3.3. Estimation routine. We estimate the parameters of the model via the Simulated Method of Moments (SMM). More specifically, we choose a set of parameters $\mathscr{P}$ that minimizes the SMM objective function

$$\left(\frac{\boldsymbol{m}_m(\mathscr{P})}{\boldsymbol{m}_d} - 1 \right)' \boldsymbol{W} \left(\frac{\boldsymbol{m}_m(\mathscr{P})}{\boldsymbol{m}_d} - 1 \right),$$

where $\boldsymbol{m}_m$ and $\boldsymbol{m}_d$ are a vector of model-simulated moments and data moments, respectively, and $\boldsymbol{W}$ is a diagonal matrix. To compute the model-simulated moments, we follow these steps:

Step 1: Given a vector of parameters $\mathscr{P}$, we find the steady state of the model. For this, we slightly modify the previous algorithm. Since in the estimation we normalize aggregate output $C = Y = 1$ and the normalized wage $\tilde{W} = 1$ (see Step 1 of the solution algorithm) with the free parameters (λ, ξ), we need to solve for only one aggregate variable, P_m.

Step 2: We simulate 100,000 firms for 150 periods and compute model moments using data from the last 25 periods. When matching moments based on the entire US economy, we use data from all simulated firms. When matching moments based on Compustat data, we impose a filter that mimics selection into Compustat based on firm age and size. On the age dimension, we restrict the simulated sample to those firms that are at least 7 years old, as in Ottonello and Winberry (2020). On the size dimension, we restrict the sample to firms with sales above 19% of the average sales in the simulated economy. This cutoff corresponds to the ratio of the 5th percentile of the sales distribution in Compustat (USD1.06 million) to the average firm sales in SUSB (USD5.7 million) in 2012.

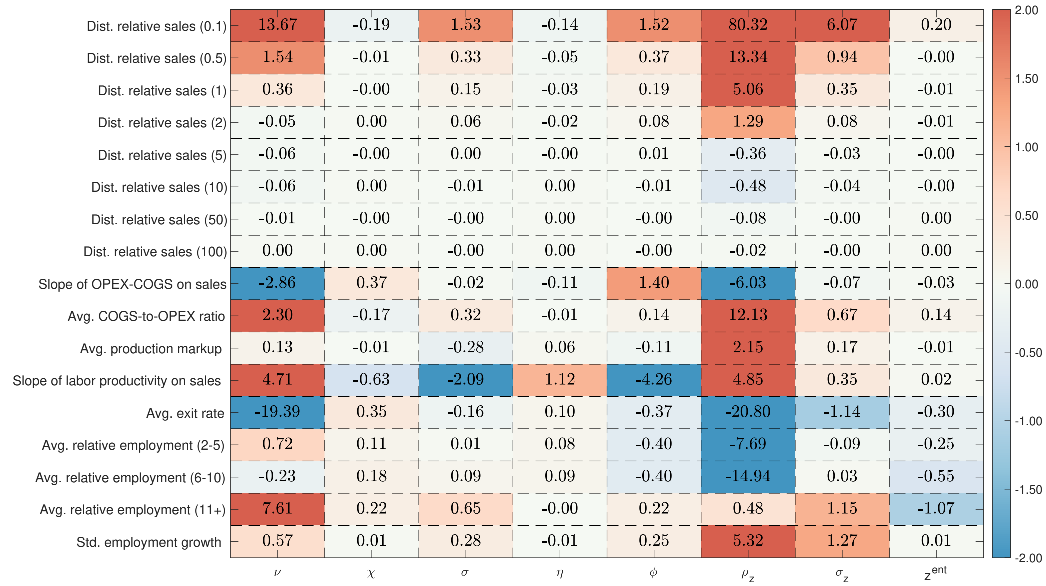
To minimize the SMM objective function and be confident of reaching the global minimum, we follow a two-step procedure in the spirit of Arnoud, Guvenen, and Kleineberg (2019). In the first step, we construct 500 quasi-random vectors of parameters $\mathscr{P}$ from a Halton sequence, which is a deterministic sequence designed to cover the parameter space evenly. After computing the SMM objective in those points, we choose the 30 parameters vectors with the lowest objective values. In the second step, we initiate a local Nelder-Mead optimizer from each of the 30 starting points and select the local minimum with the lowest objective value.

4 Additional Model Analysis

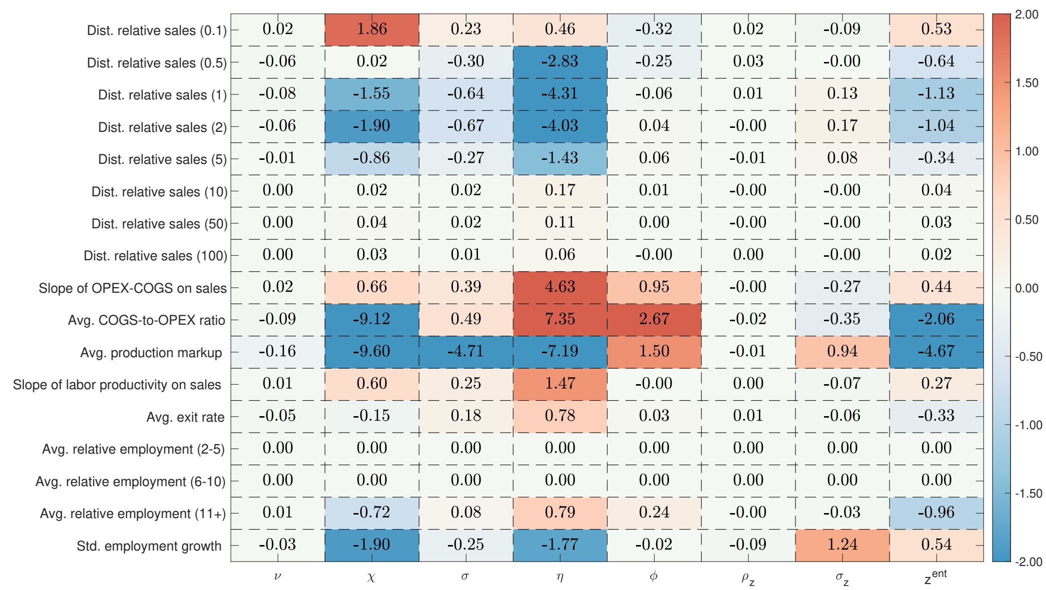
4.1. Identification of Model Parameters. In this section, we formally guide the discussion of the identification of model parameters. Panel A of Figure SM.4.1 shows the *local* elasticity of simulated moments (rows) with respect to parameters (columns), evaluated at the calibrated parameters. In general, the intuition behind the choice of targets is borne out in the model. For example, a higher exogenous exit rate $1-\nu$ mechanically increases the average exit rate. Similarly, a higher super-elasticity of demand η increases the comovement between revenue labor productivity and sales. A higher elasticity of the matching function ϕ makes the positive relationship between a firm's SGA expenses and sales stronger and reduces the comovement between labor productivity on sales, as predicted in the simple version of the model in Propositions 3 and 4. The persistence of productivity shocks ρ_z affects multiple moments, but it affects most strongly the dispersion of the sales distribution. Finally, a smaller average productivity of entrants $\bar{z}_{ent}$ increases the relative size of old firms.

We complement this discussion by analyzing the sensitivity measure developed by Andrews, Gentzkow, and Shapiro (2017), which shows the sensitivity of model parameters with respect to targeted moments. To render the numbers more comparable, we convert this measure into elasticities and plot $\left(J'(\mathscr{P})WJ(\mathscr{P})\right)^{-1}J'(\mathscr{P})Wm_m(\mathscr{P})/\mathscr{P}$, where $J(\mathscr{P})$ is the Jacobian evaluated at calibrated parameters, W is the weighting matrix, and $m_m(\mathscr{P})$ are the model moments evaluated at calibrated parameters. Panel B of Figure SM.4.1 shows that overhead cost χ is quite sensitive to the average COGS-to-OPEX ratio in the data. Similarly, the elasticity of substitution σ is sensitive to the average production markup, and the standard deviation of productivity shock σ_z is most strongly influenced by the standard deviation of employment growth.

Figure SM.4.1: Parameter Identification



(a) Sensitivity of Moments



(b) Sensitivity of Parameters

Notes: Panel A shows the sensitivity of simulated moments to parameters by computing the local elasticity of moments with respect to parameters. Panel B shows the sensitivity of calibrated parameters to moments by constructing the sensitivity measure of Andrews, Gentzkow, and Shapiro (2017) and converting it into an elasticity. Both measures are evaluated at the calibrated parameters.

4.2. Calibration Results: Goodness of Fit. Table SM.4.1 and Figure SM.4.2 show the targeted moments and their model counterparts. Overall, the model closely matches the targets. The model reproduces the average cost structure very well, but it slightly under-predicts the relationship between SGA and sales. The model matches the cost-weighted average markup well and generates a similar relationship between revenue productivity of labor and sales. Figure SM.4.2 shows that the model approximates the sales distribution of firms accurately. For example, in the data 33% of firms have sales that are lower than 10% of the average sales in the economy, and 1% of firms have sales that are larger than 10 times the average sales. In the model, these shares are 25% and 1.5%. Figure SM.4.2 also shows that the model is able to replicate the relative size of old firms: 6.07 in the data and 6.4 in the model. In Appendix 4, we show the untargeted data and model relationships between average labor productivity and SGA with sales, and the average COGS-to-OPEX ratio by firm age and size. Although we targeted specific moments that summarize these relationships in the calibration exercise, the model matches the data patterns more broadly.

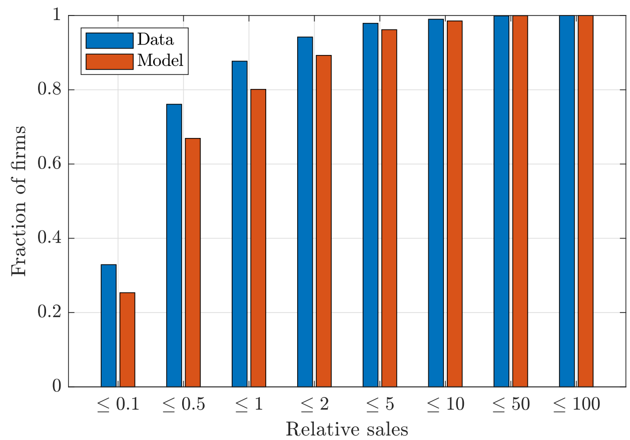
Table SM.4.1: Targeted Moments

Moment	**Data**	**Model**
Slope SGA on sales	0.521	0.490
Avg. COGS-to-OPEX ratio	0.660	0.668
Avg. cost-weighted production markup	1.250	1.274
Slope labor prod. on sales	0.036	0.031
Avg. exit rate	0.073	0.070
SD. employment growth	0.416	0.448

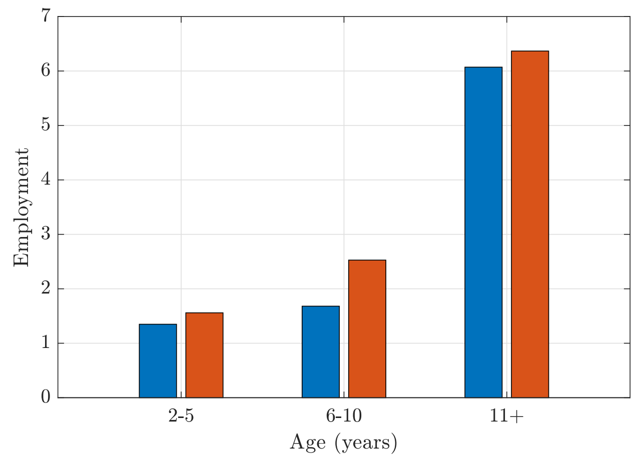
Notes: This table shows the set of moments targeted in the calibration of the model. Slope SGA on Sales refers to the OLS coefficient of the regression $SGA_{i,t} = c + \beta Sales_{i,t} + \psi COGS_{i,t} + \varepsilon_{i,t}$. Avg. COGS-to-OPEX ratio refers to the average of the ratio across firms. Avg. cost-weighted production markup corresponds to the COGS-weighted average markup from Edmond, Midrigan, and Xu (2022). These moments were computed using data from Compustat in 2012. Slope labor prod. on sales corresponds to the OLS coefficient of the sales-weighted regression of relative revenue labor productivity on relative sales from Edmond, Midrigan, and Xu (2022), restricting the sample of firms to those with relative sales above one. This moment was computed using data from the SUSB in 2012. The average exit rate was obtained from the BDS in 2012. The standard deviation of annual employment growth for continuing establishments is obtained from Elsby and Michaels (2013). The growth rate of variable x is computed as in Davis and Haltiwanger (1992): $(x_{i,t} - x_{i,t-1})/(0.5(x_{i,t} + x_{i,t-1}))$. The last column shows the model counterparts of each moment, which were obtained by simulating a panel of firms and computing each moment with the simulated data. In the model, we account for selection into Compustat by restricting the simulated sample of firms to those that are at least 7 years old and have sales above 19% of the average sales in the simulated economy (which corresponds to the ratio of the 5th percentile of the sales distribution in Compustat to the average sales in SUSB).

Figure SM.4.3 also plots the relationship between relative revenue productivity of labor and relative sales in the data (SUSB) and the model (both the raw data and a linear fit). The model is able to match the positive association between these variables well. Figure SM.4.4 shows the relationship between relative SGA expenses and relative sales in the data (Compustat) and the model. Although

Figure SM.4.2: Model Fit: Firm Size



(a) Distribution of Sales

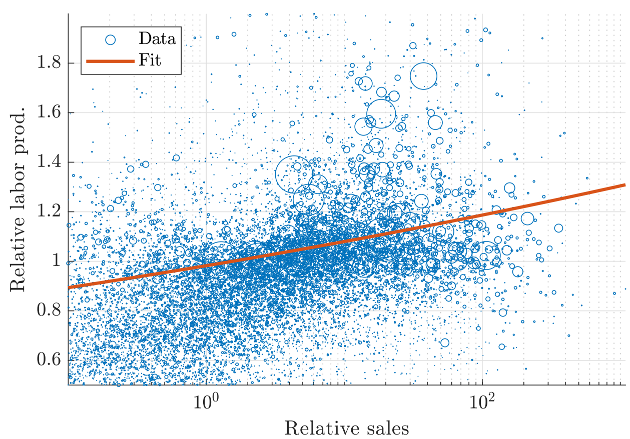


(b) Avg. Size by Age

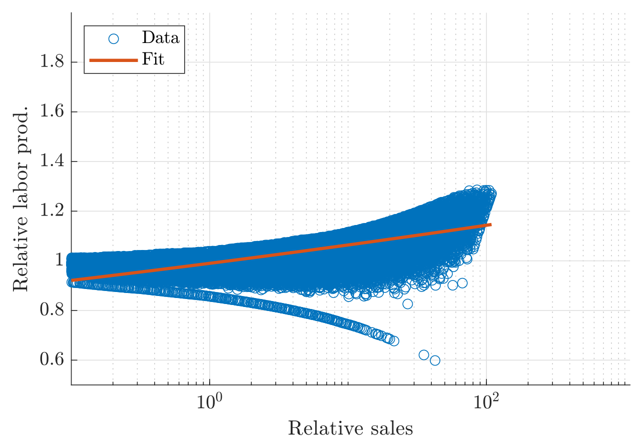
Notes: This figure shows moments targeted in the calibration of the model. Panel (a) shows the model fit of the distribution of relative sales. The distribution of relative sales is obtained from the SUSB in 2012. Panel (b) shows the model fit of average employment, relative to 1-year-old firms, by firm age. Average firm employment by age group was obtained from BDS in 2012. In the calibration exercise, we only target the relative size of firms older than 10 years.

the "Compustat-equivalent" sample from the model does not generate all of the dispersion in relative sales as in the data, the relationship with relative SGA expenses is well matched in the overlapping range of relative sales. Finally, Figure SM.4.5 plots the average COGS-to-OPEX ratio as a function of a firm's age and size in the data (Compustat) and the model. Here, age is normalized as time since entry into Compustat (which in the model occurs after the 7th year). In the model, the composition of firms' costs exhibits a strong size profile and a weak age profile, as in the data.

Figure SM.4.3: Model Fit: Labor Productivity and Sales



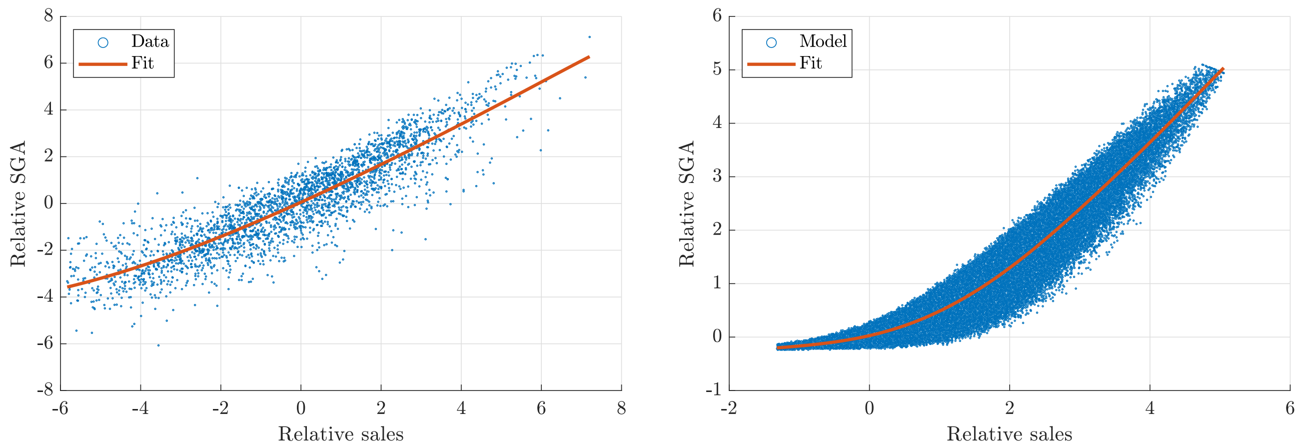
(a) Data



(b) Model

Notes: This figure plots the relationship between the relative revenue productivity of labor and relative sales. Panels A and B show the relationship obtained from the SUSB data and the model-simulated data, respectively. Each figure includes the best linear fit of the data. The x-axis is in log scale.

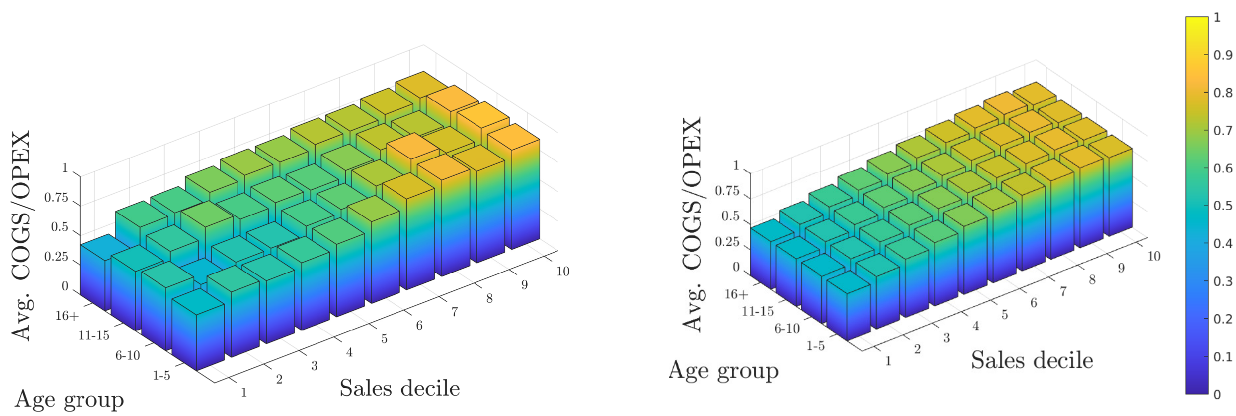
Figure SM.4.4: Model Fit: SGA and Sales



(a) Data (b) Model

Notes: This figure plots the relationship between relative spending in SGA and relative sales. Panels A and B show the relationship obtained from the Compustat data and the model-simulated data, respectively. The model data are obtained by simulating the model and restricting the sample to firms that are at least 7 years old and have sales above 19% of the average sales in the simulated economy (which corresponds to the ratio of the 5th percentile of the sales distribution in Compustat to the average sales in SUSB). Each figure includes the local linear kernel best fit of the data.

Figure SM.4.5: Steady-state COGS/OPEX by Size and Age

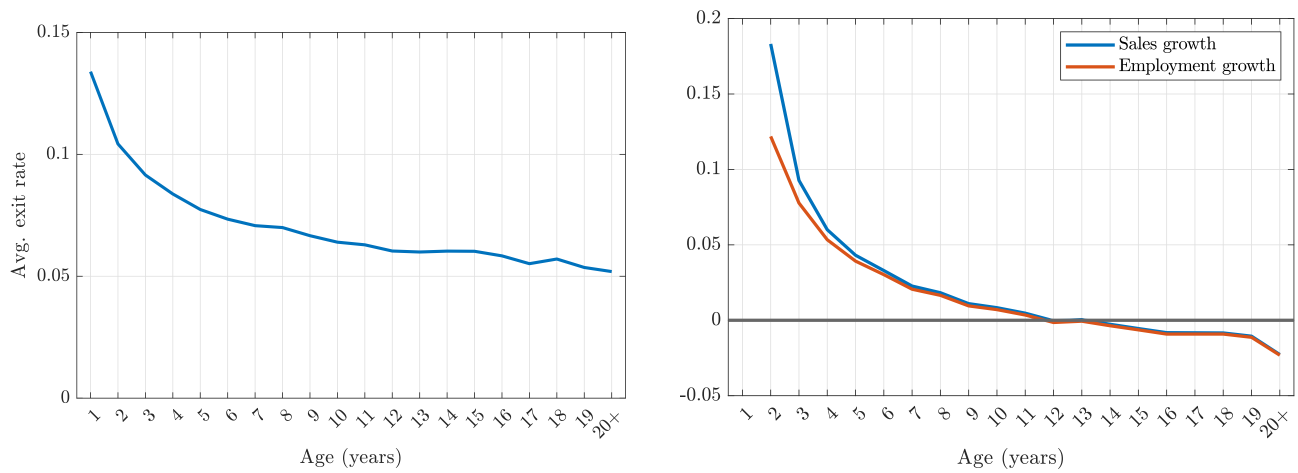


(a) Data (b) Model

Notes: This figure plots the average COGS-to-OPEX ratio as a function of a firm's age and size. Panels A and B show the relationship obtained from the Compustat data and the model-simulated data, respectively. The model data are obtained by simulating the model and restricting the sample to firms that are at least 7 years old and have sales above 19% of the average sales in the simulated economy (which corresponds to the ratio of the 5th percentile of the sales distribution in Compustat to the average sales in SUSB). Age is normalized as years since entry into Compustat, which in the model corresponds to year 7.

Untargeted Moments. Figure SM.4.6 shows two model-based moments that were not explicitly targeted in the calibration exercise: the average exit rate by age and the average employment growth by age. As Panel A shows, in the model, the average exit rate is decreasing in a firm's age (as in the data; see, e.g., Haltiwanger, Jarmin, and Miranda, 2013). The fact that entrants enter the economy with lower average productivity and no customer base makes young firms more likely to exit when faced with negative productivity shocks due to overhead costs. As firms grow larger, their larger customer base and higher productivity allow them to absorb negative productivity shocks without forcing them to exit. It is important to note that in the firm dynamics literature, the decreasing profile of the average exit rate by age is typically used as a target to indirectly calibrate the size of the overhead cost χ. Here, we took a more direct approach by matching the observed cost structure of firms. The fact that the model replicates the profile of exit rates provides additional support to the model of the composition of firms' cost structures. Panel B plots decreasing profiles of average employment and sales growth as a function of firms' age. Both patterns are consistent with the empirical evidence of Haltiwanger, Jarmin, and Miranda (2013). For example, in the data, the average net employment growth rate of 1-2 year-old and 7-8 year-old continuing firms is close to 12% and 2.5%, respectively. In the model, the average employment growth rates (computed as in Davis, Haltiwanger, Schuh, et al. (1998)) for 2- and 7-year-old firms are 12.5% and 2.1%, respectively.

Figure SM.4.6: Exit and Growth by Age



(a) Exit Rate by Age

(b) Avg. Growth by Age

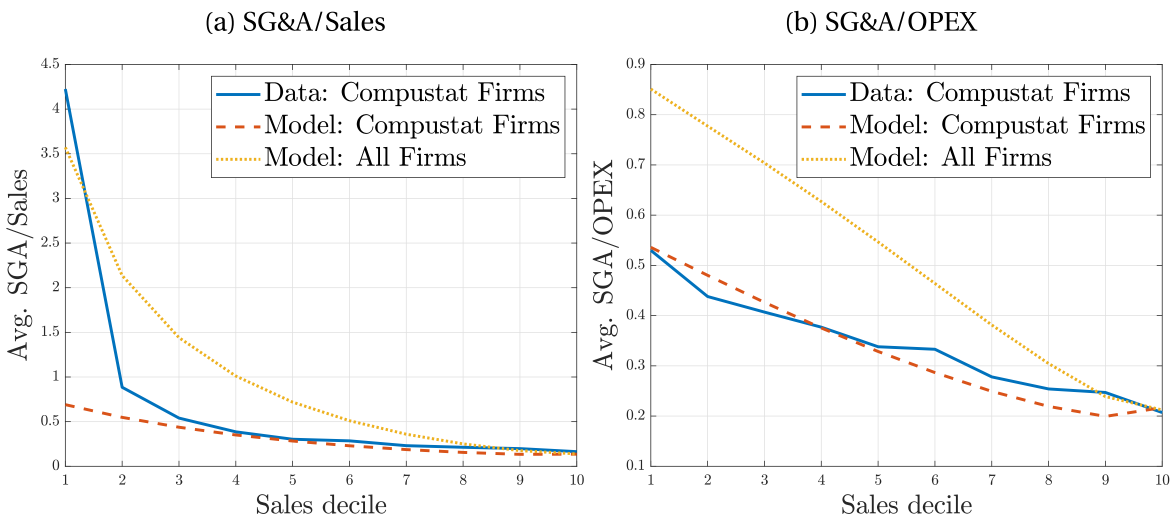
Notes: Panels (a) and (b) plot model predictions regarding two untargeted moments—the average exit rate by age and the average employment growth by age—respectively.

Validation of the Customer Acquisition Technology. In this section, we further validate the customer acquisition technology by comparing the model's predictions related to SG&A expenditures with empirical data. We focus on two key relationships previously documented in the literature: (i) the behavior of the SG&A/Sales and SG&A/OPEX ratios across the firm size distribution, and (ii) the relationship between firm age and SG&A intensity.

Regarding the first fact, Figure SM.4.7a shows the average SG&A-to-Sales ratio across firm size deciles. As documented by Cavenaile and Roldan-Blanco (2021), this ratio declines with firm size.

The figure also presents the model-simulated results both for the entire sample of firms and the subset that satisfies the Compustat filters. Figure SM.4.7b complements this by showing the average SG&A/OPEX ratio by firm size, which highlights the evolving nature of firms' cost structure. Both, in the model and in the data, the SG&A/OPEX ratio also declines in firm size. In both dimensions, the model closely replicates the empirical patterns: the ratios decline as firm size increases. This pattern arises because smaller firms in the model are disproportionately composed of entrants who must invest heavily in SG&A to build their customer base and offset its depreciation over time. In contrast, larger firms allocate SG&A primarily to maintain their existing customer base against depreciation and respond to occasional productivity shocks. As a result, when normalizing SG&A by either sales or OPEX, the model predicts a declining ratio with firm size. This is consistent with empirical observations and reflects the diminishing relative importance of customer acquisition costs as firms grow in size.

Figure SM.4.7: Firm Cost Structure by Size: Data vs. Model



Regarding the second fact, we replicate the analysis of Argente, Fitzgerald, Moreira, and Priolo (2021), who show that incumbent firms spend more on SG&A than entrants. Using model-simulated data, we estimate the following regression:

$$\ln(\text{SGA}_{it}) = \alpha + \sum_{k=2}^{5} \beta_{E,k} \cdot \mathbf{1}[\text{Entrant}_k]_{it} + \beta_I \cdot \mathbf{1}[\text{Incumbent}]_{it} + \varepsilon_{it} \tag{4.1}$$

where $\ln(\text{SGA}_{it})$ is the log of SGA expenditures for firm i at time t. A firm is considered an entrant in the first 4 years after entry; an incumbent is a firm that is at least 5 years old. The dummy variables $\mathbf{1}[\text{Entrant}_k]$ indicate that the firm that is an entrant that survived exactly k years. The indicator $\mathbf{1}[\text{Incumbent}]$ equals one if the firm is an incumbent. The baseline category is firms that survived only one year. To ensure comparability with Argente, Fitzgerald, Moreira, and Priolo (2021), who use data from 2010 to 2016, we restrict the analysis to the first seven years of the simulated panel (age ≤ 7).

Table SM.4.2 report the results. The coefficients on entrant survival dummies increase monotonically with survival duration, suggesting that longer-lived entrants spend more in SGA. The incumbent coefficient is 0.048, indicating that incumbents spend almost 5% more on SGA than the baseline entrants. While the magnitudes are not directly comparable to that of Argente, Fitzgerald, Moreira, and Priolo (2021) due to differing dependent variables, the qualitative pattern is consistent. In the model, this results from the fact that larger firms, which are typically incumbents, spend more resources (in levels) to compensate for the depreciation of their larger customer bases.

Table SM.4.2: The Lifecycle of Expansionary Activities

	(1)
Entrant $\beta_{E,2}$	0.012^{***}
	(0.004)
Entrant $\beta_{E,3}$	0.012^{***}
	(0.004)
Entrant $\beta_{E,4}$	0.025^{***}
	(0.004)
Entrant $\beta_{E,5}$	0.037^{***}
	(0.005)
Incumbent β_I	0.048^{***}
	(0.004)
Observations	158032
R^2	0.003

Notes: Standard errors are clustered at the firm level. Incumbents include only firms aged 5-7 years.

4.3. Model Mechanisms. In Section 4.3, we compared our calibrated model with a recalibrated homogeneous customer model. Here, we further highlight the mechanisms at play by providing comparative statics with the homogeneous customer model, which shuts down endogenous customer acquisition (i.e., we set $\phi \to 0$ and $\delta = 1$, while keeping the remaining parameter values fixed across models).

Implications for Concentration. Figure SM.4.8 plots the firm dynamics of two entrants that start with productivities equal to the 50th and 99th percentiles of the productivity distribution of entrants. After these initial draws, productivity follows the AR(1) process without further shocks. Panel (a) shows that in the baseline model (depicted with solid lines), firms frontload their efforts to acquire customers when young, with more productive firms spending more on advertising labor, $l_{i,s,t}$. Panel (b) shows that the sales of the more productive firm are larger than those of the less productive firm, and also that sales increase initially due to a growing stock of customers but eventually decline due to the mean-reversion of its productivity.

How do firms grow their sales? Panels (c) and (d) of Figure SM.4.8 plot the dynamics of the number of customers and sales per customer, respectively. Both firms build their customer base gradually over time due to decreasing returns to advertising. However, sales per customer are higher when firms are younger due to mean-reverting productivity and the decreasing marginal product of labor in production. Thus, over time, firms shift their sales strategy from selling more to a few customers to selling less to more customers.

How does *endogenous* customer acquisition affect these dynamics? Relative to the homogeneous customers model (depicted with dashed lines), the more productive firm is able to achieve higher sales by selling less per customer but accumulating more than twice as many customers in the first year. Since more productive firms charge lower prices but higher markups, profits per marginal customer are increasing in productivity, which induces the more productive firms to accumulate customers more rapidly. Given a fixed stock of customers, this can only be possible if firms with lower productivity accumulate fewer customers relative to the homogeneous customers model. Thus, endogenous customer acquisition increases the dispersion across firms of the number of customers and decreases the dispersion of sales per customer. Table SM.4.3 shows that the former effect dominates and the overall concentration of sales increases: Whereas in the homogeneous customers model the 5% largest firms capture 17% of sales, in our baseline model they capture 50% of sales.

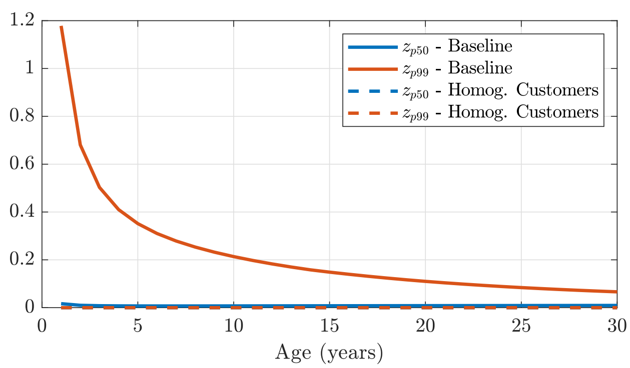
Implications for Market Power. Table SM.4.3 shows that despite higher concentration, the baseline model features a lower aggregate markup: 1.26 as opposed to 1.38 in the homogeneous customers model. To understand the sources of this difference, Figure SM.4.9 shows the histogram of markups and the scatter plot between markups and relative employment (the weights used in the construction of the aggregate markup) across the two models. These figures illustrate two forces. On the one hand, in the model with endogenous customer acquisition, the distribution of markups is more concentrated. High-productivity firms charge lower markups than firms with similar productivity in the homogeneous customers model. On the other hand, in our model, those high-productivity firms account for a larger fraction of total employment.

The following decomposition of the difference in aggregate markups, which follows from Equation (29), quantifies each force:

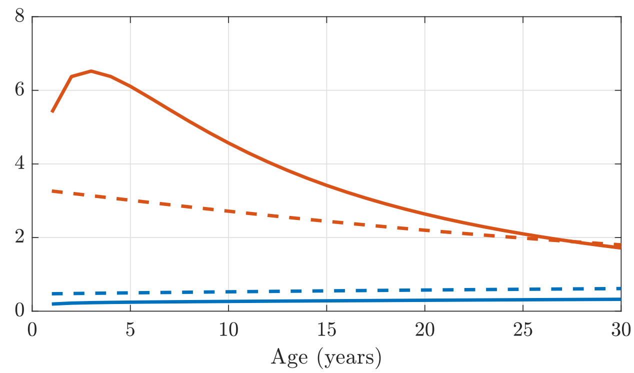
$$\underbrace{\ln(\mathcal{M}_t)-\ln(\mathcal{M}_t^{HC})}_{-9.35\%} \approx \underbrace{\ln\left(\frac{\int_{i\in N_t}\omega_{i,t}\mu_{i,t}di}{\int_{i\in N_t^{HC}}\omega_{i,t}^{HC}\mu_{i,t}di}\right)}_{\Delta\text{ Distribution: }8.91\%}+\underbrace{\ln\left(\frac{\int_{i\in N_t^{HC}}\omega_{i,t}^{HC}\mu_{i,t}}{\int_{i\in N_t^{HC}}\omega_{i,t}^{HC}\mu_{i,t}^{HC}di}\right)}_{\Delta\text{ Market power: -18.26\%}},$$

where the superscript HC denotes distributions and allocations in the homogeneous customers

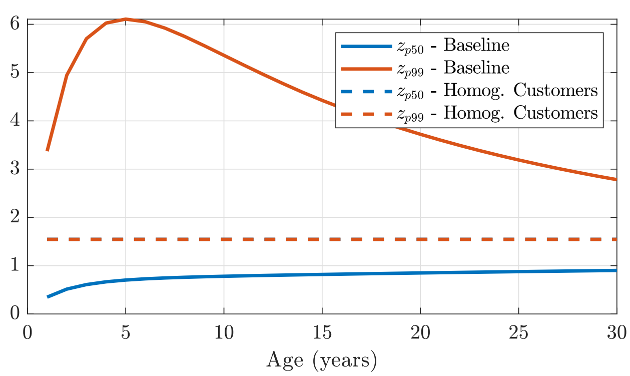
Figure SM.4.8: Average Firm Dynamics



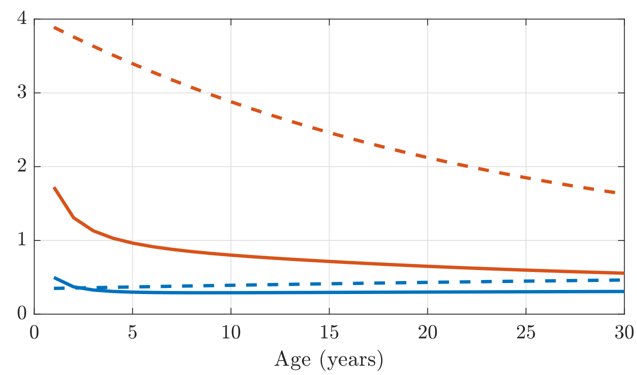
(a) $l_{i,s,t}$



(b) $p_{i,t}y_{i,t}$



(c) $m_{i,t}$



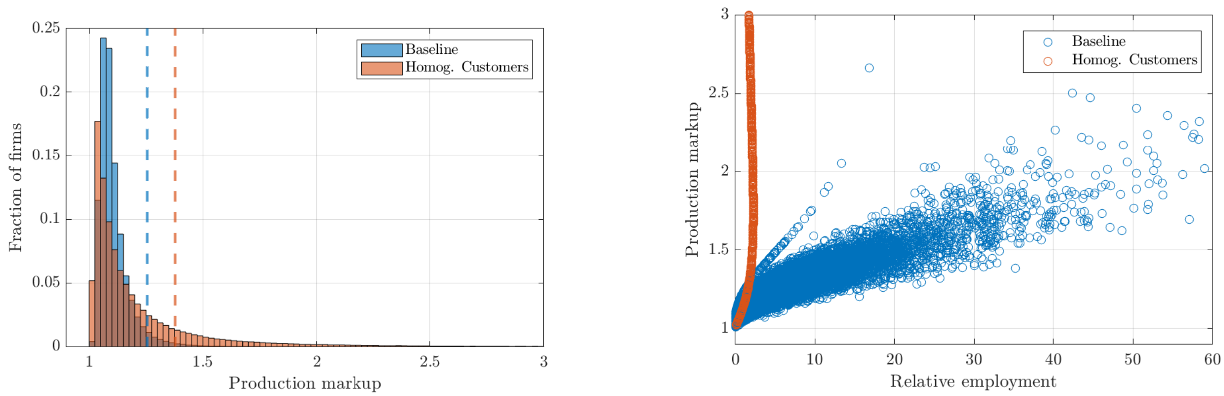
(d) $q_{i,t}$

Notes: This figure plots the firm dynamics of two new firms that start with zero customer base and productivities equal to the 50th and 99th percentiles of the distribution of productivities among entrants. After this initial draw, firms follow the AR(1) productivity process without any further shock. Solid lines correspond to the baseline model. Dashed lines correspond to the homogeneous customers model with an exogenous and identical customer base ($m_{i,t} = 1/N_t$). For the latter, we compute the general equilibrium using the calibrated parameters of the baseline model and imposing the parameter restrictions $\phi \to 0$ and $\delta = 1$. Panels (a)-(d) plot the evolution of labor devoted to customer acquisition ($l_{i,s,t}$), sales ($p_{i,t}y_{i,t}$), customer base ($m_{i,t}$), and output per customer ($q_{i,t}$), respectively.

model.[1] The first term (denoted "Δ Distribution") captures the contribution of differences in the distribution of relative employment (cost-based Domar weights) across firms while keeping firms' markups fixed at their level in the baseline model. The second term (denoted "Δ Market power") captures the contribution of differences in markups across models while keeping the distribution of relative employment fixed at its distribution in the homogeneous customers model. Since more productive firms charge higher markups, switching the distribution of relative employment from the homogeneous customers model to the baseline model *increases* the average markup by 8.91 pp. However, the contribution of lower markups in the baseline model *reduces* the aggregate markup by 18.26 pp, so the aggregate markup is on net smaller by 9.35 pp. To summarize, why does the

[1]While the equation presents the decomposition based on its approximation for expositional purposes, the numbers we present are computed based on the exact decomposition.

Figure SM.4.9: Customer Acquisition and Market Power



(a) Distribution of Markups

(b) Markups vs. Relative Employment

Notes: Panel (a) plots the distribution of production markups in the baseline and homogeneous customers models. Vertical dashed lines show the average cost-weighted production markup in each model. Panel (b) shows the scatter plot of relative employment ($l_{i,p,t}/(L_{p,t}/N_t) \propto \omega_{i,t}$) and production markups $\mu_{i,t}$. The homogeneous customers model refers to the model with an exogenous customer base ($m_{i,t} = 1/N_t$).

baseline model feature a much higher concentration but a lower aggregate markup? Because in the baseline model, firms grow through larger customer bases $m_{i,t}$ rather than higher average sales per customer $p_{i,t}q_{i,t}$, which reduces their market power (but increases their lifetime profits).

Aggregate Implications. Beyond market power, endogenous customer acquisition also affects other aggregate outcomes by concentrating consumers among higher-productivity firms. First, relative to the homogeneous customers model, it allocates more of the economy's resources toward more productive firms, which reflects itself in the higher aggregate TFP derived in Equation (26). Table SM.4.3 shows that shutting down endogenous customer acquisition in our model reduces aggregate TFP by 28.6%. This is because in the baseline model, more productive firms employ a higher share of production resources to meet the demand from their larger customer bases.

Second, the concentration of customers among more productive firms leaves fewer customers for firms at the bottom of the productivity distribution and brings them closer to the exit threshold. Therefore, when we restrict customer bases to be equal across firms, the equilibrium number of firms increases by 54%. This additional inflow of firms comes from less productive firms that can now generate positive discounted profits due to a larger (exogenous) customer base.

The lack of endogenous customer acquisition in the restricted model also implies that high productivity firms can no longer serve a large customer base and therefore opt for higher markups and lower production. As a result, fixing an aggregate wage, labor demand reallocates from highly productive firms to less productive ones and aggregate production labor demand declines by 5.6 percent, as shown in Table SM.4.3. The aggregate labor demand falls less (3.7 percent) than the production labor because more firms enter in the restricted model, which features higher overhead costs. Finally, the changes in aggregate TFP and production labor lead to an overall 32.2% decline in aggregate output.

Table SM.4.3: Aggregate Effects of Customer Acquisition

	Baseline Model	**Homog. Customers Model**
TFP		-28.6
Output		-32.2
Employment		-3.7
Production		-5.6
Number of firms		54.2
Agg. markup	1.26	1.38
Top 5% sales share	0.50	0.17

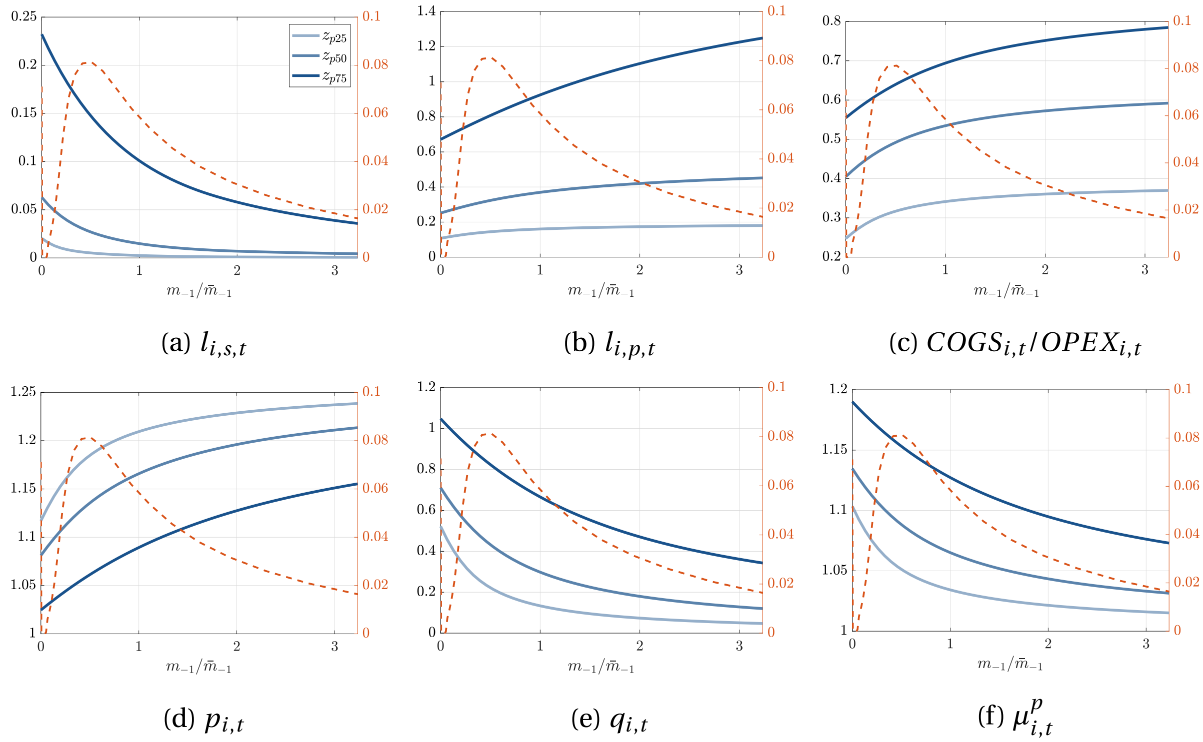
Notes: The table reports equilibrium aggregates in the baseline and homogeneous customers versions of the model. The homogeneous customers model refers to the model with an exogenous customer base ($m_{i,t} = 1/N_t$). The second column reports percentage differences with respect to aggregates in the baseline model, with the exception of the aggregate markup and the top 5% sales share, which are reported in levels.

4.4. Additional Analysis of Model in Steady State. This section further describes how the model works in the steady state. First, we describe firms' optimal policies. Then, we show the average firm dynamics, taking selection into account.

Firms' Optimal Policies. Figure SM.4.10 shows firms' steady-state optimal policy functions for three productivity levels (the 25th, 50th, and 75th percentiles of the marginal productivity distribution in steady state). The y-axis on the right plots the marginal distribution of the relative customer base.

While optimal spending in $l_{i,s,t}$ decreases with the size of the customer base, production labor is increasing in a firm's customer base. Thus, when firms have a small customer base, they spend more resources to increase it. However, due to decreasing returns to customer accumulation, firms increase their customer base gradually over time. As firms grow, they spend less on customer acquisition and more on producing goods to satisfy the growing demand. This is reflected in a firm's cost structure: The average COGS-to-OPEX ratio is also increasing in m_{-1}. As total output increases due to a larger customer base, the marginal cost of production also increases since production is subject to decreasing returns. This raises the price charged by the firm, which in turn reduces the consumption per capita $q_{i,t}$ and optimal markups. The figure also shows that for a given level of m_{-1}, spending in $l_{i,s,t}$ is increasing in a firm's productivity. A higher productivity allows firms to charge lower prices and higher markups. Thus, profits per marginal customer are increasing in productivity, which incentivizes firms to accumulate customers more quickly by spending more on $l_{i,s,t}$.

Figure SM.4.10: Firms' Optimal Policies



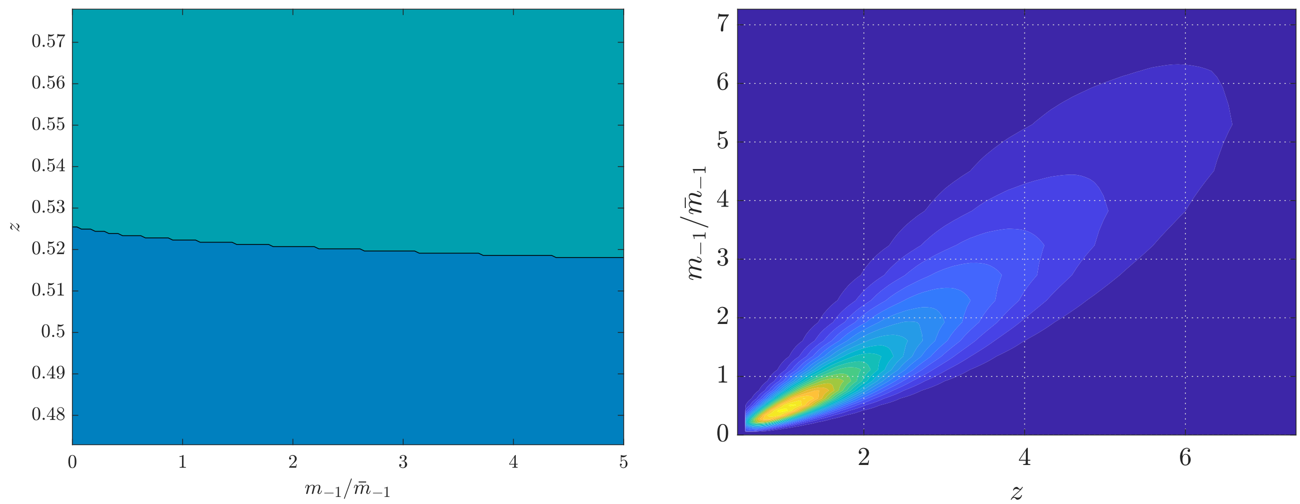
(a) $l_{i,s,t}$ (b) $l_{i,p,t}$ (c) $COGS_{i,t}/OPEX_{i,t}$

(d) $p_{i,t}$ (e) $q_{i,t}$ (f) $\mu^{p}_{i,t}$

Notes: These figures plot firms' policy functions in steady state. Each figure shows policies as a function of relative customer base for three levels of productivity: the 25th, 50th and 75th percentiles of the stationary productivity distribution. The y-axis on the right plots the stationary marginal distribution of the relative customer base.

Figure SM.4.11 plots firms' optimal exit policies and the stationary joint distribution of (m_{-1}, z). Panel A shows the threshold productivity $z^*(m_{-1})$ such that if $z < z^*(m_{-1})$, the firm optimally chooses to exit. The figure shows that $z'^*(m_{-1}) < 0$—that is, firms with larger customer bases are able to survive large productivity shocks without the need to exit the market. Although a lower productivity reduces markups and profits per customer, aggregate profits are increasing in a firm's customer base. Panel B shows that there is a positive correlation between firms' productivities and customer bases in the steady state.

Figure SM.4.11: Optimal Exit and Stationary Distribution



(a) Optimal exit policy

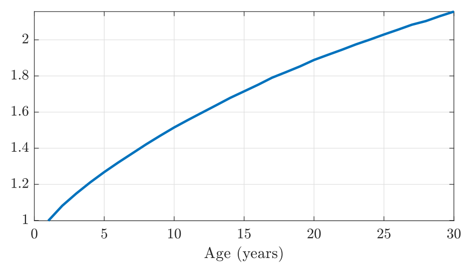
(b) Stationary distribution of (m_{-1}, z)

Notes: Panel A plots the exit threshold $z^*(m_{-1})$ such that if $z < z^*(m_{-1})$, the firm optimally chooses to exit. Panel B shows the contour plot of the stationary joint distribution of (m_1, z), censored at the 99th percentile of each variable.

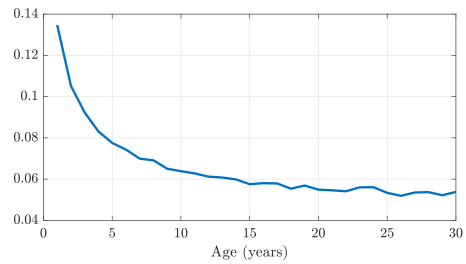
Average Firm Dynamics with Shocks. Figure SM.4.12 plots the average firm dynamics, taking selection into account. To construct this figure, we simulate a cohort of firms that starts with a zero customer base and draws productivities from the distribution of entrants. Since firms are subject to productivity shocks, some decide to exit over their lifetime. The figure plots the average of each variable across firms that survived up to a given age.

Firms start with lower productivity, which grows over time due to the calibrated lower productivity of entrants and endogenous exit. Conditional on a productivity level, firms frontload spending on customer acquisition, and the average customer base and marginal costs rise rapidly for young firms. This, in turn, increases prices and reduces average output per customer and markups. Over time, only the most productive firms survive, so the average marginal cost and price decline, and output per customer and markups increase.

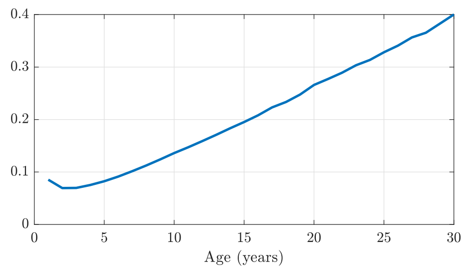
Figure SM.4.12: Average Firm Dynamics



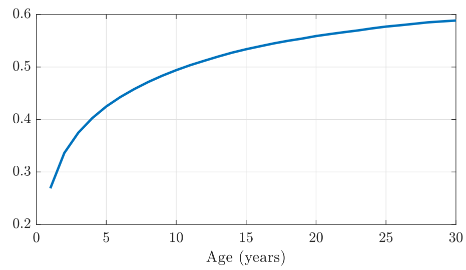
(a) Productivity



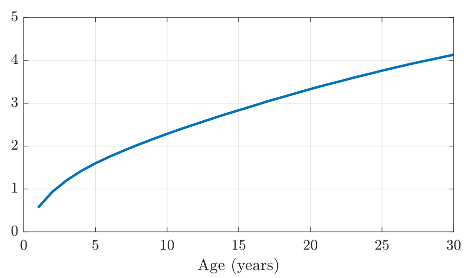
(b) Exit rate



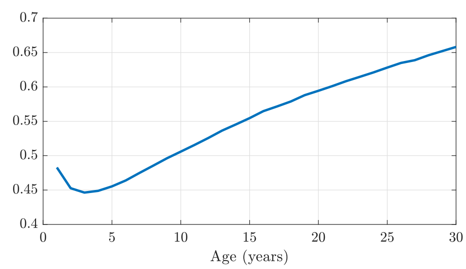
(c) $l_{i,s,t}$



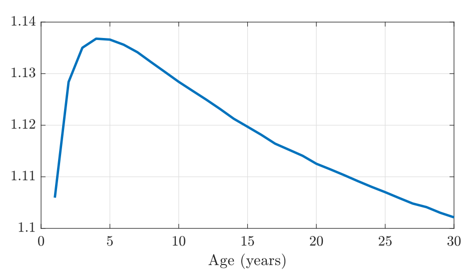
(d) $COGS_{i,t}/OPEX_{i,t}$



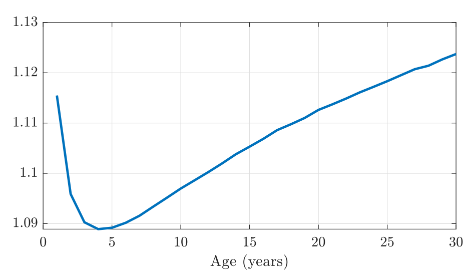
(e) $m_{i,t}$



(f) $q_{i,t}$



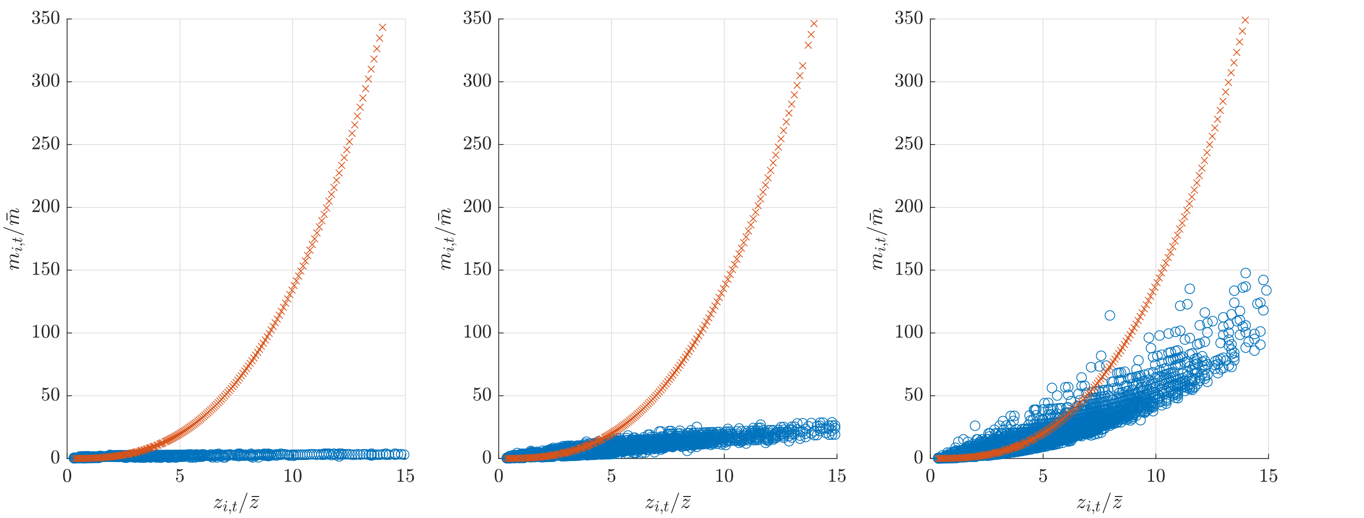
(g) $p_{i,t}$



(h) $\mu^{p}_{i,t}$

Notes: The figure plots the average firm dynamics, which are obtained by simulating a cohort of firms that start with $m_{-1} = 0$, draw z from the distribution of entrants, and experience productivity shocks over their lifetime. Each figure plots the average of a variable as a function of firms' age across firms that survived up to that age.

Figure SM.4.13: Allocation of Customers: Equilibrium vs. Efficient Allocation



(a) Low ϕ (b) Baseline ϕ (c) High ϕ

Notes: This figure shows a scatter plot between relative productivity $z_{i,t}/\bar{z}$ and relative customer bases $m_{i,t}/\bar{m}$, for both the equilibrium and the social planner's allocation. We present three plots by varying the value of ϕ, while keeping the remaining parameters fixed at the values in the baseline calibration. Low ϕ corresponds to 0.25, baseline to 0.53, and high to 0.75.

5 Calibration of the Homogeneous Customers Model

Table SM.5.1: Model Parameters: Homogeneous Customers Model

Parameter	Description	Value
	Panel A: Fixed Parameters	
β	Annual discount factor	0.960
γ	Elast. of intertemporal substitution	1.000
ψ	Frisch elasticity	1.000
α	Decreasing returns to scale	0.640
δ	Prob. of losing customer	1.000
	Panel B: Calibrated Parameters	
χ	Overhead cost	0.704
σ	Avg. elasticity of substitution	5.998
η	Superelasticity	0.559
ν	Exog. survival probability	0.960
ρ_z	Persistence of productivity shock	0.981
σ_z	SD of productivity shock	0.259
$\bar{z}_{ent}$	Mean productivity of entrants	-2.379
λ	Mass of entrants	0.123
ξ	Disutility of labor supply	2.246

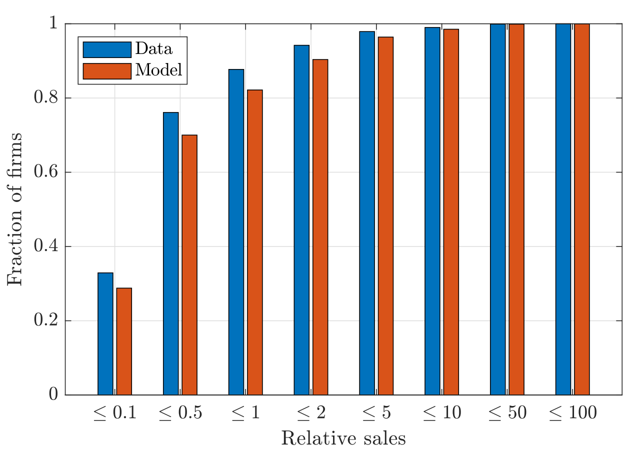
Notes: This table shows the calibration of the homogeneous customers model with an exogenous customer base. Panel A contains parameters externally chosen. Panel B contains parameters internally calibrated to match moments presented in Table SM.5.2 and Figure SM.5.1.

Table SM.5.2: Targeted Moments: Homogeneous Customers Model

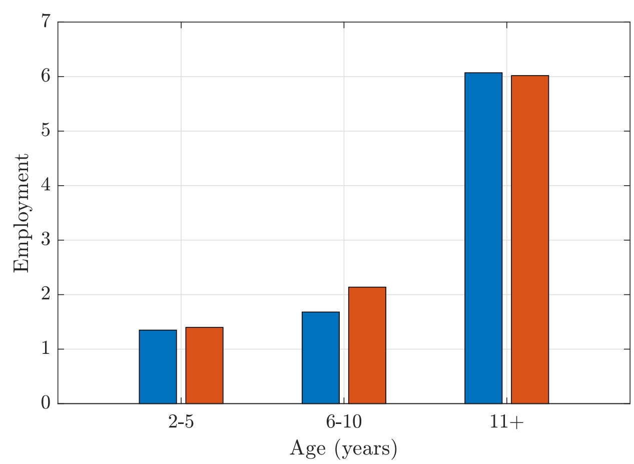
Moment	**Data**	**Model**
Avg. COGS-to-OPEX ratio	0.660	0.667
Avg. cost-weighted production markup	1.250	1.266
Slope labor prod. on sales	0.036	0.035
Avg. exit rate	0.073	0.073
SD. employment growth	0.416	0.436

Notes: This table shows the set of moments targeted in the calibration of the homogeneous customers model with an exogenous and homogeneous customer base. Avg. COGS-to-OPEX ratio refers to the average of the ratio across firms. Avg. cost-weighted production markup corresponds to the COGS-weighted average markup from Edmond, Midrigan, and Xu (2022). These moments were computed using data from Compustat in 2012. Slope of labor prod. on sales corresponds to the OLS coefficient of the sales-weighted regression of relative revenue labor productivity on relative sales from Edmond, Midrigan, and Xu (2022), restricting the sample to firms with relative sales above one. This moment was computed using data from the SUSB in 2012. The average exit rate was obtained from the BDS in 2012. The standard deviation of annual employment growth for continuing establishments is obtained from Elsby and Michaels (2013). The growth rate of variable x is computed as in Davis and Haltiwanger (1992): $(x_{i,t} - x_{i,t-1})/0.5(x_{i,t} + x_{i,t-1})$. The last column shows the model counterparts of each moment, which were obtained by simulating a panel of firms and computing each moment with the simulated data. In the model, we account for selection into Compustat by restricting the simulated sample of firms to those that are at least 7 years old and have sales above 19% of the average sales in the simulated economy (which corresponds to the ratio of the 5th percentile of the sales distribution in Compustat to the average sales in SUSB).

Figure SM.5.1: Homogeneous Customers Model Fit: Firm Size



(a) Distribution of Sales



(b) Avg. Size by Age

Notes: This figure shows moments targeted in the calibration of the homogeneous customers model with an exogenous and homogeneous customer base. Panel (a) shows the model fit of the distribution of relative sales. The distribution of relative sales is obtained from the SUSB in 2012. Panel (b) shows the model fit of average employment, relative to 1-year-old firms, by firm age. Average firm employment by age group was obtained from BDS in 2012. In the calibration exercise, we only target the relative size of firms older than 10 years.

6 Model Robustness

6.1. Higher Elasticity of Production to Labor. In our baseline calibration we set the value of α, the elasticity of output to labor, to 0.64 . This value then corresponds to a value of the labor share of 0.6 as a non-targeted moment.[2] However, in models without capital, such as ours, α also governs the degree of decreasing returns to scale in production, which is usually calibrated to a higher value. To assess the robustness of our results to that interpretation of the parameter α, we recalibrate both the baseline and our homogeneous customers models for the alternative value of $\alpha = 0.85$ and establish the robustness of our two main results:[3] (1) the degree of markup dispersion implied by each model when both target the same distribution of sales (which relates directly to the degree of misallocation), and (2) the welfare gains and its decomposition under the efficient allocation.

Tables SM.6.1 and SM.6.2 show the calibrated values of the parameters in the baseline and homogeneous customers models, respectively, when we set $\alpha = 0.85$. To assess the fit of each of two models, Tables SM.6.3 and SM.6.4 and Figures SM.6.1a and SM.6.1b show the values implied by each model for the targeted moments vs. their data counterparts.

Given these calibrated values, we next turn to examining our main results. First, Figure SM.6.2 shows the distribution of markups implied by each model. Comparing this with the same figure under the baseline calibration in Figure 3, we observe a similar pattern across the two models: once both models are calibrated to match the same distribution of firm size and the same level of cost-weighted aggregate markup, the homogeneous customers model implies a much lower level of dispersion in markups across firms, hinting that the model with endogenous customer acquisition features large efficiency losses due to misallocation.

[2] While in conventional models α maps directly to aggregate markup multiplied by the labor share, in our model, labor share also depends on the amount of labor used for advertising. Formally, we define the labor share in the model as total compensation to labor allocated towards production and advertising jointly divided by total sales of all firms.

[3] Another case that one might consider is the case of constant returns to scale, but that is not a well-defined case in our setting. It is well-known that the distribution of firm size is not well-defined in competitive economies with constant returns to scale. However, with monopolistic competition, even if production has constant returns to scale, the curvature of the demand function determines a monopolist's size as there is an optimal point where the marginal benefit of an additional unit of sale is offset by the reduction in its sale price. In our model, however, demand curvature applies only at the individual level and firms can increase their size independently through expansionary activities. Thus, with constant returns to scale in production, the curvature of demand at the individual level is not necessarily sufficient for pinning down a firm's size, given the role of the extensive margin of demand.

Table SM.6.1: Calibration of Baseline Model

Parameter	Description	Value
	Panel A: Fixed Parameters	
β	Annual discount factor	0.960
γ	Elast. of intertemporal substitution	1.000
ψ	Frisch elasticity	1.000
α	Decreasing returns to scale	0.850
δ	Prob. of losing customer	0.280
	Panel B: Calibrated Parameters	
ϕ	Elasticity matching function	0.587
χ	Overhead cost	0.460
σ	Avg. elasticity of substitution	6.342
η	Superelasticity	4.993
ν	Exog. survival probability	0.956
ρ_z	Persistence of productivity shock	0.992
σ_z	SD of productivity shock	0.085
$\bar{z}_{ent}$	Mean productivity of entrants	-1.191
λ	Mass of entrants	0.269
ξ	Disutility of labor supply	1.611

Notes: This table shows the calibration of the model with the alternative value of $\alpha = 0.85$. Panel A contains parameters externally chosen. Panel B contains parameters internally calibrated to match moments presented in Table SM.6.3 and Figure SM.6.1a in Supplemental Materials.

Table SM.6.2: Calibration of Homogeneous Customers Model

Parameter	Description	Value
	Panel A: Fixed Parameters	
β	Annual discount factor	0.960
γ	Elast. of intertemporal substitution	1.000
ψ	Frisch elasticity	1.000
α	Decreasing returns to scale	0.850
δ	Prob. of losing customer	1.000
	Panel B: Calibrated Parameters	
χ	Overhead cost	0.688
σ	Avg. elasticity of substitution	6.516
η	Superelasticity	0.672
ν	Exog. survival probability	0.958
ρ_z	Persistence of productivity shock	0.991
σ_z	SD of productivity shock	0.135
$\bar{z}_{ent}$	Mean productivity of entrants	-2.523
λ	Mass of entrants	0.929
ξ	Disutility of labor supply	1.681

Notes: This table shows the calibration of the homogeneous customers model with an exogenous customer base and the alternative value of $\alpha = 0.85$. Panel A contains parameters externally chosen. Panel B contains parameters internally calibrated to match moments presented in Table SM.6.4 and Figure SM.6.1b.

Table SM.6.3: Targeted Moments: Baseline Model

Moment	Data	Model
Slope SGA on sales	0.521	0.485
Avg. COGS-to-OPEX ratio	0.660	0.735
Avg. cost-weighted production markup	1.250	1.260
Slope labor prod. on sales	0.036	0.035
Avg. exit rate	0.073	0.074
SD. employment growth	0.416	0.336

Notes: This table shows the set of moments targeted in the calibration of the baseline model with $\alpha = 0.85$.

Table SM.6.4: Targeted Moments: Homogeneous Customers Model

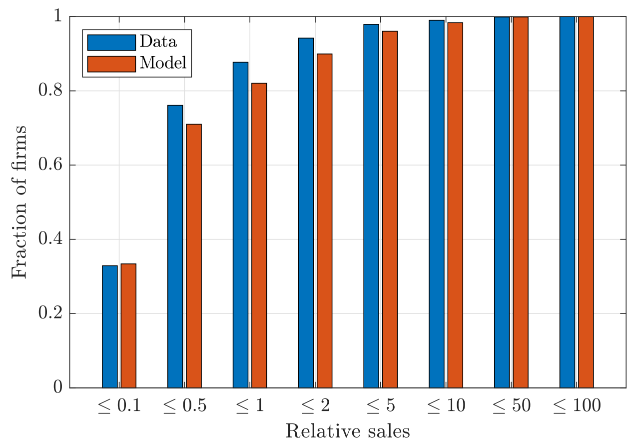
Moment	Data	Model
Avg. COGS-to-OPEX ratio	0.660	0.694
Avg. cost-weighted production markup	1.250	1.248
Slope labor prod. on sales	0.036	0.036
Avg. exit rate	0.073	0.074
SD. employment growth	0.416	0.392

Notes: This table shows the set of moments targeted in the calibration of the homogeneous customers model with $\alpha = 0.85$.

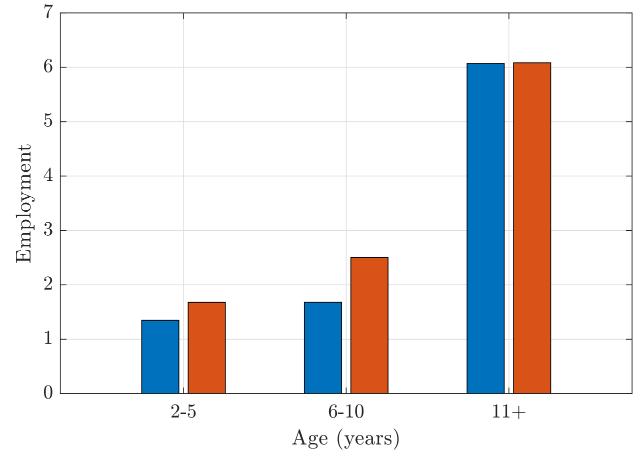
To measure the degree of misallocation, and more broadly the welfare implications under this alternative calibration, we reproduce the analog of Table 7 in the main text, as presented in Table SM.6.5. First, the welfare equivalent welfare gains under the efficient allocation are now 24.4%, close to the 27.3% that we obtained under our baseline calibration. Second, relative to the baseline calibration, we now recover a much *higher* degree of misallocation at 38.9% (as opposed to 12.2% under the baseline calibration). The reason for this larger TFP gain is that with higher α, more

Figure SM.6.1: Model Fit: Firm Size

(a) Baseline Model

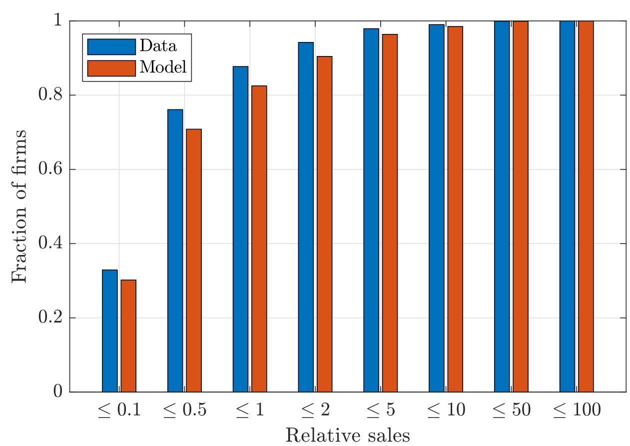


(a.1) Distribution of Sales

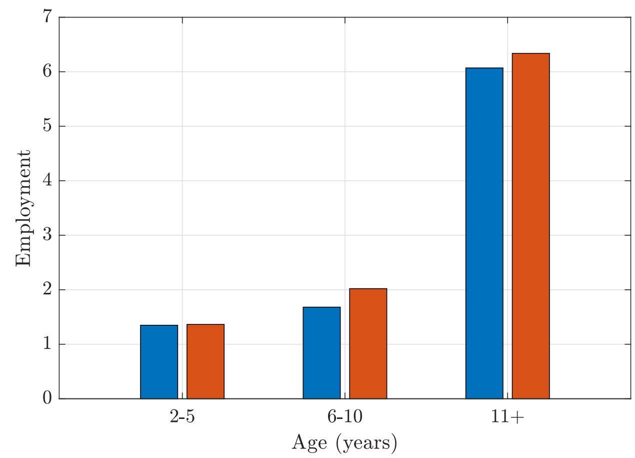


(a.2) Avg. Size by Age

(b) Homogeneous Customers Model



(b.1) Distribution of Sales

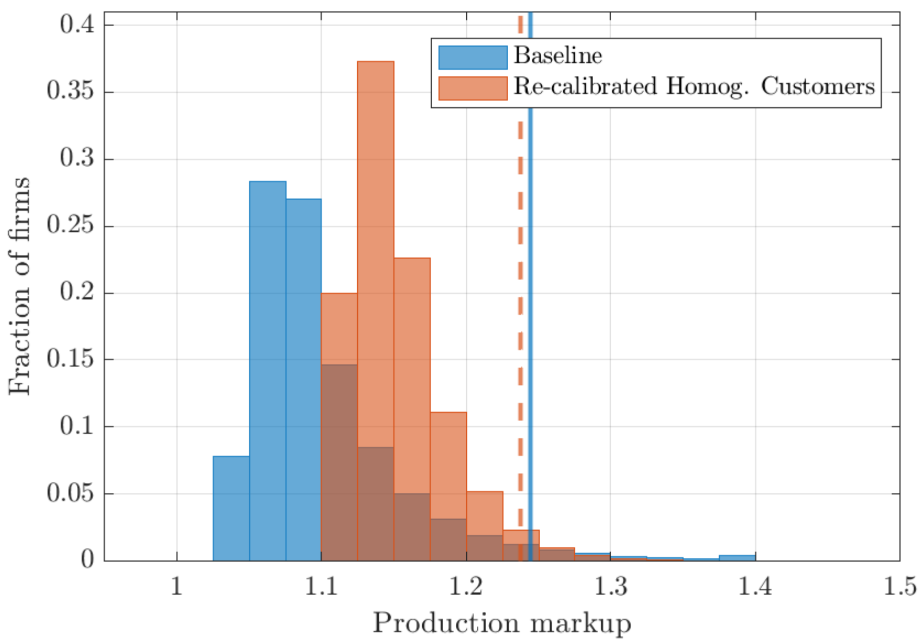


(b.2) Avg. Size by Age

Notes: Panel (a) shows moments targeted in the calibration of the baseline model with the alternative value of $\alpha = 0.85$. Figure (a.1) shows the model fit of the distribution of relative sales. The distribution of relative sales is obtained from the SUSB in 2012. Figure (a.2) shows the model fit of average employment, relative to 1-year-old firms, by firm age. Average firm employment by age group was obtained from BDS in 2012. In the calibration exercise, we only target the relative size of firms older than 10 years. Panel (b) shows the analogous figures given by the recalibrated homogeneous customers model under $\alpha = 0.85$.

productive firms have more production capacity and as a result the efficient allocation is more aggressive in concentrating customers among the more productive firms relative to the equilibrium. Interestingly, labor supply is also higher under the efficient allocation, partially mitigating the effects of large TFP gains on welfare. The reason for this is that the efficient allocation now keeps

Figure SM.6.2: Distribution of Markups: Baseline vs Homogeneous Customers Model



Notes: The figure plots the distribution of unweighted production markups in the baseline and *recalibrated* homogeneous customers models with the alternative value of $\alpha = 0.85$. The histograms are censored at markups above 1.4. The vertical dashed lines show the average cost-weighted production markup in each model.

more firms as operating firms relative to the equilibrium which increases output and labor supply substantially. The culmination of all these effects sum to a significant welfare gain of 24.4%, similar to the gains of 27.3% baseline calibration.

6.2. Correlated Production and Advertising Productivities at the Firm Level. So far we have assumed that firm productivity affects production but not advertising, in the sense that all firms have the same advertising technology. Given that, all differences in firms' customer acquisition strategies arise from the fact that they assign different values for additional customers, with more productive anticipating higher markups and investing more in customer acquisition than less productive firms.

However, a more holistic interpretation of firm productivity would be to assume that $z_{i,t}$ not only represents a firm's ability to produce goods, but also represents their ability to advertise and acquire customers. Such a model would now give a bigger boost to more productive firms' customer acquisition incentives: not only such firms assign a higher value to an additional customer (all else equal), but also, with such a technology, they need to spend less to acquire that customer relative to their less productive counterparts. Given that the efficient allocation concentrates customers more aggressively among the more productive firms relative to the equilibrium, one expects that such an

Table SM.6.5: Comparison with Efficient Allocation

	Endogenous $m_{i,t}$			
	$\phi = 0.25$	Baseline	$\phi = 0.75$	$\phi = 1.0$
TFP	61.2	38.9	24.3	14.6
Output	55.5	34.3	20.2	12.4
Number of firms	67.8	80.2	69.4	31.0
Employment	15.3	12.0	9.5	4.8
Production	8.4	9.8	10.4	12.6
Welfare	49.4	24.4	10.6	7.3
Agg. markup	-23.8	-21.9	-20.0	-17.5
Top 5% sales share	144.4	54.0	18.2	0.5

Notes: The table compares aggregate variables between the social planner's allocation and the equilibrium allocation calibrated with the alternative value of $\alpha = 0.85$. Differences are reported as percent deviations from equilibrium allocations. Three comparisons are presented by varying the value of ϕ while keeping the remaining parameters fixed at the values in the calibration with $\alpha = 0.85$.

extension would imply a more efficient equilibrium allocation of customer across firms relative to our baseline model. In this section, we quantify these differences.

To implement this idea, consider the following extension of the model in the main text where the production technology for ads is now given by

$$a_{i,t} = z_{i,t}^{\theta} l_{i,s,t}^{\phi} \implies m_{i,t} \leq (1-\delta) m_{i,t-1} + \frac{z_{i,t}^{\theta} l_{i,s,t}^{\phi}}{P_{m,t}} \tag{6.1}$$

where the new parameter $\theta \in [0,1]$ captures the degree to which productivity $z_{i,t}$ also affects advertising technology. Here, $\theta = 0$ corresponds to our baseline specification, $\theta = 1$ to the case that $z_{i,t}$ is the productivity in both technologies, and interim values of θ capturing intermediate cases where more productive firms in production are also more productive in advertising but productivity difference are more muted in advertising relative to production.

We begin our analysis with a comparative statics exercise on welfare, in which we compare the efficient and equilibrium allocations under $\theta = 0$—which is our baseline calibration as presented in the main text—and $\theta = 1$, holding other parameters the same at the $\theta = 0$ calibration. Table SM.6.6 shows these comparisons. As expected, all else equal, higher θ dampens the welfare gains of the efficient allocation by making the equilibrium allocation of customers more efficient. The consumption equivalent welfare decreases from 27.3% in the baseline to 11.6% under $\theta = 1$, with the bulk of that decline coming from a dampened TFP gain of 1.8% relative to 12.2% in the baseline, which also manifests itself in output gains of 12.4% with $\theta = 1$ relative to 25.5% in the baseline. This modest TFP gain is consistent with the fact that higher θ makes the allocation of customers more efficient by giving more productive firms another edge in acquiring customers. The fact that the changes in other aggregate variables such as number of firms and employment are similar supports the interpretation that the bulk of welfare gains differences is explained by the more efficient allocation of customers in the $\theta = 1$.

Table SM.6.6: Comparative Statics with Respect to θ

	Comp. Eq. $\theta = 0$	Comp. Eq. $\theta = 1$
TFP	12.2	1.8
Output	25.5	12.4
Number of firms	-2.6	-1.3
Employment	1.9	1.8
Production	20.9	16.4
Welfare	27.3	11.6
Agg. markup	-22.8	-18.1
Top 5% sales share	39.9	6.5

Notes: The table reports welfare gains under the efficient allocation for the baseline calibrated model ($\theta = 0$) and an alternative model where advertising and production productivity are the same ($\theta = 1$)—while holding other parameters fixed at the baseline calibration. All else equal, higher θ reduces welfare gains to 11.6% and TFP gains to 1.8% because in the $\theta = 1$ model more productive firms are also equally productive in acquiring customers—which allows them to acquire more customers and get closer to the efficient allocation of demand.

However, we must note that by changing θ from 0 to 1, the model now misses some of its key targeted moments in the calibration, and possibly overstates the impact of $\theta = 1$ on mitigating the welfare gains under the efficient allocation. Accordingly, we next turn to a recalibrated version of the model under the fixed value of $\theta = 1$. Table SM.6.7 show the calibrated values of the parameters when we set $\theta = 1$. To assess the fit of this model, Table SM.6.8 and Figure SM.6.3 show the values implied by each model for the targeted moments vs. their data counterparts.

A key observation in Table SM.6.7 is that the calibrated value of ϕ (returns to scale in advertising) under $\theta = 1$ (0.355) is smaller than its calibrated value in the baseline model with $\theta = 0$ (0.533). This is indicative of the fact that higher values of θ behave somewhat similarly to higher values of ϕ. Higher ϕ or higher θ give more productive firms more scope to acquire customers. So, once we fix $\theta = 1$, the model assigns a lower value to ϕ to match the same targeted moments, especially the relationship between SGA expenditures and sales.

Given these calibrated values, we next reexamine our main results. First, Figure SM.6.4 shows the distribution of markups implied by each model. Comparing this with the same figure under the baseline calibration in Figure 3, we observe a similar pattern across the two models: once both models are calibrated to match the same distribution of firm size and the same level of cost-weighted aggregate markup, the homogeneous customers model implies a much lower level of dispersion in markups across firms.

Second, Table SM.6.9 shows a similar exercise to Table SM.6.6 but now under the recalibrated values of model parameters under $\theta = 1$ and for a wider range of values for θ. We observe that as θ moves from 0 to 1, TFP, output, and welfare gains all decrease, but even in extreme case of $\theta = 1$—where firms have the same productivity in advertising and production—consumption equivalent welfare gain relative to the equilibrium allocation is substantial at 15.2%, arising from a 20.5% increase in production labor and meaningful TFP gains of 3.7%, which is more than twice as big as the TFP gains under the comparative statics of Table SM.6.6.

Table SM.6.7: Model Parameters

Parameter	**Description**	**Value**
	Panel A: Fixed Parameters	
β	Annual discount factor	0.960
γ	Elast. of intertemporal substitution	1.000
ψ	Frisch elasticity	1.000
α	Decreasing returns to scale	0.640
δ	Prob. of losing customer	0.280
	Panel B: Calibrated Parameters	
ϕ	Elasticity matching function	0.355
χ	Overhead cost	0.370
σ	Avg. elasticity of substitution	5.931
η	Superelasticity	7.137
ν	Exog. survival probability	0.963
ρ_z	Persistence of productivity shock	0.956
σ_z	SD of productivity shock	0.255
$\bar{z}_{ent}$	Mean productivity of entrants	-0.827
λ	Mass of entrants	0.059
ξ	Disutility of labor supply	2.093

Notes: This table shows the calibration of the model with the alternative value of $\theta = 1$. Panel A contains parameters externally chosen. Panel B contains parameters internally calibrated to match moments presented in Table SM.6.8 and Figure SM.6.3 in Supplemental Materials.

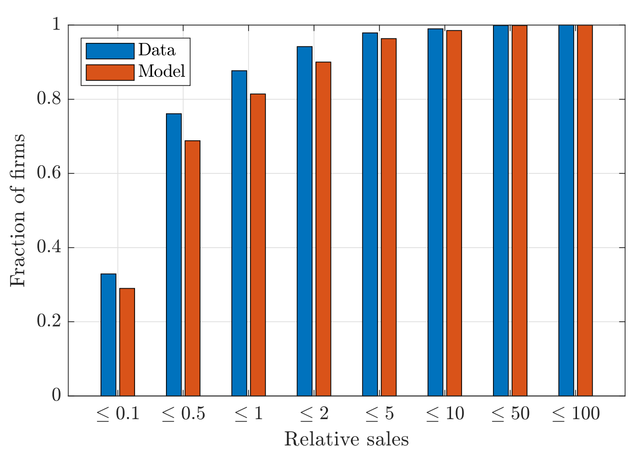
6.3. Hopenhayn Model of Entry. In our baseline, we model entry such that at each period a constant λ measure of potential entrants observe their initial productivity and decide whether to enter or not. This has two implications. First, since the entry decision is conditional on initial productivity, only potential entrants with high enough initial productivity enter. Second, λ is taken as a deep parameter—e.g., the economy can only generate λ new "ideas" every period independent

Table SM.6.8: Targeted Moments

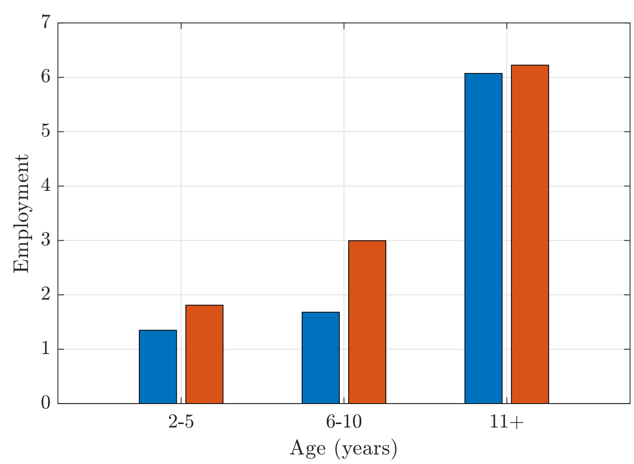
Moment	**Data**	**Model**
Slope SGA on sales	0.521	0.460
Avg. COGS-to-OPEX ratio	0.660	0.670
Avg. cost-weighted production markup	1.250	1.281
Slope labor prod. on sales	0.036	0.035
Avg. exit rate	0.073	0.073
SD. employment growth	0.416	0.489

Notes: This table shows the set of moments targeted in the calibration of the model with the alternative value of $\theta = 1$. Slope SGA on Sales refers to the OLS coefficient of the regression $SGA_{i,t} = c + \beta Sales_{i,t} + \psi COGS_{i,t} + \varepsilon_{i,t}$. Avg. COGS-to-OPEX ratio refers to the average of the ratio across firms. Avg. cost-weighted production markup corresponds to the COGS-weighted average markup from Edmond, Midrigan, and Xu (2022). These moments were computed using data from Compustat in 2012. Slope labor prod. on sales corresponds to the OLS coefficient of the sales-weighted regression of relative revenue labor productivity on relative sales from Edmond, Midrigan, and Xu (2022), restricting the sample of firms to those with relative sales above one. This moment was computed using data from the SUSB in 2012. The average exit rate was obtained from the BDS in 2012. The standard deviation of annual employment growth for continuing establishments is obtained from Elsby and Michaels (2013). The growth rate of variable x is computed as in Davis and Haltiwanger (1992): $(x_{i,t} - x_{i,t-1})/(0.5(x_{i,t} + x_{i,t-1}))$. The last column shows the model counterparts of each moment, which were obtained by simulating a panel of firms and computing each moment with the simulated data. In the model, we account for selection into Compustat by restricting the simulated sample of firms to those that are at least 7 years old and have sales above 19% of the average sales in the simulated economy (which corresponds to the ratio of the 5th percentile of the sales distribution in Compustat to the average sales in SUSB).

Figure SM.6.3: Model Fit: Firm Size



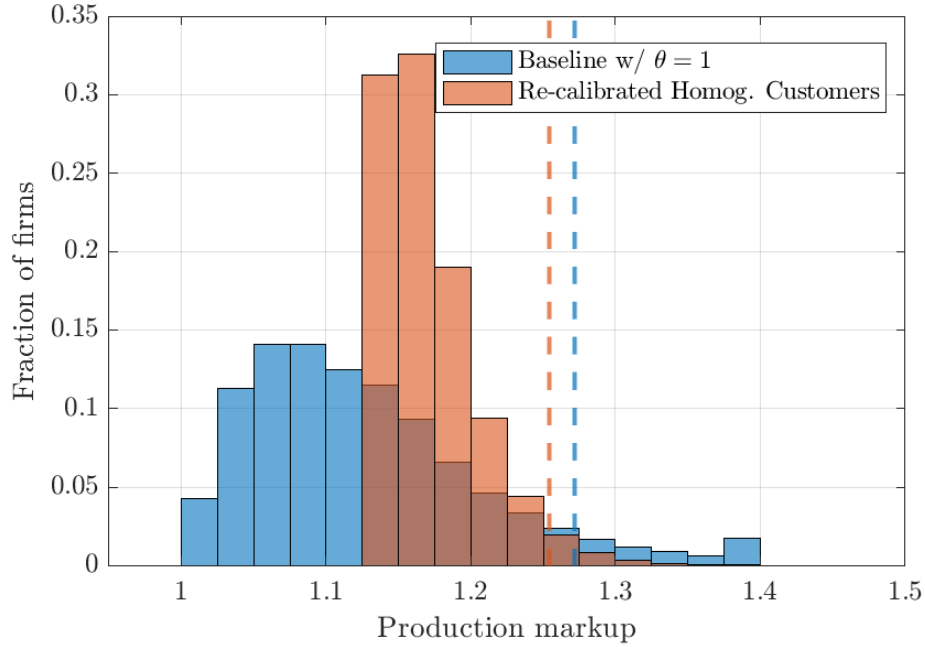
(a) Distribution of Sales



(b) Avg. Size by Age

Notes: This figure shows moments targeted in the calibration of the model with the alternative value of $\theta = 1$. Panel (a) shows the model fit of the distribution of relative sales. The distribution of relative sales is obtained from the SUSB in 2012. Panel (b) shows the model fit of average employment, relative to 1-year-old firms, by firm age. Average firm employment by age group was obtained from BDS in 2012. In the calibration exercise, we only target the relative size of firms older than 10 years.

Figure SM.6.4: Distribution of Markups: Baseline vs Homogeneous Customers Model



Notes: The figure plots the distribution of unweighted production markups in the recalibrated baseline model with the alternative value of $\theta = 1$ and the calibrated *recalibrated* homogeneous customers model. The histograms are censored at markups above 1.4. The vertical dashed lines show the average cost-weighted production markup in each model.

Table SM.6.9: Comparison with Efficient Allocation

	Endogenous $m_{i,t}$			
	$\theta = 0$	$\theta = 0.25$	$\theta = 0.75$	$\theta = 1$
TFP	21.3	15.4	6.6	3.7
Output	38.9	31.7	20.6	16.8
Number of firms	-23.4	-18.7	-14.0	-12.7
Employment	3.5	3.8	3.7	3.6
Production	27.5	25.4	21.9	20.5
Welfare	43.7	33.5	19.5	15.2
Agg. markup	-31.0	-29.2	-25.6	-24.0
Top 5% sales share	93.4	69.2	35.7	24.6

Notes: The table compares aggregate variables between the social planner's allocation and the equilibrium allocation with different values of θ, where parameters are set to the calibrated parameters at $\theta = 1$. Differences are reported as percent deviations from equilibrium allocations.

of the allocation. Here, we consider the alternative entry model of Hopenhayn (1992), where there are always an *infinite* number of potential entrants who decide whether to enter or not, subject to paying a fixed cost of entry, κ, *before* observing their initial productivity. We denominate κ in units of labor.

Formally, this model deviates from our baseline only in two aspects. First, it adds a free entry condition along with a new parameter κ:

$$\underbrace{\int v_t(0,z)dF^e(z)}_{\text{expected value after entry}} = \underbrace{\kappa W_t}_{\text{cost of entry}} \tag{6.2}$$

where $v_t(0,z)$ is the value of an operating firm with no customers and productivity z and $F^e = \mathcal{N}(\bar{z}_{ent}, \sigma_z^2)$ is the distribution from which potential entrants draw their productivity. Second, by changing the timing of when entrants observe their productivity, it changes the distribution from which they draw these productivities: In our baseline, this distribution is a truncated log-normal with the endogenous entry threshold, and in this alternative model, it is the unconditional distribution of productivity. Given that all other aspects of the model remain the same, the definition of the equilibrium for this alternative model deviates minimally by taking these two differences into account. Below, we state the definition of such an equilibrium, explicitly underlining the differences from the definition in the main text for our baseline model:

Equilibrium Definition.. An equilibrium is (a) an allocation for the households $\{(c_{i,j,t})_{j\in[0,1]}, C_t, L_t\}_{t\geq 0}$; (b) a set of exit decisions for potential entrants and incumbents $\{(\mathbf{1}_{i,t})_{i\in\Lambda_t\cup N_{t-1}}\}_{t\geq 0}$; (c) an allocation for operating firms $\{(p_{i,t}, y_{i,t}, m_{i,t}, l_{i,p,t}, l_{i,s,t})_{i\in N_t}\}_{t\geq 0}$; and (d) a sequence of aggregate prices $\{W_t, P_{m,t}\}_{t\geq 0}$ and a sequence of sets $\{N_t\}_{t\geq 0}$ such that

1. given (c) and (d), household's allocation in (a) solves their problem in Equation (4),
2. given (a) and (d), firms' allocations in (b) and (c) solve their problems in Equation (14),
3. labor and matching markets clear: $L_t = \int_{i\in N_t}(l_{i,p,t} + l_{i,s,t} + \chi)di, \quad 1 = \int_{i\in N_t} m_{i,t}di$,
4. <u>the free entry condition in Equation (6.2) holds,</u>
5. and letting Λ_t denote the set of entrants at time t, which decide on entry <u>before</u> observing their initial productivity draw, the set of operating firms, N_t, evolves according to

$$N_t = \{i \in \Lambda_t \cup N_{t-1} : \mathbf{1}_{i,t}\nu_{i,t} = 1\},\ N_{-1} \text{ given.} \tag{6.3}$$

Given this adjusted definition, we solve and calibrate this alternative model as before but now treat λ_t (the measure of entrants, or more precisely, the measure of the set Λ_t) as an endogenous variable determined by the free entry condition. Furthermore, the change in the timing also changes the law of motion for the measure and the distribution of operating firms in the equilibrium. Formally, letting $N_t : M \times Z \to [0,1]$ denote the cumulative measure of incumbent firms measured after the realization of idiosyncratic productivity shocks, but before exit decisions are made, the law of motion for this measure is given by

$$N_t(m,z') = \int_{M\times Z} F(z'|z)\mathbf{1}_{\{m_t^*(m_{-1},z)\leq m\}}\nu\mathbf{1}_{\{v_t(m_{-1},z)\geq 0\}}dN_{t-1}(m_{-1},z)$$

$$+\lambda_t \int_{M\times Z} F(z'|z)\mathbf{1}_{\{m_t^*(0,z)\leq m\}} dF^e(z),$$

where $F(z'|z)$ is the Markov chain given by the AR(1) productivity process, $F^e(z)$ is the productivity distribution of potential entrants, and $m^*(m_{-1}, z)$ denotes the optimal policy for customer acquisition.

Given the different timing of entry in the Hopenhayn model and the endogenous measure of entrants, one potential concern is that this model would have different predictions for the distributions of sales and markups relative to our baseline model. Namely, we want to reexamine the validity of the exercise the model with endogenous customers would predict a higher markup dispersion relative to the homogeneous customers model, once the two models are calibrated to match the same distribution of firm size. Tables SM.6.10 and SM.6.11 show the calibrated values of the parameters in the baseline and homogeneous customers models, respectively. To assess the fit of each of two models, Tables SM.6.12 and SM.6.13 and Figures SM.6.5a and SM.6.5b show the values implied by each model for the targeted moments vs. their data counterparts.

With these calibrated values in hand, Figure SM.6.6 shows the distribution of markups implied by each model. Similar to our baseline specification, we observe that the model with endogenous customers implies a much higher markup dispersion relative to the homogeneous customers model.

Table SM.6.10: Calibration of Baseline Model

Parameter	Description	Value
	Panel A: Fixed Parameters	
β	Annual discount factor	0.960
γ	Elast. of intertemporal substitution	1.000
ψ	Frisch elasticity	1.000
α	Decreasing returns to scale	0.640
δ	Prob. of losing customer	0.280
	Panel B: Calibrated Parameters	
ϕ	Elasticity matching function	0.521
χ	Overhead cost	0.310
σ	Avg. elasticity of substitution	6.925
η	Superelasticity	5.487
ν	Exog. survival probability	0.979
ρ_z	Persistence of productivity shock	0.964
σ_z	SD of productivity shock	0.226
$\bar{z}_{ent}$	Mean productivity of entrants	-0.592
κ	Entry cost	3.018
ξ	Disutility of labor supply	1.921

Notes: This table shows the calibration of the model with the alternative value of $\alpha = 0.85$. Panel A contains parameters externally chosen. Panel B contains parameters internally calibrated to match moments presented in Table SM.6.3 and Figure SM.6.1a in Supplemental Materials.

Table SM.6.11: Calibration of Homogeneous Customers Model

Parameter	Description	Value
	Panel A: Fixed Parameters	
β	Annual discount factor	0.960
γ	Elast. of intertemporal substitution	1.000
ψ	Frisch elasticity	1.000
α	Decreasing returns to scale	0.640
δ	Prob. of losing customer	1.000
	Panel B: Calibrated Parameters	
χ	Overhead cost	0.426
σ	Avg. elasticity of substitution	6.680
η	Superelasticity	0.798
ν	Exog. survival probability	0.984
ρ_z	Persistence of productivity shock	0.964
σ_z	SD of productivity shock	0.254
$\bar{z}_{ent}$	Mean productivity of entrants	-0.826
κ	Entry cost	2.774
ξ	Disutility of labor supply	2.097

Notes: This table shows the calibration of the homogeneous customers model with an exogenous customer base and the alternative value of $\alpha = 0.85$. Panel A contains parameters externally chosen. Panel B contains parameters internally calibrated to match moments presented in Table SM.6.4 and Figure SM.6.1b.

Table SM.6.12: Targeted Moments: Baseline Model

Moment	Data	Model
Slope SGA on sales	0.521	0.468
Avg. COGS-to-OPEX ratio	0.660	0.685
Avg. cost-weighted production markup	1.250	1.250
Slope labor prod. on sales	0.036	0.036
Avg. exit rate	0.073	0.073
SD. employment growth	0.416	0.464

Notes: This table shows the set of moments targeted in the calibration of the baseline model with $\alpha = 0.85$.

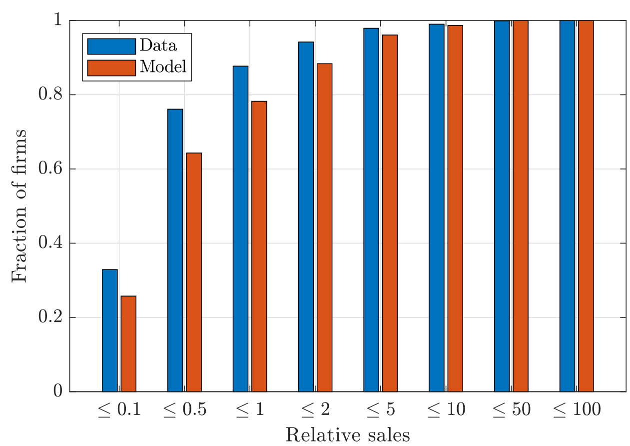
Table SM.6.13: Targeted Moments: Homogeneous Customers Model

Moment	Data	Model
Avg. COGS-to-OPEX ratio	0.660	0.672
Avg. cost-weighted production markup	1.250	1.224
Slope labor prod. on sales	0.036	0.037
Avg. exit rate	0.073	0.072
SD. employment growth	0.416	0.444

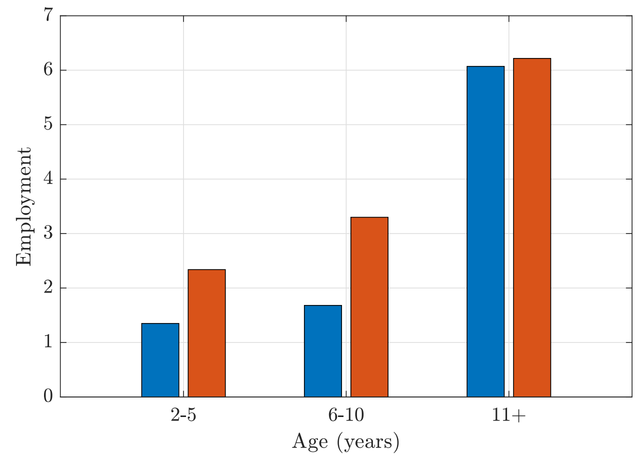
Notes: This table shows the set of moments targeted in the calibration of the homogeneous customers model with $\alpha = 0.85$.

Figure SM.6.5: Model Fit: Firm Size

(a) Baseline Model

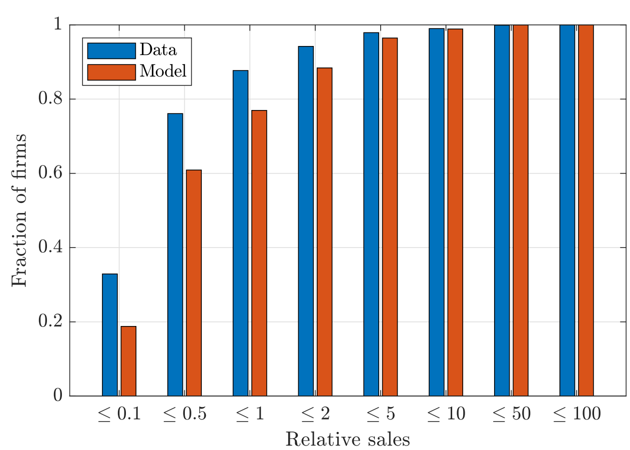


(a.1) Distribution of Sales

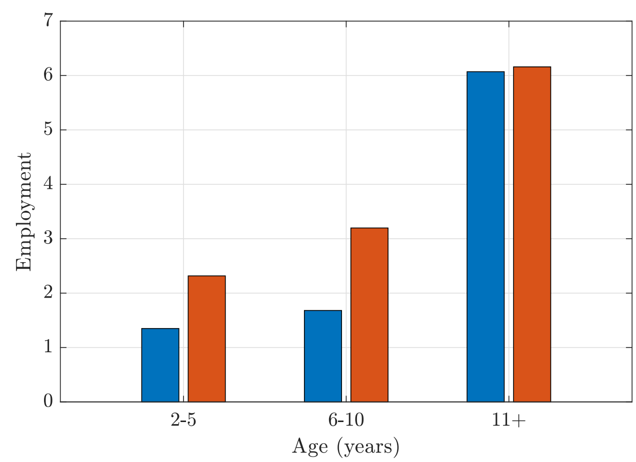


(a.2) Avg. Size by Age

(b) Homogeneous Customers Model



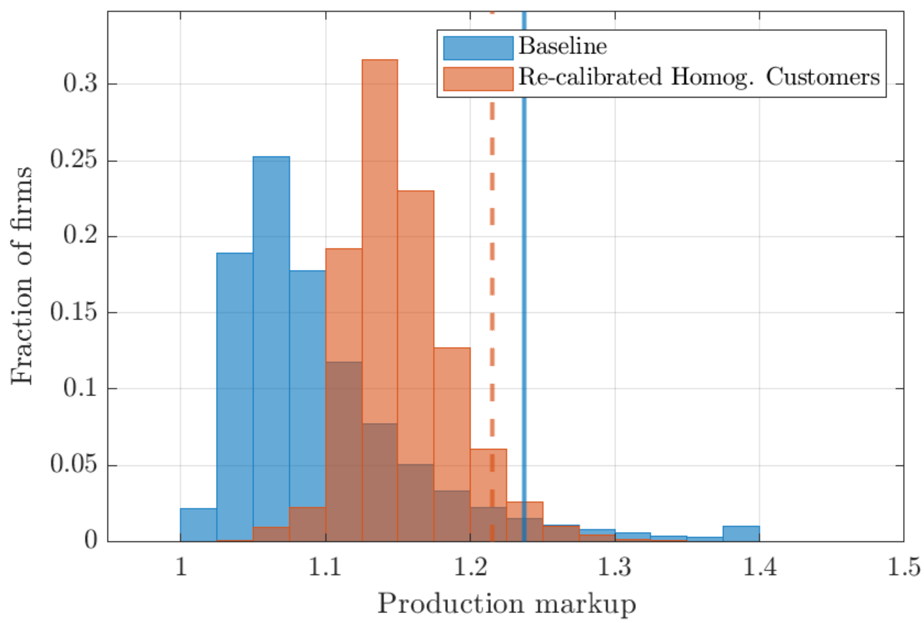
(b.1) Distribution of Sales



(b.2) Avg. Size by Age

Notes: Panel (a) shows moments targeted in the calibration of the baseline model with the alternative value of $\alpha = 0.85$. Figure (a.1) shows the model fit of the distribution of relative sales. The distribution of relative sales is obtained from the SUSB in 2012. Figure (a.2) shows the model fit of average employment, relative to 1-year-old firms, by firm age. Average firm employment by age group was obtained from BDS in 2012. In the calibration exercise, we only target the relative size of firms older than 10 years. Panel (b) shows the analogous figures given by the recalibrated homogeneous customers model under $\alpha = 0.85$.

Figure SM.6.6: Distribution of Markups: Baseline vs Homogeneous Customers Model



Notes: The figure plots the distribution of unweighted production markups in the calibrated baseline and the *recalibrated* homogeneous customers models with the alternative entry model of Hopenhayn (1992). The histograms are censored at markups above 1.4. The vertical dashed lines show the average cost-weighted production markup in each model.

References

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the Sensitivity of Parameter Estimates to Estimation Moments," *Quarterly Journal of Economics*, 132(4), 1553–1592.

Argente, D., D. Fitzgerald, S. Moreira, and A. Priolo (2021): "How Do Firms Build Market Share?," *Available at SSRN 3831706*.

Argente, D., M. Lee, and S. Moreira (2024): "The life cycle of products: Evidence and implications," *Journal of Political Economy*, 132(2), 337–390.

Arnoud, A., F. Guvenen, and T. Kleineberg (2019): "Benchmarking Global Optimizers," Manuscript.

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

Davis, S. J., and J. Haltiwanger (1992): "Gross Job Creation, Gross Job Destruction, and Employment Reallocation," *Quarterly Journal of Economics*, 107(3), 819–863.

Davis, S. J., J. C. Haltiwanger, S. Schuh, et al. (1998): "Job Creation and Destruction," *MIT Press Books*, 1.

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

Edmond, C., V. Midrigan, and D. Y. Xu (2022): "How Costly Are Markups?," Manuscript.

Elsby, M. W., and R. Michaels (2013): "Marginal Jobs, Heterogeneous Firms, and Unemployment Flows," *American Economic Journal: Macroeconomics*, 5(1), 1–48.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013): "Who Creates Jobs? Small Versus Large Versus Young," *Review of Economics and Statistics*, 95(2), 347–361.

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

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

Neiman, B., and J. Vavra (2023): "The rise of niche consumption," *American Economic Journal: Macroeconomics*, 15(3), 224–264.

Ottonello, P., and T. Winberry (2020): "Financial heterogeneity and the investment channel of monetary policy," *Econometrica*, 88(6), 2473–2502.

Stokey, N. L., and R. E. Lucas Jr (1989): *Recursive methods in economic dynamics*. Harvard University Press.

Traina, J. (2019): "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements," Manuscript.

Young, E. R. (2010): "Solving the Incomplete Markets Model With Aggregate Uncertainty Using the Krusell–Smith algorithm and non-stochastic simulations," *Journal of Economic Dynamics and Control*, 34(1), 36–41.