Online Appendix

Concentration, Market Power, and Misallocation:
The Role of Endogenous Customer Acquisition

by Hassan Afrouzi, Andres Drenik, and Ryan Kim

Table of Contents for the Online Appendix

A Additional Empirical Results

This section provides additional results to confirm the robustness of our main empirical findings.

A.1. Measurement error in $\ln m_{it}$

Table A.1: Markups, Sales per Customer, and Number of Customers: IV Approach

	(1)	(2)	(3)	(4)	(5)
ln m_{it}	-0.002	-0.002	0.001	0.001	0.002
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
ln $p_{\text{it}}q_{\text{it}}$	0.107***	0.107***	0.072***	0.071***	0.073***
	(0.037)	(0.037)	(0.023)	(0.023)	(0.026)
Observations	2012	2012	2012	2012	2012
R^2	0.062	0.061	0.033	0.031	0.032
Year FE		✓		✓	
SIC FE			✓	✓	
SIC-year FE					✓
First-stage F statistics	1.6e+04	1.6e+04	1.1e+04	1.1e+04	1.1e+04

Notes: This table replicates Table 1 by instrumenting $\ln m_{it}$ with its lagged value ($\ln m_{it-1}$). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; standard errors are clustered at the firm-level. Markups are measured as the Sales-to-COGS ratio. The variable $\ln p_{it}q_{it}$ denotes the log of the average sales per customer and $\ln m_{it}$ denotes the log number of customers. SIC industries correspond to a two-digit SIC code. All Nielsen variables are projection-factor adjusted.

Table A.2: Markups, Sales per Customer, and Number of Customers: First Stage

	(1)	(2)	(3)	(4)	(5)
ln $m_{\text{i,t-1}}$	0.970***	0.971***	0.958***	0.958***	0.963***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
ln $p_{\text{it}}q_{\text{it}}$	-0.010	-0.010	0.005	0.008	0.016
	(0.026)	(0.027)	(0.032)	(0.032)	(0.034)
Observations	2012	2012	2012	2012	2012
R^2	0.948	0.948	0.950	0.950	0.957
Year FE		✓		✓	
SIC FE			✓	✓	
SIC-year FE					✓

Notes: This table presents the first-stage results behind the IV estimation in Table A.1.

A.2. Markups, Sales per Customers, and Sales

Table A.3 replicates Table 1 by replacing the number of customers with a firm's total sales. We replace the number of customers with sales, so that our regressors have the same unit. Regardless of whether we control for total sales or the number of customers, our results show a strong correlation between markups and sales per customer, which suggests the importance of sales per customer in understanding firm-level markups.

Table A.3: Markups, Sales, and Sales per Customer

	(1)	(2)	(3)	(4)	(5)
ln $p_{it}q_{it}$	0.094***	0.093***	0.058**	0.056**	0.057**
	(0.031)	(0.032)	(0.023)	(0.023)	(0.025)
ln S_{it}	-0.002	-0.002	0.002	0.002	0.003
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Observations	2433	2433	2433	2433	2433
R^2	0.046	0.047	0.311	0.313	0.338
Year FE		✓		✓	
SIC FE			✓	✓	
SIC-year FE					✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; standard errors are clustered at the firm-level. Markups are measured as Sales-to-COGS ratio. SIC is a two-digit SIC code, and all Nielsen variables are projection-factor adjusted.

A.3. Evidence Based on Product-level Markups

This section revisits our analysis of the relationship between markups and the relative size of different margins of a firm's demand by combining two alternative approaches to measure price-cost markup and account for marginal costs. Our baseline analysis focuses on firm-level markups and public firms. Here, we show that we find similar results when expanding the scope to product-level markups and analyzing a much broader set of firms in the Nielsen data.

As a first alternative approach, we use the difference between the retail-level price and the wholesale cost at the product level to measure markups. For this, we use microdata on the UPC-market-year-level wholesale cost from the Nielsen PromoData available through the Chicago Booth Kilts Center. The PromoData record wholesale costs by UPC, market, and year and are collected from 12 national wholesalers in the period 2006-2012 (we restrict our sample to the years 2006-2011 since there is a substantial number of missing observations in 2012). In total, we use data from 45 markets (examples of markets are Chicago, Los Angeles, and Atlanta). Given that the data lack detailed sales information, we take a simple geometric average of wholesale costs by UPC, market, and year after adjusting for the package size. We then (1) combine this information with the retail-level price, sales, and sales per household from the Nielsen Homescan Panel data and (2) measure the retail-level markup at the product (UPC) level as the difference between the retail price and the wholesale cost (we drop negative values of the measured markups). A similar approach has been followed by Gopinath, Gourinchas, Hsieh, and Li (2011) and Stroebel and Vavra (2019).

This measure of markups assumes that other costs, such as wages and capital expenditures, do not confound the observed relationship among markups, average sales per customer, and the number of customers. However, there could exist sources of variation in marginal costs across retailers as well, in addition to the variation arising from differences in product-level wholesale costs. To alleviate these concerns, we exploit the rich variation in the data and control for such differences in marginal costs by incorporating various sets of fixed effects in the regression, which is the approach followed by Fitzgerald, Haller, and Yedid-Levi (2016). Our markup measure could vary at the UPC, year, market, and retailer level, so we progressively include different combinations of fixed effects at those levels to absorb differences in marginal costs that could potentially confound the relationship of interest. For example, the fact that retailers sell the same UPC in multiple locations allows us to control for differences in marginal costs at the retailer-year-UPC level. The underlying assumption is that the retailer's marginal cost of selling a given UPC is the same across markets. Similarly, we can account for any time-invariant costs by including a set of UPC-market-retailer fixed effects and any differential distribution costs by incorporating a set of UPC-market-year fixed effects. The final sample contains data from 7,802 UPCs, 45 markets, 6 years, and 167 retailers.

Table A.4 confirms our previous results regarding the relationship among markups, average sales per customer, and the number of customers reported in Table 1. Column (1) shows that markups are strongly positively correlated with average sales per customer. The relationship between markups and the number of customers is negative but an order of magnitude smaller in size. Column (2) includes UPC-, market-, and retailer-fixed effects interacted with year-fixed effects. We find similar results, although the coefficients become smaller (in absolute value). Columns (3), (4), (5), and (6) allow for additional sets of fixed effects that absorb any differences in marginal costs at those levels (including the variation in wholesale costs, which is at the UPC-market-year-level). Once we exploit the geographic variation in the data and include retailer-year-UPC fixed effects in column (3), the coefficient on the number of customers declines to -0.016, while the coefficient on the average sales per customer remains stable at 0.166. Controlling for time-invariant costs in column (4) shows similar changes in the coefficients: There is still a positive and significant relationship between markups and average sales per customer, but no economically significant relationship with the number of customers (the coefficient is estimated to be close to zero). Adding the additional sets of fixed effects in columns (5) and (6) shows similar relationships among variables. Table A.5 replaces the number of customers in Table A.4 with total sales so that our two main regressors are expressed in the same unit. These results confirm that the relevant margin for the relationship between relative size and markups is the average sales per customer.

Table A.4: UPC-market-retailer-year-level analysis

	(1)	(2)	(3)	(4)	(5)	(6)
$\ln p_{\text{urmt}} q_{\text{urmt}}$	0.550***	0.186***	0.166***	0.185***	0.185***	0.187***
	(0.035)	(0.027)	(0.023)	(0.023)	(0.021)	(0.019)
$\ln m_{\text{urmt}}$	-0.063***	-0.040***	-0.016***	-0.004*	-0.003	-0.004
	(0.018)	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)
Observations	426032	426032	426032	426032	426032	426032
R^2	0.126	0.550	0.742	0.870	0.875	0.928
UPC-year FE		✓	✓	✓	✓	
market-year FE		✓	✓	✓		
retail-year FE		✓				
retail-year-UPC FE			✓	✓	✓	✓
UPC-market-retail FE				✓	✓	✓
year-market-retail FE					✓	✓
UPC-market-year FE						✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; standard errors are three-way clustered by product group, retail, and market. Markup$_{urmt}$ is measured as the difference between retail-level price and wholesale cost, $p_{urmt}q_{urmt}$ denotes the average sales per customer, and m_{urmt} the number of customers, where the subscript u refers to a particular UPC, r the retailer, m the market, and t the year. All variables are projection-factor adjusted. We balance the sample across columns based on the tightest specification in column (6); our final sample includes 7,802 UPCs, 45 markets, 6 years, and 167 retailers.

Table A.5: UPC-market-retailer-year-level analysis

	(1)	(2)	(3)	(4)	(5)	(6)
$\ln p_{\text{urmt}} q_{\text{urmt}}$	0.612***	0.226***	0.182***	0.189***	0.188***	0.191***
	(0.046)	(0.028)	(0.025)	(0.023)	(0.022)	(0.020)
$\ln S_{\text{urmt}}$	-0.063***	-0.040***	-0.016***	-0.004*	-0.003	-0.004
	(0.018)	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)
Observations	426032	426032	426032	426032	426032	426032
R^2	0.126	0.550	0.742	0.870	0.875	0.928
UPC-year FE		✓	✓	✓	✓	
market-year FE		✓	✓	✓		
retail-year FE		✓				
UPC-market-year FE						✓
retail-year-UPC FE			✓	✓	✓	✓
year-market-retail FE					✓	✓
UPC-market-retail FE				✓	✓	✓

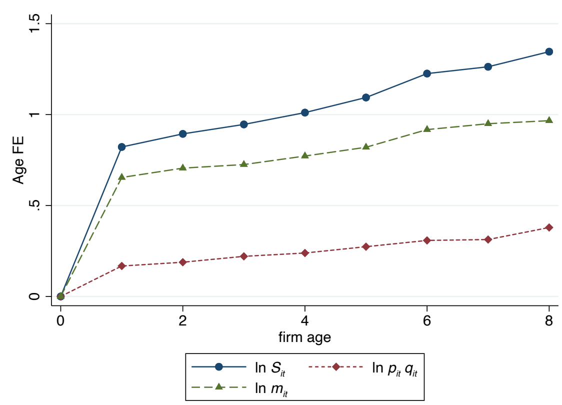
Notes: The regression specification is the same as the one estimated in Table A.4, with the exception that we replace the number of customers with total sales.

A.4. Firm Sales Growth Decomposition

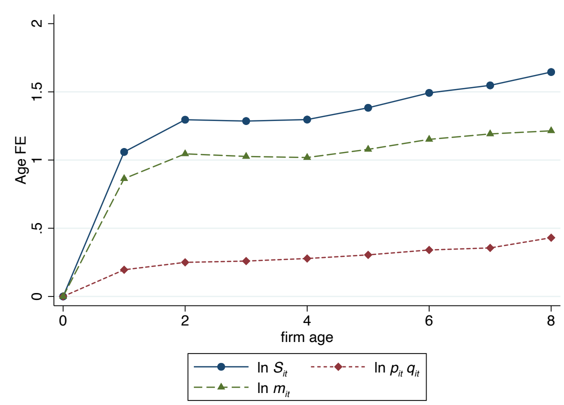
One concern in Figure 1 is that some firms might only appear temporarily in our data—not because of their actual behavior but due to sampling error. For example, it could be that households in our sample do not appear to purchase a firm's product even though the product was actually purchased and not 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 A.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.

Figure A.1: Decomposition of Firm Sales Growth by Firm Age: Survivors



(a) Survive for 3+ Consecutive Years

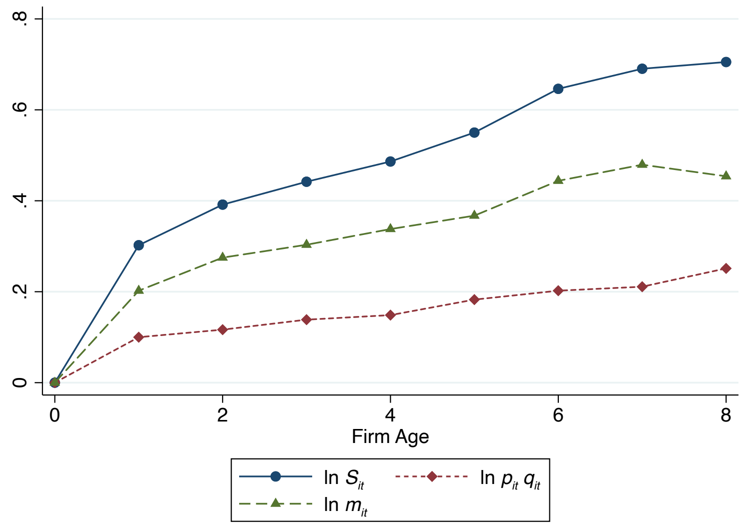


(b) Survive for 5+ Consecutive Years

Notes: Figures A.1a and Figure A.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 A.1a and 19,603 number of observations and 2,997 firm used in Figure A.1b.

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 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 A.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 A.2: Decomposition of Firm Sales Growth by Firm Age: Monthly Sales



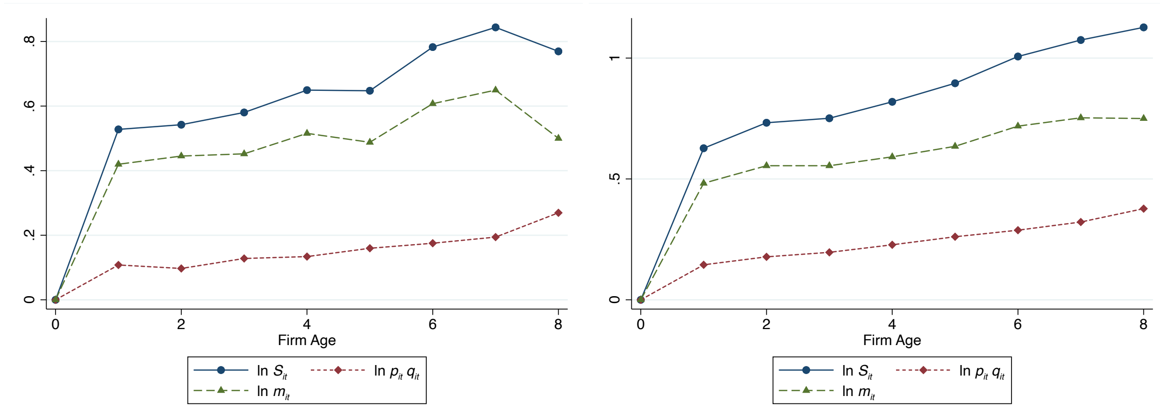
Notes: Figure A.2 replicates Figure 1 by using average monthly sales per firm and year instead of yearly sales.

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 (2019) 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 A.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 A.3: Decomposition of Firm Sales Growth by Firm Age: By Durability



(a) > p50 durability

(b) ≤ p50 durability

Notes: Figures A.3a and Figure A.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 A.3a and 29,816 observations and 7,323 firms used in Figure A.3b.

A.5. Customer Acquisition and Firms' Non-production Costs

In this section, we study the extent to which firms are able to control their growth through different sales margins. To do so, we estimate the following specification:

$$\ln S_{igt} = \gamma \ln \text{SGA}_{it} + \boldsymbol{X}'_{it}\boldsymbol{\nu} + \lambda_{ig} + \lambda_{st} + \lambda_{gt} + \varepsilon_{igt}, \tag{A.1}$$

where S_{igt} stands for sales and its components of firm i in product group g and year t, $\boldsymbol{X}'_{it}$ is a vector of firm-time-level control variables, λ_{ig} are firm-product-group fixed effects, λ_{st} are 2-digit SIC-year fixed effects, and λ_{gt} are product-group-year fixed effects.[34] The vector of controls $\boldsymbol{X}'_{it}$ includes lagged total sales and lagged total number of customers, which allow us to compare firms with similar relative sizes and customer bases. The coefficient of interest is γ, which captures the correlation between total sales (and its components) and SGA expenses.

As shown in the first column of Table A.6, firms that spend more on SGA expenses have larger sales. Moreover, the second and third columns show that approximately 95% (0.090/0.095) of the correlation between sales and SGA expenses is due to the correlation between non-production costs and the number of customers, not to the correlation with the average sales per customer. Finally, the last two columns further decompose the correlation of SGA expenses with the size of firms' customer bases into the acquisition of new customers and the retention of old customers. To measure these outcomes, we only include households that appear in the Nielsen data in at least two consecutive periods.[35] We find that while there is a strong correlation between SGA expenses and

[34] Since sales in the Nielsen data vary across both a detailed product category in the Nielsen data ("product group") and the major industry code available in Compustat data ("SIC"), we allow for both product-group fixed effects and firm-SIC-code fixed effects to compare products within fine product categories.

[35] There is a change in the number of surveyed households every year in the Nielsen data, especially in 2006 and 2007.

Table A.6: Sales and SGA: A Decomposition

	Decomposition of $\ln S_{igt}$			$\ln m_{igt}$: New vs. Old	
	(1)	(2)	(3)	(4)	(5)
	$\ln S$	$\ln pq$	$\ln m$	$\ln m^{\text{New}}$	$\ln m^{\text{Old}}$
$\ln \text{SGA}_{it}$	0.095***	0.005	0.090***	0.095***	0.016
	(0.036)	(0.014)	(0.028)	(0.032)	(0.027)
Observations	13131	13131	13131	13131	13131
R^2	0.962	0.909	0.965	0.943	0.961
Firm-year Controls	✓	✓	✓	✓	✓
Group-year FE	✓	✓	✓	✓	✓
SIC-year FE	✓	✓	✓	✓	✓
Group-firm FE	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; standard errors are two-way clustered at the firm and product-group level. S denotes total sales, pq the average sales per customers, m the number of customers, m^{New} the number of new customers, and m^{Old} the number of old customers. New customers are defined as customers who do not purchase firm i's products in group g at time $t-1$ but start to purchase those products at time t, whereas old customers are the customers who consecutively purchase firm i's products in group g in both $t-1$ and t ($m = m^{\text{New}} + m^{\text{Old}}$). We only use households that consecutively appear in $t-1$ and t in the Nielsen data when we measure m^{New} and m^{Old}. SIC industries correspond to a two-digit SIC code. All Nielsen variables are projection-factor adjusted.

the number of new customers, the regression coefficient for old customers is neither economically nor statistically significant. These results show that the non-production costs of firms are associated with the acquisition of new customers rather than retaining the existing customer base.

Robustness Below, we redo the analysis for subcomponents of SGA expenses. The results are consistent for subcomponents that are related to expansionary activities (advertising and rent expenses). We find a positive and statistically significant relationship between each of these SGA subcomponents and sales, which arises from the customer acquisition margin.[36]

Given our results in the theory, the extent to which firms can grow through customer acquisition depends on a *variable* component of SGA expenses, which has been a topic of discussion in the recent literature (see Traina, 2019, De Loecker, Eeckhout, and Unger, 2020). The correlation between sales and SGA expenses in Table A.6 is indicative of such a variable nature of these costs. However, as we show below, this contemporaneous correlation is weaker than the one between sales and COGS, which is commonly considered to be a measure of variable production costs. Therefore, total SGA expenses seem to be composed of both variable and fixed components. Although our

One concern in using these data is that the entry and exit of customers could arise from entry and exit in the sample. For example, the entry of new customers may not reflect the actual customer acquisition of a firm but may instead reflect an increase in the number of surveyed customers who already purchased this firm's products. To address this concern, new customers are defined as households that are present in the sample in periods $t-1$ and t but only buy the product at t. The exit of customers is similarly defined.

[36]On the other hand, other SGA subcomponents do not show a statistically significant relationship with sales at conventional levels, which supports the view that the correlation between sales and SGA expenses is due to firms' expansionary activities.

reduced-form empirical analyses do not allow us to separate these components, in our model we incorporate both components and provide a strategy to measure how much firms can grow through investing in their customer bases.

A.5.1. Sales and SGA Expenses: Decomposition by Durability of Products. Using the same durability measure we constructed and used in Figure A.3, we analyze the potential heterogeneity in the relationship between SGA expenses and sales based on a product's durability. Table A.7 reports the results. We find no statistically significant differences in this relationship by the durability of products.

Table A.7: Sales and SGA Expenses: Decomposition by Durability of Products

	Decomposition of ln S_{igt}			ln m_{igt}: New vs. Old	
	(1)	(2)	(3)	(4)	(5)
	ln S	ln pq	ln m	ln m^{New}	ln m^{Old}
ln SGA_{it}	0.079**	-0.006	0.085***	0.078***	0.024
	(0.036)	(0.017)	(0.030)	(0.030)	(0.038)
ln SGA_{it} x Durability	0.040	0.026	0.013	0.042	-0.018
	(0.046)	(0.028)	(0.032)	(0.039)	(0.046)
Observations	13131	13131	13131	13131	13131
R^2	0.962	0.909	0.965	0.943	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.6, with the exception that we include as additional regressors the durability measure and its interaction with the log of SGA expenses.

A.5.2. Sales and the Components of SGA Expenses. This section revisits the correlation of sales and SGA expenses by analyzing the subcomponents of the latter. In the main body of the paper, we rely on total SGA expenses as the main cost variable since it is reported for almost all the firms in the dataset. To be more explicit on the nature of these expenses, we further investigate the subcomponents of SGA expenses that are available for a subset of our sample. Among all the components reported in the data, the ones that are strongly correlated with firms' sales are related to expansionary activities: advertising (e.g., awareness of the existence of products/firms) and rent expenses (geographic proximity). We analyze these expansionary components to provide further evidence on the correlation between SGA expenses and sales.

Tables A.8 and A.9 show the correlation between sales and its components with advertising and rent expenses, respectively. Both tables show the same empirical relationship reported in Table A.6: The correlation between sales and the expansionary components of SGA costs arises from the correlation with the number of customers and, in particular, with the acquisition of new customers.

Overall, these results corroborate the empirical evidence of endogenous customer acquisition.[37]

Our analysis also reveals that other components of SGA (such as Foreign Exchange Income, Staff Expense, Receivables, and State Income Taxes) are not correlated with sales (results not reported), consistent with the view that total SGA expenses correlate with a firm's sales through their expansionary components.

Table A.8: Sales and Advertising Expenses

	Decomposition of $\ln S_{igt}$			$\ln m_{igt}$: New vs. Old	
	(1)	(2)	(3)	(4)	(5)
	$\ln S$	$\ln pq$	$\ln m$	$\ln m^{New}$	$\ln m^{Old}$
$\ln ADV_{it}$	0.073**	-0.005	0.078***	0.077**	0.019
	(0.032)	(0.012)	(0.028)	(0.031)	(0.028)
Observations	11239	11239	11239	11239	11239
R^2	0.964	0.920	0.966	0.947	0.966
Firm-year Controls	✓	✓	✓	✓	✓
Group-year FE	✓	✓	✓	✓	✓
SIC-year FE	✓	✓	✓	✓	✓
Group-firm FE	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regression specification is the same as the one estimated in Table A.6, with the exception that we replace total SGA expenses with advertising expenses.

Table A.9: Sales and Rent Expenses

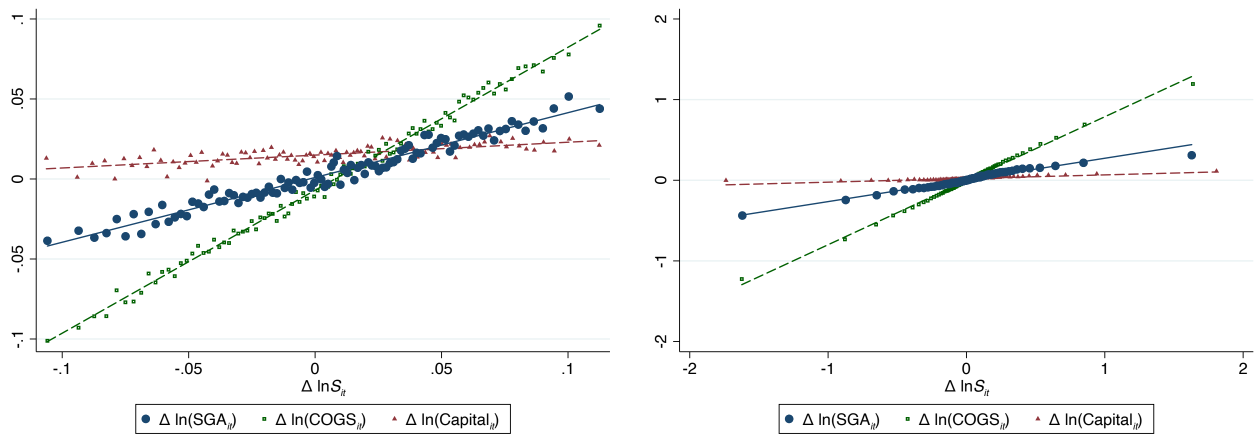
	Decomposition of $\ln S_{igt}$			$\ln m_{igt}$: New vs. Old	
	(1)	(2)	(3)	(4)	(5)
	$\ln S$	$\ln pq$	$\ln m$	$\ln m^{New}$	$\ln m^{Old}$
$\ln Rent_{it}$	0.086***	-0.004	0.090***	0.099***	0.027
	(0.030)	(0.015)	(0.025)	(0.029)	(0.030)
Observations	12552	12552	12552	12552	12552
R^2	0.962	0.907	0.965	0.943	0.960
Firm-year Controls	✓	✓	✓	✓	✓
Group-year FE	✓	✓	✓	✓	✓
SIC-year FE	✓	✓	✓	✓	✓
Group-firm FE	✓	✓	✓	✓	✓

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The regression specification is the same as the one estimated in Table A.6, with the exception that we replace total SGA expenses with rent expenses.

[37]For the particular case of R&D expenses, which could also be interpreted as an expansionary activity, the point estimates obtained are similar in magnitude to the ones reported in Tables A.8 and A.9. However, the small fraction of firms reporting these expenses decreases the precision of the estimates.

A.5.3. 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.

Figure A.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 (the 25th and 75th percentiles of the quarterly change of log sales are -8% and 11%, respectively). 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 17,168 firms in the 1964-2016 period used in this analysis.

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 A.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 A.4b shows that similar relationships hold in the full sample.

We consider other empirical specifications to confirm the semi-variable nature of SGA expenses.

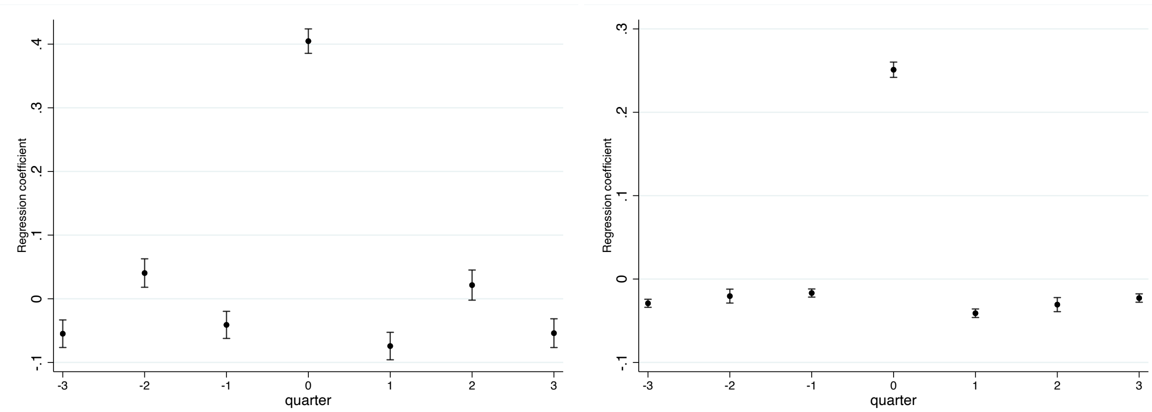
Table A.10 presents the regression results that correspond to Figure A.4. The semi-variable nature of SGA expenses is clear in this table, both with and without fixed effects. We also show that R&D expenses (a subcomponent of SGA costs) are more variable than the stock of capital but less variable than total SGA expenses. Finally, Figure A.5 reports the coefficients of a regression of SGA expenses on leads and lags of total sales, which further supports the short-run variability of SGA expenses: Although there is a strong correlation between SGA expenses and contemporaneous sales, we find no large correlation with future or past sales.

Table A.10: 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}	0.920***	0.894***	0.473***	0.405***	0.133***	0.081***	0.278***	0.200***
	(0.008)	(0.008)	(0.009)	(0.010)	(0.008)	(0.008)	(0.028)	(0.031)
R^2	0.055	0.162	0.011	0.103	0.002	0.155	0.002	0.083
Quarter FE	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	No	Yes	No	Yes	No	Yes	No	Yes
N	293962	292739	292785	291547	134743	133745	69845	69363

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 A.5: Correlation between SGA expenses and leads and lags of sales



(a) Trimmed Sample

(b) Full Sample

Notes: Figure A.5a uses the trimmed sample presented in Figure A.4b and Figure A.5b uses the full sample. 95% confidence intervals are presented for every estimate. Figure A.5a replicates column (4) in Table A.10 at quarter = 0.

B Derivations

B.1. Lemma 1

Proof. We start from the expression for the optimal price of the firm:

$$\ln\left(\frac{p_{i,t}}{D_t}\right) = \ln(\varepsilon_{i,t}) - \ln(\varepsilon_{i,t} - 1) + \ln\left(\frac{mc_{i,t}}{D_t}\right), \tag{B.1}$$

where

$$\varepsilon_{i,t} = -\frac{\partial \ln(q_{i,t})}{\partial \ln(p_{i,t})} = \frac{\sigma}{1 - \eta \ln(p_{i,t}) + \eta \ln(D_t(1-\sigma^{-1}))}. \tag{B.2}$$

The last equality in Equation (B.2) follows from the expression of demand per match in Equation (3.7). Differentiating Equation (B.1) we have

$$d\ln\left(\frac{p_{i,t}}{D_t}\right) = (1-\mu_{i,t})d\ln(\varepsilon_{i,t}) + d\ln\left(\frac{mc_{i,t}}{D_t}\right) = \frac{1}{1+\eta\sigma^{-1}\varepsilon_{i,t}(\mu_{i,t}-1)} d\ln\left(\frac{mc_{i,t}}{D_t}\right),$$

where $\mu_{i,t} \equiv \frac{\varepsilon_{i,t}}{\varepsilon_{i,t}-1} \geq 1$ is the firm's markup. Then, it follows that

$$d\ln(\mu_{i,t}) = d\ln(p_{i,t}) - d\ln(mc_{i,t}) = -\frac{\eta\sigma^{-1}\varepsilon_{i,t}(\mu_{i,t}-1)}{1+\eta\sigma^{-1}\varepsilon_{i,t}(\mu_{i,t}-1)} d\ln\left(\frac{mc_{i,t}}{D_t}\right). \tag{B.3}$$

Restricting $d\ln(D_t) = 0$ and considering partial changes in $\ln(mc_{i,t})$—e.g., moving in the cross-section of firms toward firms with higher marginal costs within a particular time when D_t is fixed—give us the expressions of interest. The sign restrictions follow from $\mu_{i,t} \geq 1$ and $\varepsilon_{i,t} \geq 0$. ■

B.2. Proposition 1

Proof. Consider the sales per match of firm i normalized by the demand index D_t, $p_{i,t}q_{i,t}/D_t$. Differentiating the log of this quantity, we have

$$d\ln\left(\frac{p_{i,t}q_{i,t}}{D_t}\right) = (1-\varepsilon_{i,t})d\ln\left(\frac{p_{i,t}}{D_t}\right) = -\frac{\varepsilon_{i,t}-1}{1+\eta\sigma^{-1}\varepsilon_{i,t}(\mu_{i,t}-1)} d\ln\left(\frac{mc_{i,t}}{D_t}\right). \tag{B.4}$$

Therefore, combining this expression with Equation (B.3), we have

$$d\ln(\mu_{i,t}) = \eta\sigma^{-1}\mu_{i,t}(\mu_{i,t}-1)d\ln\left(\frac{p_{i,t}q_{i,t}}{D_t}\right). \tag{B.5}$$

Again, restricting $d\ln(D_t) = 0$ and considering partial changes in sales per customer—e.g., moving in the cross-section of firms toward firms with higher sales per customer within a particular period when D_t is fixed—we get

$$\left.\frac{d\ln(\mu_{i,t})}{d\ln(p_{i,t}q_{i,t})}\right|_{d\ln(D_t)=0} = \eta\sigma^{-1}\mu_{i,t}(\mu_{i,t}-1) \geq 0, \tag{B.6}$$

where the sign restriction follows from $\mu_{i,t} \geq 1$. ■

B.3. Corollary 1

Proof. Recall that a firm's relative total sales within a period is given by

$$\frac{p_{i,t}y_{i,t}}{\int_{i\in N_t} p_{i,t}y_{i,t}di} = \frac{p_{i,t}y_{i,t}}{C_t} = p_{i,t}q_{i,t}m_{i,t}. \tag{B.7}$$

Now, restricting $p_{i,t}y_{i,t}/C_t = \bar{s}$, we have

$$0 = d\ln(p_{i,t}y_{i,t})\Big|_{p_{i,t}y_{i,t}/C_t=\bar{s}} = d\ln(p_{i,t}q_{i,t})\Big|_{p_{i,t}y_{i,t}/C_t=\bar{s}} + d\ln(m_{i,t})\Big|_{p_{i,t}y_{i,t}/C_t=\bar{s}} \tag{B.8}$$

$$\Rightarrow d\ln(p_{i,t}q_{i,t})\Big|_{p_{i,t}y_{i,t}/C_t=\bar{s}} = -d\ln(m_{i,t})\Big|_{p_{i,t}y_{i,t}/C_t=\bar{s}}. \tag{B.9}$$

Now, using Equation (B.6) we have

$$\frac{\partial\ln(\mu_{i,t})}{\partial\ln(m_{i,t})}\Big|_{p_{i,t}y_{i,t}/C_t=\bar{s}} = -\frac{\partial\ln(\mu_{i,t})}{\partial\ln(p_{i,t}q_{i,t})} = -\eta\sigma^{-1}\mu_{i,t}(\mu_{i,t}-1) \le 0, \tag{B.10}$$

where the sign restriction follows from $\mu_{i,t} \ge 1$. ■

B.4. Proposition 2

Proof. This relationship is obtained directly from the first-order condition of the firm's problem with respect to $m_{i,t}$. For the rest of the proof, we derive this first-order condition.

We start by showing that the firm's customer acquisition constraint always binds (meaning that the firm never disposes of its existing customers). To show this, note that it cannot be the case that $l_{i,s,t} > 0$ but $m_{i,t} < (1-\delta)m_{i,t-1} + \frac{l_{i,s,t}^{\phi}}{P_{m,t}}$ since the firm can keep the same $m_{i,t}$ with a lower $l_{i,s,t}$. Thus, if $m_{i,t} < (1-\delta)m_{i,t-1} + \frac{l_{i,s,t}^{\phi}}{P_{m,t}}$, then optimality requires that $l_{i,s,t} = 0$. Now, suppose $l_{i,s,t} = 0$ but $m_{i,t} < (1-\delta)m_{i,t-1}$. Note that in this case, the slope of the firm's "production" profit function with respect to $m_{i,t}$ is given by

$$\frac{\partial}{\partial m_{i,t}}(p_{i,t}y_{i,t} - W_t l_{i,p,t}) = (p_{i,t} - mc_{i,t})\frac{y_{i,t}}{m_{i,t}} > 0, \tag{B.11}$$

where the last equality follows from the fact that for any choice of $q_{i,t} > 0$, the firm's markup is always strictly larger than 1 and hence $p_{i,t} > mc_{i,t}$. Thus, the firm's profit is strictly increasing in $m_{i,t}$ and since $m_{i,t} < (1-\delta)m_{i,t-1}$, then the firm can increase its $m_{i,t}$ at no cost and gain more profits at time t. Moreover, this will not affect firms' profits in the future, since the firms can always dispose of the increase in $m_{i,t}$ in the next period at no cost. Hence, optimality requires that $m_{i,t} = (1-\delta)m_{i,t-1} + \frac{l_{i,s,t}^{\phi}}{P_{m,t}}$.

Now, in writing firm i's problem at time t, replace $l_{i,p,\tau} = \left(y_{i,\tau}/z_{i,\tau}\right)^{\alpha^{-1}}$, $y_{i,\tau} = m_{i,\tau}q_{i,\tau}C_\tau$, and $l_{i,s,\tau} = P_{m,\tau}^{\phi^{-1}}(m_{i,\tau} - (1-\delta)m_{i,\tau-1})^{\phi^{-1}}$ to obtain the problem as

$$\max_{\{p_{i,\tau},m_{i,\tau},q_{i,\tau}\}_{\tau\ge t}} \mathbb{E}_t \sum_{\tau\ge t}(\beta\nu)^{\tau-t}C_\tau^{-\gamma}\left(\prod_{h=t}^{\tau}\mathbf{1}_{i,h}\right)\times$$

$$\left[p_{i,\tau} m_{i,\tau} q_{i,\tau} C_\tau - W_\tau \left(\frac{m_{i,\tau} q_{i,\tau} C_\tau}{z_{i,\tau}} \right)^{\alpha^{-1}} - W_\tau P_{m,\tau}^{\phi^{-1}} (m_{i,\tau} - (1-\delta) m_{i,\tau-1})^{\phi^{-1}} - W_\tau \chi \right]$$

$$s.t. \quad q_{i,\tau} = \left[1 - \eta \ln \left(\frac{p_{i,\tau}}{D_\tau (1-\sigma^{-1})} \right) \right]^{\frac{\sigma}{\eta}}.$$

Next, if $\mathbf{1}_{i,t} = 0$, then $l_{i,s,t} = 0$. However, conditional on $\mathbf{1}_{i,t} = 1$, the FOC with respect to $m_{i,t}$ is

$$\begin{aligned} 0 = & \mathbf{1}_{i,t} (p_{i,t} - \alpha^{-1} \frac{W_t l_{i,p,t}}{y_{i,t}}) q_{i,t} C_t - \mathbf{1}_{i,t} \phi^{-1} \frac{W_t l_{i,s,t}}{m_{i,t} - (1-\delta) m_{i,t-1}} \\ & + \beta \nu (1-\delta) \mathbb{E}_t \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \mathbf{1}_{i,t+1} \phi^{-1} \frac{W_{t+1} l_{i,s,t+1}}{m_{i,t+1} - (1-\delta) m_{i,t}} \right]. \end{aligned}$$

Replacing $mc_{i,t} = \alpha^{-1} \frac{W_t l_{i,p,t}}{y_{i,t}}$ and iterating the FOC forward gives us the expression of interest. ■

B.5. Proposition 3

Proof. Recall that from Equation (3.16) the relationship between a firm's production labor share and markup is given by

$$\frac{W_t l_{i,p,t}}{p_{i,t} y_{i,t}} = \frac{\alpha}{\mu_{i,t}}. \tag{B.12}$$

Moreover, assuming $\delta = 1$, we can use the characterization of the firm's optimal marketing strategy in Equation (3.21) to write their marketing labor share of an operating firm as

$$\phi^{-1} \frac{W_t l_{i,s,t}}{m_{i,t}} = (p_{i,t} - mc_{i,t}) q_{i,t} C_t \Leftrightarrow \frac{W_t l_{i,s,t}}{p_{i,t} y_{i,t}} = \phi (1 - \mu_{i,t}^{-1}). \tag{B.13}$$

Combining these two equations, we get that

$$W_t l_{i,s,t} = \phi p_{i,t} y_{i,t} - \phi \alpha^{-1} W_t l_{i,p,t}. \tag{B.14}$$

Finally, notice that

$$SGA_{i,t} \equiv W_t \chi + W_t l_{i,s,t} = \underbrace{SGAF_t}_{=W_t \chi} + \phi \underbrace{Sales_{i,t}}_{=p_{i,t} y_{i,t}} - \frac{\phi}{\alpha} \underbrace{COGS_{i,t}}_{=W_t l_{i,p,t}}. \tag{B.15}$$

■

B.6. Proposition 4

Proof. This result can be derived from combining Equations (B.12) and (B.13):

$$\frac{W_t (l_{i,p,t} + l_{i,s,t})}{p_{i,t} y_{i,t}} = \alpha \mu_{i,t}^{-1} + \phi (1 - \mu_{i,t}^{-1}). \tag{B.16}$$

Notice that this is strictly decreasing in $\mu_{i,t}$ if and only if $\alpha > \phi$. Hence, the firm's revenue productivity of labor, the inverse of the equation above, is increasing in $\mu_{i,t}$ if and only if $\alpha > \phi$. ■

B.7. Proposition 5

Proof. Let $(C_t, L_{p,t}, L_{s,t}, L_t, N_t)_{t \geq 0}$ denote the equilibrium allocation. A log-linearization of $U(C_t, L_t) =$

$\frac{C_t^{1-\gamma}}{1-\gamma} - \xi \frac{L_t^{1+\psi}}{1+\psi}$ around this allocation gives

$$\Delta U(C_t, L_t) = C_t^{1-\gamma} \Delta \ln(C_t) - \xi L_t^{\psi} L_{p,t} (\Delta \ln(L_{p,t}) + \frac{L_{s,t}}{L_{p,t}} \Delta \ln(L_{s,t}) + \chi \frac{N_t}{L_{p,t}} \Delta \ln(N_t)) + \mathcal{O}(\|.\|^2). \tag{B.17}$$

Next, divide by $U_{c,t} C_t = C_t^{1-\gamma}$ and use the household's optimal labor supply condition $\xi \frac{L_t^{\psi}}{C_t^{-\gamma}} = W_t$ to get

$$\frac{\Delta U(C_t, L_t)}{U_{c,t} C_t} = \Delta \ln(C_t) - \frac{W_t L_{p,t}}{C_t} (\Delta \ln(L_{p,t}) + \frac{L_{s,t}}{L_{p,t}} \Delta \ln(L_{s,t}) + \chi \frac{N_t}{L_{p,t}} \Delta \ln(N_t)) + \mathcal{O}(\|.\|^2). \tag{B.18}$$

Finally, using the aggregate production function in Equation (5.2), replace $\Delta \ln(C_t) = \Delta \ln(Z_t) + \alpha \Delta \ln(L_{p,t})$, and using the definition of the aggregate markup in Equation (5.6), replace the labor share in terms of the cost-weighted markup ($\frac{W_t L_{p,t}}{C_t} = \frac{\alpha}{\mathcal{M}_t}$) to get

$$\frac{\Delta U(C_t, L_t)}{U_{c,t} C_t} \approx \Delta \ln(Z_t) + \alpha (1 - \mathcal{M}_t^{-1}) \Delta \ln(L_{p,t}) - \alpha \mathcal{M}_t^{-1} (\frac{L_{s,t}}{L_{p,t}} \Delta \ln(L_{s,t}) + \chi \frac{N_t}{L_{p,t}} \Delta \ln(N_t)). \tag{B.19}$$

■

B.8. Planner's Problem

The planner's problem for this economy is given by

$$\max_{\left\{ \begin{matrix} (c_{i,j,t})_{j \in [0,1]}, (\mathbf{1}_{i,t})_{i \in N_{t-1} \cup \Lambda_t}, \\ (\delta_{i,t}, m_{i,t}, l_{i,p,t}, l_{i,s,t})_{i \in N_t}, C_t \end{matrix} \right\}_{t \geq 0}} \sum_{t=0}^{\infty} \beta^t \left[\frac{C_t^{1-\gamma}}{1-\gamma} - \xi \frac{L_t^{1+\psi}}{1+\psi} \right] \tag{B.20}$$

subject to

$$\int_0^1 \mathbf{1}_{\{j \in m_{i,t}\}} c_{i,j,t} dj = z_{i,t} l_{i,p,t}^{\alpha}, \quad \forall i \in N_t, \tag{B.21}$$

$$\int_{i \in N_t} \int_0^1 \mathbf{1}_{\{j \in m_{i,t}\}} \Upsilon \left(\frac{c_{i,j,t}}{C_t} \right) dj di = 1 \tag{B.22}$$

$$\int_{i \in N_t} (l_{i,p,t} + l_{i,s,t} + \chi) di = L_t \tag{B.23}$$

$$\int_{i \in N_t} m_{i,t} di = 1 \tag{B.24}$$

$$N_t = \{ i \in N_{t-1} \cup \Lambda_t : \mathbf{1}_{i,t} \nu_{i,t} = 1 \}, \quad N_{-1} \text{ given.} \tag{B.25}$$

$$m_{i,t} = (1 - \delta_{i,t}) m_{i,t-1} + \frac{1 - \int_{i \in N_t} (1 - \delta_{i,t}) m_{i,t-1}}{\int_{i \in N_t} l_{i,s,t}^{\phi} di} l_{i,s,t}^{\phi}, \quad \forall i \in N_t, \tag{B.26}$$

$$\delta_{i,t} \in [\delta, 1], l_{i,s,t} \geq 0, \quad \forall i \in N_t, \tag{B.27}$$

$$\int_{i \in N_t} l_{i,s,t} di = \bar{L}_{s,t} > 0. \tag{B.28}$$

Here, Equation (B.21) requires that for every firm, their supply meets their allocated demand; Equation (B.22) is the Kimball aggregator that implicitly defines C_t given the planner's allocation of demand; Equation (B.23) requires that labor supply meets demand for labor from production, advertising and overhead costs; Equation (B.24) requires that the matching market clears; Equation (B.25) is the law of motion for the set of operating firms given an entry/exit policy by the planner; Equation (B.26) determines firm i's evolution of customers given an allocation for advertising; Equation (B.27) requires the nonnegativity of labor for advertising and the constraint that while the planner can separate customers from firms, the separation rate should be at least δ; and finally Equation (B.28) requires that the planner at least spends $\bar{L}_{s,t}$ on advertising. This last constraint is for an arbitrary but strictly positive $\bar{L}_{s,t}$—in Lemma 2 we show that the level of this quantity does not matter for the optimal distribution of customers, which is also well defined in the limit when $\bar{L}_{s,t} \to 0$.

B.9. Lemma 2

Proof. Suppose that at any given time t, a choice for N_t is fixed. Suppose next that the planner desires to allocate matches according to a rule

$$\mathscr{A} : (i \to m^*_{i,t})_{i \in N_t}. \tag{B.29}$$

Note that this can be any arbitrary allocation of matches as long as it is feasible:

$$m^*_{i,t} \geq 0, \quad \forall i \in N_t \tag{B.30}$$

$$\int_{i \in N_t} m^*_{i,t} di = 1. \tag{B.31}$$

To show that the allocation $\mathscr{A}$ is implementable on N_t for any given level of $\bar{L}_{s,t}$, we need to show that (1) it is generated by a choice of $(\delta \leq \delta_{i,t} \leq 1, l_{i,s,t} \geq 0)_{i \in N_t}$, and (2) it is feasible $\int_{i \in N_t} l_{i,s,t} di = \bar{L}_{s,t}$. We show this by construction. In particular, consider the choice

$$\left(\delta^*_{i,t} = 1, l^*_{i,s,t} = \bar{L}_{s,t} \frac{m^{*\phi^{-1}}_{i,t}}{\int_{i \in N_t} m^{*\phi^{-1}}_{i,t} di}\right)_{i \in N_t} \tag{B.32}$$

That is, first, let the planner separate all the matches from their corresponding firms ($\delta^*_{i,t} = 1$) and then reallocate them based on $\mathscr{A}$. It follows that Equations (B.30) and (B.31) from above hold by construction. Next, to verify that these values implement $\mathscr{A}$, observe that

$$m_{i,t} \equiv (1 - \delta_{i,t}) m_{i,t-1} + \left(1 - \int_{i \in N_t} (1 - \delta_{i,t}) m_{i,t-1} di\right) \frac{l^{*\phi}_{i,s,t}}{\int_{i \in N_t} l^{*\phi}_{i,s,t} di} = m^*_{i,t}. \tag{B.33}$$

Finally, to confirm feasibility, note that $\int_{i \in N_t} l^*_{i,s,t} di = \bar{L}_{s,t}$. ■

B.10. Proposition 6

Proof. To prove this Proposition, we proceed in two steps. First, we characterize the optimal demand per customer and the allocation of customers for a given set of operating firms, N_t, over time. Second, we show that the *optimal* allocation for these two objects maximizes the aggregate TFP, given N_t, subject to feasibility constraints for $(q_{i,t}, m_{i,t})_{i \in N_t}$.

Moreover, throughout this proof we rely on the result from Lemma 2 and directly characterize the distribution of matches $m_{i,t}$, ignoring the constraints in Equations (B.26) and (B.27) as well as Equation (B.28). Having characterized this distribution, we can then use the result from Lemma 2 to find the allocation of advertising labor that implements it.

Step 1: Optimal Allocation of Demand. The results in this Proposition follow from the first-order conditions of the planner's problem in Equation (B.20), fixing the planner's other choices at an arbitrary allocation. In the remainder of this proof, we characterize these first-order conditions.

Formally, for firm i and period t, let $\beta^t \varphi_{c,i,t} di$ be the shadow cost on Equation (B.21); for period t, let $\beta^t \varphi_{\Upsilon,t}$, $\beta^t \varphi_{L,t}$ and $\beta^t \varphi_{m,t}$ be the shadow costs on Equation (B.22), Equation (B.23) and Equation (B.24), respectively. Moreover, similar to the equilibrium allocation, let us define $q_{i,j,t} \equiv \frac{c_{i,j,t}}{C_t}$. It is straight forward to show that for $j \notin m_{i,t}$, $q_{i,j,t} = 0$. So hereafter, we only refer to $q_{i,j,t}$ when $j \in m_{i,t}$.

The first-order conditions with respect to $q_{i,j,t}$ are

$$\varphi_{c,i,t} C_t = \Upsilon'(q_{i,j,t}) \varphi_{\Upsilon,t}, \quad \forall j \in m_{i,t}. \tag{B.34}$$

It immediately follows that all households that are matched to a variety consume the same amount:

$$q_{i,j,t} = q_{i,t}, \quad \forall j \in m_{i,t}. \tag{B.35}$$

Replacing this result in Equation (B.21) and Equation (B.22) and taking the first-order condition with respect to $m_{i,t}$, we have

$$\varphi_{c,i,t} q_{i,t} C_t + \varphi_{m,t} = \Upsilon(q_{i,t}) \varphi_{\Upsilon,t}. \tag{B.36}$$

Notice that in deriving this first-order condition, we have ignored the constraint in Equation (B.26). The reason we can do this goes back to Lemma 2, which states that any choice of $(m_{i,t})_{i \in N_t}$ can be implemented without any loss of generality. Therefore, we can ignore the constraint in Equation (B.26) and use Lemma 2 to show that it is satisfied.

Next, replacing Equation (B.35) in Equation (B.34), multiplying it by $q_{i,t}$ and subtracting it from Equation (B.36), we have

$$\varphi_{m,t} = \left[\Upsilon(q_{i,t}) - q_{i,t} \Upsilon'(q_{i,t}) \right] \varphi_{\Upsilon,t}. \tag{B.37}$$

Since $\varphi_{\Upsilon,t} \neq 0$,[38] it follows that

$$\Upsilon(q_{i,t}) - q_{i,t} \Upsilon'(q_{i,t}) = \frac{\varphi_{m,t}}{\varphi_{\Upsilon,t}}, \quad \forall i \in N_t. \tag{B.38}$$

Notice that the left-hand-side of this equation is only a function of $q_{i,t}$ and is strictly monotonic in $q_{i,t} > 0$.[39] Moreover, the right-hand-side of the equation is only a function of time-t shadow costs

[38] To see why, suppose not. Then, by Equation (B.34), either $C_t = 0$—which is clearly not optimal since marginal utility approaches infinity as $C_t \to 0$—or $\varphi_{c,i,t} = 0$—which means that the household can freely supply infinite labor to firm i at t and is also a contradiction since it violates the positive disutility of the labor supply.

[39] Observe that $D_x[\Upsilon(x) - \Upsilon'(x)x] = -\Upsilon''(x)x > 0$.

and is independent of i. Hence, there exists a unique q_t^* such that

$$q_{i,t} = q_t^* \quad \forall i \in N_t. \tag{B.39}$$

Replacing this last equation into Equation (B.22) we have

$$\int_{i\in N_t} m_{i,t}\Upsilon(q_t^*)di = 1 \Rightarrow \Upsilon(q_t^*) = 1 \Rightarrow q_t^* = 1, \tag{B.40}$$

where the second statement uses the market-clearing condition for matches in Equation (B.24) and the last statement uses the strict monotonicity of $\Upsilon(x)$ and the fact that $\Upsilon(1) = 1$.

Given that the social planner sets $q_{i,t} = 1$ for all firms, this implies that firms' production will differ under the efficient allocation only through different numbers of customers. To determine the optimal level of production, we only need to consider the first-order condition with respect to $l_{i,p,t}$:

$$\alpha\varphi_{c,i,t} z_{i,t} l_{i,p,t}^{\alpha-1} = \varphi_{L,t}. \tag{B.41}$$

Dividing this equation by the first-order condition for $m_{i,t}$ in Equation (B.34), we get

$$z_{i,t} l_{i,p,t}^{\alpha-1} = \frac{C_t}{\alpha\Upsilon'(1)}\frac{\varphi_{L,t}}{\varphi_{\Upsilon,t}}. \tag{B.42}$$

Solving for $l_{i,p,t}$ from this equation and replacing it Equation (B.21) we have

$$\int_{j\in m_{i,t}} c_{i,j,t}dj = m_{i,t}C_t = z_{i,t}\left(\frac{C_t}{z_{i,t}\alpha\Upsilon'(1)}\frac{\varphi_{L,t}}{\varphi_{\Upsilon,t}}\right)^{\frac{\alpha}{\alpha-1}} \Rightarrow m_{i,t} = \left(\frac{z_{i,t}}{C_t}\right)^{\frac{1}{1-\alpha}}\left(\frac{\varphi_{L,t}}{\alpha\Upsilon'(1)\varphi_{\Upsilon,t}}\right)^{\frac{\alpha}{\alpha-1}}. \tag{B.43}$$

Finally, imposing the market-clearing condition for matches in Equation (B.24) we get

$$m_{i,t} = \frac{z_{i,t}^{\frac{1}{1-\alpha}}}{\int_{i\in N_t} z_{i,t}^{\frac{1}{1-\alpha}}di}, \quad \forall i \in N_t. \tag{B.44}$$

Step 2: Optimal Demand Allocation Maximizes Aggregate TFP. First, let us derive aggregate TFP. Consider an allocation of $l_{i,p,t}$ across firms in N_t. Then, aggregate production labor is given by

$$L_{p,t} = \int_{i\in N_t} l_{i,p,t}di = \int_{i\in N_t}\left(\frac{C_t\int_{j\in m_{i,t}} q_{i,j,t}dj}{z_{i,t}}\right)^{\alpha^{-1}} di \tag{B.45}$$

where the second equality follows from the fact that for every firm, demand must meet supply. Rearranging this gives us an expression for TFP as a function of demand allocation within N_t:

$$C_t = \underbrace{\left[\int_{i\in N_t}\left(\frac{z_{i,t}}{\int_{j\in m_{i,t}} q_{i,j,t}dj}\right)^{-\alpha^{-1}} di\right]^{-\alpha}}_{Z_t \equiv \text{Aggregate TFP}} \times L_{p,t}^{\alpha} \tag{B.46}$$

Now, maximizing Z_t is equivalent to minimizing $Z_t^{-\alpha^{-1}}$. Thus, the problem of maximizing Z_t subject to the Kimball aggregator becomes

$$\min_{\{q_{i,j,t},\mathbf{1}_{\{j\in m_{i,t}\}}\}} \int_{i\in N_t} \left(\frac{\int_0^1 \mathbf{1}_{\{j\in m_{i,t}\}} q_{i,j,t} dj}{z_{i,t}} \right)^{\alpha^{-1}} di \tag{B.47}$$

$$s.t. \int_{i\in N_t} \int_0^1 \mathbf{1}_{\{j\in m_{i,t}\}} \Upsilon(q_{i,j,t}) dj di = 1 \tag{B.48}$$

$$\int_{i\in N_t} \int_0^1 \mathbf{1}_{\{j\in m_{i,t}\}} dj di = 1 \tag{B.49}$$

First, fix an allocation for $m_{i,t}$. Notice that since $\alpha^{-1} > 1$, the objective function is convex in each $q_{i,j,t}$ for any $j \in m_{i,t}$ so the first-order condition for $q_{i,j,t}$ is sufficient for optimality. Allowing $\eta_{\Upsilon,t}$ and $\eta_{m,t}$ to be the multipliers on Equations (B.48) and (B.49), the first-order condition for $q_{i,j,t}, j \in m_{i,t}$ reads

$$\frac{1}{\alpha z_{i,t} \eta_{\Upsilon,t}} \left(\frac{\int_0^1 \mathbf{1}_{\{j\in m_{i,t}\}} q_{i,j,t} dj}{z_{i,t}} \right)^{\alpha^{-1}-1} = \Upsilon'(q_{i,j,t}). \tag{B.50}$$

Since the left-hand-side of this equation depends only on firm/time level variables, it follows that all the consumers in the $m_{i,t}$ consume the same $q_{i,j,t}$—i.e., $q_{i,j,t} = q^*_{i,t}$ where $q^*_{i,t}$ solves the above equation. Substituting this into the objective, we can rewrite the choice of $\mathbf{1}_{\{j\in m_{i,t}\}}$ as simply choosing the size of the customer base for firms, which with slight abuse of notation we also denote as $m_{i,t}$. Formally,

$$\int_0^1 \mathbf{1}_{\{j\in m_{i,t}\}} q_{i,j,t} dj = m_{i,t} q^*_{i,t} \tag{B.51}$$

Again, it is straightforward to see that the objective is convex in $m_{i,t}$ and the first-order condition for it reads

$$\frac{q^*_{i,t}}{\alpha z_{i,t}} \left(\frac{m_{i,t} q^*_{i,t}}{z_{i,t}} \right)^{\alpha^{-1}-1} + \eta_{m,t} = \Upsilon(q^*_{i,t}) \eta_{\Upsilon,t}. \tag{B.52}$$

Subtracting the first-order condition for $m_{i,t}$ from the first-order condition for $q_{i,j,t}$ we arrive at

$$\frac{\eta_{m,t}}{\eta_{\Upsilon,t}} = \Upsilon(q^*_{i,t}) - q^*_{i,t} \Upsilon'(q^*_{i,t}). \tag{B.53}$$

Note that the right-hand-side is a monotonic function of $q^*_{i,t}$ (since Υ is increasing and concave) and the left-hand-side does not depend on i. So $q^*_{i,t} = q^*_t$ for some q^*_t. Substituting $q_{i,j,t} = q^*_t$ in Equation (B.48) then yields that $\Upsilon(q^*_t) = 1$, which implies $q^*_t = 1$. Substituting this into the FOCs and using the constraints, we then retrieve the optimal $m^*_{i,t}$ as

$$m^*_{i,t} = \frac{z_{i,t}^{\frac{1}{1-\alpha}}}{\int_{i\in N_t} z_{i,t}^{\frac{1}{1-\alpha}} di} \tag{B.54}$$

Hence, choosing the allocation of demand to maximize the aggregate TFP gives the optimal allocation of demand under the social planner's problem. ■

B.11. Proposition 7

Proof. Our goal here is to derive the social value of a firm that enters the economy at a certain point in time. To do so, we start by writing a (partial) Lagrangian for the planner's problem and rearrange it to isolate all the terms in the Lagrangian that are related to a particular firm. In writing this Lagrangian, we employ two of the results we have established so far. First, given our result in Lemma 2, we ignore the constraints in Equations (B.26) and (B.27) as well as Equation (B.28). Instead, we can allow the social planner to directly choose the distribution of matches at any given time and use the results in the proof of Lemma 2 to substitute for the optimal advertising labor. Second, to render the comparison to the equilibrium value of firms natural, we take advantage of Equation (B.35), which establishes that the social planner chooses the same $q_{i,j,t}$ for all households that are matched to firm i at time t. Aside from these two results, however, we refrain from imposing any other optimality conditions at the firm level to maximize the comparability of the firms' social value with their equilibrium value.

Thus, we can form the following partial Lagrangian for the social planner (with the Lagrange multipliers defined as in Appendix B.10):

$$\max_{\left\{\substack{(\mathbf{1}_{i,t})_{i\in N_{t-1}\cup\Lambda_t},\\ (q_{i,t},m_{i,t},l_{i,p,t})_{i\in N_t},C_t,L_t}\right\}_{t\geq 0}} \sum_{t=0}^{\infty}\beta^t\left[\frac{C_t^{1-\gamma}}{1-\gamma}-\xi\frac{L_t^{1+\psi}}{1+\psi}+\varphi_{\Upsilon,t}\left(\int_{i\in N_t} m_{i,t}\Upsilon(q_{i,t})-1\right)\right.$$
$$\left.+\varphi_{L,t}\left(L_t-\int_{i\in N_t}(l_{i,p,t}+\bar{L}_{s,t}+\chi)di\right)+\varphi_{m,t}\left(1-\int_{i\in N_t} m_{i,t}di\right)\right] \tag{B.55}$$

subject to the constraints in Equation (B.21) (i.e., demand meets supply) and Equation (B.25), which captures the law of motion for the set of firms active in the economy. To rearrange this partial Lagrangian so that it resembles the problem of the firms in the equilibrium, consider the following two operations: First, rearrange the terms within the brackets to put all integrals of the form $\int_{i\in N_t} di$ in one bracket. Second, factor out $\varphi_{\Upsilon,t}$ from these integrals. These two operations lead to the following expression for the partial Lagrangian:

$$\max_{\left\{\substack{(\mathbf{1}_{i,t})_{i\in N_{t-1}\cup\Lambda_t},\\ (q_{i,t},m_{i,t},l_{i,p,t})_{i\in N_t},C_t,L_t}\right\}_{t\geq 0}} \mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\left[\frac{C_t^{1-\gamma}}{1-\gamma}-\xi\frac{L_t^{1+\psi}}{1+\psi}-\varphi_{\Upsilon,t}+\varphi_{L,t}\left(L_t-\bar{L}_{s,t}\right)+\varphi_{m,t}\right]$$
$$+\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t\int_{i\in N_t}\varphi_{\Upsilon,t}\left(m_{i,t}\Upsilon(q_{i,t})-\frac{\varphi_{L,t}}{\varphi_{\Upsilon,t}}(l_{i,p,t}+\chi)-\frac{\varphi_{m,t}}{\varphi_{\Upsilon,t}}m_{i,t}\right)di \tag{B.56}$$

The second line is what we are after, once we switch the order of the sum and the integration: Since this expression first integrates over the set of all operating firms at time t and then sums over time, we can isolate the contribution of every firm to this Lagrangian separately by switching the order of integration and summation. In doing so, we will take advantage of the law of motion for the set N_t, Equation (B.25), as a function of the planner's choice for the entry/exit of firms as well as endogenous exit shocks. Formally, pick a firm i that enters the economy for the first time at time t (i.e., $t = \min_{\tau\geq 0}\{i \in N_\tau\}$, with $t \equiv -1$ indicating firms that were in the initial distribution of firms in the economy). Then, i will be in any N_τ, for $\tau \geq t$ as long as $\prod_{h=t}^{\tau} \nu_{i,h}\mathbf{1}_{i,h} = 1$. Therefore, the second line in the equation above can be written as

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_{i\in N_t} \varphi_{\Upsilon,t}\left(m_{i,t}\Upsilon(q_{i,t}) - \frac{\varphi_{L,t}}{\varphi_{\Upsilon,t}}(l_{i,p,t}+\chi) - \frac{\varphi_{m,t}}{\varphi_{m,t}} m_{i,t}\right) di \tag{B.57}$$

$$= \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \int_{i\in \tilde{\Lambda}_t} \mathbb{E}_t \sum_{\tau=t}^{\infty} (\beta\nu)^{\tau-t}\left(\prod_{h=t}^{\tau} \mathbf{1}_{i,h}\right)\varphi_{\Upsilon,\tau}\left(m_{i,\tau}\Upsilon(q_{i,\tau}) - \frac{\varphi_{L,\tau}}{\varphi_{\Upsilon,\tau}}(l_{i,p,\tau}+\chi) - \frac{\varphi_{m,\tau}}{\varphi_{\Upsilon,\tau}} m_{i,\tau}\right) di, \tag{B.58}$$

where $\tilde{\Lambda}_0 = \Lambda_0 \cup N_{-1}$ and $\tilde{\Lambda}_t = \Lambda_t$ for all $t \geq 1$.

Finally, substituting Equation (B.57) into Equation (B.56), we arrive at the following reformulation of the social planner's problem:

$$\max_{\{C_t,L_t\}_{t\geq 0}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \Bigg\{ \frac{C_t^{1-\gamma}}{1-\gamma} - \xi \frac{L_t^{1+\psi}}{1+\psi} - \varphi_{\Upsilon,t} + \varphi_{L,t}\left(L_t - \bar{L}_{s,t}\right) + \varphi_{m,t}$$
$$+ \varphi_{\Upsilon,t} \int_{i\in\tilde{\Lambda}_t} \underbrace{\max_{\left\{\substack{\mathbf{1}_{i,\tau}, m_{i,\tau} \\ q_{i,\tau}, l_{i,p,\tau}}\right\}_{\tau\geq t}} \mathbb{E}_t \sum_{\tau=t}^{\infty} (\beta\nu)^{\tau-t} \left(\prod_{h=t}^{\tau} \mathbf{1}_{i,h}\right) \frac{\varphi_{\Upsilon,\tau}}{\varphi_{\Upsilon,t}} \left(m_{i,\tau}\Upsilon(q_{i,\tau}) - \frac{\varphi_{L,\tau}}{\varphi_{\Upsilon,\tau}}(l_{i,p,\tau}+\chi) - \frac{\varphi_{m,\tau}}{\varphi_{\Upsilon,\tau}} m_{i,\tau}\right) di}_{\text{Social Value of Firm } i \equiv G_{i,t}} \Bigg\} \tag{B.59}$$

$$s.t. \quad m_{i,\tau} q_{i,\tau} C_\tau = z_{i,\tau} l_{i,p,\tau}^{\alpha}, \quad \forall i \in \tilde{\Lambda}_t, \forall \tau \geq t, \tag{B.60}$$

where $G_{i,t}$ is the culmination of all of the terms in the social planner's problem in which a firm i shows up and thus captures the value of the firm for the planner.

To compare $G_{i,t}$ with the equilibrium value of the firm, the final step is to relate the Lagrange multipliers to aggregate objects that are similar to the equilibrium wage W_t and demand index D_t. Let us derive the first-order conditions for the aggregate consumption C_t and aggregate labor supply L_t. For C_t, we get

$$C_t^{*-\gamma} = \varphi_{c,i,t} \int_{i\in N_t} m_{i,t}^* q_{i,t}^* di = \varphi_{\Upsilon,t} \int_{i\in N_t} m_{i,t}^* q_{i,t}^* \Upsilon'(q_{i,t})^* di C_t^{*-1}, \tag{B.61}$$

where stars denotes that these equations hold under the optimal choice of these variables and the right-hand-side follows from substituting the FOC for $c_{i,j,t}$ from above. Moreover, note that the integral on the right-hand-side is exactly how we defined the demand index in the equilibrium. So let us define it similarly for the social planner as

$$D_t^* = \left[\int_{i\in N_t} m_{i,t}^* q_{i,t}^* \Upsilon'(q_{i,t}^*)\right]^{-1} \Rightarrow \varphi_{\Upsilon,t} = C_t^{*\,1-\gamma} D_t^*. \tag{B.62}$$

Furthermore, the FOC for L_t gives: $\xi L_t^{*\psi} = \varphi_{L,t}$. If we define the notion of the "wage" for the planner as the marginal rate of substitution between leisure and consumption, $W_t^* \equiv \xi L_t^{*\psi} C_t^{*\gamma}$, we can write

$$\frac{\varphi_{L,t}}{\varphi_{\Upsilon,t}} = \frac{W_t^*}{C_t^* D_t^*}. \tag{B.63}$$

Now, to derive an expression for $\varphi_{m,t}/\varphi_{\Upsilon,t}$, we use Equation (B.38). Multiplying both sides of that equation by $m_{i,t}$ and integrating over the set N_t, we arrive at

$$\frac{\varphi_{m,t}}{\varphi_{\Upsilon,t}} = \frac{\overbrace{\int_{i\in N_t} m_{i,t}\Upsilon(q_{i,t})di}^{=1} - \overbrace{\int_{i\in N_t} m_{i,t}q_{i,t}\Upsilon'(q_{i,t})di}^{=D_t^{*-1}}}{\underbrace{\int_{i\in N_t} m_{i,t}di}_{=1}} = 1 - D_t^{*-1}. \tag{B.64}$$

The last step in comparing $G_{i,t}$ with the equilibrium value of firms defined in Equation (3.13) is to render their units consistent. The equilibrium value of a firm entering at time t in Equation (3.13) is defined in the units of aggregate consumption at time t, where $G_{i,t}$ is in time t utils; to convert time t utils to time t aggregate consumption units we need to multiply $G_{i,t}$ by $C_t^* D_t^*$. Let $v_{i,t}^*$ denote this converted value; then it is given by

$$\begin{aligned} v_{i,t}^* &\equiv C_t^* D_t^* G_{i,t} \\ &= \max_{\left\{\substack{\mathbf{1}_{i,\tau}, m_{i,\tau} \\ q_{i,\tau}, l_{i,p,\tau}}\right\}_{\tau\geq t}} \mathbb{E}_t \sum_{\tau=t}^{\infty} (\beta\nu)^{\tau-t} \left(\prod_{h=t}^{\tau} \mathbf{1}_{i,h}\right) \left(\frac{C_\tau^*}{C_t^*}\right)^{-\gamma} \left(D_\tau^* \frac{\Upsilon(q_{i,\tau})}{q_{i,\tau}} y_{i,\tau} - W_t^*(l_{i,p,\tau} + \chi) - C_\tau^*(D_\tau^* - 1)m_{i,\tau}\right) \end{aligned} \tag{B.65}$$

$$s.t. \quad y_{i,\tau} = m_{i,\tau} q_{i,\tau} C_\tau^* = z_{i,\tau} l_{i,p,\tau}^{\alpha}. \tag{B.66}$$

■

Supplemental Materials

Concentration, Market Power, and Misallocation: The Role of Endogenous Customer Acquisition

by Hassan Afrouzi, Andres Drenik, and Ryan Kim

SM.1 Data Description

This section further describes the process to clean the Nielsen 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 Nielsen and De Loecker, Eeckhout, and Unger 2020, Traina 2019 for Compustat).

SM.1.1. Nielsen and GS1: Variables and Data Cleaning

We construct the following variables from the Nielsen Homescan Panel, Promodata, and GS1 company data:

- *UPC*: scanner-level product identifier available in Nielsen 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 Nielsen 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 Nielsen 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 (2019), 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 Nielsen 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.

SM.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.
- *Company's initial public offering date* (mnemonic `IPODATE`).
- *Age*: given that the initial public offering date was missing for a large portion of our dataset, we calculated age as the fiscal year of a given observation minus the first year we observe a firm in the dataset. According to Compustat, a firm enters the dataset after it starts providing consistent accessible annual reports trading on a U.S. exchange market, i.e., after its IPO. Following Haltiwanger, Jarmin, and Miranda (2013), we exclude the first 15 years of the dataset for all analyses using age and we group together firms older than 16 years, because we do not

know for certain a firm's IPO date for firms that were in the Compustat data since the first year.

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.

SM.1.3. Coverage of the Nielsen-Compustat Sample

Approximately 300 firms identified in Compustat can be matched with the Nielsen 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 Nielsen Homescan Panel data, as shown in Table SM.1.1.

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

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

Note: Sales is the projection-factor-weighted sales in Nielsen 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.

SM.1.4. Summary Statistics

Table SM.1.2: Summary Statistics

Variable	**N**	**Mean**	**SD**	**p10**	**p50**	**p90**
	Panel A: Nielsen-GS1, Firm-Product Group-Year Variables					
S_{igt} (in thousands $)	557,820	6,708.16	64,961.12	3.97	126.08	5,170.55
$p_{igt}q_{igt}$ (in $)	557,820	10.03	20.14	1.96	5.88	19.50
m_{igt} (in thousands)	557,820	500.79	2,789.82	0.82	19.83	639.87
m_{igt}^{New} (in thousands)	557,820	250.34	988.95	0.48	16.03	424.89
m_{igt}^{Old} (in thousands)	557,820	250.45	1,963.09	0.00	1.60	194.18
	Panel B: Nielsen-Compustat, Firm-Year Variables					
SGA$_{it}$ (in millions $)	2,101	2,009.17	4,993.82	7.63	299.09	4,882.24
COGS$_{it}$ (in millions $)	2,299	7,147.11	18,558.75	17.17	1,123.66	17,251.26
OPEX$_{it}$ (in millions $)	2,299	9,147.10	21,337.48	25.72	1,620.66	24,212.49
SGA-to-OPEX$_{it}$	2,101	0.30	0.20	0.08	0.27	0.58
Sales-to-COGS$_{it}$	2,299	1.79	1.06	1.14	1.49	2.61

Notes: The Nielsen-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 Nielsen variables are projection-factor adjusted. The Nielsen-Compustat matched data in Panel B have 332 firms in the period 2004-2016. All cost-side variables are in millions of US dollars and are deflated by the GDP deflator.

SM.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.

SM.2.1. Taste for Quantity

Consider the following extension of the Kimball aggregator in Equation (3.1):

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

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

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

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:

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

where the sign and magnitude of β_0 determine the type and strength of sorting, respectively. To test this, we aggregate the Nielsen 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_{it}	0.026***	0.025***	0.023**	0.029***	0.029***
	(0.009)	(0.009)	(0.009)	(0.002)	(0.002)
ln p_{it}	0.162***	0.180***	0.132***	0.680***	0.688***
	(0.028)	(0.037)	(0.027)	(0.017)	(0.017)
ln $p^2{}_{it}$		0.021*	0.025	-0.003	-0.003
		(0.012)	(0.015)	(0.005)	(0.005)
ln $p^3{}_{it}$			0.005	-0.001	-0.001
			(0.003)	(0.001)	(0.001)
Observations	9097076	9097076	9097076	8452707	8452707
R^2	0.086	0.096	0.102	0.851	0.852
UPC FE				✓	✓
Year FE				✓	
Product 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.

SM.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 (3.1):

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

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

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

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 $\frac{p_{i,t}}{\mathbb{E}[\zeta_{i,j,t}]D_t}$:

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

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

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

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.

SM.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.

SM.3.1. Solution Method

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 (3.13)):

$$v_t(m_{-1},z) \equiv \max_{l_s,l_p,p}\left\{py - W_t l_p - W_t(l_s+\chi) + \beta\nu\frac{U_{c,t+1}}{U_{c,t}}\mathbb{E}\left[V_{t+1}(m,z')|z\right]\right\}$$

$$\begin{aligned} s.t.\quad & q=\left[1-\eta\ln\left(\frac{p}{D_t(1-\sigma^{-1})}\right)\right]^{\frac{\sigma}{\eta}} \\ & y = mqC_t = zl_p^{\alpha} \\ & m=(1-\delta)m_{-1}+\frac{l_s^{\phi}}{P_{m,t}}, \end{aligned}$$

where $V_t(m_{-1},z) \equiv \max\{0, v_t(m_{-1},z)\}$ denotes the endogenous exit choice.

Stationary Distribution Let $\mathcal{N}_t : M\times Z \to [0,1]$ denote the cdf of incumbent firms measured after the realization of idiosyncratic productivity shocks, but before exit decisions are made. The law of motion of the distribution of firms is given by

$$\begin{aligned} \mathcal{N}_t(m,z') = & \int_{M\times Z} F(z'|z)\mathbf{1}_{\{m_t^*(m_{-1},z)\le m\}}\nu\mathbf{1}_{\{v_t(m_{-1},z)\ge 0\}}d\mathcal{N}_{t-1}(m_{-1},z) \\ & +\lambda\int_{M\times Z} F(z'|z)\mathbf{1}_{\{m_t^*(0,z)\le m\}}\mathbf{1}_{\{v_t(0,z)\ge 0\}}dF^e(z), \end{aligned}$$

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.

Solution Algorithm In steady state consumption is constant, so that $U_{c,t+1}/U_{c,t}=1$. 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:

$$\mathbb{E}\left[V_t(m, z')|z\right] = \sum_{i=1}^{50} \omega_i V_t\left(m, \exp(\rho \ln z + \varepsilon_i)\right). \tag{SM.3.1}$$

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

$$Q_M(s, m'(s)) = \left[\mathbf{1}_{m'(s)\in[m_{j-1}, m_j]} \frac{m'(s) - m_j}{m_j - m_{j-1}} + \mathbf{1}_{m'(s)\in[m_j, m_{j+1}]} \frac{m_{j+1} - y'(s)}{m_{j+1} - m_j}\right] \tag{SM.3.2}$$

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 (SM.3.2) 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

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

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 $X = (\nu \mathbf{1}_{\upsilon(\boldsymbol{s})\geq 0}\mathcal{N} + \lambda \mathbf{1}_{\upsilon(\boldsymbol{s})\geq 0} F^e)' x(\boldsymbol{s})$. 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**.

SM.3.2. 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.

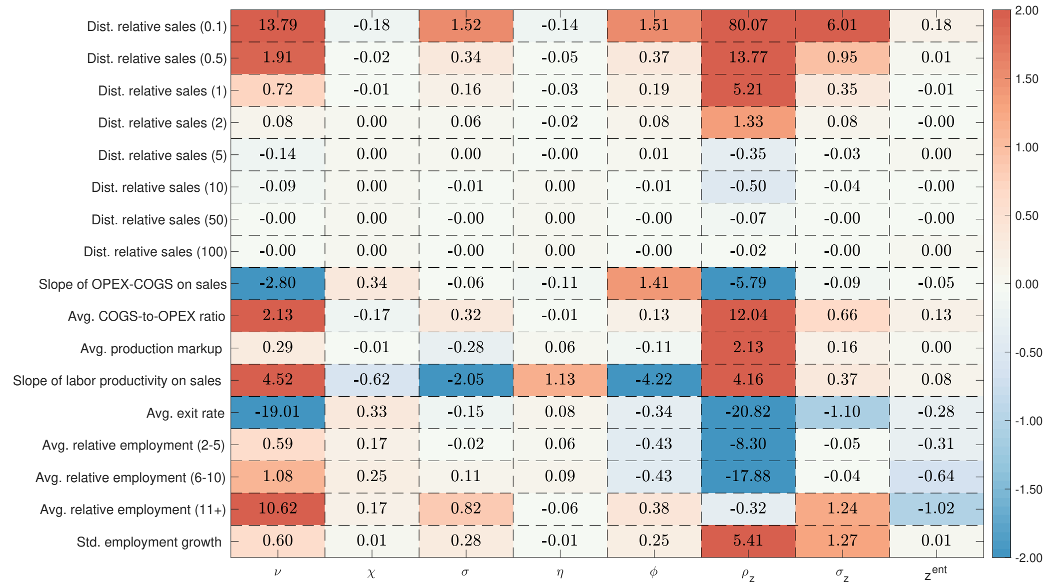
SM.4 Additional Model Analysis

SM.4.1. Identification of Model Parameters

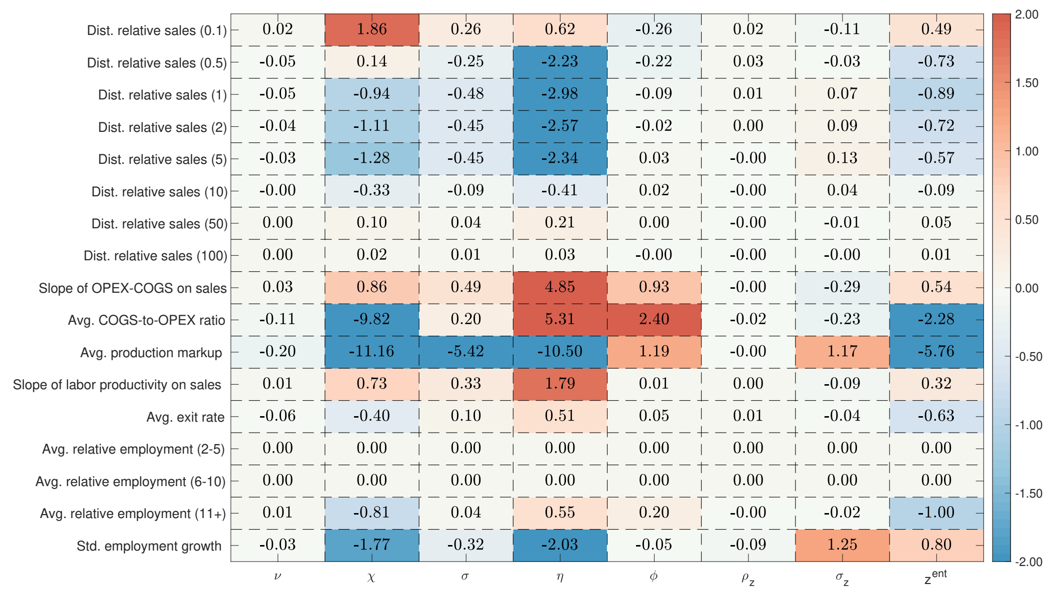
In this section, we formally guide the discussion of the identification of model parameters. Panel A of Figure SM.4.1 shows the *local* elasticity of simulated moments (rows) with respect to parameters (columns), evaluated at the calibrated parameters. In general, the intuition behind the choice of targets 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.

SM.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 SM.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.

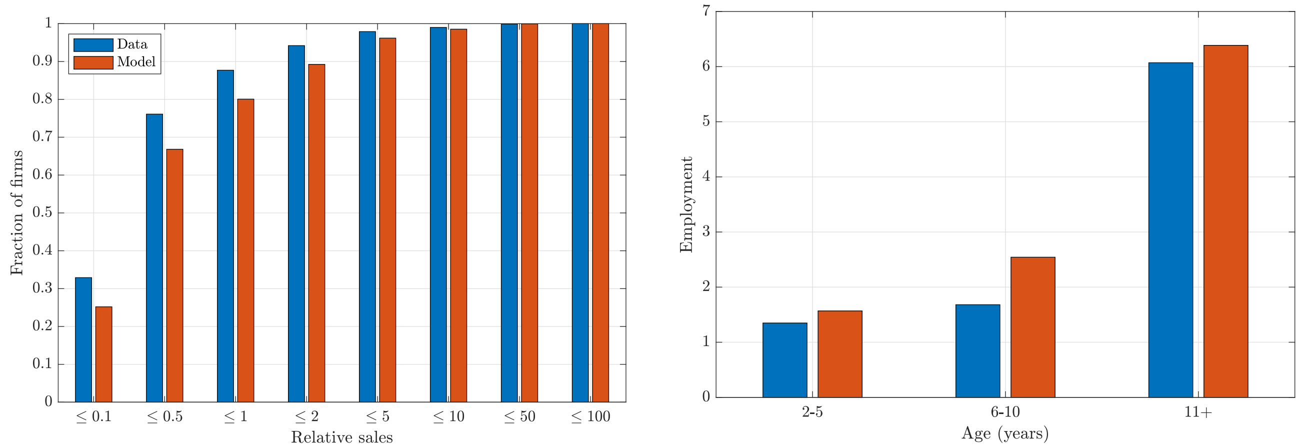
Table SM.4.1: Targeted Moments

Moment	**Data**	**Model**
Slope SGA on sales	0.492	0.474
Avg. COGS-to-OPEX ratio	0.660	0.669
Avg. cost-weighted production markup	1.250	1.275
Slope labor prod. on sales	0.036	0.033
Avg. exit rate	0.073	0.071
SD. employment growth	0.416	0.447

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

Figure SM.4.2: Model Fit: Firm Size



(a) Distribution of Sales

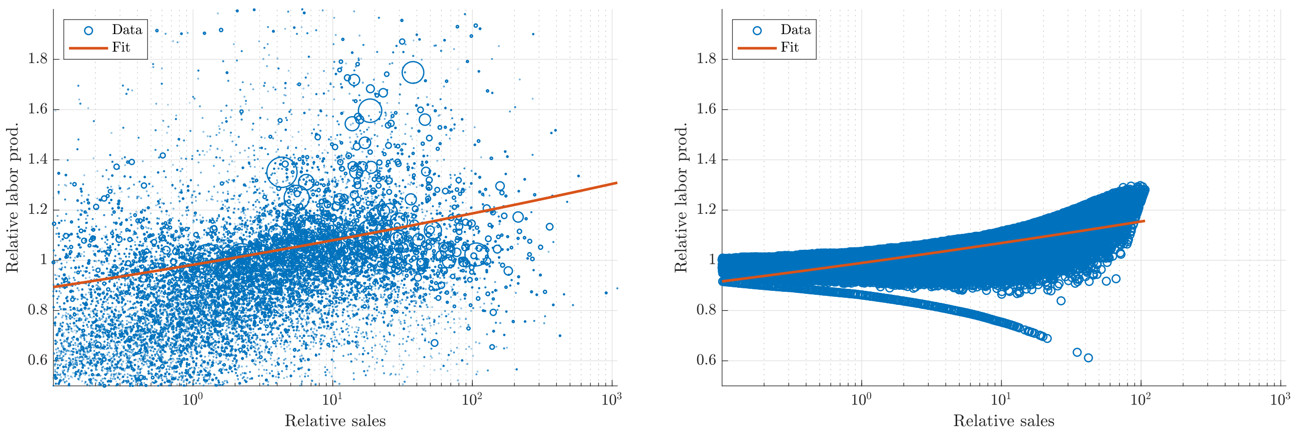
(b) Avg. Size by Age

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.

between relative SGA expenses and relative sales in the data (Compustat) and the model. Although 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.

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

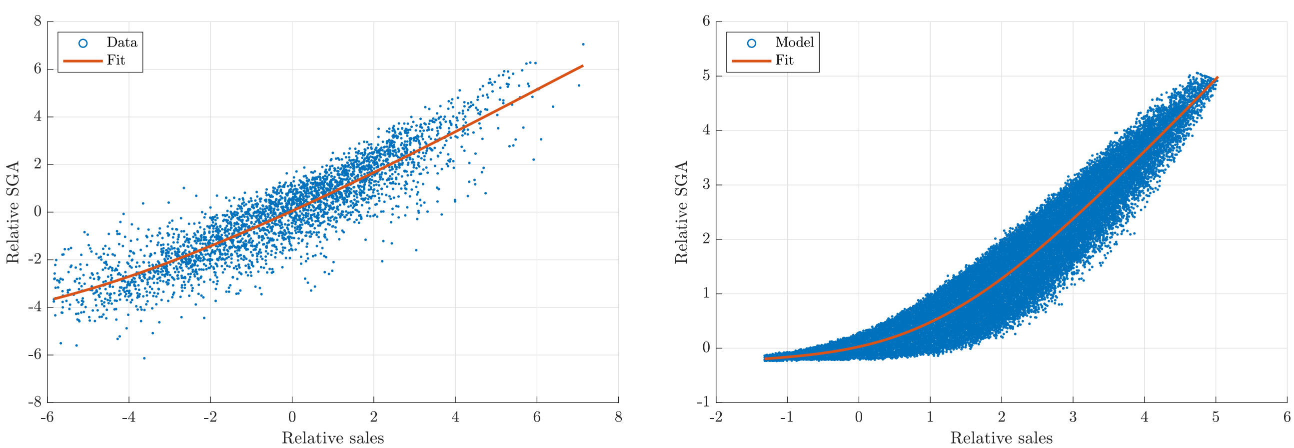
Figure SM.4.3: Model Fit: Labor Productivity and Sales



(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

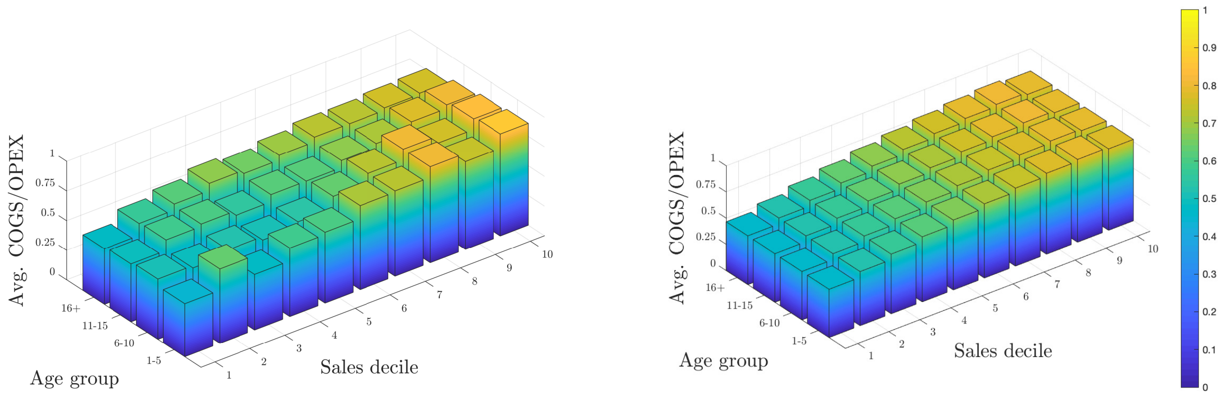


(a) Data (b) Model

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.

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.

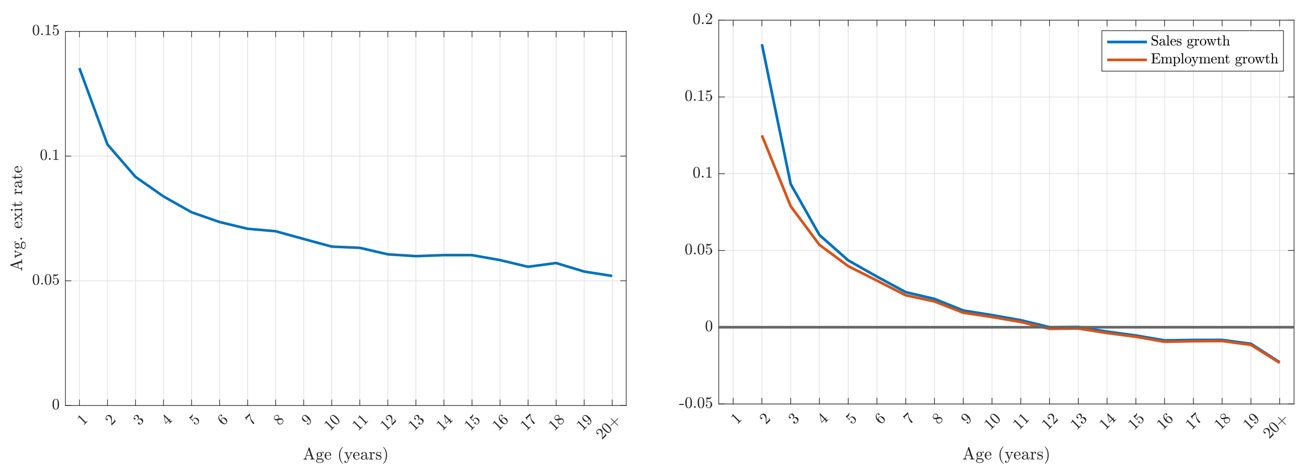
Figure SM.4.5: Steady-state COGS/OPEX by Size and Age



(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.

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.

SM.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.4.7 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.

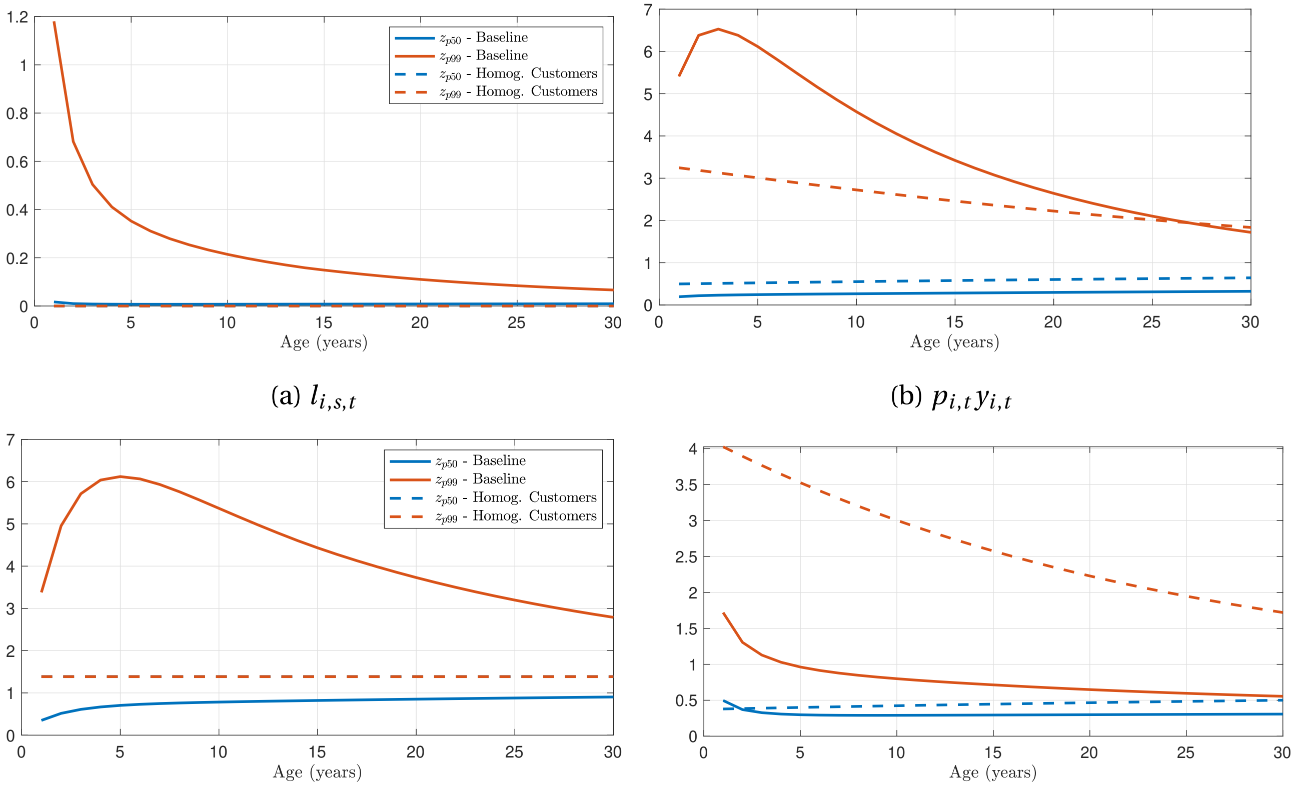
How do firms grow their sales? Panels (c) and (d) of Figure SM.4.7 plot the dynamics of the number of customers and sales per customer, respectively. Both firms build their customer base gradually over time due to decreasing returns to advertising. However, sales per customer are higher when firms are 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.4.2 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.4.2 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.4.8 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 Equa-

Figure SM.4.7: 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.

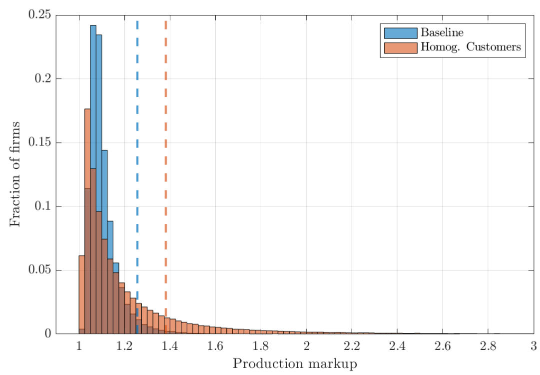
tion (5.6), quantifies each force:

$$\underbrace{\ln(\mathcal{M}_t) - \ln(\mathcal{M}_t^{HC})}_{-9.71\%} \approx \underbrace{\int_{i\in N_t} (\omega_{i,t} - \omega_{i,t}^{HC}) \ln(\mu_{i,t})\, di}_{\Delta \text{ Distribution: } 9.00\%} + \underbrace{\int_{i\in N_t^{HC}} \omega_{i,t}^{HC} \left(\ln(\mu_{i,t}) - \ln(\mu_{i,t}^{HC})\right) di}_{\Delta \text{ Market power: -18.71\%}},$$

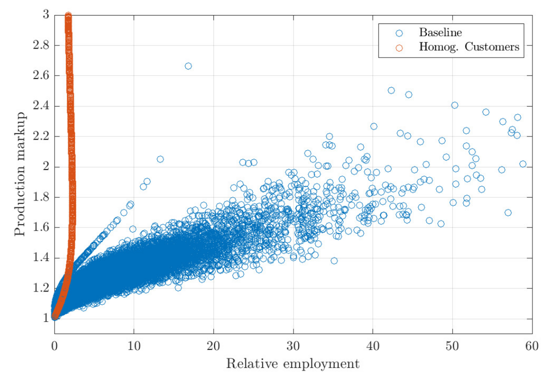
where the superscript HC denotes distributions and allocations in the homogeneous customers model.[40] The first term (denoted "Δ Distribution") captures the contribution of differences in the distribution of relative employment (cost-based Domar weights) across firms while keeping firms' markups fixed 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

[40]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.4.8: 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 ($m_{i,t} = 1/N_t$).

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 9 pp. However, the contribution of lower markups in the baseline model *reduces* the aggregate markup by 18.7 pp, so the aggregate markup is on net smaller by 9.71 pp. To summarize, why does the 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 (5.3). Table SM.4.2 shows that shutting down endogenous customer acquisition in our model reduces aggregate TFP by 28%. 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 65%. 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 higher number of firms in the homogeneous customers model also leads to higher demand for employment. Table SM.4.2 shows that in the homogeneous customers model, total employment is 8 percent larger than in the baseline model (despite the fact that there is no spending on customer acquisition in the former). Part of this difference stems from the larger number of firms, which requires higher overhead costs. Table SM.4.2 also shows that aggregate *production* labor is 6.3% larger in the homogeneous customers model. This is the result of income effects that increase labor supply due to lower aggregate consumption. Finally, the changes in aggregate TFP and production labor lead to an overall 24% decline in aggregate output.

Table SM.4.2: Aggregate Effects of Customer Acquisition

	Baseline Model	**Homog. Customers Model**
TFP		-27.9
Output		-23.9
Employment		7.9
Production		6.3
Number of firms		65.1
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.

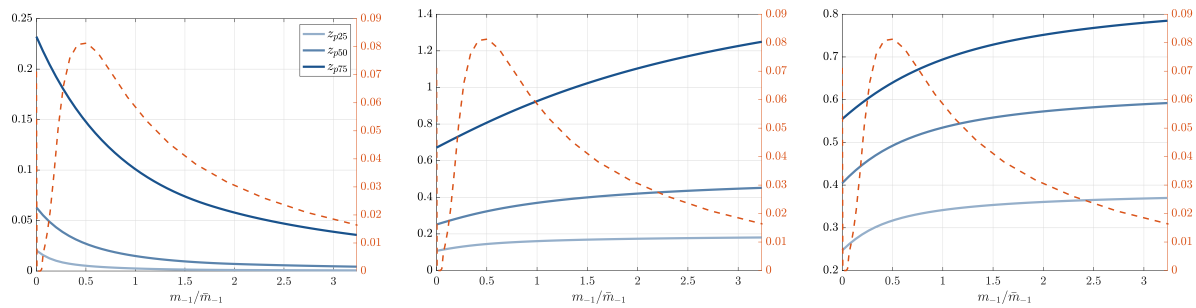
SM.4.4. Additional Analysis of Model in Steady State

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.

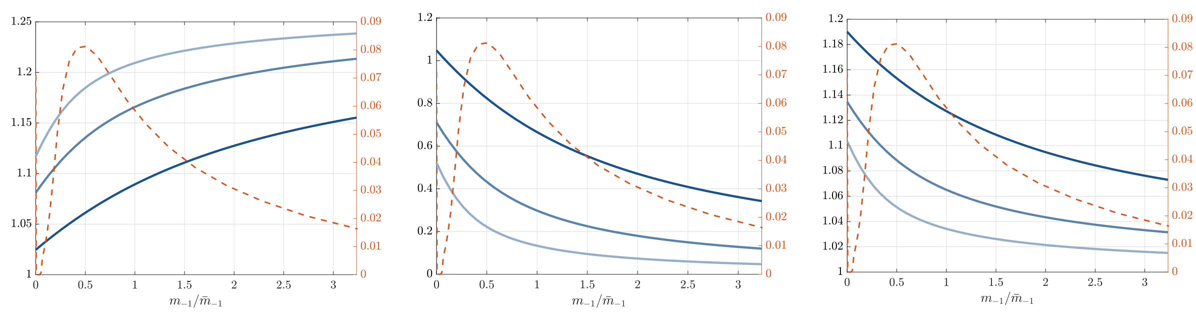
Firms' Optimal Policies Figure SM.4.9 shows firms' steady-state optimal policy functions for three productivity levels (the 25th, 50th, and 75th percentiles of the marginal productivity distribution in steady state). The y-axis on the right plots the marginal distribution of the relative customer base.

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.4.9: Firms' Optimal Policies



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

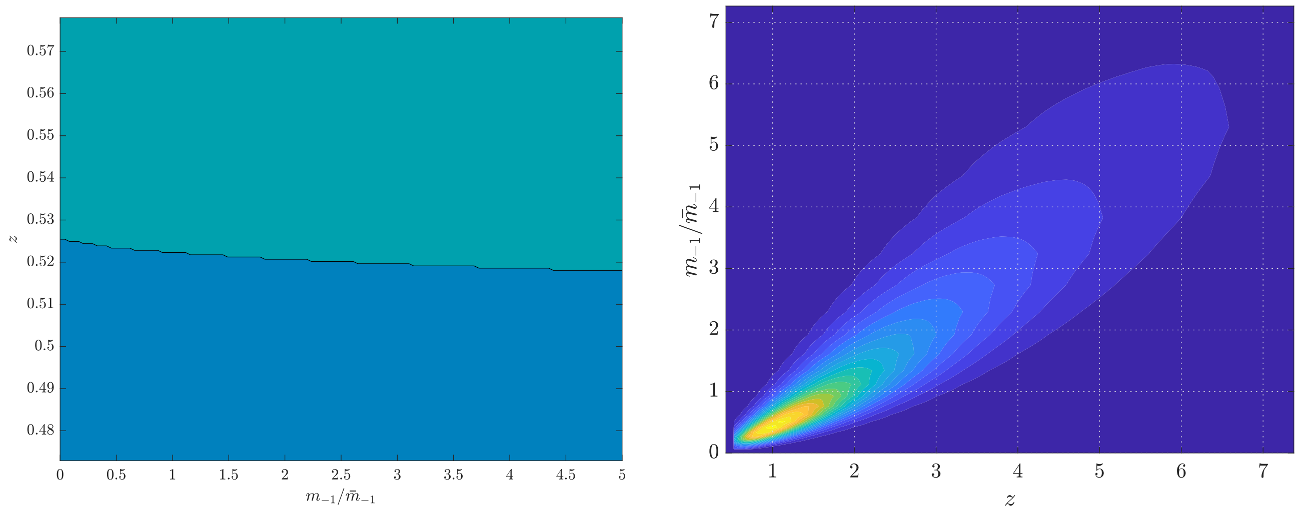


(d) $p_{i,t}$ (e) $q_{i,t}$ (f) $\mu^p_{i,t}$

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.4.10 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.4.10: Optimal Exit and Stationary Distribution



(a) Optimal exit policy

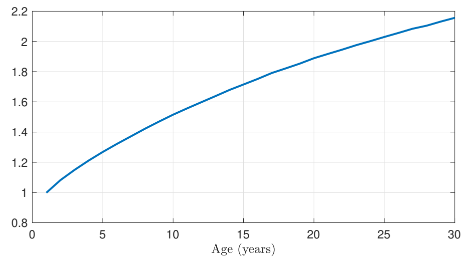
(b) Stationary distribution of (m_{-1}, z)

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.

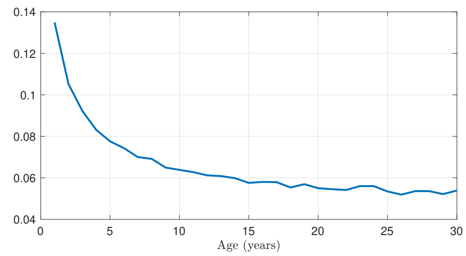
Average Firm Dynamics with Shocks Figure SM.4.11 plots the average firm dynamics, taking selection into account. To construct this figure, we simulate a cohort of firms that starts with a zero customer base and draws productivities from the distribution of entrants. Since firms are subject to productivity 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.

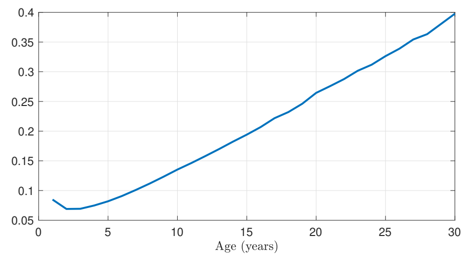
Figure SM.4.11: Average Firm Dynamics



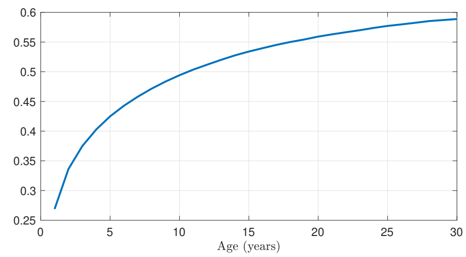
(a) Productivity



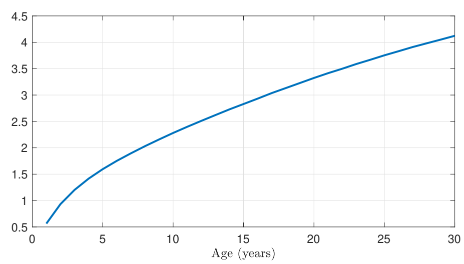
(b) Exit rate



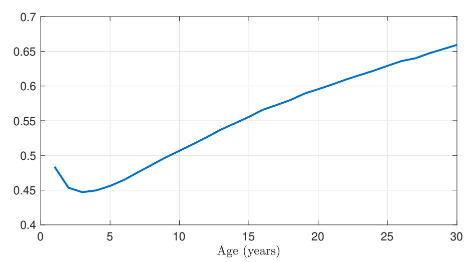
(c) $l_{i,s,t}$



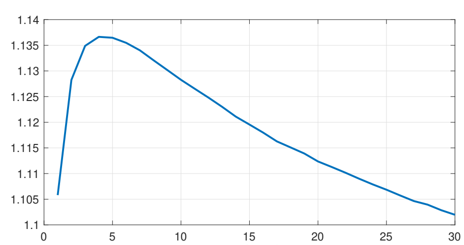
(d) $COGS_{i,t}/OPEX_{i,t}$



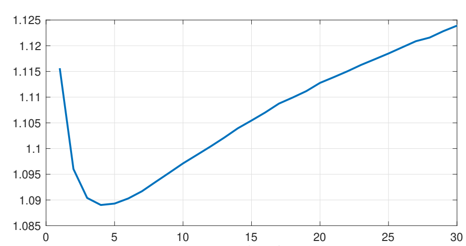
(e) $m_{i,t}$



(f) $q_{i,t}$



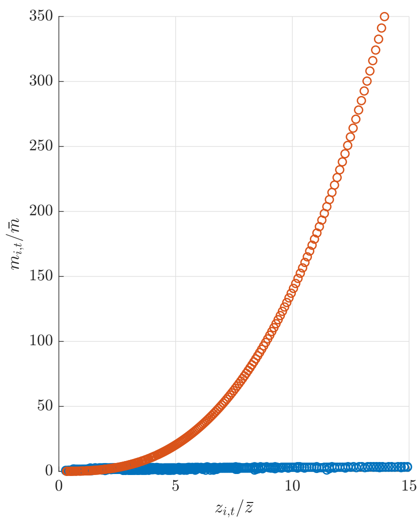
(g) $p_{i,t}$



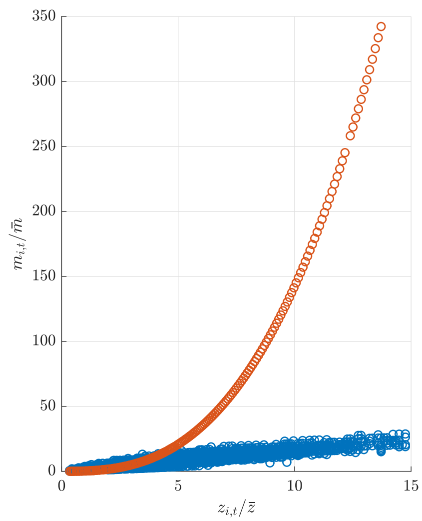
(h) $\mu^p_{i,t}$

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.

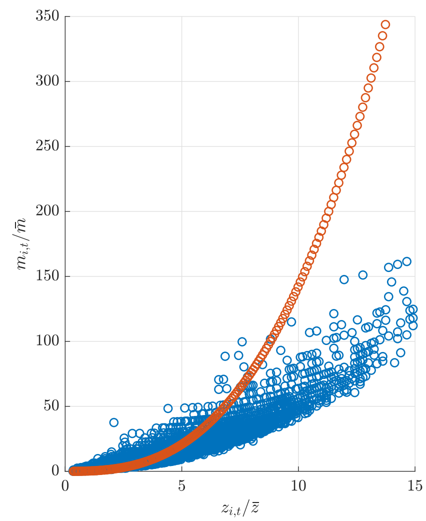
Figure SM.4.12: Allocation of Customers: Equilibrium vs. Efficient Allocation



(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.

SM.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	2.000
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

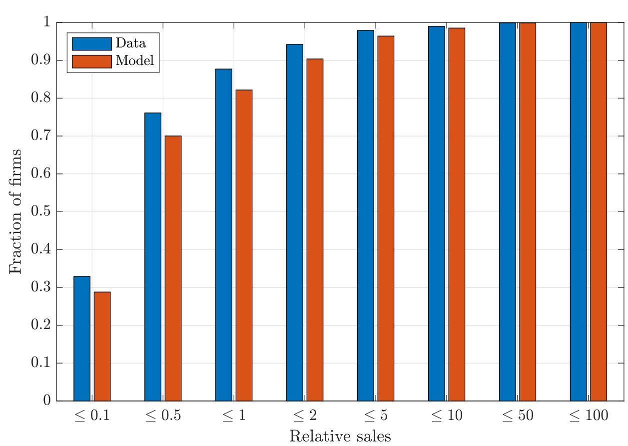
Notes: This table shows the calibration of the homogeneous customers model with an exogenous customer base. Panel A contains parameters externally chosen. Panel B contains parameters internally calibrated to match moments presented in Table SM.5.2 and Figure SM.5.1.

Table SM.5.2: Targeted Moments: Homogeneous Customers Model

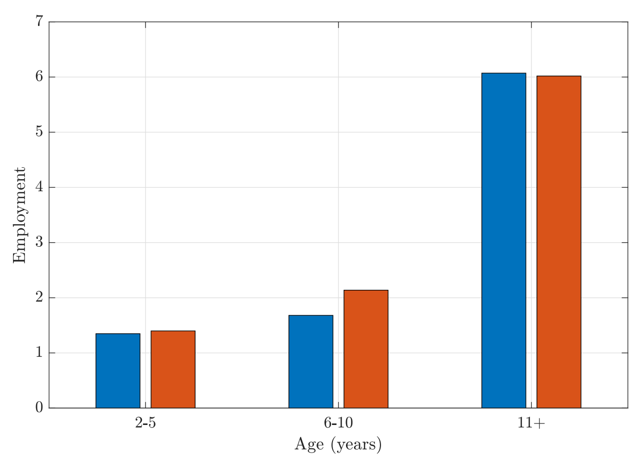
Moment	**Data**	**Model**
Avg. COGS-to-OPEX ratio	0.660	0.667
Avg. cost-weighted production markup	1.250	1.266
Slope labor prod. on sales	0.036	0.035
Avg. exit rate	0.073	0.073
SD. employment growth	0.416	0.436

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.