Online Appendix
Customers and Retail Growth

Liran Einav Peter J. Klenow

Jonathan D. Levin Raviv Murciano-Goroff*

December 6, 2021

This online appendix contains five parts. In Appendix A we discuss the main datasets we use in more detail. Appendix B reproduces many of our baseline 2016–2019 results on a longer dataset from 2007–2019 which covers only larger retailers. In Appendix C we report facts about store growth and exit by age. Appendix D provides patterns for new versus returning customers. And Appendix E provides full model derivations.

*Einav, Klenow, and Levin: Stanford University and NBER; Murciano-Goroff: Boston University.

Appendix A: Data and samples

A.1. The datasets

Visa Transaction Table

We combine several datasets for our analysis. The main dataset is a proprietary dataset by Visa Inc. covering the universe of transactions on the Visa network. This dataset is at the level of the transaction and contains transactions by both credit and debit cards. We observe transactions starting in 2007 and up to (and including) 2017.

The main variables for our analysis are the card number, a merchant identifier, the transaction ZIP code (available for brick and mortar transactions), the transaction amount and the transaction date. One limitation of this dataset is that we are unable to distinguish between different outlets of the same merchant in the same ZIP code. We will address this issue with a different Visa dataset (the GMR table discussed below). To that end, the transaction table also contains an establishment identifier (available since mid 2015) which can be linked to the GMR table.

One important limitation of the transaction data concerns the merchant identifier. While every transaction is assigned a merchant identifier, this identifier does not always allow us to infer the exact merchant. The Visa data distinguishes between two types of merchants, 'named' and 'unnamed' merchants. Roughly speaking, 'named' merchants are large chains for which Visa assigns a unique merchant ID, i.e. there is a one to one mapping between Visa's merchant id and the merchant. 'Unnamed' merchants are typically smaller chains and single establishment merchants. All 'unnamed' merchants within the same industry are assigned the same merchant id. As a consequence it is not possible to identify the actual merchant behind a Visa merchant id for 'unnamed' merchants. We will at times restrict our analysis to named merchants for the parts of the paper for which identifying the exact merchant is important. 58% of dollars in our sample are transacted at named merchants.

There are additional merchant variables in the original dataset, such as a "merchant string". This is the merchant name that would e.g. appear on a credit card statement. While this could in theory allow us to a.) distinguish between smaller (unnamed) merchants that carry the same merchant id and b.) disentangle different stores of the same merchant in the same ZIP code or, the merchant strings are in practice very fuzzy and cannot be easily linked.

Visa GMR Table

Global Merchant Repository (GMR) is an effort by Visa to create a master file of merchant information from data provided by the merchant's acquiring bank and from external data providers. All Visa transactions were linked to a GMR entity via a unique identifier starting in mid 2015. Each GMR-stamped transaction is mapped to a merchant ID and a store ID. For each store, GMR contains the mailing address and the corresponding latitude/longitude pair.

Credit Bureau Data

We have access to an additional dataset that provides cardholder-level demographics which can linked to a sample of Visa credit cards. This dataset is provided by a large credit bureau. About 50% of active credit cards in 2016 and 2017 were linked to an entity in this dataset. An additional 7% were linked to multiple rows in the dataset; we discard these records. For each cardholder matched to the credit bureau data, we observe the cardholder's age and their 9-digit billing ZIP code, as well as their estimated household income, marital status, number of children, and education level.

A.2. The sample

The Transaction Sample

We impose several sample restrictions. We focus on the transaction of Visa credit and debit cards (i.e. we discard non-Visa cards as well as Visa Pre-Paid

Cards) at U.S. merchants. Furthermore, we focus on Credit and Debit-Signature transactions only. This mainly excludes Debit-PIN transaction. Following the Durbin Amendment in 2010 (part of the Dodd-Frank Bill), Visa was not able to restrict how merchants routed Debit-PIN transactions. Therefore, starting in 2012 (when the law went into effect), the data exhibits significant fluctuations in the Debit-PIN transactions of stores. One day or hour a store is transacting with Visa and the next day it looks like they have 0 transactions and the next day they are back again. All the while, their neighbors stay steady on Visa. We hence focus on Credit and Debit-Signature transactions where merchants' network routing is fairly consistent. Transactions worth 91.5% of total dollars on the Visa network satisfy these filter.

We furthermore impose the additional restriction that cards in our sample must have transacted with at least five merchants over their lifetime. This filter was chosen to exclude cards that are only used for one merchant and gift cards (there is a large number of cards that only transact with one merchant for a total of USD 50 or USD 100). Transactions worth 87.6% of total dollars satisfy all filters combined.

Table A2 reports total sales, transaction counts, and card counts in the offline retail sample.

Table A1: Sales, Transactions, and Cards in All Data Sample

Year	Sales (in billions)	Transactions (in billions)	Number of Merchants
2016	1873.07	42.07	2,099,287
2017	2071.20	45.52	2,113,509
2018	2157.09	45.83	2,107,604
2019	2322.43	48.58	2,127,732
Avg.	2105.95	45.50	2,112,033

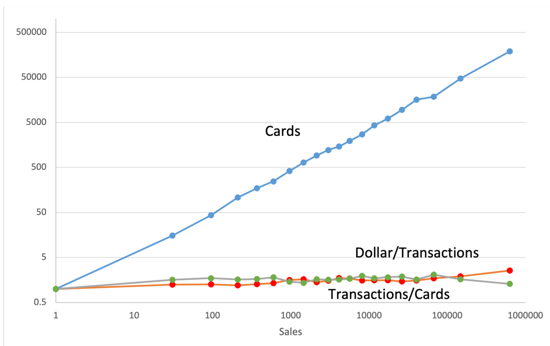
Table A2: Sales, Transactions, and Cards in Offline Retail Sample

Year	Sales (in billions)	Transactions (in billions)	Cards (in millions)	Number of Merchants
2016	1007.04	30.53	426.07	937,152
2017	1078.12	31.84	426.21	917,990
2018	1089.77	31.55	430.83	902,319
2019	1100.70	31.89	430.17	893,602
Avg.	1068.91	31.45	428.32	912,766

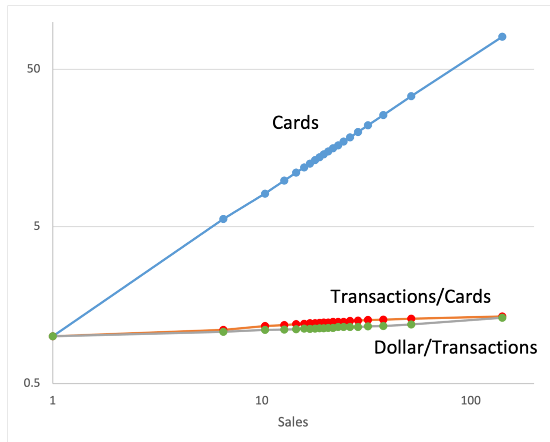
Appendix B: Robustness

Longer panel (2007-2019) on a smaller set of merchants

Figure B1: Decomposing Merchant Sales



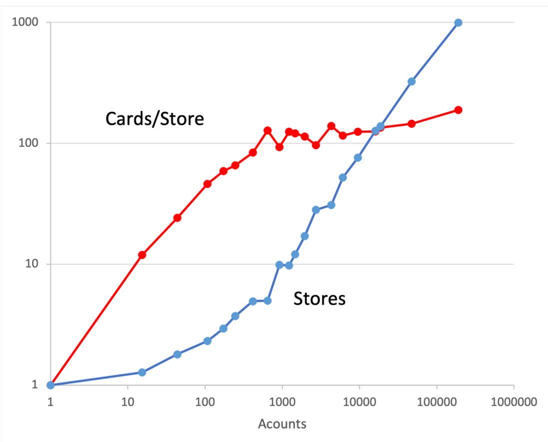
(a) across Merchants



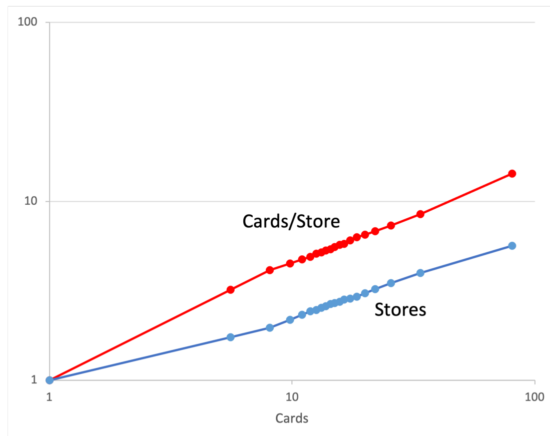
(b) within Merchants over time

Note: Panel (a) is based on a cross section of all merchants in 2019. In panel (a), we group the x-axis into 20 bins, and report averages by bin, normalizing each variable by its average for the first group. Panel (b) repeats the same exercise, but use a panel data of merchants using 2007–2019 data, and de-meaning each variable by its merchant average and its year average, so results are all within merchants and adjusted for overall trends. Both panels are plotted on (base 10) log scale.

Figure B2: Stores vs. sales per store



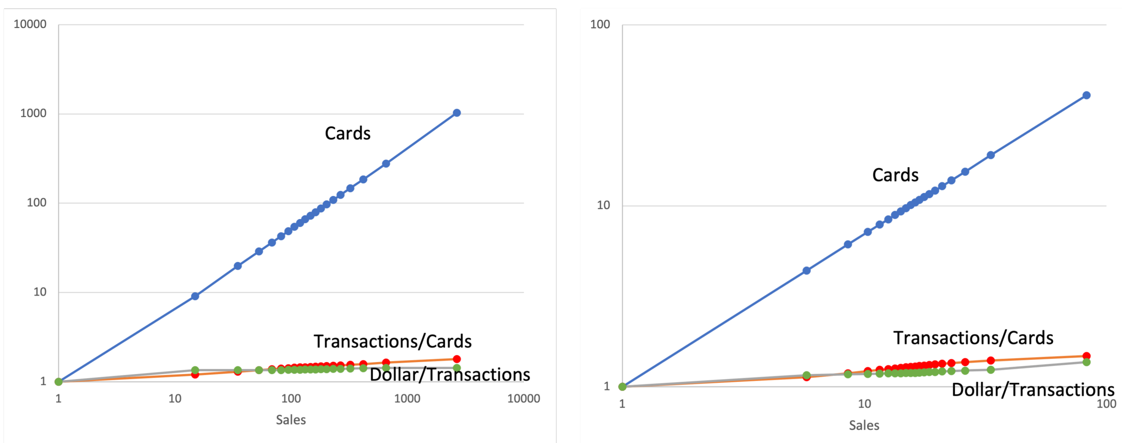
(a) across Merchants



(b) within Merchants over time

Note: Panel (a) is based on a cross section of all merchants in 2019. In panel (a), we group the x-axis into 20 bins, and report averages by bin, normalizing each variable by its average for the first group. Panel (b) repeats the same exercise, but use a panel data of merchants using 2007–2019 data, and de-meaning each variable by its merchant average and its year average, so results are all within merchants and adjusted for overall trends. Both panels are plotted on (base 10) log scale.

Figure B3: Decomposing Store Sales



(a) Across Stores within Merchants

(b) Within Stores over time

Note: Panel (a) uses a cross section of stores in 2019 and de-means each store by its merchant average. We group the x-axis into 20 bins, and report averages by bin, normalizing each variable by its average for the first group. Panel (b) repeats the same exercise, but uses a panel of stores from 2007–2019, de-meaning each variable by its store average and its year average. All panels are plotted on (base 10) log scale.

Table B1: Sales Decomposition Regressions using 2007–2019 over time

	Stores	Cards/Store	Trans/Card	Dollar/Trans
Across Merchants	0.561	0.348	0.056	0.035
$(N = 2,741)$	(0.009)	(0.009)	(0.004)	(0.007)
	[0.601]	[0.366]	[0.079]	[0.009]
Within Merchants over Time	0.337	0.542	0.057	0.063
$(N = 36,940)$	(0.002)	(0.002)	(<0.001)	(0.001)
	[0.966]	[0.948]	[0.954]	[0.952]
Across Stores within Merchants		0.868	0.075	0.057
$(N = 663,110)$		(<0.001)	(<0.001)	(<0.001)
		[0.965]	[0.737]	[0.832]
Within Stores over Time		0.829	0.088	0.083
$(N = 8,186,180)$		(<0.001)	(<0.001)	(<0.001)
		[0.988]	[0.941]	[0.938]

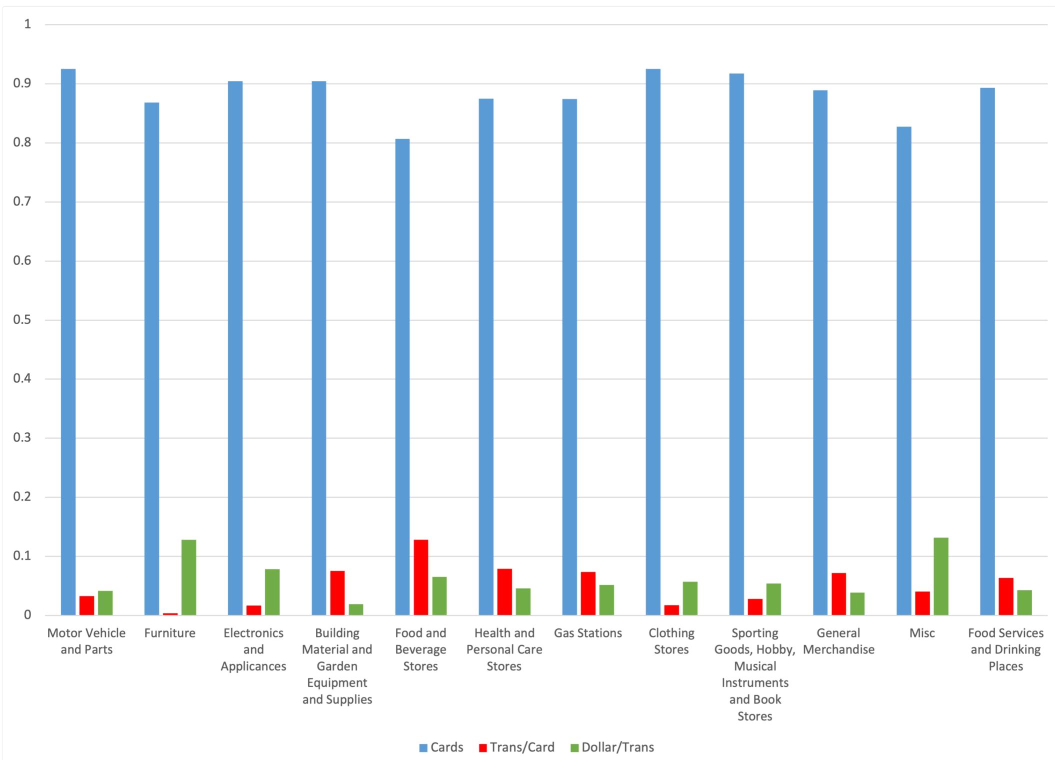
Note: Standard errors in round brackets, R-Squared in square brackets. Store is defined as a Merchant-ZIP combination. Across Merchant Decomposition and Across Store Within Merchant decompositions based on 2019 data. Within Merchants Over Time and Within Stores Over Time based on 2007–2019 data. Within Merchants over Time regressions include merchant and year fixed effects. Across Stores within Merchants regressions include merchant fixed effects. Within store over Time regressions include store and year fixed effects. See Online Appendix B for robustness with respect to the store definition in more recent years.

Table B2: Decomposing Sales in Offline Retail – Cards vs. Households

Cards	Stores	Cards/Store	Trans/Card	Dollar/Trans
A. Across Merchants	0.069	0.687	0.029	0.215
$(N = 873,592)$	[0.095]	[0.662]	[0.019]	[0.139]
B. Within Merchants over Time	0.138	0.687	0.081	0.094
$(N = 3,552,997)$	[0.792]	[0.974]	[0.906]	[0.963]
C. Across Stores within Merchants		0.816	0.079	0.105
$(N = 1,824,338)$		[0.975]	[0.798]	[0.937]
D. Within Stores over Time		0.800	0.116	0.085
$(N = 5,498,665)$		[0.993]	[0.949]	[0.983]
Households	Stores	HH/Store	Trans/HH	Dollar/Trans
A. Across Merchants	0.069	0.685	0.031	0.215
$(N = 873,592)$	[0.095]	[0.662]	[0.020]	[0.139]
B. Within Merchants over Time	0.138	0.677	0.091	0.094
$(N = 3,552,997)$	[0.792]	[0.974]	[0.910]	[0.963]
C. Across Stores within Merchants		0.811	0.085	0.105
$(N = 1,824,338)$		[0.974]	[0.800]	[0.937]
D. Within Stores over Time		0.787	0.128	0.085
$(N = 5,498,665)$		[0.993]	[0.952]	[0.983]

Note: R-Squared values are reported in square brackets. Sample is restricted to only include transactions for which an Experian HH identifier is available. Across Merchant and Across Store within Merchant decompositions are based on 2019 data. Within Merchants over Time and Within Stores over Time are based on 2016–2019 data. Within Merchants over Time regressions include merchant and year fixed effects. Across Stores within Merchants regressions include merchant fixed effects. Within store over Time regressions include store and year fixed effects.

Figure B4: Decomposing Sales Within Industries, 2007-2019



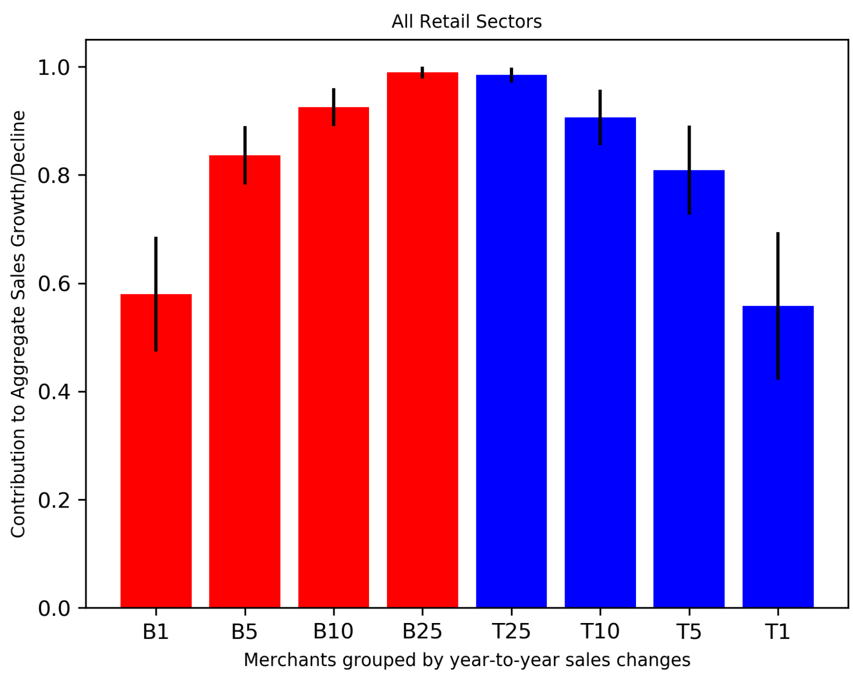
Note: This figure displays the coefficients of the "Within Merchant over time" decomposition by industry. The regressions are run with a merchant and a year fixed effect.

Table B3: Decomposing Store Growth for Entrants vs. Incumbents

	Cards	Trans/Card	Dollar/Trans
Years 1-2	0.772	0.081	0.147
$(N = 2,958,608)$	[0.992]	[0.954]	[0.947]
Years 3-5	0.810	0.085	0.105
$(N = 2,146,165)$	[0.994]	[0.977]	[0.973]
Years 6+	0.846	0.092	0.063
$(N = 3,081,407)$	[0.994]	[0.973]	[0.981]

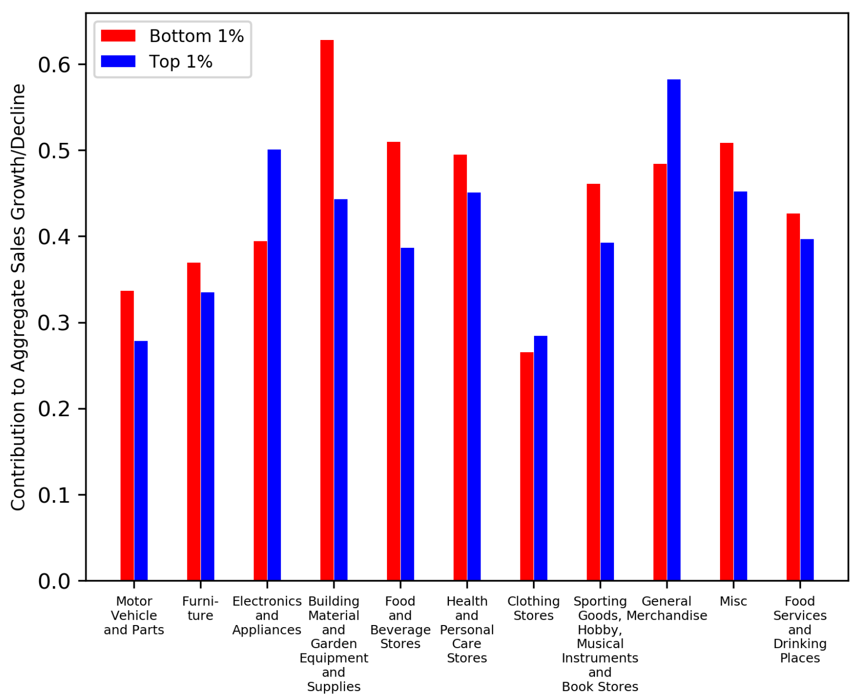
Note: Standard errors are all < 0.001. R^2 is in square brackets. Store is defined as a Merchant-ZIP combination. The observation is at a store-year level. The decomposition is based on 2007–2019 data and run with store and year fixed effects.

Figure B5: Contribution to Aggregate Sales Changes



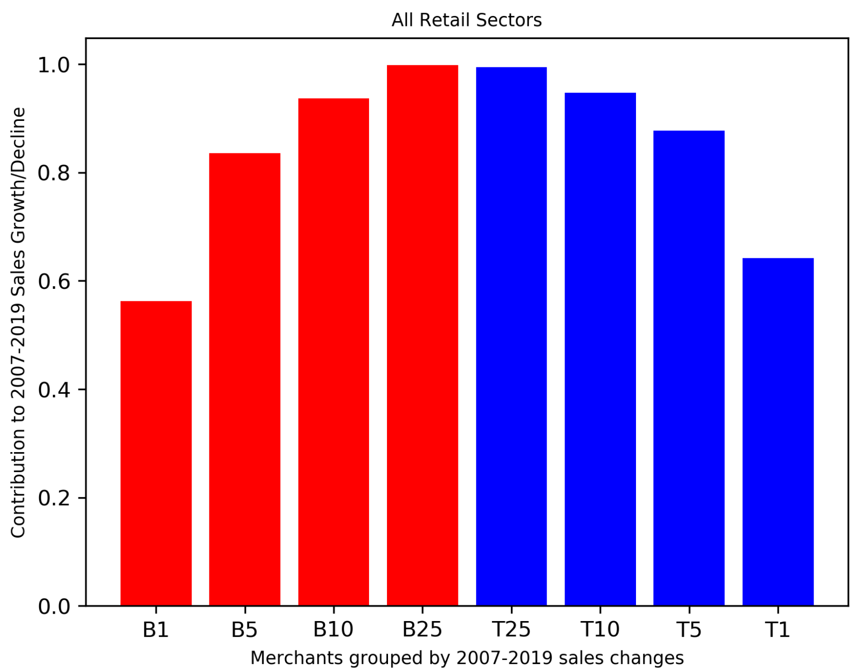
The figure reports the average contribution of each merchant group as defined in the text to aggregate sales change over year with the error bar extending one standard deviation up and down. An observation is a merchant-year and the figure uses a panel of merchants from 2007 to 2019. Each bar corresponds to a merchant group. TX refers to top X% merchants and BX refers to the bottom X% of merchants according to their absolute sales changes.

Figure B6: Contribution to Aggregate Sales Changes By NAICs



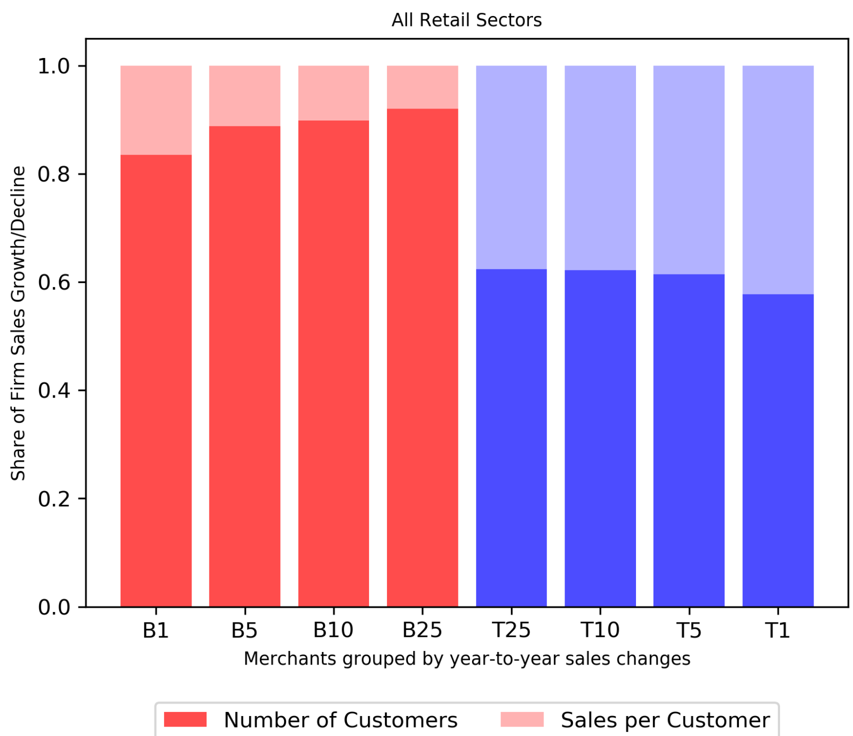
The figure reports the average contribution of top and bottom 1% merchants to within-NAICs aggregate sales change over year for each retail NAICs. An observation is a merchant-year and the figure uses a panel of merchants from 2007 to 2019. The calculation of each merchant group's contribution is described in the text.

Figure B7: Persistence of Firm Contributions



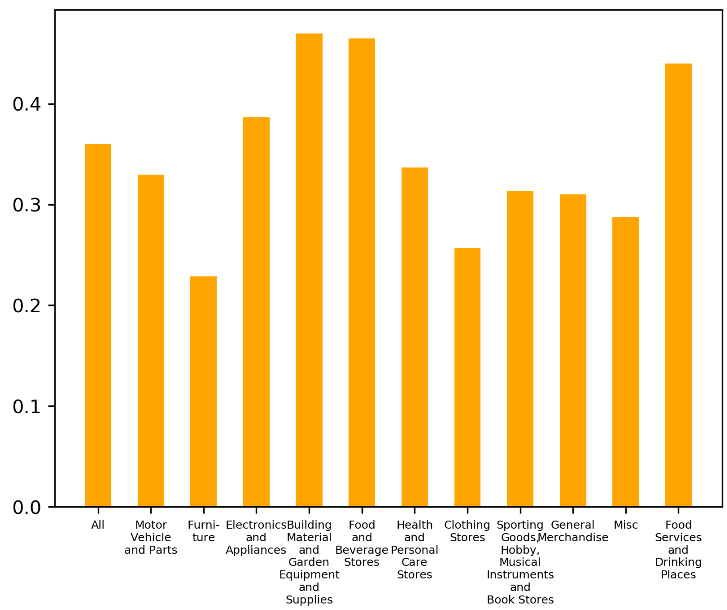
The figure reports the contribution of each firm group as defined in the text to aggregate sales change between 2007 and 2019. Each bar corresponds to a firm group. TX refers to top X% firms and BX refers to bottom X% firms by the absolute sales changes between 2007 and 2019.

Figure B8: Customers vs. sales/customer and firm sales changes



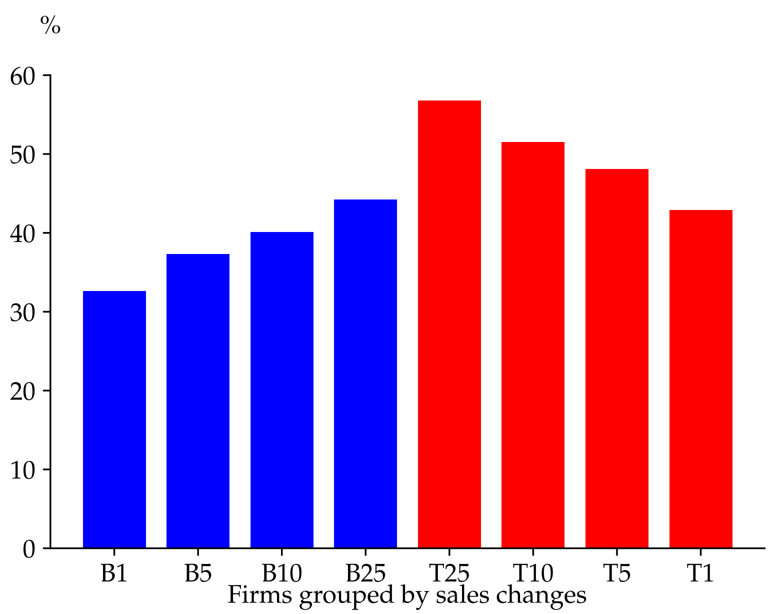
The figure reports the average share of sales changes in each firm group that can be attributed to changes in number of customers and changes in sales per customer respectively from 2007 to 2019. By construction, the two shares sum to 1. Each bar corresponds to a firm group. TX refers to top X% firms and BX refers to bottom X% firms by firms' absolute sales changes.

Figure B9: Firm Sales Growth: Customer Acquisition vs. Retention



The figure reports the coefficient in the regression of annual log change of spending per returning customer on annual log change of total sales. An observation is a store-year. The regression uses 2007-2019 data and includes a year fixed effect.

Figure B10: Firm Sales Share by Growth Contribution



The figure reports the sales share of each firm group defined by their aggregate sales change between 2016 and 2019. Each bar corresponds to a firm group. TX refers to top X% firms and BX refers to bottom X% firms by the absolute sales changes between 2016 and 2019.

Appendix C: Facts about store dynamics

Table C1: Store Growth Regressions

	Revenue Growth
Years 1-2	0.051
Years 2-3	0.068
Years 3-4	0.063
Years 4-5	0.050
Years (5-6)+	0.028

Note: Estimated coefficients are based on regressions of store growth (arc growth rates) on calendar year and store age fixed effects. The displayed coefficients are the coefficients on the store age fixed effects (adding back in the mean growth based on year fixed effects). Store is defined as a Merchant-ZIP combination and the regression observation is a store-year. Regressions are based on 2007-2019 data. To account for the fact that the first calendar year of an entrant might be partial, the growth rate from the first to the second year is based on only using the same calendar months of the second year that are available in the first year. Furthermore, the first month is dropped to avoid issues related to partial months. In particular, if a store entered in March 2010, then the arc growth rate for the second calendar year would be calculated using April-December sales in 2010 and 2011. The regressions are run using dollar weights (in particular the denominator of the arc-growth rate).

Table C2: Store Growth Standard Deviation Regressions

	Standard Deviation of Revenue Growth
Years 1-2	0.36
Years 2-3	0.33
Years 3-4	0.29
Years 4-5	0.30
Years (5-6)+	0.24

Note: Estimated coefficients are based on regressions of store growth standard deviation (arc growth rates) on calendar year and store age fixed effects. The displayed coefficients are the coefficients on the store age fixed effects (adding back in the mean growth based on year fixed effects). Store is defined as a Merchant-ZIP combination. Data from 2007-2019 is used in these calculations. The store-year data is aggregated to a age group- year level by calculating the standard deviation of store arc growth using revenue weights. The regressions are then run on the store-age year aggregates.

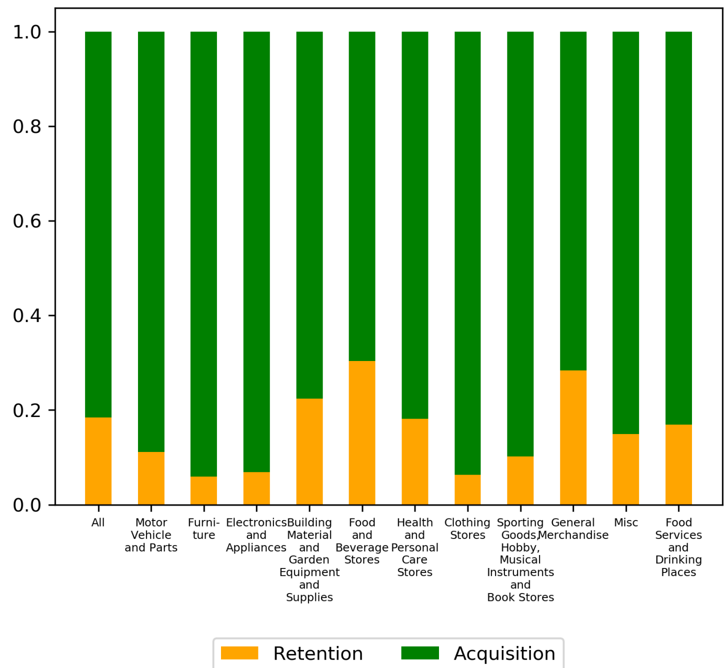
Table C3: Store Exit Rates

	Exit Rate
Years 1	0.157
Years 2	0.034
Years 3	0.015
Years 4	0.027
Years 5	0.017
Years 6+	0.005

Note: Data is at a store-month level of aggregation. Coefficients are obtained from regressing a store exit dummy on a store age fixed effect and a year fixed effect. Regressions are weighted by average monthly sales in calendar year. The coefficients on the store age fixed effects are scaled up by 12 to map monthly exit rates to annual exit rates.

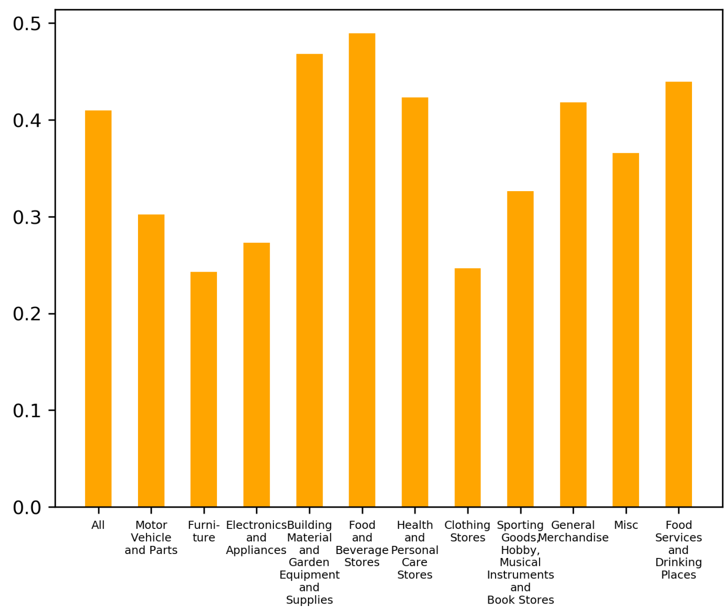
Appendix D: New vs. returning customers for stores

Figure D1: Store Sales Growth: Customer Acquisition vs Retention



The figure decomposes store sales growth into sales growth from new customers vs returning customers and reports the contribution of each component. The orange bars report the coefficient in the regression of $\frac{\Delta sales^{R}_{i,t,t-1}}{sales_{i,t-1}}$ on $\frac{\Delta sales_{i,t,t-1}}{sales_{i,t-1}}$. The green bars report the coefficient in the regression of $\frac{\Delta sales^{A}_{i,t,t-1}}{sales_{i,t-1}}$ on $\frac{\Delta sales_{i,t,t-1}}{sales_{i,t-1}}$. In both regressions, an observation is a store-year. Each bar corresponds to a retail NAICs.

Figure D2: Spending per returning customer on store sales growth



The figure reports the coefficient in the regression of annual log change of spending per returning customer on annual log change of total sales. An observation is a store-year level. The regression uses 2016-2019 data and includes a year fixed effect.

Appendix E. Model of growth with customers

Customers

Consider a unit mass of customers with identical preferences:

$$U = \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-1/\sigma}}{1-1/\sigma}.$$

Their composite consumption C is a CES aggregate of varieties:

$$C_t = \left(\int_0^1 n_{it} \left(q_{it} c_{it} \right)^{\frac{\theta-1}{\theta}} \mathrm{d}i \right)^{\frac{\theta}{\theta-1}}$$

where $n_{it} \in [0,1]$ is the probability that a customer purchases variety i and q_{it} is the quality of variety i. $\theta > 1$ is the elasticity of substitution between varieties and $0 < \beta < 1$ is the discount factor. Note there is a fixed unit measure of varieties. Finally, we assume that n_{it} is identical across consumers, so it is also the fraction of consumers who buy variety i in period t. Demand (per customer) conditional on access to variety i is given by:

$$c_{it} = \left(\frac{P_t}{p_{it}} \right)^{\theta} q_{it}^{\theta-1} C_t, \quad \forall i \in [0,1],$$

where the ideal consumer price index is:

$$P_t \equiv \left(\int_0^1 n_{it} \left(\frac{p_{it}}{q_{it}} \right)^{1-\theta} \mathrm{d}i \right)^{\frac{1}{1-\theta}}.$$

Total quantity demanded for variety i, summed across customers, is:

$$y_{it} = n_{it} c_{it}.$$

Firms

Each firm uses production labor l_{it} to produce its single variety:

$$y_{it} = l_{it}.$$

It uses marketing labor m_{it} to reach a random fraction n_{it} of customers:

$$n_{it} = \left(\frac{\gamma m_{it}}{\phi M_t^{\delta}}\right)^{\frac{1}{\gamma}} \quad \text{where} \quad M_t \equiv \int_0^1 m_{it} \mathrm{d}i.$$

Here $\gamma > 1$, $\delta > 0$ and $\phi > 0$, and M is aggregate marketing labor across all firms. δ controls whether there is a negative externality with respect to other firms' marketing efforts. Choosing labor as the numéraire, the firm's static profit maximization problem is:

$$\max_{p_{it}, m_{it}} \left(p_{it} - 1\right) y_{it} - m_{it}.$$

Assuming that firms engage in monopolistic competition, they set their price to a constant markup above unit marginal cost:

$$p_{it} = \mu \quad \text{where} \quad \mu \equiv \frac{\theta}{\theta - 1}.$$

Substituting the firm's price in its demand function yields:

$$c_{it} = \left(\frac{q_{it} P_t}{\mu}\right)^{\theta - 1} \cdot \frac{P_t C_t}{\mu}.$$

The firm's static marketing problem becomes:

$$\max_{n_{it}} n_{it} \left(\frac{q_{it} P_t}{\mu}\right)^{\theta - 1} \cdot \frac{P_t C_t}{\theta} - \frac{\phi M_t^{\delta} n_{it}^{\gamma}}{\gamma}.$$

This marketing problem yields the following first order condition:

$$n_{it} = \min\left\{\left(\frac{q_{it}P_t}{\mu}\right)^{\theta-1} \cdot \frac{P_tC_t}{\theta\phi M_t^\delta}, 1\right\}^{\frac{1}{\gamma-1}}.$$

Assuming that ϕ is large enough to ensure that no firm accesses all customers and denoting $\Gamma \equiv \frac{\gamma}{\gamma-1}$, it follows that a firm's flow profits are:

$$\pi_{it} = \left[\left(\frac{q_{it}P_t}{\mu}\right)^{\theta-1} \cdot \frac{P_tC_t}{\theta\phi M_t^\delta}\right]^{\Gamma} \cdot \frac{\phi M_t^\delta}{\Gamma}.$$

Static equilibrium

It is useful to first define an aggregate quality index as:

$$Q_t \equiv \left(\int_0^1 q_{it}^{\Gamma(\theta-1)} \mathbf{d}i\right)^{\frac{1}{\Gamma(\theta-1)}}.$$

The market clearing condition for aggregate consumption is:

$$P_tC_t = \int_0^1 \pi_{it}\mathbf{d}i + \int_0^1 l_{it}\mathbf{d}i + \int_0^1 m_{it}\mathbf{d}i.$$

Substituting in the expressions for π_{it}, l_{it} and m_{it}, we obtain:

$$P_tC_t = \theta\phi M_t^\delta \left(\frac{\mu}{P_tQ_t}\right)^{\gamma(\theta-1)}.$$

Letting S_t denote aggregate research labor, which will be defined below, the market clearing condition for labor is given by:

$$1 - S_t = \int_0^1 l_{it}\mathbf{d}i + \int_0^1 m_{it}\mathbf{d}i.$$

Using the above expression for aggregate nominal consumption and substituting in the expressions for l_{it} and m_{it}, we obtain:

$$1 - S_t = \phi M_t^{\delta}(\theta - 1 + 1/\gamma)\left(\frac{\mu}{P_t Q_t}\right)^{\gamma(\theta-1)}.$$

Using the definition of M_t with the above expression for aggregate nominal consumption, and substituting in m_{it}, we have:

$$M_t = (\phi/\gamma)^{\frac{1}{1-\delta}}\left(\frac{\mu}{P_t Q_t}\right)^{\frac{\gamma(\theta-1)}{1-\delta}}.$$

Substituting this in previous equations, we obtain the following expressions for aggregates as a function of aggregate research and re-entry labor:

$$\begin{aligned}
P_t C_t &= \frac{\gamma\theta(1 - S_t)}{\gamma(\theta - 1) + 1} \\
L_t &= P_t C_t/\mu \\
M_t &= P_t C_t/\gamma\theta \\
C_t &= [\gamma(\theta - 1)]^{\frac{\delta-1}{\gamma(\theta-1)}} \cdot (\gamma/\phi)^{\frac{1}{\gamma(\theta-1)}} \cdot L_t^{1+\frac{1-\delta}{\gamma(\theta-1)}} \cdot Q_t.
\end{aligned}$$

Defining a firm's relative quality as $z_{it} \equiv q_{it}/Q_t$, we can use the above results to solve for the fraction of customers reached and profits:

$$n_{it} = \left(\gamma z^{\Gamma(\theta-1)}/\phi\right)^{\frac{1}{\gamma}} \cdot \left(\frac{L_t}{\gamma(\theta - 1)}\right)^{\frac{1-\delta}{\gamma}} \quad \text{and} \quad \pi_{it} = \frac{L_t z_{it}^{\Gamma(\theta-1)}}{\Gamma(\theta - 1)}.$$

Innovation

A firm with absolute quality q_{it} and relative quality z_{it} that hires research labor s_{it} sees its quality follow a controlled binomial process with probability $x_{it} \in [0, 1]$:

$$q_{it+1} = \begin{cases} q_{it}e^{\Delta} & \text{w/ prob. } x_{it} \\ q_{it} & \text{w/ prob. } 1 - x_{it} \end{cases} \quad \text{and} \quad s_{it} = \lambda \log\left(\frac{1}{1 - x_{it}}\right) z_{it}^{\zeta}.$$

Here Δ, λ and ζ are all strictly positive. Δ is the step size of successful quality innovations, and x_{it} is the probability that a firm succeeds in innovating. λ is a scalar for the level of research labor and ζ quantifies how much more research labor is necessary to innovate from a higher level of relative quality. Note the knowledge spillover in this formulation: the higher the quality of other firms, the lower the cost of successfully innovating ($\zeta > 0$). A continuing firm's value function is given by:

$$v_t(z) = \pi_t(z) + \max_{x\in[0,1]} \left\{ R_t^{-1} \left[x v_{t+1}\left(ze^{\Delta - g_t}\right) + (1-x)\, v_{t+1}\left(ze^{-g_t}\right)\right] - s_t(z,x) \right\}$$

where R is the gross interest rate. The Euler equation produces the usual relationship between the growth rate g and the consumer's discount factor in the absence of aggregate uncertainty:

$$(1+g_t)^{1/\sigma} = \beta R_t.$$

The first-order condition of the firm's dynamic problem implies:

$$x_t(z) = 1 - \frac{\lambda z^{\zeta} R_t}{v_t\left(ze^{\Delta - g}\right) + v_t\left(ze^{-g}\right)}$$

and aggregate research labor is defined as:

$$S_t \equiv \int_0^{\infty} s_t(z)\, \mathrm{d}F_t(z)\,.$$

Numerical solution

To solve the firm's dynamic innovation problem, we first define initial guesses for the value function $v_0(z)$ and the relative quality distribution function $f_0(z)$ on a relative quality grid z_{grid}. Importantly, theses guesses must meet the following conditions:

- $v_0(z)$ is non-decreasing in z.

- $f_0(z)$ satisfies: $\int z^{\Gamma(\theta-1)} f_0(z) \mathrm{d}z = 1$.

Value function iteration

To start the value function iteration process, we must first find the growth rate g_n^* that is consistent with firm decisions. In fact, the firms' innovation decisions are functions of the growth rate, which is in turn a function of those innovation decisions.

With functions $v_n(z)$ and $f_n(z)$, define an initial guess for the growth rate g_n^0 to evaluate the value function for successful $v_n(ze^{\Delta - g_n^0})$ and unsuccessful $v_n(ze^{-g_n^0})$ innovations. Use the first-order condition of the firm's dynamic problem and the Euler equation to compute the innovation's probability of success:

$$x_n(z) = 1 - \frac{\lambda z^{\zeta} R_n^0}{v_n\left(ze^{\Delta - g_n^0}\right) + v_n\left(ze^{-g_n^0}\right)}.$$

Update the relative quality distribution function with the following law of motion:

$$f_n'(z_i) = x_n(z_{i-1}) f_n(z_{i-1}) + [1 - x_n(z_{i+1})] f_n(z_{i+1}) \quad \forall z_i \in z_{\text{grid}}.$$

Calculate the resulting growth rate:

$$g_n^1 = \int z^{\Gamma(\theta-1)} f_n''(z) \mathrm{d}z.$$

If that growth rate is close enough to g_n^0, proceed. If not, repeat the steps above until two successive iterations of the growth rate are close enough.

Once the growth rate convergence is achieved and we have g_n^*, adjust the relative quality distribution function for the growth in absolute quality such that it satisfies:

$$\int z^{\Gamma(\theta-1)} f_{n+1}(z)\mathbf{d}z = 1.$$

Compute aggregate research labor:

$$S_t = \int \lambda \log\left(\frac{1}{1-x_n(z)}\right) z^{\zeta} f_n(z)\mathbf{d}z.$$

Update the value function:

$$v_{n+1}(z) = \pi\left(z\right) + \left[x_n(z) \cdot v_n(ze^{\Delta - g_n^*}) + [1-x_n(z)] \cdot v_n(ze^{-g_n^*})\right]/R_n^* - \lambda \log\left(\frac{1}{1-x_n(z)}\right) z^{\zeta}.$$

If the value and relative quality distribution functions are close enough to their previous iteration, the solution is found. If not, repeat the steps above until two successive iterations of those functions are close enough.

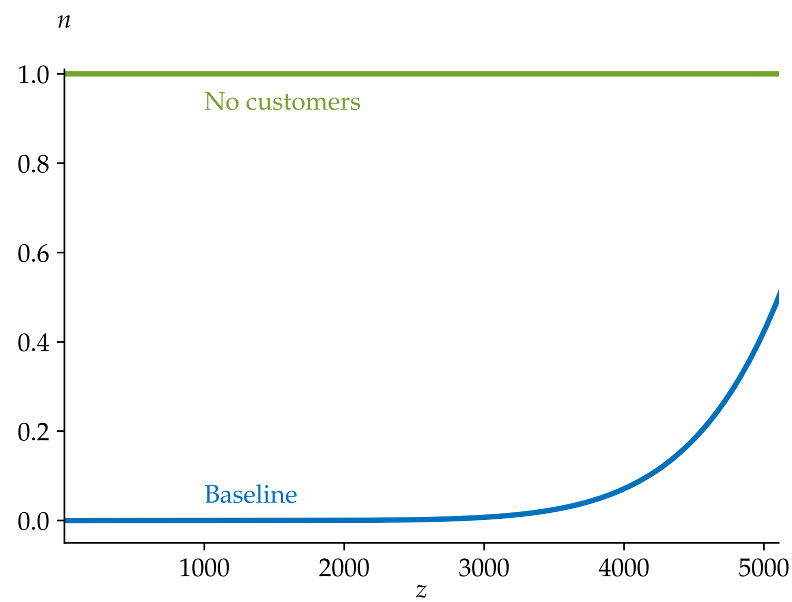
Numerical results

Figure E3 shows how n, the fraction of consumers the firm sells to, varies with the firm's relative quality z. It is log-linear with elasticity $\gamma/(\gamma-1)$ in the Baseline. This in turn makes the value of the firm much more convex with respect to z in the Baseline than in the No Customers case — see the log-log scale in Figure E4.

Because the customer margin makes variable profits increase much faster in relative quality, it induces higher quality firms to do more innovation than they would otherwise do. This can be seen in Figure E6a. A corollary is that R&D intensity (research spending as a share of sales) is slowly decreasing with respect to z in the baseline case, whereas it falls quickly with z in the model without a customer margin. As a result, the stationary distribution of relative qualities is much more dispersed with customer variation than without it (Figure E7).[1]

[1]Note that in both models, the probability of successfully innovating is equal to one for the smallest firms and zero for the largest ones, which delivers a stationary distribution of relative quality.

Figure E3: Customers and Firm Quality



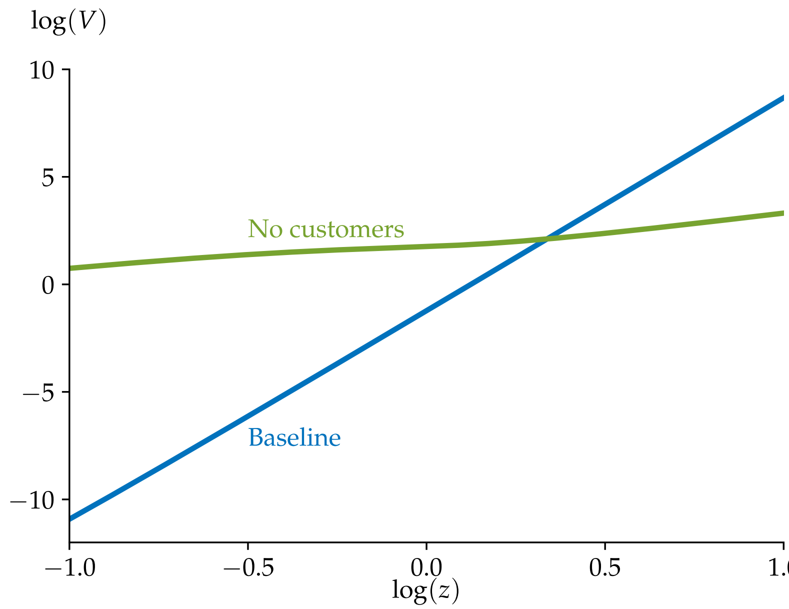
Note: This figure shows how n, the fraction of consumers the firm sells to, varies with the firm's relative quality z. The baseline features $\gamma = 1.25$ and the "no customers" version uses $\gamma = \infty$. γ is the elasticity of marketing costs with respect to customers.

The distribution of sales ends up being much more dispersed with an extensive margin for customers in Figure E5. Higher quality firms have more customers, and this endogenously induces more quality dispersion.

Figure E8 shows how R&D intensity (spending on R&D labor relative to sales) varies with firm sales. It is essentially flat in the baseline model, wherein both the value and difficulty of innovating rise in rough proportion to sales. In the model with no customers, however, R&D intensity is hump-shaped with respect to firm sales: at low levels of quality innovation is easy but not to worthwhile, at medium levels innovation is valuable, and at high levels innovation is too difficult. Marketing expenditures (not shown) are the same across all firms in the baseline model.

Just like in the data, we can calculate the contribution of the top 1% of firms (based on their sales increases) to aggregate sales increases. Recall from Figure B5 that this is over 60% in the data. As depicted in Figure E9, our baseline

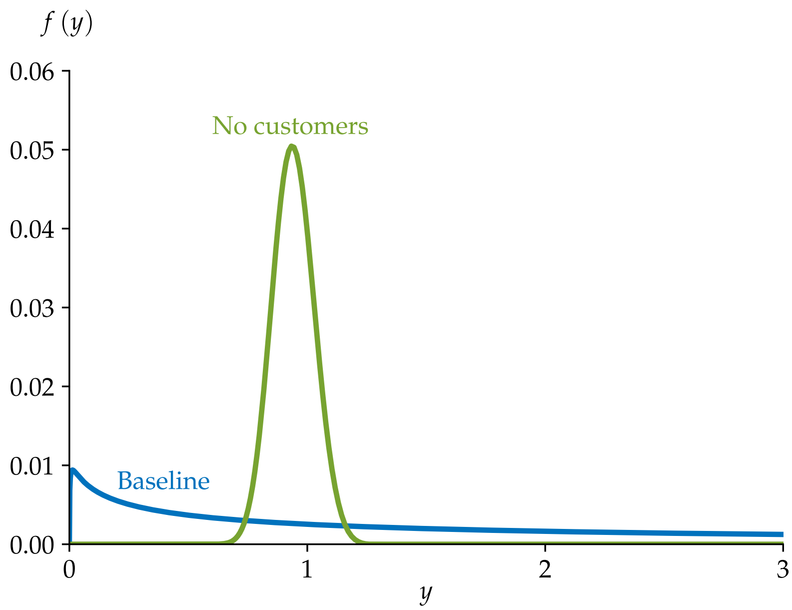
Figure E4: Customers and Firm Value



Note: This figure shows how the value of the firm v varies with the firm's relative quality z. The Baseline features $\gamma = 1.25$ and the "No customers" version uses $\gamma = \infty$, where γ is the elasticity of marketing costs with respect to customers.

model is calibrated to achieve a contribution of 70% from the top 1%. Without a customer margin, however, the top 1% of firms would account for less than 1% of all sales increases. Again, this comes from both the direct effect of acquiring customers in response to rising z, and the indirect effect of a much narrower z distribution in the absence of a customer margin.

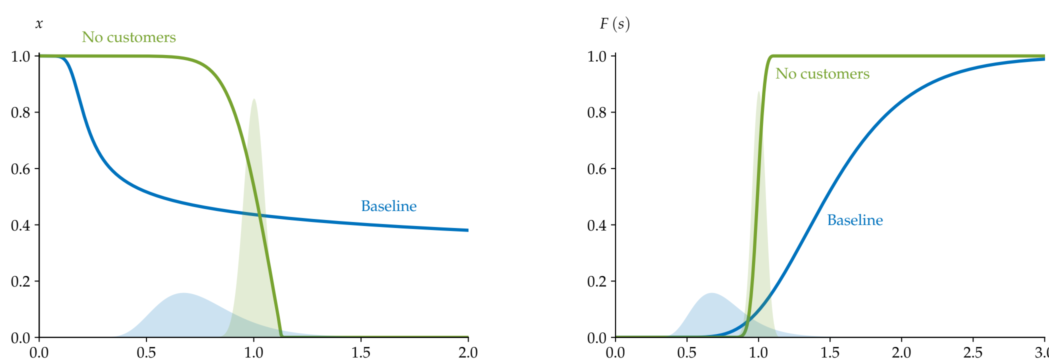
Figure E5: The Distribution of Sales



Note: This figure shows the density of firm sales y. The Baseline features $\gamma = 1.25$ and the "No customers" version uses $\gamma = \infty$, where γ is the elasticity of marketing costs with respect to customers.

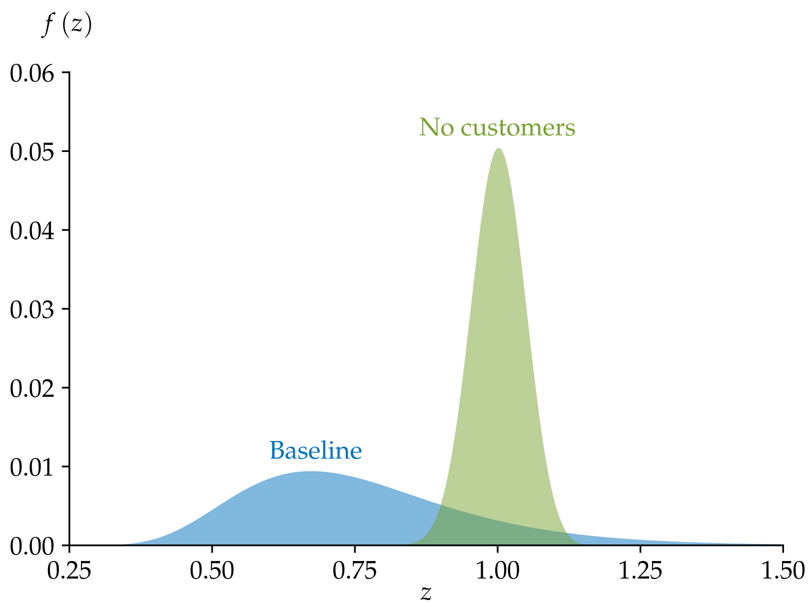
Figure E6: Innovation

(a) Innovation (b) Cumulative research



Note: This figure shows how the arrival rate of innovations x varies with the firm's relative quality z. The Baseline features $\gamma = 1.25$ and the "No customers" version uses $\gamma = \infty$, where γ is the elasticity of marketing costs with respect to customers.

Figure E7: The Distribution of Quality



Note: This figure shows the density of firm relative quality z. The Baseline features $\gamma = 1.25$ and the "No customers" version uses $\gamma = \infty$, where γ is the elasticity of marketing costs with respect to customers.

Figure E8: Research intensity

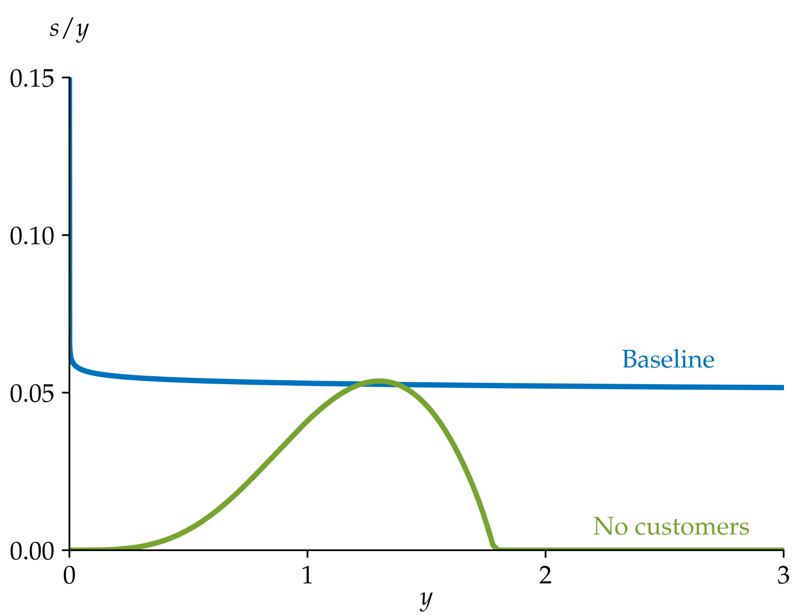


Figure E9: Firm Contributions to Aggregate Sales Changes