Preliminary and Incomplete

The Expansion of Product Varieties in the New Age of Advertising*

Salome Baslandze
Federal Reserve Bank of Atlanta

Jeremy Greenwood
University of Pennsylvania

Ricardo Marto
University of Pennsylvania

Sara Moreira
Northwestern University

February 27, 2023

Abstract

The last decades have seen large improvements in advertising technology that allowed firms to better target specific consumers. The relationship between advertising, the rise of product varieties, and economic growth is studied here. A model of advertising and product varieties is developed, where firms choose the intensity of digital ads directed at specific consumers as well traditional ads that are undirected. The calibrated model shows that improvements in digital advertising have driven the rise in product varieties over time. Causal empirical evidence, using detailed micro data on firms' products and advertising choices for the 1995-2015 period and exogenous variation in consumers' differential access to the internet, supports the suggested theoretical mechanism.

Keywords: digital (directed) advertising, traditional (undirected) advertising, specialization, targeting, varieties

JEL Nos: E1, L1, O3

* *Contact Information*: Salome Baslandze, baslandze.salome@gmail.com

1 Introduction

> Every trackable interaction creates a data-point, and every data-point tells a piece of the customer's story. *Paul Roetzer*

The new age of advertising dawned in 1994 with AT&T's "You Will" campaign that showcased the first digital advertisement.

Total spending on advertising has grown substantially since 1950 in the United States, but as Figure 1.1 (left panel) shows it has been relatively constant as a percentage of GDP. When total spending on advertising is broken down, it can be seen that there was a reallocation of spending away from traditional advertising toward to digital advertising (right panel). At the same time as spending on digital advertising rose there was an increase in the number of product varieties, as displayed in Figure 1.2.

Bar

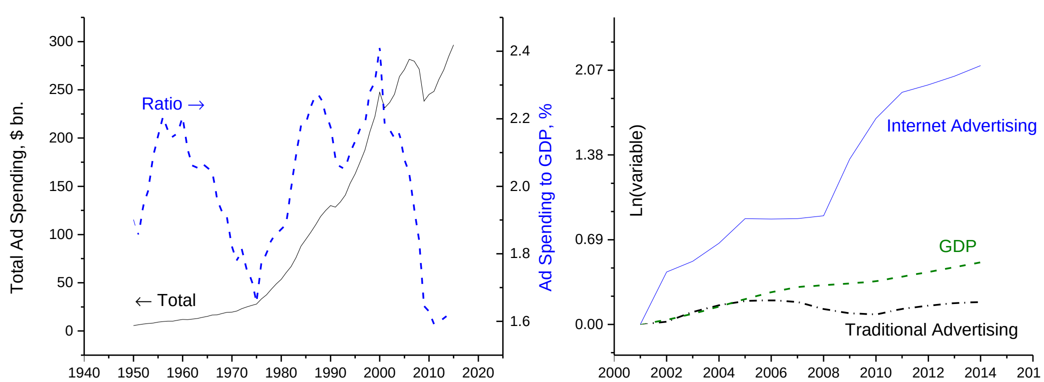

Figure 1.1: Aggregate Spending on Advertising.

Note: The left panel shows total aggregate spending on advertising, 1950-2015. The right panel shows the breakdown of total spending between traditional and digital advertising, 2001-2015.

The hypothesis entertained here is that the advent of digital advertising, which can be more precisely directed toward certain consumer attributes such as their tastes, stimulated more product variety. The impact of the new age of advertising on product variety is addressed in two ways. First, a model is developed where firms produce and sell their own product line. Within a product line there are different varieties of the

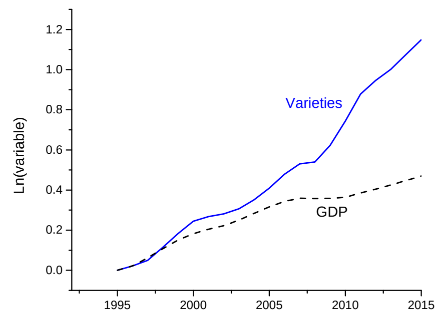

Figure 1.2: Increase in the Number of Varieties, 1995-2015.

good, each catering to the distinct tastes that consumers have. To sell its varieties a firm must advertise. There are two types of advertising, traditional and digital. Traditional advertising is broad based and applies to all of the varieties within the product line. Digital advertising aims to sell a specific variety and is directed toward consumers who have a preference for that variety. A firm chooses the number of varieties that it wishes to sell and the intensities of both types of advertising. To capture the digital advertising revolution, the efficiency of digital advertising is allowed to increase over time. As the ability improves to target more precisely the demands by groups of customers for high-value specialized varieties, the incentive to create more varieties increases.

The developed model is calibrated to match a set of stylized facts about advertising for the period 1995 to 2015. Some key facts are: the ratios of total advertising spending to GDP for 1995 and 2015; the ratios of digital to traditional advertising for 1995 and 2015; the increase in varieties over the period; the elasticity of sales with respect to advertising; and the elasticity of varieties with respect to sales. The calibrated model is then used to assess the impact of digital advertising on the number of varieties sold by a firm.

Second, the hypothesis is examined statistically using micro-level data. There is positive correlation between the growth in varieties and the growth in digital ads. Regression analysis suggests that the growth in digital advertising has caused an increase in number of varieties. To establish a causal impact, lightning strikes are used to instrument for digital advertising. Lightning strikes affected internet penetration.

Literature Review

Evans (2009) and Goldfarb (2014) make the case that digital advertising is fundamentally different than traditional advertising. The cost of targeting consumers is much lower with digital advertising. Advertisers now collect vast amounts of information about potential customers, and they use this to target consumers based on things such as the keywords used in search engines, past online behaviors, and demographic characteristics such as age, sex, location, etc.

Information-based models of digital advertising are rare in macroeconomics. Greenwood, Ma and Yorukoglu (2022) present a model where firms advertise the price of their goods. All goods are the same in their setting. Because consumers' information sets do not include ads from all firms, firms can set different prices for the same good. Digital advertising can be used to target consumers by their income levels: there is no point in sending an ad with a high price to a consumer who cannot afford to purchase the good at that price. By contrast, in the model presented here, digital advertising is used to target consumers who have preferences for specific varieties of a product line. The advent of digital advertising encourages firms to develop new varieties, something absent in the Greenwood, Ma and Yorukoglu (2022) model. At the core of the current analysis is a version of Salop (1979) well-known location model. Firms must decide where to locate their varieties on a circle vis à vis consumer preferences. Here, though, there is an added information friction. Firms must advertise on the circle to make consumers aware of their varieties while factoring in that not all consumers will receive digital and/or traditional ads.

In other information-based models of advertising, Dinlersoz and Yorukoglu (2012) show how technological progress in information dissemination favors efficient firms and leads to higher concentration. Perla (2019) studies how increased product awareness through information diffusion improves competition as an industry ages. In contrast to these papers, the focus of this study is on how digital advertising facilitates the creation of new differentiated product varieties.

Other recent papers on advertising and innovation are Cavenaile and Roldan-Blanco (2021) and Cavenaile et al. (2022). These papers develop a rich framework where firms make advertising and R&D choices, and where market structure is endogenous. R&D and advertising are substitutes at the firm level. The models, and focus of this research, differ significantly from the model presented here. While in these papers, advertising works by increasing firms' product quality shifter, here, advertising plays an information role and facilitates better matching between consumer tastes and products. Digital advertising is considered, and it is shown that digital advertising is complementary to the development of specialized varieties.

Several related studies examine the consequences of the reduction in search frictions for product markets. In Bar-Isaac et al. (2012), firms choose between niche and generic product design. A reduction in search costs (presumably due to the arrival of new information technologies) leads to both a dominance of more efficient firms but also to an increasing importance of smaller firms with specialized niche products (the so-called "long tail"). The theoretical model in Menzio (Forthcoming) explains why declining search frictions do not increase competition nor do they reduce price dispersion across firms. An increase in product specialization that helps firms differentiate their products explains these facts. The mechanics of the model presented here are quite different from these papers: firms endogenously choose the amount of targeting taking into account its cost (which parallels the exogenous reductions of search frictions in these other papers), and the model provides microfoundations for targeting and the match between consumer tastes and specialized varieties. In addition, while these models consider a choice of a niche or generic design by the single-product firms, in the current model, the number of different varieties offered by multi-product firms grows with targeting, consistent with the data.

The empirical results here relate to recent work that documents the rise in product specialization and an increase in firms' scopes. Hoberg and Phillips (2022) document an increase in a firm's product market scope for the sample of publicly traded firms over the past 30 years. Neiman and Vavra (Forthcoming) show that consumers are increasingly buying more niche products, presumably closer to their tastes, while Brynjolfsson et al. (2022) document an increase in book titles on the largest digital platform in China indicating an increased consumption of more niche titles by consumers. Gao and Hitt (2004) report a positive relationship between product varieties measured by trademarks and the information technology use by firms. While the empirical results in the current research are consistent with an increase in specialized product varieties produced by firms, the results also provide evidence that this increase in varieties is closely linked to the increasing use of digital advertising.

2 Empirical Evidence

What is the evidence supporting the hypothesis that improvements in the efficiency of digital advertising led to an increase in digital advertising spending and the number of product varieties? Two distinct empirical analyses, which explore distinct data and measures of product variety, are presented supporting the hypothesis. These data sets are detailed in Section 2.1, and aggregate data trends are discussed in Section 2.2. The first analysis uses firm-level panel data on digital advertising and product varieties. Section 2.3 shows that there is a positive relationship between the growth in digital advertising spending and the growth in product varieties, both at the firm-level and at more aggregated levels. The second analysis combines multiple data sources to build a spatial data set on household internet access and product varieties. Section 2.4 explores exogenous changes in the cost of targeting consumers through digital advertising by exploiting exogenous changes in households internet access across space and time. Spatial variations in the number of product varieties and exogenous changes in household internet access are used to study the causal impact that improvements in the efficiency of digital advertising, facilitated by better internet conditions, have on product varieties.

2.1 Data

The construction of data sets on digital advertising, internet access, and product varieties is briefly outlined here.[1] Product varieties refer to the number of different product variants within a specific product category. To ensure that the empirical findings are not driven by a specific definition of the variant, multiple definitions are employed with different levels of product aggregation. The most disaggregated level is more likely to capture differences in various attributes of a product, while a more aggregated definition differentiates only between the most important product attributes.

Firm-level advertising and varieties data set A data set is built that covers information on advertising spending and the number of varieties at the firm level over time. The data source is Kantar Media's AdSpender for the period 1995-2019. Kantar Media is a media intelligence company that systematically collects data on ads placed in different advertising media channels.[2] The different media channels are aggregated into *digital ads* (internet display, internet search, online video, and mobile web ads) and *traditional ads* (TV, magazines, newspapers, radio, and outdoor ads) and restrict attention to the product-related ads. For each advertised product, various product classifications are known, as well as which firm advertises the product, and how many units of ads are placed in different media channels for this product over time. Kantar converts units of ads into estimates of the ads expenditure. Two benchmark definitions

[1]Appendix A.1 and A.2 details the data sets.

[2]Advertising spending in Kantar data accounts for 40%-51% of aggregate advertising expenditure estimates from the US Census over time and for 30% to 36% of aggregate expenditure estimates from the IRS.

of product varieties in Kantar are used: distinct number of products and distinct number of brands. An example of a product is "Nike Air Max: Sneakers Men," and a brand for this product is "Nike Air."[3] Varieties are measured within a specific product category. Multiple definitions of product categories are also employed based on industry, major, and subcategory of the advertised product. In the example, industry is "Footwear," major is "Sport shoes," and subcategory is "Sneakers." The advertising firm is Nike.

County-level internet and varieties data set A new data set is assembled at the county $\times$ year$\times$ product-category level covering the consumer goods industry. This data set combines information on product varieties obtained from detailed scanner-level data with information on household internet access obtained from the Federal Communications Commission. An instrument in the regression analysis for internet access is taken from the National Lightning Database Network. Some additional variables used as controls in the regressions are sourced from the Bureau of Economic Analysis.

Nielsen scanner data on products sold in grocery, drugs, and general merchandise stores from 2006 to 2020 is harnessed. Two baseline definitions of product variety are used: product barcodes and brands. In the scanner-level data, products are defined as barcodes – the finest level of product disaggregation.[4] An example of a brand is "Chobani" that includes multiple barcodes with differences in flavor, form, size, package, and formula (among others). Barcodes and brands are measured within specific product categories, defined using Nielsen's product classification structure. The original data cover a wide range of products (e.g., from non-durables, such as cereals, to semi-durables like lamps), and each product is classified into one of the 1,070 product categories. To minimize concerns of potential mismeasurement of product varieties across locations, the baseline data set uses 602 product categories that have high coverage across all counties. [5]

Information about households' internet use at the county level over time is gathered from the Federal Communications Commission (FCC). The data come from the FCC

[3] There is also sub-brand information, an intermediate category between product and brand. However, because sub-brand information is often missing, it is only used for robustness.

[4] Argente et al. (2021b) discusses the advantages of defining products as barcodes. For robustness, definitions of varieties that lie somewhere between products and brands as in Kaplan and Menzio (2015) are also used.

[5] These include Dry Grocery (e.g., baby food, canned vegetables), Frozen Foods, Dairy, Deli, Packaged Meat, Fresh Produce. Robustness exercises are performed adding other product categories (including Health and Beauty Aids, Non-food Grocery, Alcohol, and General Merchandise). Moreover, the baseline sample covers a balanced set of stores across all years, so that changes in varieties cannot result from changes in the set of stores. For robustness, the entire set of stores is considered.

Form 477. This form is sent to internet service providers to request information about the types of services they offer, internet speeds, and subscribership, among other items. Data on the number of residential fixed connections (i.e., not mobile connections) per every 1,000 housing units is reported in 5 bins corresponding to quintiles of the population with residential fixed connections, where 1 stands for 0 to 200 out of 1,000 housing units, 2 is for 200 to 400 out of 1,000 housing units, and so on. The connections data are reported consistently for speeds above 200kbps at the census tract level for every year between 2008 and 2017. The data on the share of residential fixed connections reported by census blocks are aggregated into counties by computing the county means weighted by the number of housing units in each census tract as of 2010 (housing unit data is harnessed from the US Census Bureau).

Information on lightning strikes is used as an instrument in the regression analysis for differences in household internet access across counties over time. The data are obtained from the National Lightning Database Network (NLDN), which collects data on lightning strikes via ground-based sensing stations across the United States. The data are from the "County and State Summaries" and are available between 1986 and 2020, with records of the number of lightning strikes by county for every individual day of the year. This data set is combined with data on the size of US counties from the Census Bureau to get measures of lightning strikes per square mile per year at the county $\times$ year level.

Table 2.1 provides the summary statistics of the two baseline data sets. The firm-level data includes data from 1995-2015, with 332,190 firm-year observations, and 110,916 distinct firms. The spatial data set covers the period of 2008 and 2017 (where all variables are measured) for 2,259 counties and 602 distinct product categories. The key advantage of the firm-level data set is that it includes advertising spending. One of its disadvantages is that it only captures advertised product varieties, which are not necessarily the same as the total number of varieties offered by firms in the market. Appendix A.1 argues that the data capture well all of the advertised product varieties in the consumer goods sector and shows that the number of advertised product varieties is strongly correlated with the number of all varieties offered by firms. While the spatial data set does not have advertising measures (digital advertising is not available at the local level), it contains the measures of consumed varieties and additional variables employed to implement the causal analysis.

Table 2.1: Summary Statistics

		Mean	Median	St.Dev
	FIRM × YEAR LEVEL			
Varieties	Number Unique products	2.628	1	11.740
	Number Unique brands	1.845	1	5.986
	Number Unique sub-brands	3.341	1	9.228
Product Cat.	Number of subcategories	1.474	1	2.538
	Number of majors	1.261	1	1.263
	Number of industries	1.150	1	0.680
Advertising	Digital ad spending ($1,000s)	35.43	0.3	659
	Traditional ad spending ($1,000s)	221.00	4.0	1,575
	Total ad spending ($1,000s)	245.70	10.5	1,779
	COUNTY × YEAR × PRODUCT-CATEGORY LEVEL			
Varieties	Number of Unique Barcodes	39.96	10	99.71
	Number of Unique Brands	11.37	4	23.04
	Number of Products Aggregation 1	39.34	10	97.07
	Number of Products Aggregation 2	35.63	9	89.81
	Number of Products Aggregation 3	29.42	7	70.20
	Number of Firms	6.69	3	10.23
Internet	Household Internet Use	3.51	3.58	0.79
	Lightning Strikes per sq. mi. 5-year lag	10.85	9.78	7.75
	Lightning Strikes per sq. mi. 3-5 years lag	10.25	9.85	6.48
Other	Population (1,000s)	134.74	41.12	372.89
	Income Per Capita ($1,000s)	30.71	29.21	15.31

Note: Summary statistics for the two data sets. The "Firm×Year Level" data set uses Kantar and include ads only for products. Ads related to services and amusement, retail (store promotions), automotive dealers, financial, government/politics/organizations, schools, restaurants, hotels, and other services, as well as general ads about corporate promotions and recruiting are excluded. The "Varieties" variables from the "County×Year×Product Category Level" data set are from Nielsen RMS. The distinct definitions of products (Aggregations 1-3) follow Kaplan and Menzio (2015). The variables "Internet" and "Other" are at the county×year level using data from the FCC, NLDN, and BEA. Additional details on the variables are in the Data Appendix.

2.2 Aggregate Trends: Increase in Varieties and Decline in Prices of Digital Advertising

The empirical analysis begins by establishing that an increase in product varieties over time, documented earlier in Figure 1.2, is a robust fact that does not depend on the specific data set or definition of product variety used. Figure 2.1 illustrates this fact using the two data sets. The left panel shows the evolution of the (normalized) log number of advertised products and brands (as well as sub-brands and firms, for robustness) from Kantar. You can see that the number of distinct products offered in

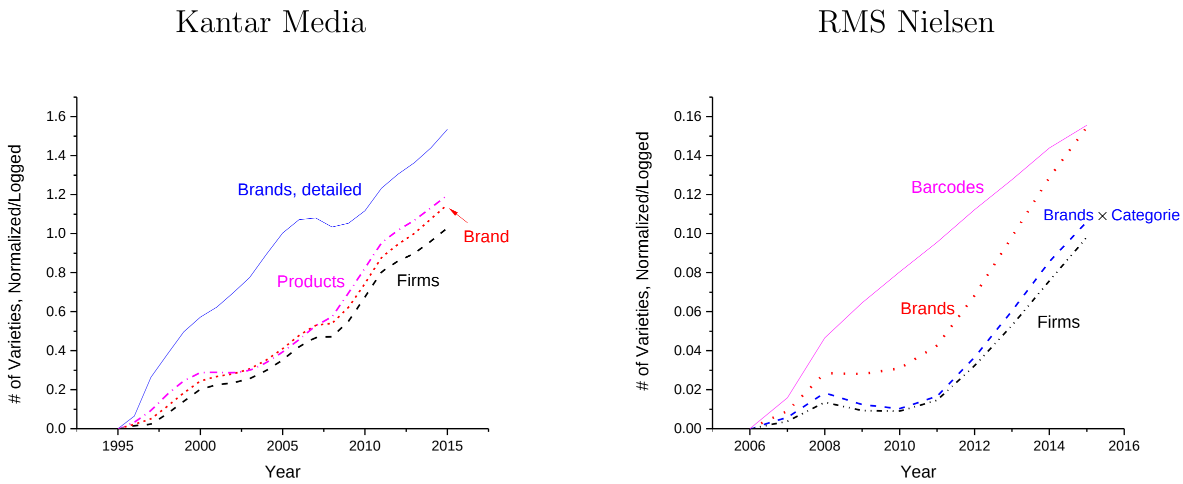

Figure 2.1: Product Varieties over Time

Note: The left panel shows trends in the normalized log number of product varieties over time from Kantar data. Product variety is defined based on the number of products, brands, detailed brands, and firms. The right panel shows trends in the normalized log number of product varieties over time from RMS Nielsen data. Product variety is defined based on the number of barcodes, brands, brands × category (module), and firms.

2015 is 3.2 times larger than the number of distinct products in 1995. The right panel displays this trend using the RMS Nielsen county-level data set. The sample period for this data set begins later but shows a similar trend: the number of distinct barcodes and brands (also, brand × categories and firms, for robustness) increases steadily from 2006.

At the same time, with advent of internet and improvements in targeting technologies, digital advertising became cheaper. Figure 2.2 shows time fixed effects of firm-level log prices by media type from 2001 to 2015. While advertising prices in traditional media – TV and newspapers, are stable or grow, digital advertising prices – here, captured by internet display ad prices, sharply dropped, consistent with technological improvements in digital ads.

2.3 Relationship Between Growth in Digital Advertising and and Varieties

To show the relationship between digital advertising and product varieties, first a visual plot is presented of the correlation between digital ads growth and product variety growth in the Kantar data. Figure 2.3 scatterplots the log change in the number of distinct product varieties and the log change in the digital-ads spending across product categories from the first year with digital ads, 2001, to the last year in the sample,

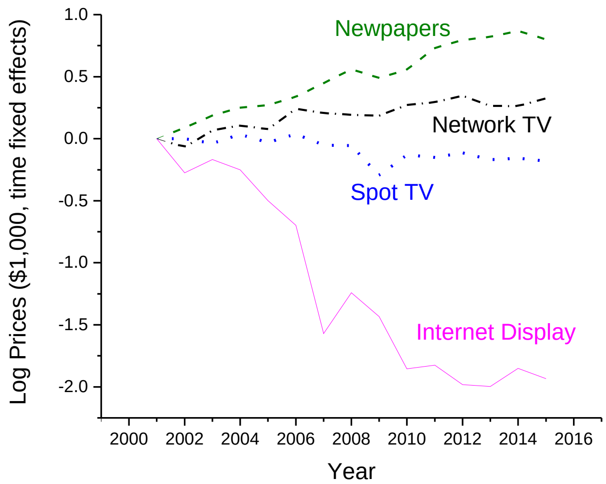

Figure 2.2: Advertising prices by media type

Note: Ad prices by media type come from Kantar Media. Prices are defined as total ad spending ($1,000) divided by the number of ads of a certain media type by a firm. The lines represent the estimated time fixed effects of firm-level log prices of each type of advertising using data from 2001 to 2015.

2015. Product variety is defined by distinct products and product category is defined using subcategories. You can see that product categories using more digital ads over time also increase the number of varieties. Similar scatterplots obtain when alternative definitions of product variety and product category are used.

These correlations are evaluated using industry-level and firm-level variation over time in Table 2.2. Panel A shows regressions of the year-to-year log change in the number of distinct products and brands on the log change in the digital-ads spending across product categories. Since the number of distinct product varieties offered in a product category depends on traditional ad spending and, mechanically, on the number of firms in product categories, the regressions control for traditional ads and for the number of firms. Product category and year fixed effects filter out product-category specific constant characteristics and annual common demand and supply shocks. Panel B looks at similar regressions at the firm level, controlling for firm size, traditional ad spending, year, and product line or firm fixed effects. In all specifications, you can see, other things equal, that growth in digital-ads spending is associated with growth in the number of distinct product varieties. Tables B.1 and B.2 in Appendix B show robustness of these associations using different specifications, namely using regressions in log levels instead of changes, and using digital ads relative to traditional ads as the main control.

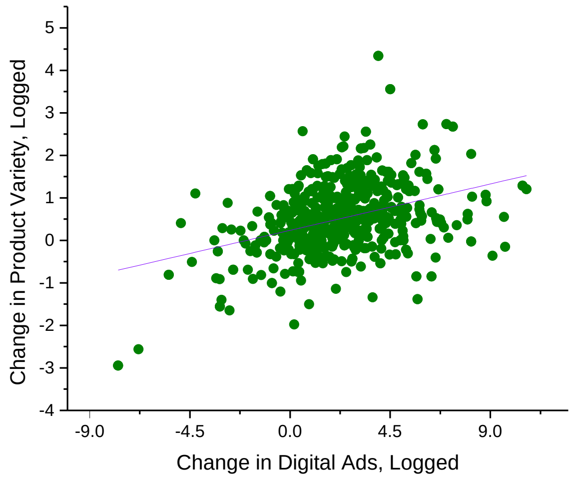

Figure 2.3: Correlation Between Growth in Digital Ads and Growth in Varieties, 2001-2015

Note: The scatterplot displays the log change in the number of distinct product varieties (products) against the log change in the digital-ads spending across product lines (subcategories) from 2001 to 2015. The red line is a linear fit. Data are from Kantar Media. Data on digital ads starts from 2001.

Table 2.2: Product Variety and Digital Ads

Panel A: Category-level	Δ *Log Products*		Δ *Log Brands*	
	Subcategory	Major	Subcategory	Major
ΔLog Digital Ads	0.024***	0.014***	0.021***	0.015***
	(0.002)	(0.004)	(0.002)	(0.004)
R^2	0.240	0.267	0.240	0.254
Observations	11,658	2,996	11,658	2,996
Panel B: Firm-level	Δ *Log Products*		Δ *Log Brands*	
	Cross-firms	Within-firms	Cross-firms	Within-firms
ΔLog Digital Ads	0.042***	0.042***	0.030***	0.031***
	(0.002)	(0.002)	(0.002)	(0.002)
R^2	0.096	0.186	0.052	0.127
Observations	17,931	16,920	17,931	16,920

Note: Panel A shows regressions of the growth in product varieties on growth in digital-ads spending in product categories over time. All regressions control for log number of firms and log traditional-ads spending in product categories over time, product line, and year fixed effects. Product variety: products and brands. Product categories: subcategory, major, and industry. Robust standard errors in parentheses. Panel B shows regressions of the growth in product varieties on the growth in digital-ads spending in firms over time. All regressions control for firm's log employment, log traditional-ads spending, year fixed effects, and product line/firm fixed effects in the "Cross-firms"/"Within-firms" columns, respectively. Product variety: products and brands. Product category: subcategory. Robust standard errors in parentheses. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

2.4 The Causal Impact of Advertising on Varieties

The previous section establishes a positive correlation between digital advertising and product varieties. However, this correlation may be driven by other factors not related to the improved targeting of consumer preferences through digital advertising. To investigate the hypothesis at hand, an ideal experiment would involve exogenously changing the cost of targeting consumers through digital advertising and examining the resulting effect on the number of product varieties offered by firms.

The spirit of this ideal experiment is captured here by exploiting exogenous changes in households' internet access and measuring its impact on the number of product varieties offered by firms. When households do not have access to the internet, firms cannot use digital advertising to target consumers' tastes (Evans, 2009; Goldfarb, 2014). As a such, households' internet access is a crucial determinant of the cost of targeting consumers through digital advertising.[6] To obtain exogenous variation in households' internet access, lightning strikes are used as an instrument for residential internet access across different locations and time. The work of Andersen, Bentzen, Dalgaard and Selaya (2012) and Guriev, Melnikov and Zhuravskaya (2021) motivates this.

2.4.1 Empirical strategy

The empirical strategy uses spatial and time-series variation, together with a lightning-strike instrument, to estimate the causal relationship between household internet access and product varieties. A data set is built at the county (l) $\times$ year (t) $\times$ product category (j) level with information on household internet access, I_{lt}, product varieties, N_{ltj}, and several other variables (including lightning strikes, Z_{lt}).[7]

There are substantial differences in households' internet access and product varieties across locations in the United States, with these differences changing over time. The empirical strategy relies on spatial heterogeneity in households' internet access and on spatial heterogeneity in firms' decisions to advertise products and introduce varieties at the local level. [8] Figure 2.4 (Panel A) shows, based on retail scanner data, that

[6] The use of households' internet access instead of households' digital ads consumption is akin to the intent-to-treat empirical strategy (Hoynes and Schanzenbach, 2009). In the current case, households' internet access proxies for the propensity of being "treated" by digital ads.

[7] The product categories with consistent spatial coverage are considered to ensure that the results are not driven by compositional differences in stores covering distinct product types across locations.

[8] While firm's digital advertising decisions at local level are not observed (only aggregated), Argente, Fitzgerald, Moreira and Priolo (2021a) use data on some types of traditional advertising and show that multi-locations firms make advertising decisions at the local level.

there is substantial variation in the county-to-nationwide ratio of the number of product and brands (plotted for first and last year of the data set). The map becomes darker as more varieties are being sold in a county. Table B.3 in Appendix B shows that even multi-market firms offer a differing number of product and brand varieties across locations and that these differences cannot be fully explained by population and income differences.

Figure 2.4 (Panel B) displays the heterogeneity in household residential fixed connections across locations in 2008 and 2017 (the first and last year of the sample). A value between 4 and 5 means that more than 80% of the population in the location has a residential fixed internet connection, while a value between 0 and 1 means that less than 20% of the population in that location has a residential fixed internet connection. In 2008, internet was fully diffused in some locations, while in other areas it was still in its early stages. Ten years later, most households in the United States had access to internet. Nevertheless, differences across locations still persist.

Differences in households' internet access across locations are instrumented using the frequency of lightning strikes per square mile. Prior studies have shown that the frequency of lightning strikes affects the diffusion of digital technologies due to an increase in the expected costs associated with voltage spikes and dips (Andersen, Bentzen, Dalgaard and Selaya, 2012). The technology needed for residential internet, including ADSL and Cable, is particularly sensitive to electrical surges caused by lightning strikes, which can lead both to immediate damage and to quicker depreciation of equipment over time.[9] Figure 2.4 (Panel C) shows a large spatial variation in lightning strikes across locations and time.

Given that the endogenous variable and instrument vary across locations and time, the following first-stage equation is estimated:

$$I_{lt} = \gamma Z_{lt-k} + \eta X_{lt} + e_{lt}, \tag{2.1}$$

where I_{lt} represents a categorical variable measuring the fraction of households with a residential fixed internet connection in location l and year t, $Z_{l,t-k}$ is the number of lightning strikes per square mile in location l and year $t-k$, and X_{lt} is a vector of characteristics that includes detailed fixed effects and other controls such as population and income. Because infrastructure investments made by companies supplying residential fixed internet connections might take a long time, there may be a lagged relationship

[9] Power surge protection can partially alleviate the problem, but it is expensive and not always effective.

between I_l and Z_l. In the baseline specification, k is set to 5 because the association is strongest with that lag. In robustness tests, alternative lags are explored.

In the second stage, the association between number of product varieties and predicted household internet access is estimated using the following specification:

$$N_{ltj} = \beta \hat{I_{lt}} + \alpha X_{ltj} + \epsilon_{ltj}, \tag{2.2}$$

where $\hat{I_{lt}}$is the predicted fraction of households with internet access in location l and year t from regression (2.1), N_{ltj} is the number of product varieties sold in location l at year t of product category j, and X_{ltj} includes detailed fixed effects and other controls such as population and income that are also used in the first-stage regression. By using exogenous variation from lightning strikes, the analysis filters out the variation from external forces that might impact both the demand for household internet access in an area and firms' product introduction decisions in the same area. Moreover, because households' internet access is measured at customers' locations rather than at firms' production facilities, concerns are alleviated about the possibility that the mediating channel between internet access and product variety decisions operates through changes in firms' operational costs (other than costs of digital advertising).

2.4.2 Results

Table 2.3 presents the main results for the instrumental variable specification.[10] The baseline estimates are for varieties measured either as the logarithm of the number of distinct products (columns 1-3) or as the logarithm of the number of distinct brands (columns 4-6). The specification of column 1 and 4 includes year × category fixed effects that control for common trends in internet access and varieties over time. Columns 2 and 5 further tighten the empirical specification by including year × category fixed effects as well as year × county controls (population and income per capita) that account for time-varying heterogeneity across counties and absorb many differences across locations over time in potential demand factors for product varieties. Finally, the last set of results is akin to a difference-in-differences specification (columns 3 and 6). This specification controls for county-specific differences in the number of product varieties in different categories and household internet access. It identifies the main effect from the differential growth of varieties and internet across locations over the nine years covered by the data set.

[10] Table B.4 presents the first-stage results. The overall F-statistic for the excluded instruments is large, especially when it captures differences across counties.

Table 2.3: Household internet access and product varieties

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(4)	(5)	(6)
Household Internet	1.098***	1.371***	0.268*	0.812***	1.005***	0.404***
	(0.040)	(0.074)	(0.158)	(0.029)	(0.054)	(0.147)
Observations (1,000s)	9,905	9,905	9,884	9,905	9,905	9,884
Year × Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × County Controls	No	Yes	Yes	No	Yes	Yes
County × Category FE	No	No	Yes	No	No	Yes
1st stage F-stat	702.4	504.6	15.2	702.4	504.6	15.2

Notes: The estimated regression coefficients for equation (2.2). The dependent variable in columns 1--3 is barcodes (in logs) and in column 4–5 is brands (in logs). The variables are described in Section 2.1. The year × county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at county × year level and shown in parentheses. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

There is a statistically significant relationship between households' internet access and product varieties across all specifications. An exogenous increase in internet access leads to more varieties. The results are statistically significant for the three alternative specifications employed in the analysis and for the two alternative measures of product variety. The magnitude of the effect of household internet access on product varieties is economically large. Using the difference-in-differences specification, an increase of 20 percentage points in the share of population with residential internet access generates an increase in the number of product varieties of about 27%-40%. Table B.5 in Appendix B documents equivalent OLS specifications. The results of the instrumental variables and OLS specifications are qualitatively similar, and the magnitudes are somewhat larger in the instrumental variables estimation.

The results are qualitatively similar when either other measures of product varieties are used (Appendix Table B.6) or when alternative lagged relationships between product varieties and internet access are considered (Appendix Table B.7). Appendix B presents results where the baseline set of product categories is expanded to include additional categories (Table B.8) and alternative samples of retail stores (Table B.9). Finally, alternative lags for the lightning strike instrument are explored. The baseline specification (2.1) considers a 5-year lagged relationship between household internet access and lightning strikes, and Table B.10 shows robustness with the average lightning strikes in the previous three to five years.

2.4.3 Discussion of alternative mechanisms

The results are predicated on two key identification assumptions. The first assumption is the relevance of the lightning strikes for household internet access. This assumption was asserted by the strong first-stage results. The second assumption is that the frequency of lightning strikes affects varieties only through its effect on household internet access, conditional on all other covariates. A potential concern with this exclusion restriction is that the lightning strikes may correlate with the firms' use of information and communication technologies (ICT), too. Improvements in ICT may, in turn, make these firms more productive and lead to an increase in their product offerings.[11],[12] The use of spatial variation addresses this concern: consumers' internet access can be distinguished from firms' internet access in the locations where they produce and operate their establishments. Consider a simple example. Suppose that a firm is located in region P but sells in locations H and L. Household internet penetration is high for location H but low for location L. This difference means that consumers in H can be easily targeted via digital advertising but not consumers in L. As a result, firms are relatively more likely to target consumers with digital advertising in location H than consumers in location L. Because in the data the majority of firms sell in multiple markets, but produce in just one or a few locations (Argente, Fitzgerald, Moreira and Priolo, 2021a), variation in the ability to target consumers differently across locations can be used.

To this end, information on the location of firms is gathered, and additional measures of varieties are defined that exclude firms co-located where their products are being sold. Thus, information is used on where firms sell products and where they produce these goods to select firms whose general internet access conditions differ from those of its consumers. In particular, two measures are built. First, for each location, products are selected that are sold by firms whose headquarters are in another state and that sell to multiple states. Second, products are selected that are sold by national (as opposed to regional) firms. National firms are defined as firms in the top quartile of the distribution of the number of states they sell in.[13]

Table 2.4 presents the results for the difference-in-differences specification (equivalent to columns (3) and (6) of Table 2.3) for the measures of varieties based on the number

[11] For example, several recent papers argue that advancements in ICT may have facilitated the expansion of firms; e.g., Aghion et al. (2019), De Ridder (2019). Hsieh and Rossi-Hansberg (2020), and Lashkari, Bauer and Boussard (2018).

[12] The theoretical model presented later accommodates this alternative channel: the firm's marginal cost of more product offerings declines over time, inducing an increase in varieties.

[13] Details on these variables are in Appendix A.2.

Table 2.4: Household internet access and product varieties: Robustness

	Log Products		*Log Brands*	
	HQ other Multi-state	HQ other National	HQ other Multi-state	HQ other National
Household Internet	0.245	0.256*	0.394***	0.414***
	(0.153)	(0.156)	(0.143)	(0.147)
Observations (1,000s)	9,790	9,223	9,790	9,223
Year × Category FE	Yes	Yes	Yes	Yes
Year × County Controls	Yes	Yes	Yes	Yes
County × Category FE	Yes	Yes	Yes	Yes
1st stage F-stat	15.7	15.4	15.7	15.4

Notes: The estimated regression coefficients for equation (2.2). The dependent variables are either barcodes or brands (in logs) in a county x year x category sold by firms whose headquarters are in another state and sell in more than one state (multi-state) or in the top quartile distribution of number of states (national). The year × county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at county × year level and shown in parentheses. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

of products and brands. These alternative specifications generate the main coefficients of interest that are in magnitude very similar to the baseline results, although the statistical significance is somewhat weaker in column 1.[14]

[14] Other measures of varieties are computed that account for concerns about retail chains making product offering decisions based on the same internet conditions as their consumers. Table B.11 in Appendix B shows the robustness of the results excluding local retail chains.

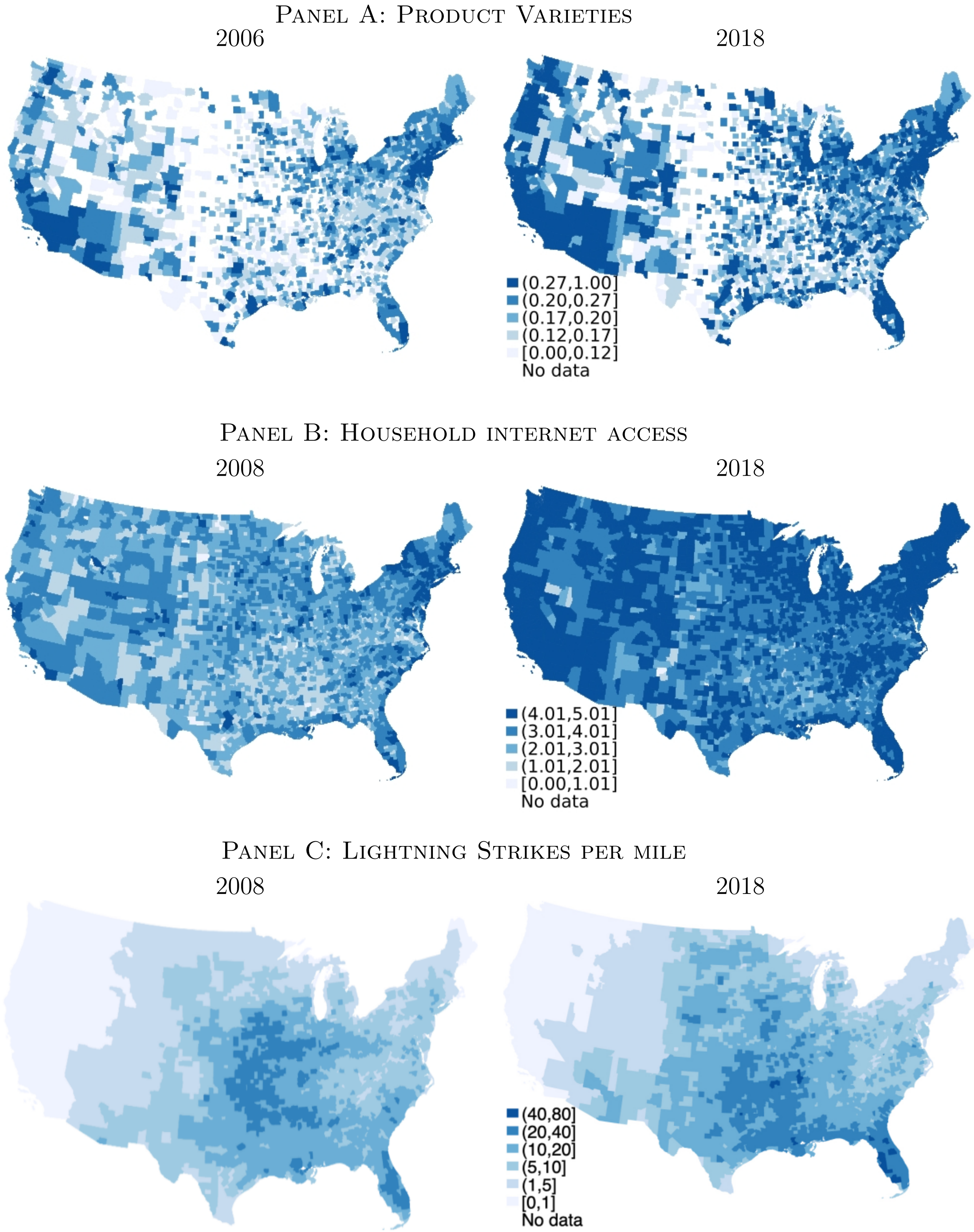

Figure 2.4: Spatial variation in product varieties, household internet penetration and lightning strikes

Note: Panel A shows for each county the share of product varieties of category j out of total varieties sold nationwide, weighting each product category j by its revenue. Panel B illustrates the quintile of population with residential fixed internet connections at the census tract level. Panel C displays lightning strikes per square mile.

3 Model

Imagine a world with a unit mass of consumers. A consumer has two sources of income, wages and profits. Wages derive from the one unit of labor that individuals exogenously supply and profits accrue from the portfolio of firms that they own. Their income is used to consume a generic good and products from a spectrum of specialized product lines. Within a specialized product line there are different varieties. A consumer prefers varieties that are more closely matched with their tastes. To know about a product line, and the varieties contained within it, a consumer must have received advertisements. The generic good is not advertised and all consumers know about it.

Firms produce generic goods and specialized products using labor supplied by consumers. They distribute any profits to the consumers. A specialized-product firm is associated with a specific product line that may contain many varieties. There is free entry into the specialized products sector. The number of varieties that a firm offers is endogenous. In order to sell its product line, a firm must advertise. There are two types of advertisements, digital and traditional. Traditional advertising makes consumers aware of a firm's product line, but it is not oriented toward consumers' specific tastes over varieties. Digital advertising is more focused and is geared toward matching a consumer's tastes to a specific variety within the product line. A specialized products firm can use both types of advertising. Specialized product lines are not perfect substitutes so firms possess some market power. In contrast, generic goods are perfect substitutes. The generic goods sector is perfectly competitive.

Over time digital advertising becomes more efficient relative to traditional advertising. This increases a firm's incentive to produce more varieties. To study this, the analysis will focus on comparing two static, symmetric equilibria.

3.1 Consumers

An individual consumes a generic good and a single variety from each specialized product line in the consumption set $\mathcal{M}_i$ specialized product lines. The utility for a specialized product line $j \in \mathcal{M}_i$ depends upon how close the variety consumed matches the consumer's tastes. This dependence is denoted by the function $S_i(j)$. A variety within product line j costs $p(j)$ while the price of the generic good is normalized to one. A person earns w in labor income and π in profits.

3.1.1 Utility Maximization Problem

Each person solves the problem

$$\max_{c,\{q(j)\}}\{\theta \ln c + (1-\theta)\int\limits_{j\in\mathcal{M}} S(j)^{\kappa}\frac{q(j)^{1-\kappa}}{1-\kappa}dj\},\ \text{with } 0<\kappa<1, \tag{3.1}$$

subject to

$$c + \int\limits_{j\in\mathcal{M}} p(j)q(j)dj = w + \pi \equiv y,$$

where c is their consumption of the generic good, $q(j)$ is their consumption of a variety within product line j, and where for expositional purposes the subscript i is dropped. The utility weight on generic goods is θ, and the inverse of the price elasticity of demand for specialized products is $0<\kappa<1$. It is easy to calculate that their consumption of generic goods is given by

$$c = \theta\hat{y}, \tag{3.2}$$

with

$$\hat{y} \equiv \frac{y}{\theta + (1-\theta)\int_{j\in\mathcal{M}} S(j)^{\kappa} q(j)^{1-\kappa}dj}. \tag{3.3}$$

Their consumption of a variety within specialized product line j is

$$q(j) = S(j)\left[\frac{(1-\theta)\,\hat{y}}{p(j)}\right]^{1/\kappa}, \tag{3.4}$$

which is linear in the quality of the match between the consumer's tastes and the variety as represented by $S(j)$. The derivations are in Appendix C.

3.1.2 Tastes for a Variety within a Product Line

How are consumer's tastes for varieties within a product line determined? Let an individual's tastes for a variety be represented by a circle with a circumference of one. The circle represents a particular product line. Now, suppose that the circle is split up into n varieties equally spaced around the circumference. So variety 1 is located at point

$1/n$, variety 2 at $2/n$, and variety n at $n/n = 1$. The consumer's tastes are situated at some point i on the circle. This point is randomly distributed across product lines. It also differs by consumer. Assume that i is uniformly distributed over the circle across consumers. Represent the distance between a particular consumer tastes, represented by the point i, and the location of variety m, represented by the point m/n, by the arc length $d(i, m/n)$. How well this particular variety matches the consumer's tastes is given by

$$\sigma\big(d(i, m/n)\big) = \chi - \lambda d(i, m/n). \tag{3.5}$$

The utility realized by purchasing this variety is

$$\sigma\big(d(i, m/n)\big)^{\kappa} q(j)^{1-\kappa}/(1-\kappa),$$

where q is the quantity purchased at price p of the variety located at point m/n. This solution for q is

$$q = \sigma\big(d(i, m/n)\big)\left[\frac{(1-\theta)\,\hat{y}}{p(i)}\right]^{1/\kappa}, \tag{3.6}$$

The situation is portrayed in Figure 3.1.

Note that σ represents the taste factor in terms of distance between the consumer's tastes and a variety while S defines the taste factor in terms of a product's index, j. The consumer buys a variety within a product line if and only if they received an ad for the product line. So, $\mathcal{M}$ is the set of specialized product lines for which the consumer got an ad. Hence, for $j \in \mathcal{M}$, $S(j) = \sigma\big(d(i, m/n)\big)$ when the consumer buys the variety, m/n, contained in product line j that is a distance $d(i, m/n)$ from their tastes, i. For $j \notin \mathcal{M}$, $S(j) = 0$.

3.2 Specialized Product Firms

There are N monopolistically competitive firms selling specialized product lines. The number of firms is endogenous and increases over time due to growth in the economy. Each specialized product line is sold by a unique firm, which may produce many different varieties. There is free entry into the specialized products sector. To produce the firm must incurEach variety is sold at the unit price p. To sell specialized products a firm

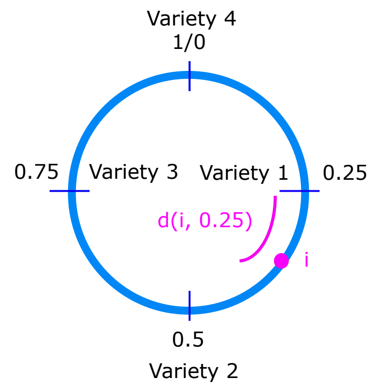

Figure 3.1: Taste over Varieties within a Given Product Line.

Note: The figure shows a situation where there are 4 varieties, located at the distances 0.25, 0.50, 0.75, and 1.0 when measuring clockwise from the top. The consumer has tastes located at the point i. The distance between his tastes and the variety 2 is measured by the arc length $d(i, 0.25)$. If the person consumed this variety, then $\sigma = \chi - \lambda d(i, 0.25)$.

must advertise. The firm uses two types of advertising, traditional and digital. It chooses the intensities for both types of advertising. Traditional advertising is generic in nature. Think about it as advertising the whole product line and not being directed toward specific consumers with tastes for particular varieties. Digital advertising is directed at selling a particular variety to a consumer who has tastes for that variety.

Suppose a consumer receives a traditional ad but no digital advertisement. The consumer is alerted to the product line. They will then buy a variety located at some random point k within that line. The distance between the location of the consumer's tastes, i, and the variety located at the point k is $d(i, k)$, implying that $\sigma(d(i, k)) = \chi - \lambda d(i, k)$. The situation is portrayed in Figure 3.2. This implies that the consumer will spend $\sigma(d(i, k))[(1-\theta)\hat{y}/p]^{1/\kappa}$ on the product line. Average spending from people who just receive a traditional ad will depend on the average taste factor as outlined in Lemma 1.

Lemma 1. *(Average taste factor for traditional advertising) The average $\sigma\big(d(i, m/n)\big)$ over all consumers who just receive a traditional ad is given by*

$$\sigma_t = \chi - 0.25\lambda.$$

Proof. Take any variety on the product circle. The maximal distance between a consumer and this variety is 0.5. Now, consumers are uniformly distributed around the circle. Therefore, consumers' distances are uniformly distributed on the interval $[0, 0.5]$. So, the average distance is just 0.25. Last, the odds of picking any variety are the same. Therefore, the average distance over all varieties is also 0.25. $\square$

The important thing to note here is that this average taste parameter, σ_t, is not a function of n. The quantity demanded from people who just receive a traditional ad is therefore

$$q_t = \sigma_t[(1-\theta)\hat{y}/p]^{1/\kappa}, \text{ cf. eq. (3.4)}.$$

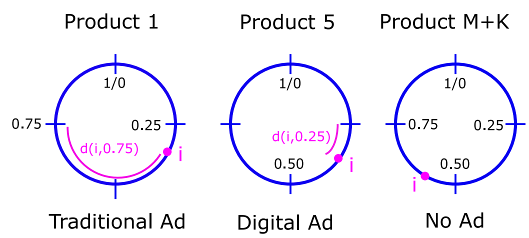

Figure 3.2: Advertising.

Note: The diagram shows the situation for 3 product lines, numbers 1, 5, and $M + K$. Each product line has 4 varieties, equally spaced around the circle. The point i marks a consumer's tastes within each product line, which differs across lines. For product line 1 the consumer got a traditional ad but no digital ad. They randomly choose the variety located at the point 0.75, which has an arc length of $d(i, 0.75)$ from their tastes. Product line 5 illustrates what happens when the consumer got a digital ad catering to consumers with tastes lying in between the first and second varieties. Here they choose the second variety located at the point 0.25 with an arc length of $d(i, 0.25)$ from their tastes. The consumer got no ad for product line $M + K$. Hence, they do not consume this product, which lies outside of the products that they consume; i.e., $M + K > M$.

Alternatively, consider a digital advertisement that is targeted to the consumers with tastes that lie within the range between two adjacent product varieties. When there is more than one variety this range will have an arc length of $1/n < 1$. Hence, traditional advertising applies to the whole circle while digital advertising targets just a segment. If a consumer with tastes positioned at i buys one of these varieties, located at the point l, then their taste parameter will given by $\sigma(d(i,l)) = \chi - \lambda d(i,l)$. Again, Figure 3.2 illustrates the situation. Suppose that consumers pick the closest adjacent variety. Lemma 2 specifies the average taste factor for consumers who receive a digital ad.

Lemma 2. *(Average taste factor for digital advertising) The average taste factor across consumers receiving digital advertisements is*

$$\sigma_d(n) = \chi - \frac{0.25\lambda}{n}.$$

Proof. A variety spans an arc with distance of $1/n$. The maximal distance between the consumer and the closest variety is $1/(2n)$. Consumers are uniformly distributed over distances on the interval $[0, 1/(2n)]$ with varieties being equally spaced. Therefore, the average distance is $0.25/n$. $\square$

The average taste factor for digital advertising is a function of the number of varieties, n. The important thing to note is that the average taste factor decreases in number of varieties, n. So long as $n \geq 2$, digital advertising will on average lead to customers buying varieties that better match their tastes. The average quantity purchased from individuals receiving digital ads is

$$q_d(n) = \sigma_d(n)[(1-\theta)\hat{y}/p]^{1/\kappa}, \text{ cf. eq. (3.4)}.$$

Last, a person may get no ads. In this case they won't buy the specialized product–see 3.2. This case can be simply represented by $\sigma_n = 0$.

As long as more than one variety is produced, Lemmas 1 and 2 establish that consumers' tastes are better matched to varieties with digital advertising, which is directed, than with traditional advertising which is undirected. The upshot of the lemmas is illustrated in Figure 3.3 which plots the average taste factor with digital and traditional advertising. As can be seen, digital advertising always results in a closer match and the gain from digital advertising increases with number of varieties.

The intensities of traditional and digital advertising are denoted by a_t and a_d. These represent the probabilities of receiving a traditional and digital advertisement. The odds of a consumer buying a digitally advertised product are a_d. A person will only buy a product based on a traditional ad if they did not receive a digital ad. This transpires because a digital ad delivers a variety that is catered to a consumer's tastes. Therefore the chance of a consumer buying a product that is traditionally advertised is $a_t(1 - a_d)$. The probability of an individual not buying a variety within the product

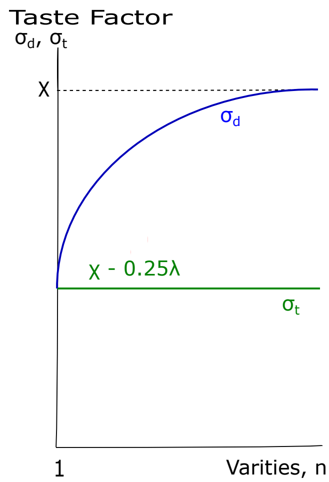

Figure 3.3: Digital versus Traditional Advertising.

Note: With traditional advertising the average taste factor, σ_t, is not a function of the number of varieties. With digital advertising the average taste factor, $\sigma_d(n)$, increases with n as consumer tastes are better matched with varieties due to the fact that now ads for varieties can be directed toward consumers with specific tastes.

line is $1 - a_d - a_t(1 - a_d)$. A specialized goods firm will sell to consumers that received digital and/or traditional ads. Its sales will be

$$p[a_d q_d(n) + a_t(1 - a_d)q_t],$$

where $q_d(n)$ and q_t represents the quantities demanded from consumers solicited from digital and traditional advertising.

The cost functions, in terms of labor, for digital and traditional advertising are

$$Aa_d^{\zeta}/\zeta \text{ and } Ba_t^{\nu}/\nu. \tag{3.7}$$

Digital advertising becomes more efficient over time as A declines. The firm's manufacturing costs in terms of labor for its product line are given by

$$\Xi o_s n^{\eta}/\eta,$$

where $o_s = a_d q_d(n) + a_t(1 - a_d)q_t$ is total output and n is the number of varieties that it is producing. Operating costs increase in the number of varieties, n. As the number of varieties increases so does the organizational cost of selling the product line. A specialized product firm must also incur a fixed cost ϕ in terms of labor. Technological progress in production of specialized products occurs Ξ when increases over time.

3.2.1 Profit Maximization Problem

The firm chooses the intensities of digital and traditional advertising, a_t and a_d, the number of varieties, n, and its price, p, to maximize its profits, π. Its maximization problem reads

$$\begin{aligned}\Pi = \max_{a_d,a_t,n,p} \Big\{ & p a_d \sigma_d(n)[(1-\theta)\hat{y}/p]^{1/\kappa} + p a_t(1-a_d)\sigma_t[(1-\theta)\hat{y}/p]^{1/\kappa} \\ & - wAa_d^{\zeta}/\zeta - wBa_t^{\nu}/\nu \\ & - w\Xi\{a_d\sigma_d(n)[(1-\theta)\hat{y}/p]^{1/\kappa} + a_t(1-a_d)\sigma_t[(1-\theta)\hat{y}/p]^{1/\kappa}\}n^{\eta}/\eta - w\phi \Big\}. \end{aligned} \tag{3.8}$$

The first line in the maximization problem is the revenue the firm realizes from its sales. The second line is the cost of advertising, while the the last line is its manufacturing costs.[15] The maximal level of profits earned by the firm is Π. In equilibrium, firms will keep entering with their own unique product lines until this is driven down to zero. The solution to the firm's problem is presented in Section C.

3.3 Generic Goods

Generic goods firms are perfectly competitive. Generic goods are produced according to the production function

$$o_g = xl^{\alpha},$$

where o_g is output and l is the amount of labor hired. Firms hire labor up to the point where the marginal product of labor equals the wage rate so that

$$w = \alpha x l^{\alpha-1}. \tag{3.9}$$

The demand for labor by the generic goods sector therefore reads

$$l = (\frac{\alpha x}{w})^{1/(1-\alpha)}.$$

Think about generic goods as using a fixed factor, say land. There is one unit of this fixed factor in economy. The profits accruing from this fixed factor, $(1-\alpha)xl^{\alpha}$, are rebated back to consumers. The productivity factor x may increase over time due to technological progress.

[15] What happened to n in the above maximization problem? The firm is selling n varieties. But, each variety spans an arc length of $1/n$ so the sales for a variety should be multiplied by $1/n$. As a result total sales should be multiplied by $n \times (1/n) = 1$; hence, n disappears.

3.4 Equilibrium

The focus is on a static symmetric equilibrium. While individuals consume different varieties, in different amounts, from different product lines, they all have the same distribution of consumption over varieties. To understand this, think about ordering variety consumptions from the lowest to the highest amount. These quantities will span the interval $[\sigma(0.5)[(1-\theta)\hat{y}/p]^{1/\kappa}, \sigma(0)[(1-\theta)\hat{y}/p]^{1/\kappa}]$ with no holes. There will be a mass of varieties at each point on this interval. Take combinations of quantities and masses to form a distribution over variety consumption quantities. While the varieties at each point will differ across consumer, this distribution will be the same for all consumers. The cardinality of the set of products lines for which a variety is consumed is the same for all consumers. That is, the number of product lines consumed by a person, M_i, is given by $M_i = M = |\mathcal{M}_i|$ for all i. Similarly, while firms sell different quantities, of different varieties, to different customers, they all have the same quantity sold distribution over customers.

The number of product lines consumed by individuals, M, is less than the number of specialized firms, N. Denote by M_d the number of product lines consumed by individuals matched through digital ads and M_t the number of product lines consumed by individuals matched through traditional ads. It transpires that

$$M_d = Na_d \text{ and } M_t = Na_t(1-a_d), \tag{3.10}$$

which gives

$$M = M_d + M_t = N[a_d + a_t(1-a_d)], \tag{3.11}$$

where a_d is the probability of a customer receiving a digital ad and $a_t(1-a_d)$ is the odds of getting a traditional ad and no digital ad.

By substituting (3.4) into (3.3), it is easy to see that in a symmetric equilibrium

$$\hat{y} \equiv \frac{y}{\theta + (1-\theta)^{1/\kappa}\,(\hat{y}/p)^{(1-\kappa)/\kappa}\int_{j\in\mathcal{M}} S(j)dj}.$$

Now, the fractions of specialized products purchases arising from digital and traditional advertising are $a_d/[a_d+a_t(1-a_d)]$ and $a_t(1-a_d)/[a_d+a_t(1-a_d)]$, while the average taste factors for digital advertising and traditional advertising are $\sigma_d(n)$ and σ_t. Therefore, the above expression can be rewritten as

$$\hat{y} \equiv \frac{y}{\theta + (1-\theta)^{1/\kappa} (\hat{y}/p)^{(1-\kappa)/\kappa} M[a_d\sigma_d(n) + a_t(1-a_d)\sigma_t]/[a_d + a_t(1-a_d)]}$$
$$= \frac{y}{\theta + (1-\theta)^{1/\kappa} (\hat{y}/p)^{(1-\kappa)/\kappa}[M_d\sigma_d(n) + M_t\sigma_t]}, \text{ using (3.10) and (3.11).} \tag{3.12}$$

The labor market must clear. Recall that an individual inelastically supplies one unit of labor. The labor-market-clearing condition is

$$N\{[a_d\sigma_d(n) + a_t(1-a_d)\sigma_t]\left[\frac{(1-\theta)\,\hat{y}}{p}\right]^{1/\kappa} \Xi\frac{n^\eta}{\eta} + \phi + A\frac{a_d^\zeta}{\zeta} + B\frac{a_t^\nu}{\nu}\}$$
$$+ (\frac{\alpha x}{w})^{1/(1-\alpha)} = 1. \tag{3.13}$$

The first line of the above expression is the labor hired by the N specialized products producing firms. This is distributed over operating costs, the fixed entry cost, and the costs of digital and traditional advertising. The term $[a_d\sigma_d(n) + a_t(1-a_d)\sigma_t](1-\theta)[\hat{y}/p]^{1/\kappa}$ is the physical quantity of specialized products sold by a firm. The left hand side of the second line is the amount of labor hired by firms in the generic goods producing sector. The sum of labor hired by specialized and generic goods producing firms must sum to one, or the right hand side of the second line.

Finally, since there is free entry into the specialized product sector each firm will earn zero profits so that

$$\Pi = 0, \text{ cf. (3.8).} \tag{3.14}$$

This free-entry condition regulates the number of specialized products firms, N. The consumer does earn profits from generic goods production in the amount

$$\pi = (1-\alpha)xl^\alpha = (1-\alpha)x(\frac{\alpha x}{w})^{\alpha/(1-\alpha)}. \tag{3.15}$$

Definition. (Equilibrium) A symmetric equilibrium consists of a solution for: a representative individual's consumption of generic goods, c; consumption of specialized products, $\{q(j)\}_{j=1}^{M}$; the intensities of digital and traditional advertising, a_d and a_t; the number of product lines consumed by a person, $M_d, M_t,$ and M; the number of varieties, n; the price of a variety, p; the number of product lines sold, N; the wage rate, w; and profits, π. These allocations are determined such that:

1. Given prices, p, profits, π, wages, w, and the consumption set, $\mathcal{M}$, the consumers solve problem (3.1). This determines c and $\{q(j)\}_{j=1}^{M}$ where $M = |\mathcal{M}|$ and $\hat{y}$ is

determined by (3.12).

2. Given w and $\hat{y}$, specialized product firms solve problem (3.8), yielding a solution for a_d, a_t, n, p, and Π.

3. Wages, w, are given by (3.9).

4. The profits accrue to a consumer, π, in line with (3.15).

5. The representative consumer's income reads $y = w + \pi$.

6. The number of product lines purchased by a consumer, $M_d, M_t,$ and M, are given by (3.10) and (3.11).

7. The labor market clears in accordance with (3.13).

8. The free-entry condition (3.14) holds.

4 Calibrating the Model to US Data

The analysis focuses on two years, 1995 and 2015. There are three sources of technological progress in the analysis. First, the cost of digital advertising falls, as reflected by a decline in A. Second, generic goods production becomes more efficient, which is captured by an increase in x. Third, the operating cost for specialized products production declines or there is a drop in Ξ.

The model has 14 parameters to determine. Three of these parameters are set exogenously. The inverse of the price elasticity of demand, κ, is set at 0.2, which corresponds to an average markup of 1.25 within the specialized sector in both 1995 and 2015. The generic firm's productivity in 1995 is normalized to one; i.e., $x_{1995} = 1.0$. The value for 2015 is set to match the accumulated growth of per-capita GDP per capita in the United States between 1995 and 2015. This yielded $x_{2015} = 1.978$. The other parameters are selected to match some data targets that are discussed now.

4.1 Data Targets

There are 6 categories of data targets. Some of these categories have two observations, others just one. They are:

A Data Appendix

A.1 Firm-level advertising and varieties data set

A.1.1 Kantar Data Coverage

Kantar Media collects ads placed in different media channels. Media coverage grows over time, mainly because new media channels are created (e.g., internet search, mobile web), and certain platforms within those media become more important (e.g., new TV stations, new high-traffic websites). Media coverage increases over time, and the media channels covered are: network/cable/syndication/spot TV ('95); magazines/Sunday magazines ('95); national ('95)/local ('99) newspapers; network ('00)/national spot ('95) radio; outdoor ('95); internet display ('01); internet search ('10); online video ('13); mobile web ('15). The numbers refer to the year the data becomes available. For TV ads, Kantar Media now monitors 8 networks, 92 cable TV networks, 1,058 spot TV stations in all 210 US designated market areas (DMA), and 64 syndicators. For radio ads, 5 radio networks are monitored, and ten representation firms are reporting for national spot radio in 205 markets. For magazines, 136 national consumer magazines (including geographic and demographic editions), 31 local magazines, and national Sunday magazines (American Profile, NY Times Magazine, Parade, Relish, Spry, T Magazine) are monitored. For newspapers, WSJ, NY Times, USA Today, and 128 local newspapers (including Sunday Supplements and free-standing inserts) are watched. Outdoor ads are derived from surveying out-of-home advertising, including billboards, bulletins, painted walls, transit/bus shelters, in-store displays, convenience stores, shopping malls, airports, taxi displays, and truck/mobile advertising in 412 markets, mapped to the top 188 DMAs. In terms of internet display ads, the company uses a spider/bot technology, operating in a standard browser environment, to systematically collect internet display advertising (ad creatives, occurrences, impressions, and spend) on over 4,200 main domains, subdomains, and content pages. To obtain internet paid search ad expenditures, information on ad creatives, spend, keywords, and clicks from 20,000 URLs (Google US) are collected. In addition, ad creatives, occurrences, impressions, and spend from 2,430 mobile sites are collected, too. Figure A.1 shows total advertising expenditure in Kantar by media channel. Newspapers and magazines show the steepest decline. TV spending is growing but at a declining pace, while digital ads expenditure is on the rise.

Figure A.2 compares aggregate ad expenditure derived from Kantar to other government-

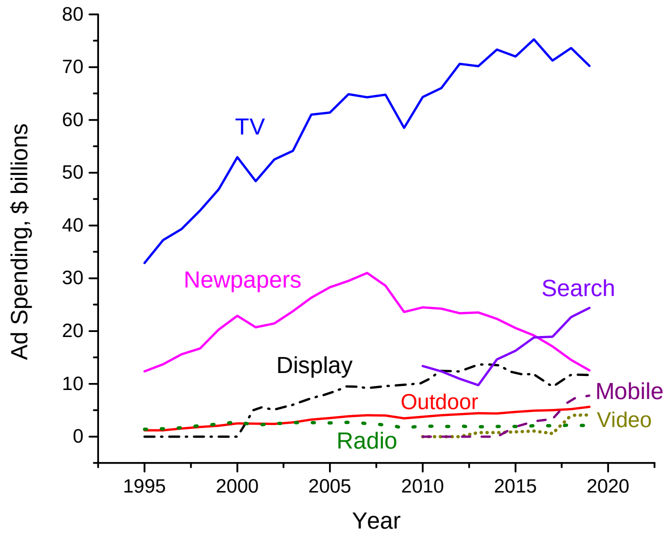

Figure A.1: Advertising Spending by Media Channel

Note: Total advertising spending by type of media channel. AdSpend data from Kantar Media, 1995-2019.

based statistics. As seen, official estimates of advertising expenditure vary based on the data source: the US Census or IRS. The Census estimates are revenue-based and are derived from the US Census Annual Survey where establishments report their revenues. The IRS estimates are cost-based and come from the advertising deductions reported on the annual tax returns filed by companies. The figure also plots the advertising series compiled by Robert J. Coen. Coen has been compiling and publishing high-quality historical advertising data that has been widely used in official government reports. Coen collected data from private sources, such as various companies, bureaus, publishers, and advertising associations. As seen, Coen's series is consistent with the IRS and Census-based estimates and are very close to the cost-based estimates from the IRS. The Census, IRS, and Coen series are discussed and shared by Douglas Galbi on his *Purplemotes* blog.[16] Figure A.2 plots these series together with the series assembled here from Kantar data for the overlapping time period in 1998-2006. Kantar data amount to 40% to 51% of aggregate advertising expenditure estimates from the US Census over time, and for 30% to 36% of aggregate expenditure estimates from IRS.

[16] https://www.purplemotes.net/2009/05/10/.

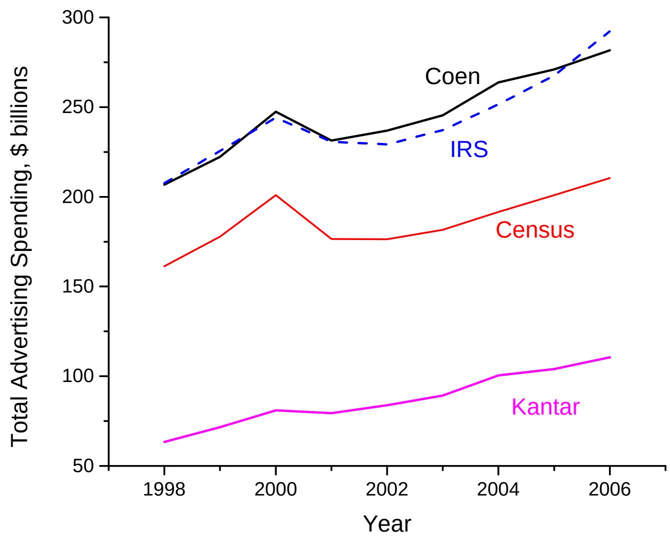

Figure A.2: Kantar Advertising Spending relative to the Aggregate US Statistics

Note: Total advertising spending in AdSpend by Kantar Media compared to the aggregate advertising expenditure estimates from the US Census, the IRS, and Coen. The Census series comes from the US Census Annual Survey; the IRS series is based on data from business deductions of ad expenses; and the Coen series is compiled by Robert J. Coen from various private sources. The series are plotted for the overlapping time period in 1998-2006.

A.1.2 Product Varieties in Kantar

Kantar data provide information on the number of advertised product varieties, but this number is lower than all the varieties companies offer in the market for two reasons. First, Kantar data does not capture all of the varieties advertised in the economy. Nevertheless, the above description of the media coverage makes it clear that almost all advertised *consumer products* are captured at the extensive margin, while the full intensive margin of ad spending on those products is not.[17] Data has lower coverage for those sectors that are not the focus of the current analysis: the data miss business-to-business advertising and have lower coverage for ads not related to consumer products, such as ads by various service providers, entertainment, government, and education.

Second, even if advertised product varieties are well captured, not all the varieties (products, brands) that firms sell are advertised. In what follows, it is shown that the number of advertised product varieties is positively correlated with the number of all varieties sold, supporting the use of Kantar-based variety measures as a proxy for all varieties offered by firms. To evaluate this correlation between varieties and

[17]Still, very niche consumer products advertised on unpopular websites or very specialized media outlets are not captured along the extensive margin.

Table A.1: Relationship Between Varieties and Advertised Varieties

	Advertised Product Varieties							
	———Levels———				———Changes Δ———			
	Brands	Brands	ProdTypes	ProdTypes	Brands	Brands	ProdTypes	ProdTypes
Brands	0.379***		0.249***		0.113***		0.067***	
	(0.011)		(0.008)		(0.023)		(0.017)	
Barcodes		0.217***		0.147***		0.061***		0.040***
		(0.007)		(0.005)		(0.015)		(0.011)
R^2	6,506	6,506	6,506	6,506	2,280	2,280	2,280	2,280
Observations	0.147	0.118	0.138	0.117	0.010	0.008	0.007	0.006

Note: Regression of the log number of advertised product varieties on the log number of all product varieties by firms in the CPG sector for the period 2010-2015. The measures of advertised product varieties (brands and product types) are available on Ad Intel; the measures of product varieties (brands and barcodes) are available on Nielsen RMS. The left panel uses cross-sectional averages over 2010-2015; the right panel uses log changes between 2010 and 2015. Standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

advertised varieties, data is needed that contain both advertised and non-advertised product varieties by firms. RMS scanner data from Nielsen is used and matched with Ad Intel Data (covering 2010-2015) from the Kilts-Nielsen Data Center using a firm name procedure as in Argente et al. (2021a). As a result, for each firm and year for the period 2010-2015, measures are obtained for all varieties (measured in RMS as the number of distinct "barcodes" and "brands" in RMS) and advertised varieties (measured in AdIntel as the number of distinct "brands" and "product-types" advertised in Ad Intel).

Table A.1 shows association between the number of product varieties from RMS and advertised product varieties from Ad Intel. The left panel uses cross-sectional variation by regressing each advertised measure of varieties (in logs) on the measure of sold varieties. You can see that cross-sectional correlation is positive, with elasticities ranging from 0.15 to 0.38. The right panel reports regressions in log changes between these 2010 and 2015. Again, you can see that the number of varieties advertised grows when the number of varieties sold by firms increases.

A.1.3 Matching Kantar to NETS

The Kantar data is combined with data on firms' employment and sales from NETS. NETS provides establishment-level longitudinal microdata covering at least three quarters of all US private sector employment for the period 1989-2017 (Barnatchez, Crane and Decker, 2017).

Matching Kantar to NETS involves defining relevant company names in each data and then using name-matching routines to link names across the two data sets. In Kantar, each product is associated with various company name variables: ultimate owner, parent name, subsidiary name, advertiser. In many cases, these names coincide, but whenever they do not the following strategy is adopted. Take the ultimate owner name as the primary company name for the advertised product. If the ultimate owner is missing, use the parent name. If this is missing, harness the subsidiary or an advertiser's name as the company name. Next choose relevant company names from NETS. NETS employment and sales data is at the establishment level, not the firm level. The data set contains the establishment id, establishment name, and the ultimate headquarter id. In NETS 97% of firms are associated with only one establishment id and one establishment name. For multi-establishment firms, the median number of establishments is nine, while the median number of establishments with different names is two. Since matching is based on company names, the preferred company definition is based on the establishment name. Hence, the aggregate employment and sales of different establishments that fall under the same establishment name are aggregated.[18] In the last step, clean the company names from Kantar and NETS using the company name cleaning routines in Argente, Baslandze, Hanley and Moreira (2020) and perform an exact name matching on the cleaned company names.

In 2015, 53% of companies with product-related ads in Kantar match to NETS (corresponding to 35,552 unique firms and 52,462 unique observations). The Kantar-NETS matched data set provides information about the number of product varieties advertised (products, brands, sub-brands), product categories operated (industry, major, subcategory), traditional and digital ads expenditures, and employment and sales for each company during the years with overlapping coverage in Kantar and NETS. Table A.2 summarizes the data.

A.2 County-level internet and varieties data set

A.2.1 RMS Nielsen

Nielsen RMS raw data from 2006 to 2020 is harnessed [details on the data set are available in Argente et al. (2021b)]. For each county $\times$ year $\times$ product category (referred to as the product module in the original data set), the following variables are computed

[18] Alternative matching strategies are entertained with aggregated sales and employment at the headquarter level in NETS. In Kantar ad expenditures of different company names that fell under the same headquarter identifier are aggregated and then matched with NETS. Although the resulting matches are not very different, the match based on the establishment names is cleaner, so this is adopted as the baseline match.

Table A.2: Additional Summary Statistics of the Data

	Mean	Median	St.Dev
FIRM × YEAR LEVEL - KANTAR MATCHED NETS			
Sales ($1,000s)	131,448	2,895	1,494,346
Employment	607	23	5,875
Digital ad spending ($1,000s)	223.7	0.15	3,170
Traditional ad spending ($1,000s)	3,084.0	3.20	41,228
Total ad spending ($1,000s)	3,235.0	14.60	42,039

Note: Summary statistics for the two data sets. The firm × year data set uses Kantar and include ads only for products. Ads are excluded related to services and amusement, retail (store promotions), automotive dealers, financial, government/politics/organizations, schools, restaurants, hotels, and other services, as well as general ads about corporate promotions and recruiting.

capturing product varieties (from detailed to aggregated):

- distinct number of barcodes (the combination of UPC + UPC version),
- distinct number of combinations of characteristics as in Kaplan and Menzio (2015)[gr8 (firm × brand × same observable characteristics × UPCdesc), gr5 (firm × brand × same observable characteristics), gr4 (firm × brand and same observable characteristics, except size)],
- distinct number of brands [referred as gr13=group(firm × brand × module)],
- distinct number of firms [this is not provided by Nielsen, coming from GS1 company data merged into RMS. referred as gr14=group(firm × module)].

The procedure to create product varieties variables involves two steps. First, build a products data set at the barcode level with information on product characteristics (e.g. size, brand, firm ownership,...). The final products data set has 1,966,044 observations, and 86 variables from Nielsen and GS1data.

Second, use the movement files to collapse by county × module × year, while counting distinct varieties under the different definitions, as well as total sales and quantity. The location of the stores is used to determine the location of the varieties sold. Compute totals across all stores and observations (unbalanced) and across a balanced set of stores in the data throughout 2006–2020.

In second step, construct information on chains (multi-state and national) and firms (headquarters state, multi-state firm, national firms), and only count varieties for those sets. The interest is to measure varieties that exist because of consumers' exposure to local internet conditions (ideally to digital advertising). Hence, exclude varieties that are being offered because of store's and firm's potential exposure to local internet conditions (assuming these are correlated with household internet access). A chain/firm's "degree of exposure" to local conditions is measured based on whether they sell in multiple states. The degree of exposure is captured with four statistics: multi-state, above the median number of states, above the 75th percentile in the number of states, and above the 90th percentile in the number of states. For firms, another measure of exposure to internet is based on headquarters location.

RMS Nielsen for US maps of varieties

A data set is built measuring product varieties at the county $\times$ year level, where products are aggregated across product categories. Let y_{jlt} be a measure of varieties in module j and location l at time t, and Y_{j0} be varieties in module j across the entire United States in the baseline year 0. Then, define

$$s_{lt}^{(1)} = \sum_{j=1}^{J} \left(\omega_j \frac{y_{jlt}}{Y_{j0}} \right), s_{lt}^{(2)} = \sum_{j=1}^{J} \left(\omega_{jl} \frac{y_{jlt}}{Y_{j0}} \right), s_{lt}^{(3)} = \sum_{j=1}^{J} \left(\omega_{jlt} \frac{y_{jlt}}{Y_{j0}} \right),$$

where ω_j is the revenue share of module j across all years and regions, ω_{jl} is the revenue share of module j in county l across all years, ω_{jlt} is the actual revenue share in that county $\times$ year $\times$ module. The plots use $s_{lt}^{(2)}$ and the patterns are similar across versions.

A.2.2 Federal Communications Commission (FCC)

The baseline internet data come from the FCC Form 477. This is a form sent to internet service providers that asks for reporting on items including the types of services they offer, speeds, and subscribership. The amount of data collected by Form 477 varies over time, with reforms in 2004, 2008, and 2014, all of which increased the level of detail in the reported data (geographically and in terms of speed tiers).

Two data sets are considered. The first contains data on the number of residential fixed connections (i.e., not mobile connections) per every 1,000 housing units.[19] This information is reported at the census tract level (73k in US) by year for every year from 2008 through 2018 (as of February 2022). The numbers per 1,000 are not raw but

[19] https://www.fcc.gov/general/fcc-form-477-additional-data

rather reported in bins of 0 to 200, 200 to 400 and so on. This variable is used in the baseline regressions. Its key advantage is a consistent way of reporting throughout the period 2008-2018. A Freedom of Information Act (FOIA) request was made to obtain additional information on the number of residential fixed connections per every 1,000 housing units for different maximum speed levels. This is used in robustness exercises.

The second data set is "Fixed Broadband Deployment Data" on the advertised speeds (for residential and commercial use) offered by providers.[20] This is at the technology-provider-census block level (11 million blocks in the United States), with technology falling under categorizations made by the FCC (different types of DSL, cable, satellite, fiber, et cetera). Collection of raw speed data was initiated in 2014 and so the availability of this data set begins in 2014 and has (as of Feb 2022) been released through 2020. This information is only used for robustness exercises.

Census blocks are aggregated into counties. For each county the mean and median, as well as the mean and median weighted by number of housing units in 2010, are computed (using housing units data from Census Bureau relationship files). The baseline measure is the average number of residential fixed connections, weighted by number of housing units in 2010. The alternative measures are used for robustness. Additionally, the Census Bureau file is used to harmonize data using pre-2010 census boundaries to 2010 census boundaries.

A.2.3 National Lightning Database Network (NLDN).

The lightning data originates from the NLDN, an organization under the National Oceanic and Atmospheric Administration (NOAA). NLDN collects data on lightning strikes via ground-based sensing stations across the United States. Data is sourced from the County and State Summaries released on this webpage. The data begins in 1986 and is available for every year until 2012 (inclusive). It records the number of lightning strikes by county for every individual day of the year. This data is combined with data on the size of US counties from the Census Bureau to get measures of lightning strikes per square mile per year.

Through negotiations between Sean Chen and Steve Ansari from the NOAA, a data set was obtained at the county-level containing lightning strikes from 1986-2020. This data has been downloaded to /Data/NLDN/2020boundaries. As the name suggests, the data uses 2020 county boundaries, which are adapted for the analysis.

[20] https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477

A.2.4 Additional data sets.

Data is sourced from the Bureau of Economic Analysis (BEA) regional economic accounts to obtain various measures on personal income, population, and wages at the county level for every year in 1969-2020. Since population by county is not collected annually, BEA makes its own imputations.

B Additional Empirical Results

Table B.1: Product Variety and Digital Ads, Product-Category Level. Robustness

	Log Products			*Log Brands*		
Panel A	subcat.	major	industry	subcat.	major	industry
Log Digital Ads	0.022***	0.017***	0.039***	0.016***	0.016***	0.030***
	(0.001)	(0.003)	(0.011)	(0.001)	(0.003)	(0.008)
R^2	0.974	0.988	0.992	0.981	0.991	0.993
Observations	13,123	3,215	733	13,123	3,215	733
Panel B						
	Log Products			*Log Brands*		
	subcat.	major	industry	subcat.	major	industry
Digital Ads Ratio	0.011***	0.010***	0.032***	0.009***	0.012***	0.026***
	(0.001)	(0.003)	(0.011)	(0.001)	(0.003)	(0.008)
R^2	0.981	0.989	0.992	0.985	0.991	0.993
Observations	15,447	3,236	733	15,447	3,236	733

Notes: Regressions of product varieties on digital-ads spending in product categories over time controlling for log traditional-ads spending in Panel A. Regressions of product varieties on the ratio of digital-ads spending to total ads spending in Panel B. All regressions control for log number of firms in product categories over time, product category, and year fixed effects. Product variety: products and brands. Product categories: subcategory, major, and industry. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.10.

Table B.2: Product Variety and Digital Ads, Firm Level. Robustness

Panel A				
	Log Products		*Log Brands*	
	Cross-sectional	Within-firms	Cross-sectional	Within-firms
Log Digital Ads	0.133***	0.058***	0.130***	0.044***
	(0.002)	(0.002)	(0.002)	(0.002)
R^2	0.552	0.913	0.498	0.901
Observations	26,958	25,159	26,958	25,159
Panel B				
	Log Products		*Log Brands*	
	Cross-sectional	Within-firms	Cross-sectional	Within-firms
Digital Ads Ratio	0.018***	0.008***	0.041***	0.013***
	(0.002)	(0.002)	(0.002)	(0.001)
R^2	0.551	0.889	0.475	0.872
Observations	67,820	64,469	67,820	64,469

Notes: Regressions of product varieties on digital-ads spending in firms over time controlling for log traditional-ads spending in Panel A. Regressions of product varieties on the ratio of digital-ads spending to total ads spending in Panel B. All regressions control for firm's log employment, year fixed effects, and product category/firm fixed effects in the "Cross-firms"/"Within-firms" columns, respectively. Product variety: products and brands. Product category: subcategory. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.10.

Table B.3: Variation of product and brand varieties across regions

Panel A: Log Products	Variance	R-Squared			
All products	2.869	0.593	0.815	0.720	0.815
Only multi-state chains	2.863	0.594	0.815	0.718	0.815
Only many states chains	2.873	0.597	0.808	0.707	0.808
Only multi-state firms & no HQ location	2.843	0.595	0.811	0.716	0.811
Only many states firms & no HQ location	2.824	0.598	0.816	0.715	0.815
Controls	-	No	No	Yes	Yes
Year FE	-	Yes	Yes	Yes	Yes
Category FE	-	Yes	Yes	Yes	Yes
County FE	-	No	Yes	No	Yes
Panel B: Log Brands	Variance	R-Squared			
All Brands	1.720	0.595	0.801	0.719	0.801
Only multi-state chains	1.715	0.597	0.800	0.717	0.800
Only many states chains	1.712	0.605	0.794	0.707	0.794
Only multi-state firms & no HQ location	1.698	0.599	0.798	0.715	0.798
Only many states firms & no HQ location	1.622	0.608	0.801	0.716	0.801
Controls	-	No	No	Yes	Yes
Year FE	-	Yes	Yes	Yes	Yes
Category FE	-	Yes	Yes	Yes	Yes
County FE	-	No	Yes	No	Yes

Notes: The variance and R^2 of regressions for the main dependent variables used in the paper. The specifications control for population (in logs) and income per capita (in logs) as well as the fixed-effects.

Table B.4: First Stage Results

	Household Internet		
	(1)	(2)	(3)
Lightning Strikes	-0.018***	-0.011***	0.002***
	(0.000)	(0.000)	(0.000)
Observations (Thousands)	9,905	9,905	9,884
Year × Category FE	Yes	Yes	Yes
Year × County Controls	No	Yes	Yes
County × Category FE	No	No	Yes

Notes: The estimated coefficients for regression (2.1). The dependent variable is household internet access. The variables are described in Section 2.1. The year × county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at county × year level and shown in parentheses. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

Table B.5: Household internet access and product varieties: OLS results

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(4)	(5)	(6)
Household Internet	0.586***	0.174***	0.005	0.446***	0.127***	0.005*
	(0.008)	(0.009)	(0.004)	(0.005)	(0.007)	(0.003)
Observations (thousands)	9,905	9,905	9,884	9,905	9,905	9,884
R^2	0.625	0.679	0.961	0.624	0.681	0.949
Year × Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × County Controls	No	Yes	Yes	No	Yes	Yes
County × Category FE	No	No	Yes	No	No	Yes

Notes: OLS results. The dependent variable in columns 1--3 is barcodes (in logs) and in column 4–5 is brands (in logs). The variables are described in Section 2.1. The year × county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at county × year level and shown in parentheses. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

Table B.6: Household Internet and Product Varieties: Alternative definitions of products varieties

	Log Products (1)	Log Agg1 (2)	Log Agg2 (3)	Log Agg3 (4)	Log Brands (5)	Log Firms (6)
Household Internet	0.268*	0.261*	0.304*	0.363**	0.404***	0.424***
	(0.158)	(0.157)	(0.158)	(0.160)	(0.147)	(0.142)
Observations (thousands)	9,884	9,884	9,884	9,884	9,884	9,857
Year × Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × County Controls	Yes	Yes	Yes	Yes	Yes	Yes
County × Category FE	Yes	Yes	Yes	Yes	Yes	Yes
1st stage F-stat	15.2	15.2	15.2	15.2	15.2	15.2

Notes: The estimated coefficients for regression (2.2) for all definitions of varieties. Columns (2), (3), (4) and (6) report estimates of alternative definitions of product variety. Columns (1) and (5) corresponds to the estimates in Table 2.3, for reference. The variables are described in Section 2.1. The year × county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at county × year level and shown in parentheses. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

Table B.7: Household internet access and product varieties: Lagged

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(4)	(5)	(6)
Household Internet	1.104***	1.399***	0.746	0.825***	1.036***	1.595
	(0.040)	(0.075)	(1.188)	(0.029)	(0.055)	(2.165)
Observations (thousands)	9,587	9,587	9,553	9,587	9,587	9,553
Year × Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × County Controls	No	Yes	Yes	No	Yes	Yes
County × Category FE	No	No	Yes	No	No	Yes
1st stage F-stat	671.0	478.0	0.6	671.0	478.0	0.6

Notes: The estimated coefficients for regression (2.2) with the variable for household internet lagged one year. The dependent variable in columns 1--3 is barcodes (in logs) and in column 4–5 is brands (in logs). The variables are described in Section 2.1. The year × county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at county × year level and shown in parentheses. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

Table B.8: Household Internet Access and Product Varieties: Alternative Samples

Panel A: Food and Health & Beauty Products						
	Log Products			*Log Brands*		
Household Internet	0.984***	1.209***	0.002	0.722***	0.878***	0.165
	(0.034)	(0.062)	(0.129)	(0.025)	(0.045)	(0.105)
R^2	0.091	0.060	0.001	0.099	0.086	-0.024
Observations	13,469	13,469	13,445	13,469	13,469	13,445
Panel B: All Product Categories						
	Log Products			*Log Brands*		
Household Internet	0.975***	1.195***	-0.061	0.720***	0.877***	0.120
	(0.033)	(0.058)	(0.120)	(0.024)	(0.043)	(0.094)
R^2	18,661	18,661	18,626	18,661	18,661	18,626
Observations						
Year × Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × County Controls	No	Yes	Yes	No	Yes	Yes
County × Category FE	No	No	Yes	No	No	Yes

Notes: The estimated coefficients of for regression (2.2) for two alternative samples. Panel A uses a selected sample that includes Nielsen RMS products modules of food and health and beauty. Panel B includes all product modules in Nielsen RMS. The variables are described in Section 2.1. The year × county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at county × year level and shown in parentheses. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

Table B.9: Household internet access and product varieties: Unbalanced Sample

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(4)	(5)	(6)
Household Internet	0.661***	0.781***	0.330	0.505***	0.586***	0.422**
	(0.030)	(0.049)	(0.213)	(0.022)	(0.035)	(0.182)
Observations (thousands)	11,555	11,555	11,527	11,555	11,555	11,527
Year $\times$ Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Year $\times$ County Controls	No	Yes	Yes	No	Yes	Yes
County $\times$ Category FE	No	No	Yes	No	No	Yes
1st stage F-stat	837.5	567.9	13.6	837.5	567.9	13.6

Notes: The estimated coefficients for regression (2.2) for all stores in Nielsen RMS (as opposed to the balanced sample of stores used in the benchmark analysis). The dependent variable in columns 1-3 is barcodes (in logs) and in column 4-5 is brands (in logs). The variables are described in Section 2.1. The year $\times$ county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at county $\times$ year level and shown in parentheses. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

Table B.10: Household Internet access and product varieties: Alternative instrument

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(4)	(5)	(6)
Household Internet	1.102***	1.367***	0.893*	0.816***	1.002***	1.253**
	(0.036)	(0.066)	(0.476)	(0.026)	(0.048)	(0.592)
Observations (thousands)	9,905	9,905	9,884	9,905	9,905	9,884
Year $\times$ Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Year $\times$ County Controls	No	Yes	Yes	No	Yes	Yes
County $\times$ Category FE	No	No	Yes	No	No	Yes
1st stage F-stat	894.7	647.9	5.0	894.7	647.9	5.0

Notes: The estimated coefficients for regression (2.2). The dependent variable in columns 1--3 is barcodes (in logs) and in column 4–5 is brands (in logs). The variables are described in Section 2.1. The year $\times$ county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at county $\times$ year level and shown in parentheses. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

Table B.11: Household internet access and product varieties: Robustness for restricted retail chains

	Log Products		*Log Brands*	
	HQ other Multi-state	HQ other National	HQ other Multi-state	HQ other National
Household Internet	0.426**	0.689**	0.394***	0.414***
	(0.187)	(0.284)	(0.177)	(0.281)
Observations (1,000s)	9,844	9,056	9,844	9,056
Year $\times$ Category FE	Yes	Yes	Yes	Yes
Year $\times$ County Controls	Yes	Yes	Yes	Yes
County $\times$ Category FE	Yes	Yes	Yes	Yes
1st stage F-stat	14.6	10.2	14.6	10.2

The estimated coefficients regression (2.2). The dependent variables are either barcodes or brands (in logs) in a county $\times$ year x category sold by retail chain that sell in more than one state (multi-state) or in the top quartile distribution of number of states (national). The year $\times$ county controls are population (in logs) and income per capita (in logs). Standard errors are clustered at the county $\times$ year level and shown in parentheses. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. The ***, **, and * represent statistical significance at 1%, 5%, and 10% levels, respectively.

C Calibration

C.1 The Consumer's Problem

Let $\hat{y}$ be the Lagrange multiplier attached to the budget constraint for the consumer's problem (3.1). The first-order conditions are

$$\theta/c = \hat{y}$$

and

$$(1-\theta)S(j)^{\kappa}q(j)^{-\kappa} = \hat{y}p(j), \text{ for all} j.$$

This implies the solutions (3.2) and (3.4). Multiplying the last equation by $q(j)$ and integrating then gives

$$\int (1-\theta)S(j)^{\kappa}q(j)^{1-\kappa}dj = \hat{y}\int p(j)q(j)dj.$$

Summing this equation with the first one and solving for $\hat{y}$ yields (3.3) or

$$\hat{y} = \frac{y}{\theta + \int (1-\theta)S(j)^{\kappa}q(j)^{1-\kappa}dj}.$$

C.2 The Firm's Problem

The solution to the firm's maximization problem (3.8) is characterized by the first-order conditions for the intensities of digital and traditional advertising, a_d and a_t, the number of varieties, n, and output price, p.

The first-order condition for the intensity of digital advertising equates the marginal cost of digital advertising to the firm's marginal revenue net of unit operating costs,

$$\underbrace{\left(p - w\Xi\frac{n^{\eta}}{\eta}\right)(q_d(n) - a_t q_t)}_{\substack{\text{Marginal revenue net of}\\ \text{marginal operating cost}\\ \text{related with digital ads}}} = \underbrace{wAa_d^{\zeta-1}}_{\substack{\text{Marginal cost of}\\ \text{digital advertising}}}.$$

Similarly, the intensity of traditional advertising satisfies the following

$$\underbrace{\left(p - w\Xi\frac{n^{\eta}}{\eta}\right)(1-a_d)\,q_t}_{\substack{\text{Marginal revenue net of}\\ \text{marginal operating cost}\\ \text{related with traditional ads}}} \quad = \quad \underbrace{wBa_t^{\nu-1}}_{\substack{\text{Marginal cost of}\\ \text{traditional advertising}}}.$$

For the number of varieties the following must hold

$$\underbrace{\left(p - w\Xi\frac{n^{\eta}}{\eta}\right)a_d\frac{\partial q_d(n)}{\partial n}}_{\substack{\text{Marginal revenue net of}\\ \text{marginal operating cost}\\ \text{related with varieties}}} = \underbrace{w\Xi n^{\eta-1}\left[a_d q_d(n) + a_t(1-a_d)\,q_t\right]}_{\substack{\text{Marginal operating cost}\\ \text{of additional varieties}}},$$

where the marginal effect of the number of varieties produced on consumers' quantity demanded is given by

$$\frac{\partial q_d(n)}{\partial n} = \frac{q_d(n)}{\sigma_d(n)}\frac{0.25\lambda}{n^2}.$$

Last, the first-order condition for output price is characterized by a markup over marginal operating cost,

$$p = \underbrace{\left(\frac{1/\kappa}{1/\kappa - 1}\right)}_{\text{Markup}} \quad \underbrace{w\Xi\frac{n^{\eta}}{\eta}}_{\text{Marginal operating cost}}.$$

References

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J Klenow, and Huiyu Li, "A theory of falling growth and rising rents," 2019.

Andersen, Thomas Barnebeck, Jeanet Bentzen, Carl-Johan Dalgaard, and Pablo Selaya, "Lightning, IT diffusion, and economic growth across US states," *Review of Economics and Statistics*, 2012, *94* (4), 903–924.

Argente, David, Doireann Fitzgerald, Sara Moreira, and Anthony Priolo, "How Do Firms Build Market Share?," 2021.

_ , Munseob Lee, and Sara Moreira, "The Life Cycle of Products: Evidence and Implications," *Available at SSRN 3163195*, 2021.

_ , Salomé Baslandze, Douglas Hanley, and Sara Moreira, "Patents to Products: Product Innovation and Firm Dynamics," FRB Atlanta Working Paper 2020-4, Federal Reserve Bank of Atlanta April 2020.

Bar-Isaac, Heski, Guillermo Caruana, and Vicente Cu ñ at, "Search, Design, and Market Structure," *American Economic Review*, April 2012, *102* (2), 1140–60.

Barnatchez, Keith, Leland D. Crane, and Ryan A. Decker, "An Assessment of the National Establishment Time Series (NETS) Database," Finance and Economics Discussion Series 2017-110, Board of Governors of the Federal Reserve System (U.S.) November 2017.

Brynjolfsson, Erik, Long Chen, and Xijie Gao, "Gains from Product Variety: Evidence from a Large Digital Platform," NBER Working Papers 30802, National Bureau of Economic Research, Inc December 2022.

Cavenaile, Laurent and Pau Roldan-Blanco, "Advertising, Innovation, and Economic Growth," *American Economic Journal: Macroeconomics*, July 2021, *13* (3), 251–303.

_ , Murat Alp Celik, Pau Roldan-Blanco, and Xu Tian, "Style Over Substance? Advertising, Innovation, and Endogenous Market Structure," Working Paper 2022.

Dinlersoz, Emin M. and Mehmet Yorukoglu, "Information and Industry Dynamics," *American Economic Review*, April 2012, *102* (2), 884–913.

Evans, David S., "The Online Advertising Industry: Economics, Evolution, and Privacy," *Journal of Economic Perspectives*, September 2009, *23* (3), 37–60.

Gao, Guodong Gordon and Lorin M. Hitt, "Information Technology and Product Variety: Evidence from Panel Data," in "International Conference on Interaction Sciences" 2004.

Goldfarb, Avi, "What is different about online advertising?," *Review of Industrial Organization*, 2014, *44* (2), 115–129.

Greenwood, Jeremy, Yueyuan Ma, and Mehmet Yorukoglu, ""You Will":A Macroeconomic Analysis of Digital Advertising," Working Paper 28537, National Bureau of Economic Research March 2022.

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya, "3g internet and confidence in government," *The Quarterly Journal of Economics*, 2021, *136* (4), 2533–2613.

Hoberg, Gerard and Gordon M. Phillips, "Scope, Scale and Concentration: The 21st Century Firm," NBER Working Papers 30672, National Bureau of Economic Research, Inc November 2022.

Hoynes, Hilary W. and Diane Whitmore Schanzenbach, "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program," *American Economic Journal: Applied Economics*, October 2009, *1* (4), 109–139.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg, "The Industrial Revolution in Services," 2020.

Kaplan, Greg and Guido Menzio, "The morphology of price dispersion," *International Economic Review*, 2015, *56* (4), 1165–1206.

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard, "Information technology and returns to scale," *Available at SSRN 3458604*, 2018.

Menzio, Guido, "Optimal Product Design: Implications for Competition and Growth under Declining Search Frictions," *Econometrica*, Forthcoming.

Neiman, Brent and Joseph S. Vavra, "The Rise of Niche Consumption," Technical Report, American Economic Journal: Macroeconomics Forthcoming.

Perla, Jesse, “A Model of Product Awareness and Industry Life Cycles,” Working Paper 2019.

Ridder, Maarten De, “Market power and innovation in the intangible economy,” 2019.

Salop, Steven C., “Monopolistic Competition with Outside Goods,” *The Bell Journal of Economics*, 1979, *10* (1), 141–156.