FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

The Expansion of Varieties in the New Age of Advertising

Salomé Baslandze, Jeremy Greenwood, Ricardo Marto, and Sara Moreira

Working Paper 2023-15
September 2023

Abstract: The last decades have seen large improvements in digital advertising technology that allowed firms to better target specific consumer tastes. This research studies the relationship among digital advertising, the rise of varieties, and economic welfare. We develop a model of advertising and varieties where firms choose the intensity of digital ads directed at specific consumers as well as traditional ads that are undirected. The calibrated model shows that improvements in digital advertising have driven the rise in varieties over time. Empirical evidence is presented using detailed micro data on firms' products and advertising choices for the 1995–2015 period. Causal analysis using exogenous variation in consumers' differential access to the internet supports the suggested mechanism.

JEL classification: E13, L15, I31, M37, O14, O31

Key words: digital (directed) advertising, traditional (undirected) advertising, specialization, targeting, internet, varieties, welfare

https://doi.org/10.29338/wp2023-15

The authors are grateful to Tom Hubbard, Ben F. Jones, Joseph Vavra, Galina Vereshchagina, and participants at seminars and conferences for their feedback. They would also like to thank Andrew Ahn, Kai Xiao (Sean) Chen, and Jiawei (Charon) Yang for excellent research assistance. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Salome Baslandze, Federal Reserve Bank of Atlanta, Baslandze.salome@gmail.com; Jeremy Greenwood, University of Pennsylvania; Ricardo Marto, Federal Reserve Bank of St. Louis; or Sara Moreira, Northwestern University.

1 Introduction

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

Advertising has long been an essential tool used by firms. 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. The new age of advertising dawned in 1994 with AT&T's "You Will" campaign that showcased the first digital advertisement. The advent of the internet and technological progress in digital advertising led to a drastic reallocation of spending away from traditional advertising toward digital advertising (right panel). This shift from traditional to digital advertising implied significant improvements in consumer targeting since digital ads, due to the vast amount of information collected on consumer characteristics and their online behavior, allow firms to more precisely infer consumer tastes and target them with products they would like (Goldfarb, 2014). At the same time as spending on digital advertising rose, there was an increase in the number of product varieties tailored to diverse tastes and needs, as displayed in Figure 1.2.[1] This raises the question of whether technological progress in digital advertising, which implied better consumer targeting, played a role in the expansion of varieties.

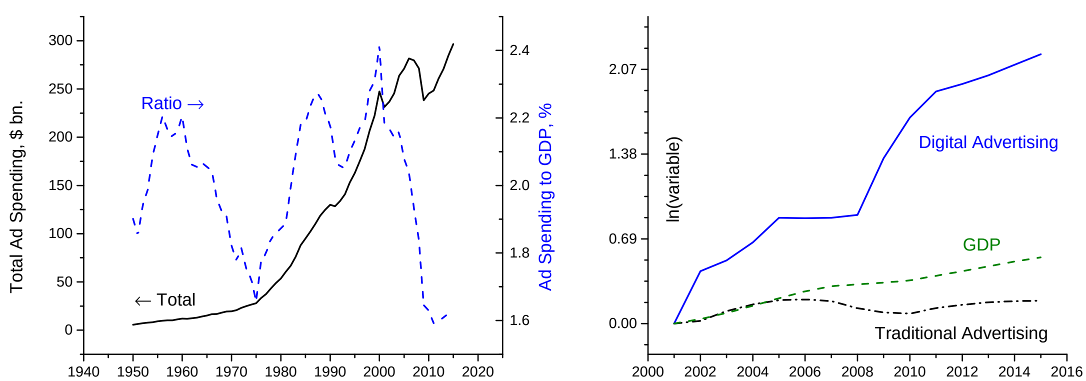

Figure 1.1: Spending on Advertising in the United States.

Note: The left panel shows total spending on advertising in U.S. dollars and as a fraction of GDP, 1950-2015. The right panel shows the breakdown of total spending between digital (internet display, search, online video, and mobile web) and traditional (TV, radio, newspapers, magazines, and outdoor exhibit) advertising, 2001-2015, using three-year moving-averages normalized to zero in 2001. Sources: The advertising data series in the left panel are from Robert J. Coen extended with Statista series after 2007; The series in the right panel are from the Kantar Media's AdSpender.

[1]Think about the wide variety of movie titles, shows, podcasts, and music catering to the diverse tastes, interests, and consumer backgrounds offered on Netflix and Spotify. According to the Food Marketing Institute, a typical supermarket now carries over 40 thousand different items, compared to 9 thousand in the 1970s.

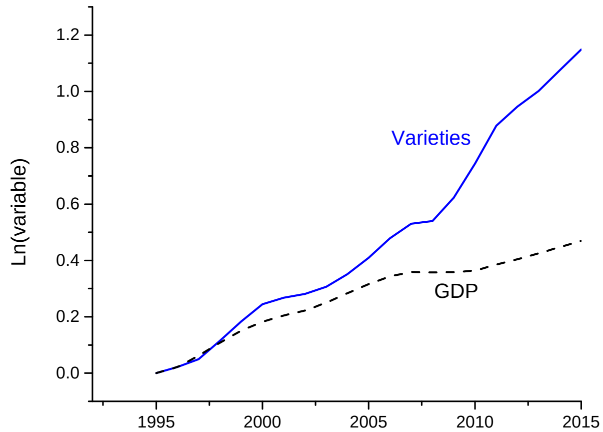

Figure 1.2: Increase in the Number of Varieties, 1995-2015.

Note: The growth of a measure of product varieties in comparison with GDP. Later, it is shown that the patterns are qualitatively similar for various measures of varieties across multiple data sets. Source: Kantar Media.

The impact of the new age of advertising on varieties is addressed in two ways. First, the hypothesis is examined empirically using micro-level data from multiple sources. To start with, Kantar Media AdSpender data is used to construct firm-level data on the number of varieties (products, brands, sub-brands), as well as digital (internet display, internet search, online video, and mobile web) and traditional (TV, radio, newspapers, magazines, and outdoor exhibit) advertising spending over time from 1995 to 2015. In terms of aggregate trends, the number of varieties, defined using multiple definitions, grows over time; digital ads become cheaper; and firms spend increasingly more on digital ads. At the micro level, there is a positive relationship between the growth in varieties and the growth in digital ads. This relationship holds both at the product category level, as well as across and within firms, conditional on traditional advertising spending, fixed effects, and other controls.

To get closer to the causal impact of digital advertising, an empirical approach is adopted that exploits exogenous changes in households' internet access and examines its impact on the number of varieties offered by firms. When households do not use the internet, firms cannot use digital advertising to target consumers' tastes (Evans, 2009; Goldfarb, 2014). As such, households' internet access is a crucial determinant of firms' abilities to target consumers through digital advertising. To obtain exogenous variation in households' internet access, lightning strikes are used as an instrument. The idea is that the frequency of lightning strikes affects the diffusion of digital technologies due to an increase in the expected infrastructure costs associated with voltage spikes and dips (Andersen, Bentzen, Dalgaard and Selaya, 2012; Guriev, Melnikov and Zhuravskaya, 2021). Indeed, the first-stage regression results using households' internet access data from the Federal Communications Commission and lightning strikes data from the National Lightning Database Network confirm the negative relationship between households' internet access and lightning strikes both across US counties and within counties over time. In the second stage, various measures of the number of varieties (barcodes, brands) using Nielsen scanner data on products sold in grocery, drug, and general merchandise stores is regressed on the instrumented households' internet access. The results show a positive relationship between varieties and the instrument, conditional on other controls (e.g., demographics, income, fixed effects) in line with the prediction that the use of digital advertising affects firms' choices of varieties. Further robustness analysis is conducted to lessen the concerns that the mediating channel between internet access and variety decisions works through changes in firms' or retail chains' operation costs.

Second, a model is developed with the goal of outlining a mechanism linking technological progress in digital advertising to the expansion of varieties. The model provides a platform for various quantitative exercises and welfare calculations. In the model, firms produce and sell their own product lines. Within a product line, there are different varieties, each catering to consumers' distinct tastes. 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. If consumers learn about a product line with a traditional ad, they buy a random variety within that product line. Digital advertising is targeted and alerts consumers to the varieties that are most suited to their tastes. If a consumer learns about a product line with a digital ad, they choose a variety closest to their taste within that line. As a result, it is shown theoretically that digital advertising allows for a better match between consumer tastes and product varieties than traditional advertising. A firm chooses the number of varieties it wishes to sell, the price of each variety, and the intensities of both types of advertising. To capture the digital advertising revolution, the cost efficiency of digital advertising is allowed to increase over time. As a firm's ability improves to target consumers more precisely, the demands by customers for specialized varieties that better match their tastes grows, and firms' incentives to create more varieties increase.

The developed model is calibrated to two static equilibria in 1995 and 2015, allowing for the improved efficiency of digital advertising, as well as process innovation in the production of varieties, technological progress in the production of non-specialized (generic) goods, and higher entry costs over time. The model matches a set of stylized facts about advertising for the period 1995 to 2015. Some key targets 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 product lines and the varieties contained within them over the period, and the elasticity of sales with respect to advertising. The estimates from the causal empirical analysis are used to discipline another target, the elasticity of varieties with respect to digital advertising.

The sway of digital advertising on the growth in total varieties (product lines multiplied by varieties within product lines) is then assessed using the calibrated model. Digital advertising increases the number of product lines and varieties available to consumers, explaining 39% of the rise in total varieties over the 20 years. Process innovation in the production of varieties explains 21% of the rise in total varieties in the economy: it increases the number of product lines but has no influence on the number of varieties within a product line. This transpires because process innovation does not affect the ability to obtain a better match between a consumer's tastes and a variety but directly translates into lower prices. The increased efficiency of digital ads and process innovation also synergize with each other, jointly explaining almost all of the growth in total varieties from the data. Changes in entry costs have little effect on the number of varieties as the measured increase in these costs over the period was small. Finally, the implications of technological progress in the production of non-specialized (generic) goods are considered. Such technological progress in the economy drives up wages and increases the price of specialized varieties. Yet, the number of product lines and varieties remains constant. People just consume less of each variety and more generic goods.

The advent of digital advertising results in consumers buying from more product lines and consuming varieties closer to their tastes. This leads to a significant gain in welfare (an equivalent variation of 1.25% measured in terms of generic goods consumption). Without digital advertising, firms would have to rely on traditional ads. Traditional advertising lowers the quality of the matches between product varieties and consumers' tastes. As a consequence, consumer welfare decreases. Additionally, this lowers firms' incentives to produce different varieties, thereby reducing the number of products available for consumption, further reducing the welfare of consumers. These welfare estimates do not consider other channels through which the rise in digital advertising could negatively impact consumers, such as the effect on firms' market power, competition, and concentration. The analysis here also abstracts from the potentially persuasive role of advertising and focuses on its informative role. These other effects of advertising are not specific to digital ads and have been studied in the prior literature.[2] The focus of this paper is on a new channel through which digital advertising – because of better targeting – has implications for the product varieties offered to consumers.

Literature Review

Evans (2009) and Goldfarb (2014) make the case that digital advertising is fundamentally

[2] See Bagwell (2007) for an extensive review of the economic analysis of advertising.

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, which they use 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. The focus of this paper is to highlight a novel implication of better targeting: firms are encouraged to produce more varieties.

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 within 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 the well-known Salop (1979) 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. Extending this framework, Dinlersoz, Goldschlag, Yorukoglu and Zolas (2023) incorporate trademark decisions. Cavenaile, Celik, Perla and Roldan-Blanco (2023) study how increased product awareness through information diffusion via ads impacts competition. Although digital ads improve consumer-firm matches, they also allow firms to obtain higher market power because of consumer segmentation. In contrast to these papers, the focus of this study is on how digital advertising facilitates the creation of new specialized product varieties.

Other recent papers on advertising and innovation are Cavenaile and Roldan-Blanco (2021) and Cavenaile, Celik, Roldan-Blanco and Tian (2022). These papers develop a rich framework where firms make advertising and R&D choices and where the market structure is endogenous. Their mechanism differs significantly from the model presented here. They do not distinguish between digital and traditional advertising. While in these papers advertising works as a demand shifter by increasing firms' effective product qualities, here, digital advertising plays an information role and facilitates better matching between consumer tastes and products. In addition, it is shown that digital advertising is conducive to the development of specialized varieties.

Several related studies examine the consequences of the reduction in search frictions for product markets. In Bar-Isaac, Caruana and Cuñat (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 (2023) 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 while taking into account its cost, which parallels the exogenous reductions of search frictions in these other papers. The framework microfounds the notion of targeting as improving the match between consumer tastes and specialized varieties. In contrast to these models, which have single-product firms choosing a niche or generic design, here the number of different varieties offered by multi-product firms grows with targeting, consistent with the data. Last, Ma (2022) analyzes both theoretically and empirically how a reduction in search costs and better patent protection have resulted in more specialization in firms' production, while Shen (2023) shows how the declining cost of marketing since the 1990s contributed to increasing concentration and declining productivity growth.

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 a 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, Chen and Gao (2022) document an increase in book titles on the largest digital platform in China indicating an increased consumption of more niche titles. Gao and Hitt (2004) report a positive relationship between product varieties measured by trademarks and the use of information technology by firms. While the empirical analysis in this study also shows an increase in varieties produced by firms, the results presented here also provide evidence that this increase in varieties is closely linked to the proliferation 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 varieties? The empirical analysis proceeds along two distinct tracks, which support the hypothesis. In

Section 2.1, firm-level panel data on digital advertising and product varieties is used. The results indicate that there is a positive relationship between the growth in digital advertising spending and the growth in varieties, both at the firm level and at more aggregated levels. In Section 2.2, spatial heterogeneity in household internet access and varieties is used to provide causal evidence on the link between digital advertising and the varieties sold by firms using an instrumental variable (IV) strategy.

2.1 Relationship Between Digital Advertising and Varieties

2.1.1 Data: Firm-Level Advertising and Varieties

The construction of a data set on digital advertising and varieties is briefly outlined here.[3] 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.[4] The different media channels are aggregated into *digital ads* (internet display, internet search, online video, and mobile web) and *traditional ads* (TV, magazines, newspapers, radio, and outdoor exhibits), and attention is restricted to the product-related ads (e.g., ads related to services, campaigns, corporate promotions are excluded). For each advertised product, various product descriptions (product name, brand, sub-brand, industry, major, subcategory) 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 Media converts units of ads into estimates of ads expenditure.

Product descriptions in Kantar Media are used to define varieties in different ways. Two benchmark definitions of *varieties* are used based on product names and brand names. An example of a product name is "Nike Air Max: Sneakers Men." A brand name for this product is "Nike Air."[5] Varieties are classified into various product categories. 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.

As a result, for each firm and product category (industry/major/subcategory), the number of varieties (products/brands) and digital- and traditional-ad spending are known over time.

[3] Appendix A.1 details the data sets.

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

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

Table A.1 in Data Appendix A provides summary statistics for the baseline data set. The firm-level data comprises 110,916 distinct firms with 332,190 firm-year observations over the period 1995 to 2015.[6] The key advantage of this firm-level data set is that it includes advertising spending and its split into digital and traditional media. While it only captures advertised varieties, there is a high correlation between all varieties offered by firms and advertised varieties. Last, to measure firm size, the Kantar Media data are combined with data on firms' employments from the National Establishment Time Series (NETS) Database.

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

An increase in varieties over time, documented earlier in Figure 1.2, is a robust pattern that does not depend on the specific data set or definition of a variety used. Figure 2.1 illustrates the evolution of the (normalized) log number of advertised products and brands (as well as sub-brands, for robustness) from Kantar Media. You can see that the number of distinct products offered in 2015 is 3.2 times larger than the number of distinct products in 1995.[7]

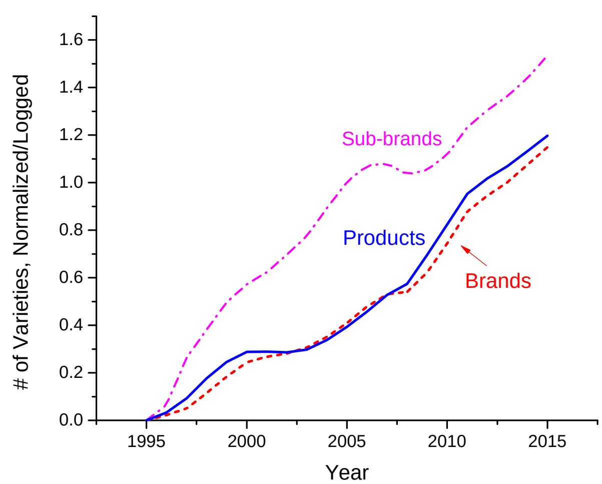

Figure 2.1: Product Varieties over Time.

Note: Trends in the normalized log number of product varieties over time. Product variety is defined based on the number of products, brands, and sub-brands. Source: Kantar Media,

At the same time, with the advent of the internet and improvements in targeting technologies, digital advertising became cheaper. Figure 2.2 shows time fixed effects from regressions of

[6]The last few years of the AdSpender data from 2016 to 2019 are not used because of data quality concerns for some variables from 2016 on.

[7]Figure C.1 in Appendix C also shows a similar increasing trend in product varieties using the RMS Nielsen data set that is described later.

product-level log prices by media type on year dummies from 2001 to 2015. Product-level prices by media type are obtained by dividing the total media-specific ad spending on a product by the number of ads in that media channel. While advertising prices in traditional media – TV and newspapers – are stable or grow, digital advertising prices – here, captured by internet display ad prices – dropped sharply, consistent with technological improvements in digital ads.[8]

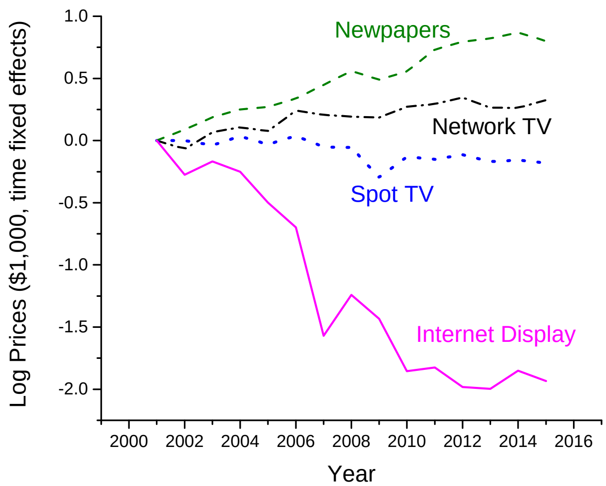

Figure 2.2: Advertising Prices by Media Type.

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

2.1.3 Relationship Between Growth in Digital Advertising and Varieties

To show the relationship between digital advertising and product varieties, a scatterplot is presented of the correlation between digital ads growth and variety growth. Figure 2.3 plots the change in the log number of varieties and the change in log digital-ad spending across product categories from the first year with digital ads, 2001, to the last year in the sample, 2015. Here, variety is defined based on product names, and product category is defined based on subcategories in Kantar Media. You can see that product categories using more digital ads over time also have an increasing number of varieties. Similar scatterplots are obtained when alternative definitions of variety and product category are used.

These correlations are evaluated using product-category-level and firm-level variation over time in Table 2.1. Panel A shows regressions of the year-to-year change in the log number of

[8]Internet display ads first appear in the data set in 2001 and hence have the longest coverage among digital ads. Internet search, online video, and mobile web entered the data in later years.

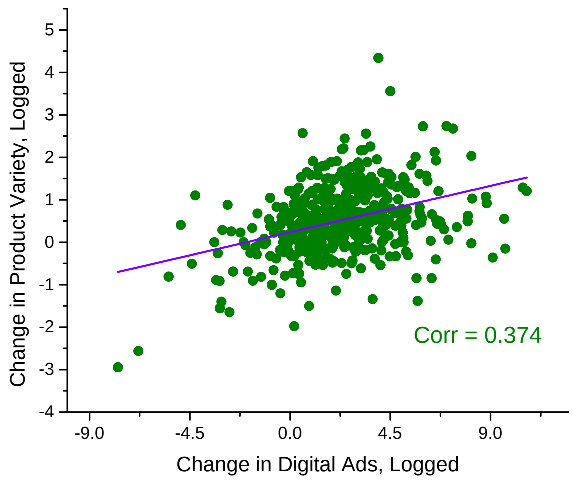

Figure 2.3: Correlation Between Growth in Digital Ads and Growth in Varieties, 2001-2015.
Note: The change in the log number of varieties (product names) is scatterplotted against the change in the log digital-ad spending across product categories (subcategories) from 2001 to 2015. The straight line is a linear fit. Data on digital ads starts from 2001. Source: Kantar Media.

Table 2.1: Varieties and Digital Ads

Panel A: Category-level	*Δ Log Products*		*Δ Log Brands*	
	Subcategory	Major	Subcategory	Major
	(1)	(2)	(3)	(4)
Δ Log Digital Ads	0.018***	0.013***	0.017***	0.015***
	(0.002)	(0.003)	(0.002)	(0.003)
R^2	0.244	0.263	0.243	0.258
Observations	8,567	2,300	8,567	2,300
Panel B: Firm-level	*Δ Log Products*		*Δ Log Brands*	
	Cross-firms	Within-firms	Cross-firms	Within-firms
	(5)	(6)	(7)	(8)
Δ Log Digital Ads	0.036***	0.036***	0.028***	0.029***
	(0.002)	(0.002)	(0.002)	(0.002)
R^2	0.085	0.184	0.045	0.132
Observations	11,070	10,298	11,070	10,298

Note: Panel A shows regressions of the growth in the number of varieties on the growth in digital-ad spending in product categories over time. All regressions control for the log number of firms and log traditional-ad spending in product categories over time, product category, and year fixed effects. Regressions are weighted by the number of firms in product categories over time. Varieties are defined based on product names (columns (1) and (2)) and brand names (columns (3) and (4)). Product categories are defined as subcategory (columns (1) and (3)) and major (columns (2) and (4)). Panel B shows regressions of the growth in varieties on the growth in digital-ad spending by firms over time. All regressions control for a firm's log employment, log traditional-ad spending, year fixed effects, and product category fixed effects (columns (5) and (7)) and firm fixed effects (columns (6) and (8)). Varieties are defined based on product names (columns (5) and (6)) and brand names (columns (7) and (8)). Product category is defined as subcategory. Robust standard errors are in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

varieties (products and brands) on the change in log digital-ad spending across product categories (subcategory and major). Since the number of 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-ad spending and for the number of firms. Product category and year fixed effects filter out product-category specific characteristics and annual common demand and supply shocks. Regressions are weighted by the number of firms in product categories. Panel B displays similar regressions at the firm level, controlling for firm size, traditional-ad spending, year, and product category fixed effects (columns "cross-firms") or firm fixed effects (columns "within-firms"). In all specifications, the table exhibits that, other things equal, the growth in digital-ad spending is associated with the growth in the number of varieties. Tables C.1 and C.2 in Appendix C show the robustness of these associations using different specifications, namely using regressions in log levels instead of changes and using digital-ad relative to traditional-ad spending as the main control.

2.2 Causal Evidence

2.2.1 Empirical Strategy and Data

Empirical Strategy. A positive correlation between digital advertising and varieties is established in the previous section. 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 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 varieties offered by firms. When households do not use the internet, firms cannot employ digital advertising to target consumers' tastes (Evans, 2009; Goldfarb, 2014). As such, households' internet access is a crucial determinant of the cost of targeting consumers through digital advertising.[9]

The main challenge with using households' internet access is that internet adoption is endogenous and depends on demand-side factors that may themselves correlate with firms' decisions about the number of varieties (e.g., the age or income distribution at a given location might affect both internet adoption and the number of varieties firms decide to offer). To obtain exogenous variation in households' internet access, the analysis explores supply-side factors associated with the wireline internet infrastructure and proposes lightning strikes as

[9] Argente, Fitzgerald, Moreira and Priolo (2021a) use data on some types of traditional advertising and show that multi-location firms make advertising decisions at the local level. This suggests that firms' digital advertising decisions might be targeted at the local level, too. Although firms' digital ads at the local level are not observed, local variation in digital ads is induced by local variation from households' internet access. The use of households' internet access instead of households' digital ads cognizance is hence 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.

an instrument for residential internet access across different locations and time. 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; Guriev, Melnikov and Zhuravskaya, 2021). The wireline infrastructure needed for residential internet, including DSL 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. Internet providers (e.g. Comcast, EarthLink) are thus less likely to make investments into adding new infrastructure that enable access and fast internet in locations more prone to lightning and where lightning may be getting more frequent over time. Dealing with frequent and increasing lightning strikes may be (partly) addressed, but only at a cost. The acquisition of surge protectors and the adoption of a wireless internet connection will increase the user cost through the price of investment. Hence, whether the equipment is left unprotected or not, more lightning-prone areas should face higher internet user cost.

Overall, this implies that there will be a negative association between lightning strikes and internet access specified as

$$I_{lt} = \gamma Z_{lt-1} + \eta X_{lt} + e_{lt}, \tag{2.1}$$

where I_{lt} represents a measure of households' internet access in location l at time t, Z_{lt-1} is the measure of the intensity of lightning strikes in location l at time $t-1$, and X_{lt} is a vector of characteristics that includes detailed fixed effects and other controls such as population, income, demographics, and urban-rural status. Because infrastructure investments made by companies supplying internet connections may take a long time, the baseline analysis uses long time periods and evaluates a lagged relationship between I_l and Z_l. Equation (2.1) constitutes the first-stage regression for the empirical strategy.

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

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

where N_{ltj} is the measure of number of varieties sold in location l at time t from product category j, $\hat{I_{lt}}$ is the predicted measure of household internet access in location l at time t from regression (2.1), and X_{ltj} includes detailed fixed effects and other controls that are also used in the first-stage regression. Varieties are measured at a more detailed level – county (l) × time (t) × product category (j) – to ensure that the results are not driven by compositional differences in the coverage of types of product categories across locations once time-category fixed effects are included. By using exogenous variation from lightning strikes, the analysis filters out the variation from other forces that might impact both the demand for households' internet access in a location and the number of varieties sold by firms in the same location. Moreover, as further discussed in Section 2.2.3, because households' internet access is measured at customers' locations rather than at firms' production facilities, concerns are lessened about the possibility that the mediating channel between internet access and variety decisions operates through changes in firms' operational costs (other than the costs of digital advertising).

Data Sources and Variables. The empirical strategy uses spatial and time-series variation, together with a lightning-strike instrument, to estimate the causal relationship between households' internet access and product varieties. Since the data used in Section 2.1 do not provide the spatial variation necessary for this analysis, a new data set is constructed at the county (l) $\times$ time (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}. Data cover every year for the period 2008-2018 and are aggregated into 5-year periods in the baseline analysis and yearly in the robustness analysis.

Measures of the number of varieties are built using Nielsen scanner data on products sold in grocery, drug, and general merchandise stores from 2008 to 2018.[10] The baseline definition of *variety* used throughout are products (barcodes) and brands. In the scanner-level data, barcodes correspond to the finest level of product disaggregation.[11] An example of a brand is "Chobani" that includes multiple products (barcodes) with differences in flavor, form, size, packaging, and formula, among others. The most disaggregated level (barcodes) is more likely to capture differences in various attributes of a product, while a more aggregated definition (brand) differentiates only between the most important attributes. Varieties (barcodes and brands) are classified into specific product categories, defined using Nielsen's product classification structure (1,070 product modules).

The original Nielsen data cover a wide range of products – from non-durables, such as cereals, to semi-durables, such as lamps. The data set includes information on the stores where products are sold. The coverage of product categories varies across stores/locations. To minimize concerns about the potential mismeasurement of varieties across locations, the baseline data set uses 602 product categories that have high coverage across all locations (defined here as

[10]Empirical results are calculated based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

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

counties). 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 the number of varieties do not result from changes in the set of stores. For robustness, the analysis is also done using the data with the entire set of stores.

Information about households' internet access at the county level covers the 2008-2018 period and is gathered from the Federal Communications Commission (FCC). The main source is the FCC Form 477, sent to internet service providers requesting information about the types of services they offer, internet speeds, and subscribership, among other items. The main measure of *households' internet access* is built using the number of residential broadband fixed connections for speeds above 200kbps, reported in five bins corresponding to quintiles of the share of households with residential fixed connections.[12]

Additional analysis also explores information on the type of internet technology and speeds available in different locations. Internet service providers of fixed connections vary in the type of technology they use – listed here in order from the slowest to the fastest: satellite, Digital Subscriber Line (DSL), cable modem, and fiber.[13] The FCC "Fixed Broadband Deployment Data" has information on the technology of the providers operating in different locations, allowing the building of variables at the county level that capture the share of households with access to at least one provider for each type of technology.[14] For information on the different speed levels available across locations, the analysis relies on data obtained via a Freedom of Information Act (FOIA) request to the FCC (connections with downstream speed of at least 200kbps, 768 Kbps, 3 Mbps, and 10Mbps during the period 2008-2012) and via the publicly available FCC Form 477 data (connections with downstream speed of at least 200kbps, 10Mbps, 25Mbps, and 100Mbps during the period 2014-2018).

The information on *lightning strikes* is 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

[12]Data on the share of residential fixed connections is originally reported by Census tracts. For the analysis, Census tracts are aggregated into counties by computing the county averages weighted by the number of housing units in each Census tract as of 2010 (housing unit data are harnessed from the US Census Bureau).

[13]Satellites in geostationary orbits deliver and receive data to and from almost any earthly location fitted with a "dish," which communicates with the satellite. DSL service is a retrofit on top of telephone lines permitting them to carry data. Cable modem service involves the addition of switches and modems consistent with Data Over Cable Service Interface Specification (DOCSIS), which adds data services to existing cable television systems. Fiber typically involves newly laid lines of fiber optic wire to the customer.

[14]Because the data on the types of technology are only available for the period 2014-2018, which is the last 5-year period of the baseline sample, only cross-sectional variation during that period can be used. Also, note that the variables capture potential access (not the use) of a type of technology. There is no comprehensive information on the type of technology actually used by internet subscribers.

available since 1986 (the baseline analysis covers 2003-2018), 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 at the county $\times$ year level.

The analysis also uses additional variables from the Bureau of Economic Analysis, the US Census Bureau, and US Department of Agriculture as controls in the regressions. Appendix B provides information on these variables and summary statistics.

Spatial Variation. The empirical strategy relies on spatial heterogeneity in varieties, households' internet access, and lightning strikes at the local level. It is thus crucial for the empirical strategy that there is variation in varieties across locations in the United States, with this variation changing over time. Let N_{ltj} be the number of varieties (barcodes or brands) in location (county) l at time t within product category j, and N_{0j} be the number of varieties in product category j across the entire United States in the baseline year 2008. Then, define the weighted county-to-nationwide share of varieties as

$$n_{lt} = \sum_{j=1}^{J} \Big(\omega_{lj} \frac{N_{ltj}}{N_{0j}} \Big),$$

where ω_{lj} is the revenue share of product category j in county l across all years.[15]

Figure 2.4 (Panel A) shows the map of the weighted county-to-nationwide share n_{lt} in the baseline year, 2008, and last year of the data set, 2018, using brands as a measure of varieties.[16] The map becomes darker as a higher share of varieties are sold in a county. The map shows that there is substantial variation in the amount of varieties across counties and that over time there was a differential increase in varieties across locations. To further illustrate variation across regions in different measures of varieties, Table C.3 in Appendix C shows the R-squared of the regression of varieties N_{ltj} (in logs) after ijoncluding various fixed effects and controls. The results show that after accounting for product category and time fixed effects, adding county fixed effects explains a large proportion of the variation in the data. Importantly, even when controlling for differences in population and income across counties over time, adding county fixed effects still explains a significant proportion of the variation, which supports the view that there are differences across regions in product varieties consumed.

Figure 2.4 (Panel B) displays heterogeneity in household residential fixed connections across locations in 2008 and 2018 (the first and last years of the sample). A value between 4 and 5

[15] Alternative measures are built using different revenue shares: (a) a location-invariant ω_j – a revenue share of product category j across all years and regions; and (b) a time-varying revenue share ω_{ltj} – a revenue share in each county $\times$ year $\times$ product category. The patterns are very similar.

[16] The patterns are similar when using barcodes as a measure of varieties.

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, the internet was fully diffused in a few locations, while in most areas it was still in its early stages. Ten years later, most households in the United States had access to the internet. Nevertheless, differences across locations still persist. The uneven geographic supply of internet infrastructure is one of the main reasons why some areas have high rates of non-adoption of internet (Anderson, 2019; Greenstein, 2020). Other factors include demographic features of users such as older age, low income, and lower education. The supply of internet infrastructure is typically smaller in rural or low-density locations, where internet infrastructure is either nonexistent or old. The costs of supplying internet service to a given geographic areas may reflect economies of scale, preexisting infrastructures of telephone lines and TV cables, and climate factors (such as lightning).

Differences in households' internet access across locations are instrumented using the frequency of lightning strikes per square mile. Figure 2.4 (Panel C) shows a large spatial variation in lightning strikes across locations, where the Southeast and many counties in the Midwest are particularly impacted. In fact, county fixed effects alone explain two-thirds of the variation in lightning strikes, and year fixed effects explain only about five percent.[17] Nonetheless, there is still some variation over time: some regions become unexpectedly more impacted in recent years (e.g., counties in the Southwest region) while others get fewer lighting strikes over time (e.g., some counties in the Northeast region).

[17] Andersen, Bentzen, Dalgaard and Selaya (2012) studies the evolution of lightning strikes from 1906 to 2005 across states and documents that there are no strong time trends, suggesting that global warming-induced climate changes might not be impacting the level of lightning strikes, contrary to other climate phenomena.

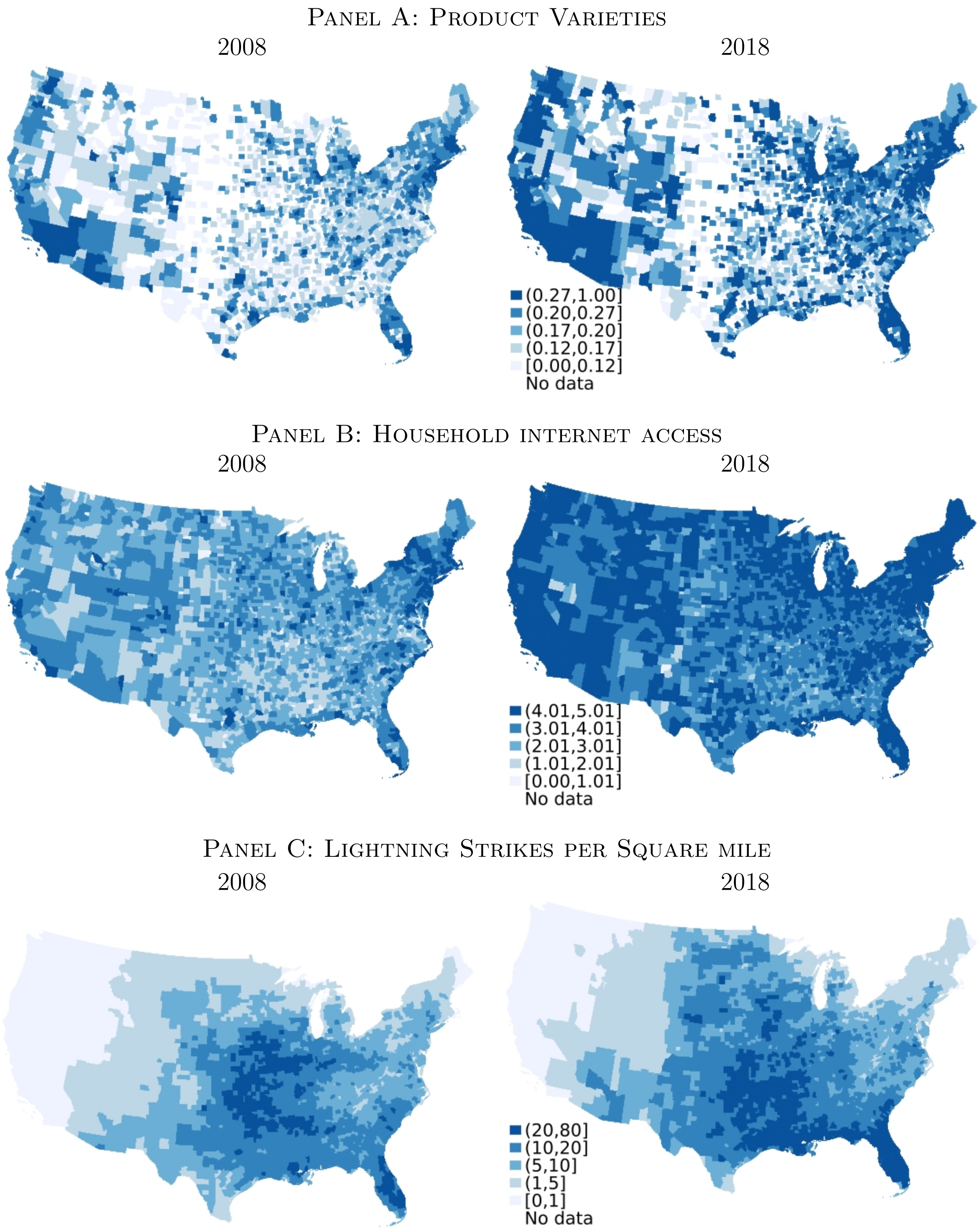

Figure 2.4: Spatial Variation in Product Varieties, Household Internet Penetration, and Lightning Strikes.

Note: Panel A exhibits for each county the weighted county-to-nationwide share of varieties, as defined in the main text. Panel B presents the share of households with residential fixed internet connections for each county, reported in five equal bins. Panel C displays lightning strikes per square mile for each county. Sources: Nielsen, Federal Communications Commission, and National Lightning Database Network.

2.2.2 Results

First Stage: Internet and Lightning Strikes. The proposed empirical strategy relies on the presumption that in areas prone to lightning strikes, everything else equal, it will be more costly to build and maintain internet infrastructure, and thus these areas will be less likely to have internet access. To test this hypothesis, three different specifications for the controls X_{lt} in equation (2.1) are considered. Specification 1 includes time fixed effects that control for common time trends in internet access and lightning strikes. Specification 2 further tightens the empirical specification by including time $\times$ county controls such as (log) population and (log) income per capita that account for time-varying heterogeneity across counties. It also includes county-level time-invariant controls that account for differences in age, education, and population density. Finally, specification 3 is akin to a difference-in-difference specification, by also controlling for county-specific differences in household internet access and lightning strikes.

There are different trade-offs to these different specifications. Specifications 1 and 2 are mostly identified out of the cross-sectional variation across counties (unconditional and conditional on various county-level controls), while specification 3 identifies the main effect from the variation in the growth of household internet access and lightning strikes across locations. Specification 2 is preferable over specification 1 because the additional controls subsume many determinants of internet demand and supply associated with economies of scale and population density. The main advantage of specification 3 relative to specification 2 is that it uses a non-parametric way to control for potentially unobserved factors impacting differences in lightning strikes and household internet access across counties. The main disadvantage is that the identification relies on differential time variation across locations in lightning strikes and that is not strongly supported by the data.

Infrastructure investments made by companies supplying residential fixed internet connections might take a long time. Thus, it is important to first understand the nature of the lag between lightning strikes and internet access. Table C.4 in Appendix C shows the results of estimating equation (2.1) for the three specifications with yearly data and including 1-to-10 year lags of lightning strikes. The results using specifications 1 and 2 show that lightning strikes are negatively associated with future internet access for at least a decade. Specification 3 indicates that, within a county, growth in lightning strikes over time is associated with lower internet access with a 5 to 10 year lag.

Because of these long lags, to estimate the first-stage equation (2.1), the yearly data is grouped into three periods: 2003-2007, 2008-2012, and 2013-2018. Table 2.2 shows the results using the data set with internet access and varieties averaged over the 2008-2012 and 2013-2018 periods, and lightning strikes averaged with a one-period lag (i.e., the periods 2003-

Table 2.2: Internet and Lightning Strikes: First-Stage Results

	Household Internet		
	(1)	(2)	(3)
Lightning Strikes (lagged)	-0.028***	-0.015***	-0.003***
	(0.000)	(0.000)	(0.000)
Observations (1,000s)	1,978	1,974	1,822
Time FE	Yes	Yes	Yes
Time × County Controls	No	Yes	Yes
County FE	No	No	Yes
1st stage F-stat	123,767	88,678	669

Note: The estimated coefficients from regression (2.1) with county-time-level data. Time is defined in 5-year periods: the 2008-2012 and 2013-2018 periods for the dependent variable, and the 2003-2007 and 2008-2012 periods for the lagged independent variable. The time × county controls are time-varying population (in logs) and income per capita (in logs), and time-invariant controls; viz., the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas. The variables are described in Sections 2.2.1 and Appendix B. Robust standard errors are shown in parentheses. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

2007 and 2008-2012). Using averages over several years exploits the strongest association in the data and reflects the lagged relationship between lightning strikes and investments in internet infrastructure. As seen, the first stage results are statistically significant, displaying the expected negative relationship between internet access and lightening strikes.[18]

To evaluate further the channel linking lightning strikes to internet access, information on the nature of the internet technology (DSL, cable, satellite and fiber) is explored. The DSL and cable wireline technologies are particularly sensitive to electrical surges caused by lightning strikes. Thus areas more prone to lightning strikes are expected to have lower access to the providers of these technologies relative to satellite or fiber providers. Households in (almost) every location in the United States can access the internet through satellite (by installing a "dish" at home which communicates with the satellite), but they may not be able to access internet through wireline options such as DSL, cable, or fiber. Access through satellite is typically more expensive and provides lower speeds than alternative technologies, so when available, households often choose wireline technologies. The FCC data show that there is substantial variation across counties in terms of the different types of internet technologies available. On average, there is a high penetration of DSL and cable providers, while only a few counties have access to fiber over the period. Table C.7 in Appendix C shows that counties that are more prone to lightning strikes have lower access to DSL and cable providers. Having access to DSL and cable providers, in turn, is associated with a larger number of households

[18] The results are qualitatively robust (although with statistically insignificant coefficients in the third columns of the second-stage) when using yearly data from 2008 to 2018 and the instrument lagged. Tables C.5 and C.6 in Appendix C.6 show the results.

with internet access (conditional on demographic and socioeconomic factors impacting the demand and supply of internet sevices). The results also show that lightning strikes do not negatively impact access to fiber providers, consistent with the fact that electrical surges travel through DSL and cable lines, but not through fiber lines.

Second Stage: Varieties and Internet. Having established that lightning strikes is a good instrument for household internet access, the next step is to use it to evaluate the causal impact of internet access on the number of available varieties. Table 2.3 presents the second-stage results for the IV specification in equation (2.2). The baseline estimates are for varieties measured as the logarithm of either the number of barcodes or brands. The table includes the same set of controls as the first-stage specifications: columns 1 includes year $\times$ category fixed effects; columns 2 adds county controls (time-varying and time-invariant); and columns 3 further adds county $\times$ category fixed effects. There is a statistically significant relationship between instrumented households' internet access and product varieties across all specifications for both measures of product varieties.

An exogenous increase in internet access leads to more varieties. The magnitude of the effect is economically large. Specification 2, which uses cross-county variation, indicates that a 20 percentage point increase in the share of population with residential internet access more than doubles the number of varieties (columns 2). Using the difference-in-difference specification, the same increase in the share of population with residential internet access generates an increase in the number of barcodes and brands by about 78% and 10%, respectively (columns 3). Table C.8 in Appendix C documents the results from equivalent OLS regressions. The results of the IV and OLS regressions are qualitatively similar, and the magnitudes are larger in the IV estimation.

Appendix C presents additional results with specifications for different internet speed levels, alternative measures of varieties, and different samples. A crucial aspect of the analysis relies on using a good measure of internet access (which is critical for digital advertisers), and having high-speed internet is more likely to lead households to use the internet more intensively. The baseline measure of internet access through the residential broadband connection captures high-speed access that is always on and is faster than the dial-up access or other traditional services. Nevertheless, to have a consistent variable for the entire period of the analysis, the baseline measure is defined as broadband access with downstream speed of at least 200kbps, which is slow by today's standards. Table C.9 in Appendix C shows that the positive causal association between internet access and varieties is significant (and even stronger) for higher internet speeds. In addition, the internet data are based on fixed connections and do not include mobile connections. While in more recent years, there has been a large increase in digital advertising distributed through mobile devices, fixed connections were crucial in

Table 2.3: Household Internet and Varieties: Second-Stage Results

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(1)	(2)	(3)
Household Internet	0.956***	1.077***	0.578***	0.718***	0.811***	0.093***
	(0.004)	(0.008)	(0.050)	(0.003)	(0.006)	(0.040)
Observations (1,000s)	1,978	1,974	1,822	1,978	1,974	1,822
Time× Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Time × County Controls	No	Yes	Yes	No	Yes	Yes
County × Category FE	No	No	Yes	No	No	Yes

Note: The estimated regression coefficients from equation (2.2) with county-product category-time-level data. Time is defined using 5-year periods: 2008-2012 and 2013-2018. The dependent variables capturing varieties are either barcodes (in logs) or brands (in logs). "Household Internet" is the instrumented variable from the first-stage regression. The time-varying county controls used in specifications 2 and 3 are population (in logs) and income per capita (in logs). The time-invariant county controls used in specification 2 are the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas. The variables are described in Section 2.2.1 and Appendix B. Robust standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

the earlier periods of internet diffusion.[19] Also note that the IV relies on lightning strikes impacting wireline access to fixed connections and, as such, this IV strategy would not be suited for instrumenting mobile connections. Table C.11 in Appendix C presents baseline results with an expanded set of product categories (including more durable products), and Table C.12 in Appendix C shows results with alternative samples of retail stores. The source data cover varieties in brick-and-mortar stores and do not capture varieties sold solely online. The latter are likely to have an even stronger association with digital advertising, suggesting that in more recent years, the link between internet access and varieties might be even stronger. Finally, Table C.10 in Appendix C shows that the results are qualitatively similar for the alternative measures of varieties.

2.2.3 Alternative Mechanisms

The results above are predicated on two key identification assumptions. The first assumption is the relevance of the lightning strikes for household internet access. This assumption was validated by the robust first-stage results. The second assumption relates to the exclusion restriction, which 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 lightning strikes may correlate with 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.[20] The

[19] In most recent periods, fixed and mobile internet access have a high spatial correlation (Anderson, 2019).

[20] 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

use of spatial variation helps to address 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' headquarters and all markets where their products are sold is gathered. Then, additional measures of varieties available to consumers are defined that exclude the varieties offered by firms co-located with where their products are sold. In particular, a distinct set of variables is built by counting solely the number of barcodes and brands that are sold by firms whose headquarters are in another state and that sell to many states.[21] As a consequence, these variables are more likely to include variation from large firms that make location-specific product offerings. These variables have the advantage of likely capturing the varieties offered by firms whose general internet access conditions differ from the conditions faced by their consumers. Thus, the effect of households' internet access on these varieties is not confounded by the impact that a firm's use of ICT may have on its productivity and product offering decisions. Table 2.4 presents baseline second-stage results for these alternative measures of varieties (the first-stage is same as in the baseline). The main coefficients of interest are qualitatively similar and the magnitudes are only marginally smaller than the baseline results. This further supports the idea that the empirical strategy captures the causal impact of households' internet access on product varieties. In a similar fashion, 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 C.13 in Appendix C shows the robustness of the results excluding local retail chains.

and Boussard (2018). 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 the number of product lines offered.

[21] Multi-state firms are defined as firms in the top quartile of the distribution of the number of states they sell in. Details on these variables are in Appendix B.

Table 2.4: Household Internet and Varieties: Excluding Local Firms

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(1)	(2)	(3)
Household Internet	0.874***	0.933***	0.541***	0.618***	0.650***	0.043
	(0.004)	(0.008)	(0.056)	(0.003)	(0.006)	(0.046)
Observations (1,000s)	1,863	1,858	1,713	1,863	1,858	1,713
Time× Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Time × County Controls	No	Yes	Yes	No	Yes	Yes
County × Category FE	No	No	Yes	No	No	Yes

Note: The estimated regression coefficients for equation (2.2) with county-product category-time-level data. Time is defined using 5-year periods: 2008-2012 and 2013-2018. The dependent variables are either barcodes or brands (in logs) in a county × time × category sold by firms whose headquarters are in another state and sell in many states (in the top quartile of the distribution of the number of states). "Household Internet" is the instrumented variable from the first stage. The time-varying county controls used in specifications 2 and 3 are population (in logs) and income per capita (in logs) The time-invariant county controls used in specification 2 are the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas. The variables are described in Section 2.2.1 and Appendix B. Robust standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

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 inelastically supply and profits accrue from the portfolio of firms that they own. Their income is used to consume a generic good and specialized varieties from a spectrum of 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 varieties 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.[22] There is free entry into the specialized-products sector, subject to incurring a fixed entry cost. 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-product firm can use both types of advertising. Product lines are not perfect substitutes so firms possess some market power. In contrast, generic goods are perfect substitutes. The generic goods

[22]To map the model into the data, think about Nike footwear as being an example of a product line, and Nike Air Max men's sneakers as an example of a variety.

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 quantitative analysis will focus on comparing two static, symmetric equilibria calibrated to the years 1995 and 2015. The developed model also allows for process innovation in the production of specialized varieties as well as technological progress in the generic goods sector between the two periods. Additionally, the fixed entry cost associated with introducing a product line are allowed to change. The effects of each of these changes in the economic environment are studied.

3.1 Consumers

An individual i consumes a generic good and a single variety from each specialized product line in the consumption set $\mathcal{M}_i$. The utility for a specialized product line $j \in \mathcal{M}_i$ depends upon how close the variety 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

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

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 notational simplicity the subscript i is dropped. It needs to be emphasized, though, that individual consumers will differ in the ads that they receive and hence in the quantities of particular varieties that they may consume out of the total number of varieties available in the economy – more on this Section 3.4. The utility weight on generic goods is θ and the inverse of the price elasticity of demand for specialized products is κ. It is easy to calculate that an individual's 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 D.

3.1.2 Tastes for a Variety within a Product Line

How are consumer's tastes for varieties within a product line determined? Represent an individual's tastes for a variety 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. Denote the distance between a particular consumer's tastes, represented by the point i, and the location of variety m, represented by the point m/n, by the arc length $d(i, m/n)$. The situation is portrayed in Figure 3.1.

How well this particular variety matches the consumer's tastes is given by

$$\sigma\big(d(i,m/n)\big) = \chi - \lambda d(i,m/n), \text{with } \chi, \lambda > 0. \tag{3.5}$$

The utility realized by purchasing this variety is

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

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

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

Note that σ represents match quality in terms of distance between the consumer's tastes,

i, and the variety, m, while S defines match quality in terms of the product line 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 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, contained in product line j that is at a distance $d(i, m/n)$ from their tastes, i. A consumer may not get an ad for some product lines; for $j \notin \mathcal{M}$, $S(j) \equiv 0$.

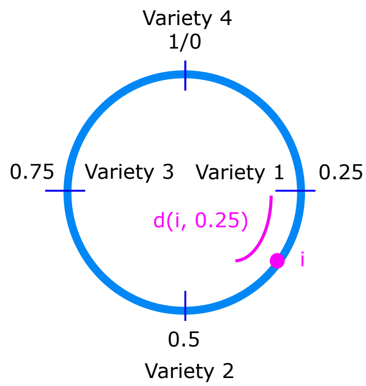

Figure 3.1: Taste over Varieties within a Given Product Line.

Note: In this example, 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 variety 1 is measured by the arc length $d(i, 0.25)$. If the person consumed this variety, then $\sigma(d(i, 0.25)) = \chi - \lambda d(i, 0.25)$.

3.2 Specialized-Product Firms

There are N monopolistically competitive firms each selling their own product line for an aggregate total of N 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 incur a fixed entry cost in the amount $w\phi$. Each variety is sold at the unit price p. To sell specialized products, a firm 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 match quality is $\sigma(d(i,k)) = \chi - \lambda d(i,k)$. 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 average match quality as specified in Proposition 1.

Proposition 1. *(Average match quality for traditional advertising) The average match quality over all consumers who just receive a traditional advertisement 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. 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. □

The important thing to note here is that average match quality, σ_t, is not a function of n. Hence, traditional advertising generates the same average match quality irrespective of the number of varieties produced. The quantity demanded from people who just receive a traditional ad is therefore

$$q_t(p) = \sigma_t[(1-\theta)\hat{y}/p]^{1/\kappa}, \text{ cf. eq. (3.4)}. \tag{3.7}$$

Digital advertising alerts consumers to the varieties that are most suited to their tastes. Specifically, suppose that ads for a variety located at position l are sent to consumers with tastes in the range $[l - 1/(2n), l + 1/(2n)]$, centered around l. Hence, traditional advertising applies to the whole circle while digital advertising targets just a segment. If a consumer with tastes positioned at i buys the variety located at the point l, then their taste parameter is given by $\sigma(d(i,l)) = \chi - \lambda d(i,l)$. Again, Figure 3.2 illustrates the situation. Proposition 2 specifies average match quality for consumers who receive a digital ad.

Proposition 2. *(Average match quality for digital advertising) The average match quality across consumers who receive a digital advertisement is*

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

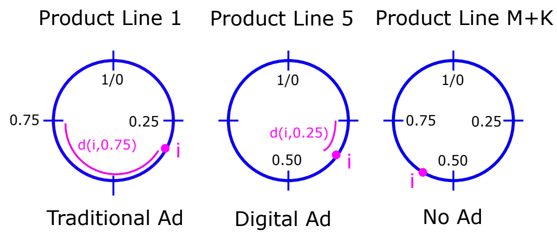

Figure 3.2: Advertising.

Note: In this example, there are 3 product lines, numbers 1, 5, and $M+K$. Each product line has 4 varieties, equally spaced around the circle. The varieties are numbered clockwise with the first variety being located at the point 0.25. 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 variety 3 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 with digital advertising. Variety 1 in product line 5 is advertised over the range $[0.25-1/8, 0.25+1/8]$. Here a person with tastes positioned at point $i \in [0.25-1/8, 0.25+1/8]$ will have arc length of $d(i, 0.25)$ between their tastes and the variety. The consumer got no ad for product line $M+K$. Hence, they do not consume this product, which lies outside of the set of products that they consume; i.e., $M+K \notin \mathcal{M}$.

Proof. The ads for each variety span an arc with distance of $1/n$ centered around the variety's location. The maximal distance between a consumer and their closest variety is $1/(2n)$. Consumers who receive an ad for a variety are uniformly distributed over distances on the interval $[0, 1/(2n)]$. Varieties are equally spaced around the circle. Therefore, the average distance is $0.25/n$. □

Average match quality for digital advertising is a function of the number of varieties, n. The important thing to note is that average match quality increases in the number of varieties, n. So long as $n \geq 1$, digital advertising will on average lead to customers buying varieties that better match their tastes than compared with traditional advertising. The quantity purchased from individuals receiving digital ads is

$$q_d(n,p) = \sigma_d(n)[(1-\theta)\hat{y}/p]^{1/\kappa}, \text{ cf. eq. (3.4).} \tag{3.8}$$

Last, a person may get no ads for a product line. In this case, they will not buy any variety in that product line – see Figure 3.2. This case is represented by $\sigma_n \equiv 0$.

As long as more than one variety is produced, Propositions 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 propositions is illustrated in Figure 3.3 which plots average match quality with digital and traditional advertising. As can be seen, digital advertising, on average, always results in a closer match. Now, consumers demand higher quantities, at a given price, when the variety is a closer match. This gives firms some scope to raise their prices as well as sell more, the breakdown of which depends on the price elasticity of demand. Thus, the advent of digital advertising allows firms to raise their profits by introducing more varieties.

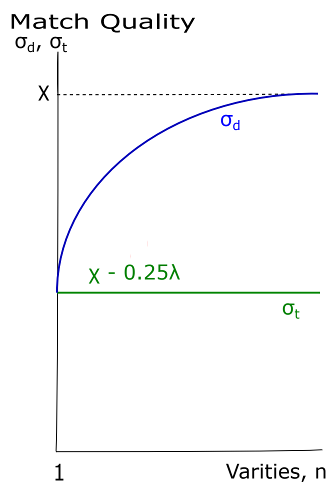

Figure 3.3: Digital versus Traditional Advertising.

Note: With traditional advertising average match quality, σ_t, is not a function of the number of varieties. With digital advertising average match quality, $\sigma_d(n)$, 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. When there are perfect matches average match quality is χ.

The intensities of traditional and digital advertising are denoted by a_t and a_d. These represent the probabilities of a consumer receiving a traditional and a digital advertisement, respectively. 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 line is $1-a_d-a_t(1-a_d)$. A specialized products firm sells to consumers that received digital and/or traditional ads. Its sales read

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

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

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

$$A a_d^{\zeta}/\zeta \text{ and } B a_t^{\nu}/\nu. \tag{3.9}$$

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 $o_s = a_d q_d(n,p) + a_t(1-a_d)q_t(p)$ is total output, and n is the number of varieties that the firm is producing. As the number of varieties, n, increases, so does the organizational cost of selling the product line. A specialized-product firm must also incur a fixed cost entry ϕ in terms of labor. Technological progress in the production of specialized products occurs when Ξ decreases over time. Think of this as process innovation. The fixed entry cost ϕ may rise due to an increase in the barriers to entry.

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

$$\begin{aligned}\Pi = \max_{a_d,a_t,n,p} \Big\{ & p a_d \sigma_d(n)[(1-\theta)\hat{y}/p]^{1/\kappa} + p a_t(1-a_d)\sigma_t[(1-\theta)\hat{y}/p]^{1/\kappa} \\ & - wAa_d^\zeta/\zeta - wBa_t^\nu/\nu \\ & - w\Xi\{a_d\sigma_d(n)[(1-\theta)\hat{y}/p]^{1/\kappa} + a_t(1-a_d)\sigma_t[(1-\theta)\hat{y}/p]^{1/\kappa}\}n^\eta/\eta - w\phi\Big\}. \qquad (3.10)\end{aligned}$$

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 last line includes its manufacturing and entry costs.[23] The maximal level of profits earned by the firm is Π. In equilibrium, firms will keep entering with their own unique product lines until Π is driven down to zero.

3.3 Generic Goods

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

$$o_g = xl^\alpha,$$

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

[23] Why aren't sales multiplied by 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.

$$w = \alpha x l^{\alpha - 1}. \tag{3.11}$$

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 the 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.

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 $\left[\sigma(0.5)[(1 - \theta)\hat{y}/p]^{1/\kappa}, \sigma(0)[(1-\theta)\hat{y}/p]^{1/\kappa}\right]$ with no holes. There will be a mass of varieties at each point on this interval. Take the combinations of quantities and masses at each point on this interval to form a distribution over variety consumption quantities. While the varieties at each point will differ across consumers, this distribution of quantities consumed is the same for all consumers. The cardinality of the set of product lines for which a variety is consumed is the same for all consumers. That is, the number of product lines consumed by person i, 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

$$M_d = N a_d \text{ and } M_t = N a_t(1 - a_d), \tag{3.12}$$

which gives

$$M = M_d + M_t = N[a_d + a_t(1 - a_d)], \tag{3.13}$$

where a_d is the probability of a customer receiving a digital ad and $a_t(1 - a_d)$ are 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 product 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 match qualities for digital advertising and traditional advertising are $\sigma_d(n)$ and σ_t. Therefore, the above expression can be rewritten as

$$\begin{aligned}\hat{y} &\equiv \frac{y}{\theta + (1-\theta)^{1/\kappa} (\hat{y}/p)^{(1-\kappa)/\kappa} M[a_d\sigma_d(n) + a_t(1-a_d)\sigma_t]/[a_d + a_t(1-a_d)]} \\ &= \frac{y}{\theta + (1-\theta)^{1/\kappa} (\hat{y}/p)^{(1-\kappa)/\kappa}[M_d\sigma_d(n) + M_t\sigma_t]}, \text{ using (3.12) and (3.13).} \end{aligned} \tag{3.14}$$

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

$$\begin{aligned} N\{[a_d\sigma_d(n) + a_t(1-a_d)\sigma_t]\left[\frac{(1-\theta)\,\hat{y}}{p}\right]^{1/\kappa} \Xi\frac{n^\eta}{\eta} + \phi + A\frac{a_d^\zeta}{\zeta} + B\frac{a_t^\nu}{\nu}\} \\ + (\frac{\alpha x}{w})^{1/(1-\alpha)} = 1. \end{aligned} \tag{3.15}$$

The first line of the above expression is the labor hired by the N specialized-product 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 product and generic goods producing firms must equal the supply of labor, or the right hand side of the second line.

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

$$\Pi = 0, \text{ cf. (3.10).} \tag{3.16}$$

This free-entry condition regulates the number of specialized-product firms, N. Consumers earn profits from generic goods production in the amount

$$\pi = (1-\alpha)xl^\alpha = (1-\alpha)x(\frac{\alpha x}{w})^{\alpha/(1-\alpha)}. \tag{3.17}$$

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

1. Given variety prices, p, profits, π, wages, w, and the consumption set, $\mathcal{M}$, consumers solve problem (3.1). This determines c and $\{q(j)\}_{j=1}^{M}$ where $M = |\mathcal{M}|$ and $\hat{y}$ is determined by (3.14).

2. Given w and $\hat{y}$, specialized-product firms solve problem (3.10), yielding a solution for $a_d, a_t, n, p,$ and Π.

3. Given wages, w, the generic goods sector hires labor, l, in accordance with (3.11). Additionally, the profits from generic goods production, π, accrue to a consumer as specified by (3.17).

4. The number of product lines purchased by a consumer, $M_d, M_t,$ and M, are given by (3.12) and (3.13).

5. The free-entry condition (3.16) holds. This governs the number of product lines, N.

6. The labor market clears in accordance with (3.15). This determines wages, w.

Deriving theoretical propositions about the equilibrium is difficult. It will be established theoretically in Section 5, however, that process innovation, or reductions in Ξ, while increasing the number of product lines, N, does not influence the number of varieties per product line, n. Technological progress in the generic goods sector, or increases in x, has no effect on either the number of product lines or the number of varieties per product line. This highlights the observation in the quantitative analysis that the primary driving force behind an expansion in the number of varieties per product line is technological progress in digital advertising, or a fall in A.

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, the operating cost for specialized products production declines, or there is a drop in Ξ reflecting process innovation. Third, generic goods production becomes more efficient, which is captured by an increase in x. Additionally, the fixed entry cost, ϕ, associated with specialized product production is allowed to rise.

The model has 17 parameters to determine. Some of these parameters are chosen based on a priori information. Others are calibrated by fitting the model to match data targets. The parameters selected to match data targets are calibrated using a two-step procedure. In the first step, a theory-based identification scheme is employed. In particular, a set of parameter values is backed out using the model's structure to hit a set of data targets exactly. In the second step, the remaining parameter values are chosen to minimize the model's prediction error with respect to another set of data targets. There is a functional dependence of the first step on the second step. The calibration strategy takes this into account.

4.1 Parameters Set Using A Priori Information

Two parameters are set exogenously. Marto (2023) reports that the average markup in the US economy between 1995 and 2015 is 1.24. Assume that this markup applies to both the generic and specialized varieties sector. This observation can then be used to pin down the output elasticity of labor in the generic sector, α, and the (inverse of the) price elasticity elasticity of demand for specialized products, κ. See Appendix F for the details.

4.2 Parameter Values Calibrated Based on an Exact Fit

In the first step, the parameter vector $\rho_f \equiv (\Xi_{1995}, \Xi_{2015}, A_{1995}, A_{2015}, B, \eta, \phi_{1995}, \phi_{2015}, x_{1995}, x_{2015})$ is calibrated to match exactly a set of 7 targets and 3 restrictions on the initial values of some of the model's variables. The 7 data targets are the total-advertising-to-GDP ratio, the ratio of digital-to-traditional advertising spendings (both in 1995 and 2015), the growth rates in the number of varieties per product line and product lines between 1995 and 2015, and the growth rate in per-capita income over that period. The first restriction normalizes the initial wage rate w_{1995} to be one. The next two restrictions set the initial number of product lines and varieties to also be one; i.e., $N_{1995} = n_{1995} = 1$. The first step makes use of the first-order and equilibrium conditions from the model, evaluated at the data targets, to back out 10 parameter values such that the 10 calibrating conditions hold. This amounts to solving a system of 10 nonlinear equations represented by

$$\rho_f = R(\rho_m). \tag{4.1}$$

The parameter vector ρ_f is a function of the parameter vector ρ_m chosen in the second step (discussed next). Again, details are in Appendix F.

4.3 Parameter Values Calibrated to Minimize the Model's Prediction Error

In the second step, the parameter vector $\rho_m \equiv (\theta, \chi, \lambda, \zeta, \nu)$ is chosen to minimize the model's prediction error with regard to the size of the specialized-products sector in 1995 and 2015, the elasticity of sales with respect to total advertising, and the elasticity of the number of total varieties with respect to digital advertising. The exponents on the cost functions for digital and traditional advertising are assumed to be the same; i.e., $\zeta = \nu$. Hence, the second step can be represented as solving the following minimization routine

$$\min_{\rho_m} \quad \sum_k \left[\frac{\mathbf{D}_k - \mathbf{M}_k(\rho_m, \rho_f)}{\mathbf{D}_k}\right]^2, \tag{4.2}$$

subject to $\rho_f = R(\rho_m)$. Here $\mathbf{D}_k$ denotes the k'th data target in the second step and $\mathbf{M}_k(\rho_m, \rho_f)$ the model's solution for this target. Each observation is weighted equally. The minimization routine internalizes how the choice of the parameter vector ρ_m in the second step affects the determination of the parameter vector ρ_f in the first step.

4.4 Data Targets

There are 10 categories of data targets (including some restrictions on initial values). Some of these categories have two observations corresponding to 1995 and 2015, others just one. The data targets are classified below according to the fitting criteria employed.

Exact fit

- *The total advertising-to-GDP ratio.* The advertising-to-GDP ratio in the United States has been roughly constant over long periods of time as the historical series compiled by Coen establishes. The average advertising-to-GDP ratio from 1995 to 2007 (the last year provided in Coen's data) is 2.2%, which is taken to be a target for both periods. So, this category has two targets.

- *The ratio of digital-to-traditional advertising.* To obtain the split between digital and traditional advertising spending, the ads are classified as digital and traditional ads as described in Section 2. Then the average traditional ads per sales and digital ads per sales of firms in the Kantar-NETS data set are calculated. These ratios provide the average digital-to-traditional advertising ratio in 1995 and 2015.[24] This ratio rose from

[24] The digital ads data are available in Kantar from 2001. The digital-to-traditional ads ratio from 2001 is extrapolated back to 1995 by assuming constant growth over the period 1995-2015.

2.3% to 96.6% over the two time periods.

- *The growth in total varieties between 1995 and 2015.*[25] As explained in Section 2, the Kantar and RMS data sets provide information about product varieties. Since product varieties can be defined differently within and across these data sets, the average growth rate of the total number of varieties using different definitions and across both data sets is used.[26] This is estimated to be 151%. This amounts to one target.

- *The growth in the number of product lines between 1995 and 2015.* Product lines are mapped in the data to establishments, with the idea that firms produce similar products within their establishments, which is in line with what is assumed in the model. By using data from the US Census Business Dynamics Statistics, this gives one target of 17%.[27]

- *The growth rate in income per capita between 1995 and 2015.* Using data from the World Bank, the growth rate of GDP per capita in purchasing power parity terms in the United States was 35.7%.

- *Restrictions on initial values for 1995*: $w = N = n = 1$.

Minimized prediction error

- *The size of the specialized-products sector in both years.* The sector (SIC 3-digit) is classified as a generic sector if the sales-weighted share of firms doing advertising is lower than 10%. The remaining sectors are classified as specialized. The sales share of specialized sectors in the Kantar-NETS data is 53% in 1995 and 59% in 2015. Hence, there are two targets here.

- *The elasticity of sales with respect to total advertising.* Many studies have tried to estimate the elasticity of firm sales to advertising spending. The metastudy by Henningsen et al. (2011) provides a wide range of elasticity estimates from the literature. Their set of estimates is narrowed to one that maps more closely with the model (for example, focusing on the carryover elasticity as opposed to the one-year elasticity, or focusing on

[25] The number of total varieties corresponds to $N \times n$ in the model.

[26] To estimate the growth between the initial and last period in each data set, only the product categories that are present throughout the whole period are used. Since the RMS data starts in 2006, the implied growth rate in the RMS sample is extrapolated back to cover the entire 1995-2015 period, assuming a constant yearly growth rate.

[27] Note that given the *growth in the number of total varieties ($N \times n$)* and the *growth in the number of product lines (N)*, a value for the *growth in the number of varieties per product line* (n) between 1995 and 2015 can be retrieved.

aggregate advertising elasticity as opposed to TV or other specific media advertising). This elasticity gives a single target of 0.2.

- *The elasticity of total varieties with respect to digital advertising.* This single target comes from the causal empirical evidence discussed in Section 2.2. The estimated coefficients using the sample with all product categories and using the most stringent specifications from Table C.11 in Appendix C are used (the average of columns 3 and 6 in Panel B). The resulting elasticity is 0.84.

Table 4.1 lists the parameter values for the model.

Table 4.1: Parameter Values

Parameter	**Value**	**Description**	**Fitting Criteria**
Tastes–generic			
θ	0.549	weight	Eq (4.2)
Tastes–specialized			
κ	0.196	exponent	a priori info
χ	0.975	constant, match quality	Eq (4.2)
λ	2.027	slope, match quality	Eq (4.2)
Production function, generic			
x_{95}, x_{15}	1.052, 1.392	TFP, 1995/2015	
α	0.804	exponent	a priori info
Production costs, specialized			
Ξ_{95}, Ξ_{15}	0.013, 0.010	constant	Eq (4.1)
η	0.050	exponent	Eq (4.1)
ϕ_{95}, ϕ_{15}	0.089, 0.091	fixed cost, 1995/2015	Eq (4.1)
Advertising–digital			
A_{95}, A_{15}	1.758, 0.818	cost shifter, 1995/2015	Eq (4.1)
ζ	3.851	exponent	Eq (4.2)
Advertising–traditional			
B	0.136	cost shifter	Eq (4.1)
ν	3.851	exponent	Eq (4.2)

5 Model Results

The results of the calibration exercise are shown in Table 5.1. As can be seen, the model matches exactly the total-advertising-to-GDP ratio, the ratio of digital-to-traditional advertising spendings (both in 1995 and 2015), the growth rates in both the number of varieties per product line and product lines between 1995 and 2015, and the growth rate in per capita income. It does a nice job matching the size of the specialized sector in 1995 and 2015 and mimicking the elasticity of sales with respect to total advertising. It does a reasonable job replicating the elasticity of varieties with respect to digital advertising.

Some other miscellaneous features of the model are now reported. As the number of varieties increases between 1995 and 2015, so do prices. Prices of specialized products move up by 9% because varieties now match better consumers' tastes and wages have risen. From the solution to the firm's problem presented in Appendix E –see equation (E.4) – the price for a specialized variety is a markup over marginal costs given by

$$p = (\frac{1/\kappa}{1/\kappa - 1}) w \Xi (\frac{n^{\eta}}{\eta}). \tag{5.1}$$

The increase in varieties per product line, n, causes prices, p, to rise by 4% due to the higher production costs of offering more varieties. The growth in varieties occurs because firms are now better able to target consumers' tastes. Marginal costs also increase because labor becomes more expensive due to a rise in wages, w. This causes prices to move up by 33%. Offsetting this is process innovation as reflected by the decline in Ξ. On this account, prices fall by 28%. These three factors taken to together result in prices moving up by 9%. The price of a variety measured in terms of time, p/w, drops due to process innovation.

The ratio of profits (before entry costs in the specialized sector and accounting for profits in the generic sector) to GDP is 17% across both time periods. The ratio of entry costs plus advertising to specialized products sales in the model is 20% across both periods compared with 16% for listed firms in the United States for selling, general, and administrative expenses (which includes advertising) for 1995 and 2015.

As a prelude to the thought experiments undertaken in Section 5, the calibration exercise suggests that there was a huge reduction in the cost of digital advertising, as measured by the 53% drop in A over the 20 year period. While formally the parameter values are jointly determined by all of the targets, intuitively the drop in A is identified by the increase in the ratio of digital-to-traditional advertising. There was also process innovation in the production of varieties as reflected by the 24% fall in Ξ. Entry costs into specialized products production rose, however; that is, ϕ increased by 3%. The identification of Ξ and ϕ follows from the increase in the number of varieties and product lines, respectively. Last, there was technological progress in the generic goods sector as shown by the 32% increase in x. This fact is pinned down by the observed increase in income.

The analysis now addresses the impact of the four drivers of the transition between 1995 and 2015: (i) technological progress in digital advertising, (ii) process innovation in specialized varieties production, (iii) heightened entry costs into product lines, and (iv) technological progress in generic goods production.

Table 5.1: Results, Data vs Model

Variable	Data		Model	
	1995	**2015**	**1995**	**2015**
Ad Spending-to-GDP, %	2.2	2.2	2.2	2.2
Digital-to-Traditional Ad Spending, %	2.3	96.6	2.3	96.6
Size of Specialized Sector, %	53.0	59.0	52.0	61.2
	1995-2015		**1995-2015**	
Growth in Varieties per Product Line (n), %	115		115	
Growth in Product Lines (N), %	17		17	
Sales Elasticity	0.20		0.17	
Total Variety Elasticity	0.84		0.78	

5.1 The Impact of Digital Advertising

Technological progress in digital advertising is modeled as a reduction in A. This lowers the cost of digital advertising, as can be seen from (3.9). To analyze the effect of the advent of digital advertising, suppose that A remains at its 1995 level. The other parameters remain at their values in the baseline calibration. The results of this thought experiment are reported in Table 5.2. First, the ratio of digital-to-traditional ad spending drops significantly to 36%, relative to the baseline level of 97%. This is not surprising, because the cost advantage of digital advertising has been severely reduced. The advertising-to-GDP ratio rises slightly to 2.3%. Traditional advertising expands to fill the gap left by the cut in digital advertising spending.

The growth in varieties per product line is much more muted, about 67%. Still, what drives this growth in varieties? There are other factors at work, process innovation and technological progress in the production of generic goods. The demand for a given variety increases with a consumer's income, via $\hat{y}$, as can be seen from equations (3.7) and (3.8). Other things equal, this increases a firm's profits from selling a variety and encourages it to introduce new varieties within its product line. It also stimulates entry by other firms. The growth in new product lines is now slightly smaller, about 15% relative to 1995. Prices of specialized products rise by roughly 7% as opposed to 9% in the benchmark model, due to the lower number of varieties produced.

Table 5.2: Experiment, No Technological Progress in Digital Advertising

Variable	**Model, 2015**	
	Benchmark	**Fixed** A
Ad Spending-to-GDP, %	2.2	2.3
Digital-to-Traditional Ad Spending, %	96.6	36.3
Size of Specialized Sector, %	61.2	60.8
Growth in Varieties per Product Line (n), %	114.5	67.4
Growth in Product Lines (N), %	17.0	15.0
Growth in Prices, %	8.9	7.3
Growth in Wages, %	38.4	38.2
Growth in Generic Consumption, %	9.7	10.6
Growth in Consumption per Variety, %	35.5	35.2
Equivalent Variation, %	1.25	

What is the welfare gain from the advent of digital advertising? Welfare, W, is given by

$$W = \theta \ln c + (1-\theta) \int_0^M S(j)^\kappa \frac{q(j)^{1-\kappa}}{1-\kappa} dj,$$

where c, $q(j)$, $M \equiv |\mathcal{M}|$, and $S(j)$ are the allocations arising from the model's general equilibrium under some particular scenario. In a symmetric equilibrium welfare can be written as

$$\begin{aligned} W &= \theta \ln c \\ &\quad + \frac{(1-\theta)^{1/\kappa}}{1-\kappa} N(\hat{y}/p)^{(1-\kappa)/\kappa} \Big[a_d 2n \int_0^{1/(2n)} \sigma(\omega) d\omega + a_t(1-a_d) 2 \int_0^{1/2} \sigma(\omega) d\omega \Big], \end{aligned}$$

where equations (3.6), (3.12), and (3.13) have been used. The second line gives the utility from specialized product varieties. With probability a_d, a person's consumption from specialized varieties derives from a digital ad. Match quality for this component of varieties consumption is uniformly distributed on the interval $[0, 1/(2n)]$. From Proposition 2, the average match quality associated with digital advertising is $2n \int_0^{1/(2n)} \sigma(\omega) d\omega = \sigma_d(n)$. Similarly, with probability $a_t(1-a_d)$, varieties consumption arises from a traditional ad. In this case, match quality is distributed uniformly on the interval $[0, 1/2]$ with an average match quality of $2 \int_0^{1/2} \sigma(\omega) d\omega = \sigma_t$, as stated in Proposition 1.

Denote the level of welfare obtained in the 2015 benchmark equilibrium by W^A. To calculate the welfare gain from digital advertising, suppose that there was no technological progress in digital advertising. That is, set $A_{15} = A_{95}$, and keep the remaining parameters at their values in the baseline calibration. Denote the level of welfare under this scenario by W^B. Now, by what (net) scaling factor would you have to boost generic consumption in regime B to make the person as well off as in A? This is the equivalent variation, $\texttt{ev}$; it must solve

$$\texttt{ev} = \exp(W^A - W^B) - 1.$$

The equivalent variation is estimated to be 1.25% of generic goods consumption. So, the advent of digital advertising improved consumer welfare.

5.2 Process Innovation in Specialized Products Production

In the analysis, the operating cost of producing specialized products falls due to technological progress. Specifically, Ξ drops over time. Theoretically speaking, process innovation does not change a firm's advertising strategy or the number of varieties that it sells. This suggests that technological progress in digital advertising is a key driver of the evolution of a_d, a_t, and n. Not surprisingly, the quantities consumed of a variety increase with process innovation. This transpires because their time price, p/w, becomes less expensive due to the reduction in marginal costs of production.

Proposition 3. *(Neutrality of process innovation on advertising and targeting) Let Ξ rise to $\xi\Xi$. Then, the demands for specialized varieties resulting from digital and traditional ads, q_d and q_t, fall by a factor of ξ; i.e., q_d declines to q_d/ξ and q_t shrinks to q_t/ξ. The time price of varieties increases from p/w to $\xi p/w$. The variables a_d, a_t, and n remain constant.*

Proof. See Appendix E. □

Entertain now the counterfactual that there is no process innovation in specialized products production. The results are displayed in Table 5.3. The number of product lines now increases by just 2% between 1995 and 2015. By substituting the pricing equation (5.1) into the demand for a variety equations (3.7) and (3.8), it can be seen that

$$q_d(n,p) = \sigma_d(n)\{(1-\theta)\hat{y}(\frac{1/\kappa - 1}{1/\kappa})/[w\Xi(\frac{n^\eta}{\eta})]\}^{1/\kappa} \text{ and } q_t(p) = \sigma_t\{(1-\theta)\hat{y}(\frac{1/\kappa - 1}{1/\kappa})/[w\Xi(\frac{n^\eta}{\eta})]\}^{1/\kappa}.$$

Thus, process innovation, characterized by a fall in Ξ, stimulates the demand for a variety through lower prices. When process innovation is turned off, other things equal, demand falls and firms earn less profits. The zero-profit condition (3.16) then implies that the number of firms selling product lines will drop. Other things are not equal, though. Incomes still rise due to technological progress in the generic goods sector. This drives up wages so that there is now a much larger rise in prices of 38% relative to 1995, due both to the resulting higher labor costs and a the lack of process innovation. This stifles the increase in product lines due to higher incomes. As a result of turning off process innovation, the share of the specialized-products sector in GDP falls to 52%, much smaller than in the 2015 benchmark economy. Interestingly, the growth in varieties remains the same. This is the result of technological innovation in digital advertising. The ratio of digital-to-traditional ad spending does not change, staying at 97%.

While Proposition 3 establishes that process innovation by itself has no impact on the number of varieties per product line, n, there is a synergy effect between A and Ξ. Technological process in digital advertising alone causes explains 41% of the growth in the number of varieties per product line.[28] Turning off process innovation does not affect the number of varieties. Both forms of technological progress taken together explain 99% of the increase in varieties. So, the combined effect of both forms of technological progress is much bigger than the sum of them independently.

Table 5.3: Experiment, No Process Innovation in Specialized Products Production

Variable	**Model, 2015**	
	Benchmark	**Fixed Ξ**
Ad Spending-to-GDP, %	2.2	1.9
Digital-to-Traditional Ad Spending, %	96.6	96.6
Size of Specialized Sector, %	61.2	52.4
Growth in Varieties per Product Line (n), %	114.5	114.5
Growth in Product Lines (N), %	17.0	2.1
Growth in Prices, %	8.9	37.7
Growth in Wages, %	38.4	32.5
Growth in Generic Consumption, %	9.7	31.3
Growth in Consumption per Variety, %	35.5	2.6

5.3 Increase in the Fixed Entry Cost of Specialized Products Production

The model's calibration suggests that the fixed entry cost associated with specialized products production increased; that is, ϕ rose. Suppose that this factor is shut down – see Table 5.4.

[28] The contribution of a variable y to the growth in a variable z is measured as

$$\frac{z_{2015}^{\text{Benchmark}} - z_{2015}^{\text{Fixed } y}}{z_{2015}^{\text{Benchmark}} - z_{1980}^{\text{Benchmark}}} \times 100\%,$$

where $z_{2015}^{\text{Fixed } y}$ denotes the 2015 level of z when y is held fixed. Here $z = n$ and $n \times N$ and $y = A, \Xi, x,$ and ϕ.

Not surprisingly, the number of product lines now climbs over time by 20%. It costs less to enter into the production of new product lines. There is a negligible reduction in the growth of varieties per product line relative to the 2015 benchmark economy. The share of the specialized-products sector in GDP rises slightly relative to the 2015 baseline. While a person buys more varieties, they consume less on average of each product line relative to the benchmark, a 32% increase relative to 35% in the benchmark model between 1995 and 2015. There is no synergistic effect between A and ϕ on n, unlike with Ξ.

Table 5.4: Experiment, Change in Fixed Entry Costs

Variable	**Model, 2015**	
	Benchmark	**Fixed** ϕ
Ad Spending-to-GDP, %	2.2	2.2
Digital-to-Traditional Ad Spending, %	96.6	96.4
Size of Specialized Sector, %	61.2	61.4
Growth in Varieties per Product Line (n), %	114.5	114.2
Growth in Product Lines (N), %	17.0	20.3
Growth in Prices, %	8.9	9.0
Growth in Wages, %	38.4	38.6
Growth in Generic Consumption, %	9.7	9.3
Growth in Consumption per Variety, %	35.5	32.5

5.4 Technological Progress in the Generic Goods Sector

For the last experiment, shut down technological progress in the generic goods sector; i.e., the increase in x. Before proceeding, it is worthwhile noting that technological progress in the generic goods sector is essentially neutral. All it really does is to increase the consumption of generic goods while having no impact on the allocations in the specialized-products sector. Wages and prices will rise however. Again, this illustrates the primacy of technological advance in digital advertising (and to a lesser extent process innovation) for determining the evolution of advertising and the number of varieties.

Proposition 4. *(Neutrality of technological progress in the generic goods sector on advertising, the number of product lines, and varieties per product line) Suppose that due to technological progress in the generic goods sector, x rises to ξx. Then, z increases to ξz, for $z = c, p, w, \hat{y}$. The variables a_d, a_t, n, M, N, and the q's all remain constant.*

Proof. See Appendix E. ☐

Consider now the experiment where x is kept at its 1995 level – Table 5.5. The price of specialized varieties falls dramatically, about 18% relative to 1995, because the growth in wages is now much lower driving down the marginal cost of production. As a consequence of lower wages, people are poorer. Both of these factors lead to a dramatic decline in generic goods consumption. Little else changes, however. It may seem odd that the share of the specialized-products sector in GDP remains fixed but note that the relative price of a variety falls by the amount needed to keep the size of the sector constant. Unlike with process innovation, there is no synergistic effect between A and x on n.

Table 5.5: Experiment, No Technological Progress in Generic Goods Production

Variable	**Model, 2015**	
	Benchmark	**Fixed x**
Ad Spending-to-GDP, %	2.2	2.2
Digital-to-Traditional Ad Spending, %	96.6	96.65
Size of Specialized Sector, %	61.2	61.2
Growth in Varieties per Product Line (n), %	114.5	114.5
Growth in Product Lines (N), %	17.0	17.0
Growth in Prices, %	8.9	-17.6
Growth in Wages, %	38.4	4.7
Growth in Consumption of Generic Goods, %	9.7	-17.1
Growth in Consumption per Variety, %	35.5	35.5

5.5 Synopsis

A summary of the thought experiments is presented in Figure 5.1. The upper panel shows the separate impact of each of the four sources of technological change, A, Ξ, x, and ϕ, on the number of total varieties, $n \times N$.[29] As can be seen, the reduction in the cost of digital advertising, as reflected by a fall in A, has the biggest impact on the number of total varieties by increasing both the number of varieties per product line, n, and product lines, N. Process innovation also has a significant effect on the number of total varieties by reducing manufacturing costs. A fall in Ξ causes $n \times N$ to move up solely through its impact on the number of product lines, N – Proposition 3. The number of total varieties is invariant to technological progress in the generic goods sector – Proposition 4. Last, an increase in the fixed entry cost associated with producing, ϕ, has a minor influence, mostly through a decline in N. The numbers do not add up to 100% because there is a synergy between the four sources of technology change, particularly between A and Ξ.

[29] See footnote 28 for how this is measured.

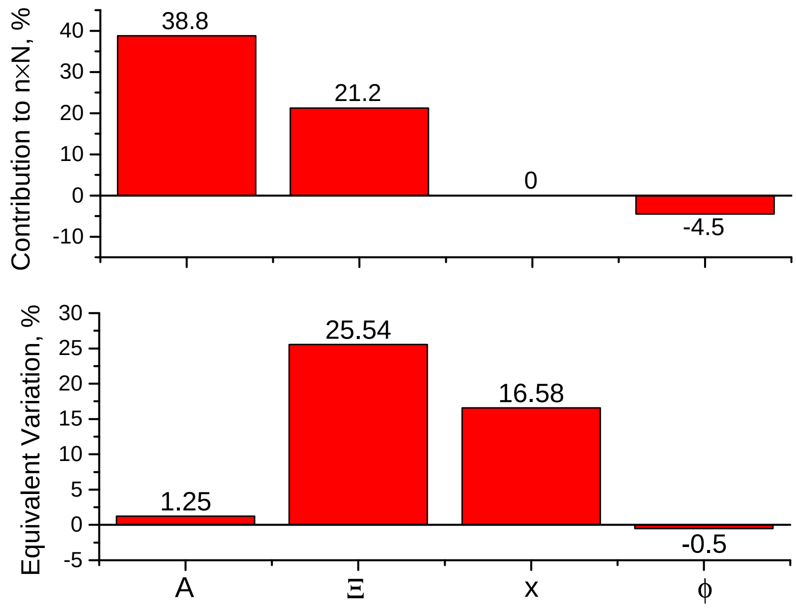

Figure 5.1: Upshot of Thought Experiments.

Note: The upper panel exhibits the individual contribution of each source of technological change to the increase in the number of total varieties, $n \times N$. The lower panel displays the equivalent variation associated with each form of technological change.

The lower panel shows the welfare gains resulting from each source of technological change. Process innovation has the biggest benefit for consumers, as measured by the equivalent variation. It reduces the production cost of varieties. Recall that Ξ dropped by 24%. From Proposition 3 this would lead to prices falling by 24% and the consumption of each variety increasing by the same amount, ceteris paribus. So, you would expect process innovation to have a large effect on welfare. For similar reasons, so does technological progress in generic goods sector, as reflected by an increase in x. Recall that x rose by 32%. Proposition 4 states that, other things equal, this would cause wages and the consumption of generic goods to rise by 32%. Again, you would expect this to have a large impact on welfare. The small increase of 3% in the fixed entry cost, ϕ, has a negligible impact on welfare.

As was mentioned, the digital advertising revolution has a benefit for consumers because it stimulates growth in both the number of varieties and product lines. This results from the improved ability of firms to offer varieties that are better aligned with consumer tastes. The effect of technological progress in digital advertising on welfare has a much smaller impact than either process innovation or technological advance in the generic goods sector.

Recall that there is strong synergistic effect between technological progress in digital advertising, A, and process innovation, Ξ, on the number of varieties per product line, n, and hence on $n \times N$. The synergistic effect on welfare is absent, though. Process innovation alone has an equivalent variation of 25.5%. Adding technological progress in digital advertising only raises this number to 27.0%.

The model developed here does not consider the channels through which the rise in digital advertising could negatively impact welfare. One such channel could operate through digital advertising's effect on firms' market power, competition, and concentration. Also the analysis does not consider advertising's use as a persuasive tool, but focuses on its informative role. Congestion in advertising is not studied either. Although not in the context of the digital advertising, prior literature has highlighted these effects of advertising on firms and consumers (Bagwell, 2007). The focus of this paper is on a new channel through which digital advertising has implications for the product varieties offered to consumers, and hence the welfare numbers should be interpreted as such.

6 Conclusion

The information age ushered in an advertising revolution. Information technology allows advertising to be more precisely targeted toward consumers who place a higher value on particular products sold by firms. As targeting becomes more refined, firms' incentives increase to produce more varieties of a product that can be sold at higher prices relative to more generic versions. This is the hypothesis entertained here.

The hypothesis is addressed on two fronts. First, an econometric investigation is undertaken using multiple micro-level data sets. There is a positive relationship between changes in the number of varieties and changes in digital ads, controlling for various factors. To establish causality, lightning strikes are used as an instrument in the regression analysis. Lightning strikes affect household internet access and hence the "treatment" by digital ads. The regression analysis suggests a statistically significant and economically meaningful causal impact of digital advertising on the number of varieties.

Second, a model is developed where firms sell specialized varieties to consumers. To sell its varieties, a firm must advertise either through digital or traditional ads. Traditional advertising is broad-brushed, painting all varieties within a firm's product line. Digital advertising has a narrower scope and targets customers with preferences for specific varieties. Two propositions are presented, establishing that digital advertising, on average, allows consumer preferences to be better matched with varieties than traditional advertising, and this advantage increases when more varieties are offered. In the model, digital advertising becomes more efficient over time due to improvements in information technologies. The developed model is calibrated to the years 1995 and 2015 by targeting several stylized facts regarding the new advertising age.

The analysis suggests that directed digital advertising led to an increase in the total varieties offered by firms. Economic welfare increased as well. The outcomes of process innovation, rising entry costs in specialized varieties production, and technological progress in generic goods production are also addressed.

This study explores a new channel through which the new age of advertising has an effect on the economy: the expansion of varieties offered by firms. Going forward, it would be interesting to understand how this channel interacts with various firm characteristics. Do younger firms with low market shares benefit most from the improved ability to target with digital advertising? How does digital advertising affect the pricing of products? Digital advertising might spur greater price competition, but it can also increase firms' market powers. What are the efficiency properties of models with digital advertising, and what is the optimal tax/subsidy policy? How should online data privacy concerns be weighed against the gains online advertisers realize by learning more about consumers? These are just some of the questions economists and regulators need to understand entering into a more and more data-driven digital era.

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.

Anderson, Monica, "Mobile technology and home broadband 2019," 2019.

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.

Bagwell, Kyle, "The Economic Analysis of Advertising," in M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, Vol. 3, Amsterdam: Elsevier B.V., 2007, pp. 1702–1844.

Bar-Isaac, Heski, Guillermo Caruana, and Vicente Cu nat, "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, Jesse Perla, and Pau Roldan-Blanco, "A Model of Product Awareness and Industry Life Cycles," Working Paper 2023.

— , — , **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.

Dinlersoz, Emin, Nathan Goldschlag, Mehmet Yorukoglu, and Nikolas Zolas, "On The Role of Trademarks: From Micro Evidence to Macro Outcomes," Working Papers 23-16, Center for Economic Studies, U.S. Census Bureau March 2023.

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.

Greenstein, Shane, "The basic economics of internet infrastructure," *Journal of Economic Perspectives*, 2020, *34* (2), 192–214.

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.

Henningsen, Sina, Rebecca Heuke, and Michel Clement, "Determinants of Advertising Effectiveness: The Development of an International Advertising Elasticity Database and a Meta-Analysis," *BuR Business Research Journal*, 2011, *4* (2), 193–239.

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.

Ma, Yueyuan, "Specialization in a Knowledge Economy," *Available at SSRN 4052990*, 2022.

Marto, Ricardo, "Structural Change and the Rise in Markups," *Working paper, mimeo*, 2023.

Menzio, Guido, "Optimal Product Design: Implications for Competition and Growth Under Declining Search Frictions," *Econometrica*, 2023, *91* (2), 605–639.

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

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.

Shen, Shihan, "Customer Acquisition, Rising Market Concentration and US Productivity Dynamics," Working Paper 2023.

A Data Appendix: Firm-level Advertising and Varieties

A.1 Kantar Media 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 new or existing media channels become more important (e.g., new TV stations, new high-traffic websites). 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 exhibit ('95); internet display ('01); internet search ('10); online video ('13); mobile web ('15). The numbers in the parentheses refer to the year the data became 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 report 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 captured by 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 Media by media channel. Newspapers and magazines show the steepest decline over time. 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 Media to other government-based statistics. As seen, official estimates of advertising expenditure vary based on the data source: the US Census Bureau or IRS. The Census estimates are revenue-based and are collected by the US Census Annual Survey where establishments (including advertising agencies

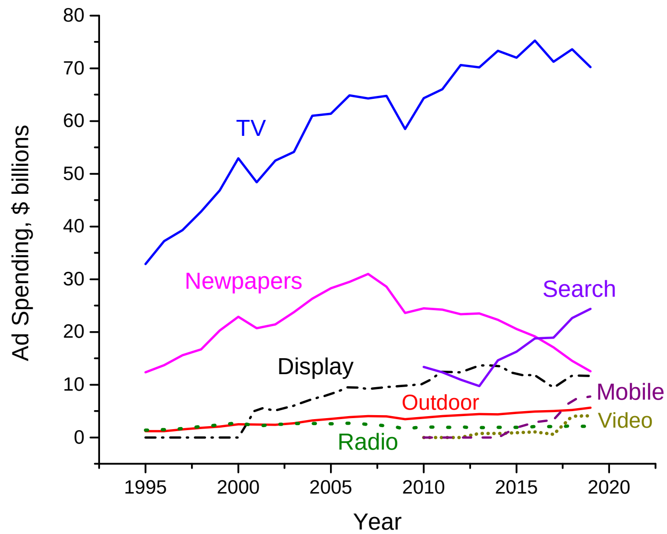

Figure A.1: Advertising Spending by Media Channel.

Note: Total advertising spending by type of media channel. Source: Kantar Media Ad Spend Data,. 1995-2019.

and platforms) 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 the *Purplemotes* blog.[30] Figure A.2 plots these series together with the series assembled here from Kantar Media data for the overlapping time period, 1998-2006. Kantar Media data constitute 40% to 51% of aggregate advertising expenditure estimates from the US Census Bureau over time, and 30% to 36% of aggregate expenditure estimates from the IRS.

A.2 Product Varieties in Kantar Media

Kantar Media data provide information on the number of advertised varieties. The number of advertised varieties is lower than the number of all varieties companies offer in the market for two reasons. Kantar Media data do not capture all advertised varieties in the economy. Data have lower coverage for sectors that are not the focus of the current analysis: business-to-business advertising and ads not related to consumer products, such as ads by various

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

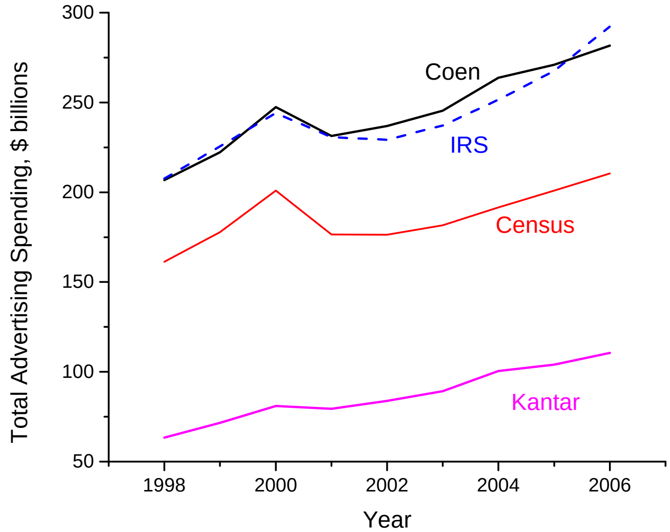

Figure A.2: Kantar Advertising Spending Relative to the Aggregate US Statistics.

Note: Total advertising spending in AdSpend by Kantar Media compared to the aggregate advertising expenditure estimates from the US Census Bureau, 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, 1998-2006. Sources: Coen, Kantar Media, the IRS, and the US Census Bureau.

service providers, entertainment, government, and education.[31]

A.3 Matching Kantar Media to NETS

The Kantar Media data are combined with data on firms' employments and sales from the National Establishment Time Series (NETS) Database. 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 Media to NETS involves defining relevant company names in each data set and then using name-matching routines to link names across the two data sets. In Kantar Media, 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.

[31]The above description of the media coverage makes it clear that almost all advertised products are captured at the extensive margin, while the full intensive margin of ad spending on those products is not. Still, very niche consumer products advertised on unpopular websites or very specialized media outlets are not captured along the extensive margin.

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 identifier, establishment name, and the ultimate headquarter identifier. In NETS, 97% of firms are associated with only one establishment identifier 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.[32] In the last step, clean the company names from Kantar Media 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.[33]

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 Media-NETS matched data set provides information about the number of advertised varieties (product names, brands, sub-brands), product categories operated (industry, major, subcategory), traditional- and digital-ad expenditures, and employment and sales for each company during the years with overlapping coverage in Kantar Media and NETS. Table A.1 summarizes the data.

B Data Appendix: Causal Analysis

B.1 RMS Nielsen

Nielsen RMS raw data from 2006 to 2020 are harnessed, and data for the period 2008-2018 are used.[34] For each county $\times$ year $\times$ product category (referred to as the product module in the original data set), the following measures of the number of varieties are computed (in the order from the most detailed to more aggregated variety definition):

- the number of barcodes (the combination of UPC + UPC version);
- the number of characteristics combinations as in Kaplan and Menzio (2015) (Agg1 – firm $\times$ brand $\times$ same observable characteristics $\times$ UPCdesc; Agg2 – firm $\times$ brand $\times$ same observable characteristics; Agg3 – firm $\times$ brand and same observable characteristics, except for size);

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

[33] For computational reasons, the match is performed for every 5-year periods (1995, 2000, 2005, 2010, 2015), and where necessary, employment is interpolated in gap years.

[34] For more detail about the data, see Argente et al. (2021b).

Table A.1: Summary Statistics

		Mean	*Median*	*St. Dev.*
	FIRM × YEAR LEVEL – KANTAR MEDIA			
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
FIRM × YEAR LEVEL - KANTAR MEDIA 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 Kantar Media and Kantar Media-NETS data sets. Kantar Media data include ads only for products. Ads related to services and amusement, retail (store promotions), automotive dealers, financial, government/political/organizations, schools, restaurants, hotels, and other services, as well as general ads about corporate promotions and recruiting are excluded. Additional details on the variables are in Data Appendix A.

- the number of brands.

To build measures of the number of varieties, first a products data set at the barcode level with information on product characteristics (e.g. size, brand, firm ownership) is obtained. The data set has 1,966,044 observations. Second, data are collapsed at the county × product category (module) × year level, while counting distinct varieties under the three definitions above, as well as computing total sales and quantities. The location of the stores is used to determine the location of the varieties sold. The baseline data use varieties across a balanced set of stores present throughout all years. However, for robustness, the unbalanced data with varieties across all stores are also considered.

For the robustness analysis, information on retail chains and firm locations are also used to construct alternative measures of product varieties. First, firms/chains are defined as multi-state firms/chains if the number of states in which they sell is in the top quartile of the distribution of the number of states firms/chains sell in. Next, the number of varieties in a county, excluding local firms, is measured as the number of varieties offered by multi-state firms whose headquarters are not located in the county. Similarly, the number of varieties in a county, excluding local retail chains, is measured as the number of varieties offered by multi-state chains.

B.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 them to report the type 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) with speeds above 200kbps per every 1,000 housing units.[35] This information is reported at the Census tract level (73k in the United States) by year for every year from 2008 to 2018 (as of February 2022) and aggregated to the county level by computing the county averages weighted by the number of housing units in each Census tract as of 2010. The share of households with residential fixed connection is reported in five bins: [0,0.20], [0.21-0.40], [0.41-0.60], [0.61-0.80], and [0.81-1]. Hence, households' internet access variable used in the baseline regressions is a categorical variable ranging from 1 to 5. Its key advantage is a consistent way of reporting internet access throughout the 2008-2018 period. 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 the robustness exercises.

The second data set is "Fixed Broadband Deployment Data" on the advertised speeds (for residential and commercial use) offered by providers.[36] The data are at the technology-provider-Census block level (11 million Census blocks in the United States), with technology falling under the categorizations made by the FCC (different types of DSL, cable, satellite, fiber, etc). Collection of raw speed data was initiated in 2014 and hence the availability of this data set begins in 2014 and has been released through 2020 (as of February 2022). 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 the 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 the number of housing units in 2010. The alternative measures are used for robustness. Additionally, the information from the Census Bureau is used to harmonize pre-2010 Census boundaries to 2010 Census boundaries.

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

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

B.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 are sourced from the County and State Summaries. The data begin in 1986 and are available for every year until 2012 (inclusive). Data record the number of lightning strikes by county for every individual day of the year. These data are combined with data on the size of US counties from the Census Bureau to get measures of lightning strikes per square mile per year.

B.4 Additional Data Sets

Data on various measures of personal income and population are sourced from the Bureau of Economic Analysis (BEA) regional economic accounts. Data on some demographics (the share of teenagers, the share of young, and the share of seniors) and the average population density per square foot are defined at the county level and are sourced from the 2010 US Census. Finally, data on the population share with college degrees or higher and categorical variables for urban-rural status are defined at the county level, sourced from the US Department of Agriculture (USDA).

Table B.1 provides summary statistics for the baseline data set. The data cover the period 2008-2017.

Table B.1: Summary Statistics (cont)

Total Observations				1,978,180
Unique counties				2,259
Unique Periods				2
Unique product categories				602
	Source	*Mean*	*Median*	*St. Dev.*
		——County × Period——		
Household internet access	FCC	3.54	3.60	0.73
Lightning Strikes per Sq. Mile one-period lagged	NLDN	9.29	8.94	5.72
Population (1,000s)	BEA	137.26	42.04	376.67
Income Per Capita ($1,000s)	BEA	37.94	35.89	10.77
Share of Households in Urban Areas (time-invariant)	BEA	0.50	0.50	0.29
Share of Teenagers (time-invariant)	Census	0.19	0.19	0.03
Share of Young (time-invariant)	Census	0.13	0.13	0.03
Share of Senior (time-invariant)	Census	0.15	0.15	0.04
Share with College or More (time-invariant)	USDA	0.23	0.20	0.10
Average Density of Population per Sq. (time-invariant)	Census	316.23	69.55	2,015.27
Categorical Variables of Urban-rural Status (time-invariant)	USDA	4.28	4	2.44
Households with Access to Technology DSL (time-invariant)	FCC	0.83	0.86	0.14
Households with Access to Technology Cable (time-invariant)	FCC	0.69	0.75	0.27
Households with Access to Technology Fiber (time-invariant)	FCC	0.19	0.09	0.24
	——County × Period × Product-Category——			
Number of Products Barcodes	RMS	40.16	10	99.64
Number of Products Aggregation 1	RMS	39.53	10	96.98
Number of Products Aggregation 2	RMS	35.85	9	89.76
Number of Products Aggregation 3	RMS	29.61	7	70.11
Number of Brands	RMS	11.44	4	23.05

Note: Summary statistics for the baseline data set covering the period 2008-2017. The variables at the county×year use data from the FCC, NLDN, and BEA. The variables at the county×year×product category are from Nielsen RMS. The distinct definitions of products follow Kaplan and Menzio (2015). Additional details on the variables are in the Data Appendix.

C Additional Empirical Results

Table C.1: Varieties and Digital Ads, Product-Category Level. Robustness

	Log Products			Log Brands		
Panel A						
	subcat.	major	industry	subcat.	major	industry
Log Digital Ads	0.013***	0.007***	0.024**	0.011***	0.008***	0.013**
	(0.001)	(0.003)	(0.010)	(0.001)	(0.002)	(0.007)
R^2	0.983	0.990	0.992	0.988	0.993	0.994
Observations	9,800	2,514	577	9,800	2,514	577
Panel B						
	Log Products			*Log Brands*		
	subcat.	major	industry	subcat.	major	industry
Digital Ads Ratio	0.009***	0.003	0.020**	0.008***	0.005**	0.010
	(0.001)	(0.003)	(0.010)	(0.001)	(0.002)	(0.006)
R^2	0.983	0.991	0.993	0.988	0.993	0.994
Observations	10,114	2,518	577	10,114	2,518	577

Note: Regressions of varieties on digital-ad spending in product categories over time controlling for log traditional-ad spending in Panel A. Regressions of varieties on the ratio of digital-ad spending to total ad spending in Panel B. All regressions control for the log number of firms in product categories over time, product category, and year fixed effects. Regressions are weighted by the number of firms in product categories over time. Product variety: products and brands. Product categories: subcategory, major, and industry. Robust standard errors are in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.2: Varieties and Digital Ads, Firm Level. Robustness

	Log Products		Log Brands	
Panel A				
	Cross-firms	Within-firms	Cross-firms	Within-firms
Log Digital Ads	0.128***	0.048***	0.131***	0.040***
	(0.003)	(0.002)	(0.003)	(0.002)
R^2	0.571	0.927	0.514	0.919
Observations	16,492	14,834	16,492	14,834
Panel B				
	Log Products		*Log Brands*	
	Cross-firms	Within-firms	Cross-firms	Within-firms
Digital Ads Ratio	0.029***	0.009***	0.056***	0.016***
	(0.003)	(0.002)	(0.002)	(0.002)
R^2	0.588	0.915	0.506	0.900
Observations	35,880	27,536	35,880	27,536

Note: Regressions of varieties on digital-ad spending in firms over time controlling for log traditional-ad spending in Panel A. Regressions of varieties on the ratio of digital-ad spending to total ad 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 are in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

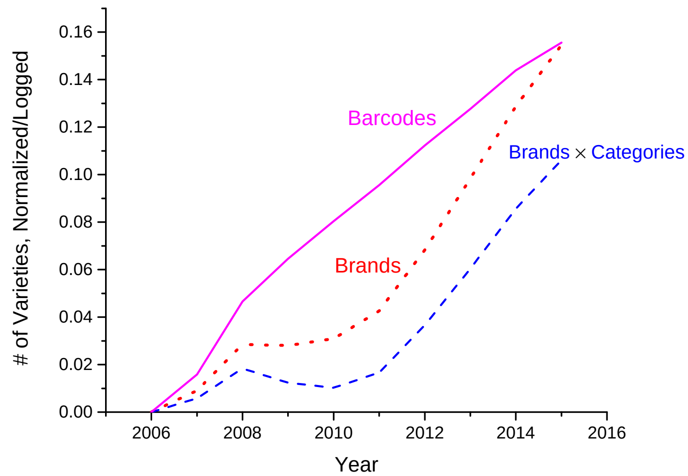

Figure C.1: Product Varieties over Time - RMS Nielsen

Note: 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, and brands × category (module).

Table C.3: Variation in Varieties Across Counties

Panel A: Log Products	*Variance*	*R-Squared*			
All products	2.869	0.593	0.815	0.720	0.815
Products, excluding local chains	2.873	0.597	0.808	0.707	0.808
Products, excluding local firms	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
Brands, excluding local chains	1.712	0.605	0.794	0.707	0.794
Brands, excluding local firms	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

Note: The variance and R^2 from regressions running the number of varieties (products in Panel A and brands in Panel B) in a product category-year-county on year, category, and county fixed effects. The specifications control for population (in logs) and income per capita (in logs). Measures of varieties excluding local firms and local chains are described in Section B.1.

Table C.4: Relationship Between Internet and Lightning: Exploring Lags

	Household Internet		
	(1)	(2)	(3)
Lightning Strikes (t-1)	-0.005***	-0.002***	0.000
	(0.001)	(0.001)	(0.001)
Lightning Strikes (t-2)	-0.006***	-0.002***	-0.000
	(0.001)	(0.001)	(0.001)
Lightning Strikes (t-13)	-0.003***	-0.001	-0.000
	(0.001)	(0.001)	(0.001)
Lightning Strikes (t-4)	-0.002**	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Lightning Strikes (t-5)	-0.002**	-0.001	0.001
	(0.001)	(0.001)	(0.000)
Lightning Strikes (t-6)	-0.004***	-0.002***	-0.001
	(0.001)	(0.001)	(0.000)
Lightning Strikes (t-7)	-0.004***	-0.003***	-0.002***
	(0.001)	(0.001)	(0.0010)
Lightning Strikes (t-8)	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.000)
Lightning Strikes (t-9)	-0.002**	-0.002***	-0.001**
	(0.001)	(0.001)	(0.000)
Lightning Strikes (t-10)	-0.002***	-0.001**	-0.002***
	(0.001)	(0.001)	(0.000)
R^2	0.282	0.638	0.880
Observations	24,697	24,653	24,697
Year FE	Yes	Yes	Yes
Year × County Controls	No	Yes	Yes
County FE	No	No	Yes

Note: The estimated coefficients from regressing household internet access at the county × year level on lagged lightning strikes (up to 10 lags) and various controls. Each column varies the set of controls. Column 1 controls for year fixed effects. Column 2 has year fixed effects and county controls (time varying and time invariant). Column 3 includes year and county fixed effects and time-varying country controls. The year× county controls are time-varying population (in logs) and income per capita (in logs). The time-invariant controls are: the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas. The variables are described in Sections 2.2.1 and Data Appendix B. Robust standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.5: First-Stage Results with Yearly Data

	Household Internet		
	(1)	(2)	(3)
Lightning Strikes (lagged)	-0.019***	-0.011***	-0.001***
	(0.000)	(0.000)	(0.000)
Observations (1,000s)	10904	10,879	10,882
Year× Category FE	Yes	Yes	Yes
Year × County Controls	No	Yes	Yes
County × Category FE	No	No	Yes

Note: The estimated coefficients from regression (2.1) with county-year-level data from 2008 to 2018. The instrument is seven-year lagged lightning strikes (see Table C.4). The time × county controls are time-varying population (in logs) and income per capita (in logs). The time-invariant controls are: the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas. The variables are described in Sections 2.2.1 and Data Appendix B. Robust standard errors are shown in parentheses. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.6: Household Internet Access and Varieties. Second-Stage Results with Yearly Data

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(1)	(2)	(3)
Household Internet	1.082***	1.283***	0.140	0.803***	0.949***	-0.072
	(0.036)	(0.070)	(0.174)	(0.027)	(0.051)	(0.129)
Observations (1,000s)	10,904	10,879	10,882	10,904	10,879	10,882
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

Note: The estimated regression coefficients from equation (2.2) with county-product category-level data for the period 2008-2018. The dependent variable in columns 1--3 is barcodes (in logs), and in column 4–5 is brands (in logs). "Household Internet" is an instrumented variable from the first stage (Table C.5). The time-varying county controls are population (in logs) and income per capita (in logs), as used in specifications 2 and 3. The time-invariant county controls are the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas, as used in specification 2. The variables are described in Sections 2.2.1 and B. Robust standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.7: Relationship Between Internet and Lightning: IV Regression Exploring Role of Technology

1st Stage	*Households' Access to Internet Technology*		
	DSL	*Cable*	*Fiber*
Lightning Strikes	-0.001***	-0.003***	-0.000
	(0.000)	(0.001)	(0.001)
Observations	2,255	2,255	2,255
County Controls	Yes	Yes	Yes
1st stage F-stat	11.14	27.04	0.00
2nd Stage	*Household Internet*		
	DSL	*Cable*	*Fiber*
Households with Access to Technology	13.103***	5.998***	1,714
	(3.795)	(1.118)	(119,624)
Observations	2,255	2,255	2,255
County Controls	Yes	Yes	Yes

Note: The instrumental variable estimates from regressing household internet access in a county on the variables measuring access to the different technologies (DSL, cable, and fiber), instrumented by lightning strikes. The regression uses data for 2,255 counties, and the variables are averaged for the period 2013-2018. Each column corresponds to a different technology. The regression includes the following controls: population (in logs), income per capita (in logs), the share of teenagers, share of young, share of seniors, share with college or more, average density of population per square foot, categorical variables of urban-rural status and the share of households in urban areas, as used in specification 2. Standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.8: Household Internet and Varieties: OLS Results

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(1)	(2)	(3)
Household Internet	0.636***	0.106***	0.010***	0.046***	0.074***	0.009***
	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.001)
R^2	0.663	0.717	0.990	0.584	0.723	0.986
Observations (1,000s)	1,978	1,974	1,822	1,978	1,974	1,822
Time × Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Time × County Controls	No	Yes	Yes	No	Yes	Yes
County × Category FE	No	No	Yes	No	No	Yes

Note: The estimated regression coefficients of varieties on household internet access with county-product category-time-level data. Time is defined using 5-year periods: 2008-2012 and 2013-2018. The dependent variable capturing varieties is either the number of products (in logs) or the number of brands (in logs). The variables are described in Section 2.2. The time-varying county controls are the population (in logs) and income per capita (in logs) variables, as used in specifications 2 and 3. The time-invariant county controls are the share of teenagers, share of young, share of senior, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas, as used in specification 2. Standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.9: Household Internet and Varieties: By Downstream Speed Level

Panel A								
	Log Products				*Log Brands*			
Speed	>200kb	>768kb	>3mb	>10mb	>200kb	>768kb	>3mb	>10mb
Household Internet	2.186***	1.650***	3.954***	4.418***	1.588***	1.198***	2.872***	3.209***
	(0.027)	(0.018)	(0.083)	(0.099)	(0.020)	(0.013)	(0.061)	(0.072)
Observations (1,000s)	928	928	928	928	928	928	928	928
Category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period	08-12	08-12	08-12	08-12	08-12	08-12	08-12	08-12
1st stage F-stat	12,226	20,663	2,728	2,308	12,226	20,663	2,728	2,308

Panel B								
	Log Products				*Log Brands*			
Speed	>200kb	>10mb	>25mb	>100mb	>200kb	>10mb	>25mb	>100mb
Household Internet	1.620***	2.026***	1.268***	0.003***	1.261***	1.577***	0.987***	0..002***
	(0.019)	(0.027)	(0.015)	(0.000)	(0.015)	(0.021)	(0.011)	(0.000)
Observations (1,000s)	1,029	1,029	1,029	1,029	1,029	1,029	1,029	1,029
Category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period	14-18	14-18	14-18	14-18	14-18	14-18	14-18	14-18
1st stage F-stat	18,089	11,266	22,575	700	18,089	11,266	22,575	700

Note: The estimated coefficients from the second-stage regressions similar to equation (2.2), but with county-product category-level data for two separate periods in Panels A and B. "Household Internet" is based on connections at various speeds presented in different columns. Panel A relies on data obtained via a Freedom of Information Act (FOIA) request at the FCC (connections with downstream speed of at least 200kbps, 768 Kbps, 3 Mbps, and 10Mbps during the period 2008-2012); Panel B uses publicly available FCC Form 477 data (connections with downstream speed of at least 200kbps, 10Mbps, 25Mbps, and 100Mbps during period 2014-2018). The dependent variables capturing varieties are either the number of barcodes (in logs) or the number of brands (in logs). The variables are described in Section 2.2. The county controls are population (in logs), income per capita (in logs) variables, the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas. Standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.10: Household Internet and Varieties: Alternative Definitions of Varieties

	Log Agg1		*Log Agg2*		*Log Agg3*	
	(1)	(2)	(1)	(2)	(1)	(2)
Household Internet	1.070***	0.582***	1.033***	0.518***	0.991***	0.360***
	(0.008)	(0.050)	(0.008)	(0.048)	(0.007)	(0.045)
Observations (1,000s)	1,974	1,822	1,974	1,822	1,974	1,822
Time × Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Time × County Controls	Yes	Yes	Yes	Yes	Yes	Yes
County × Category FE	No	Yes	No	Yes	No	Yes
1st stage F-stat	88,678	669	88,678	669	88,678	669

Note: The estimated regression coefficients from equation (2.2) with county-product category-time-level data. Time is defined using 5-year periods: 2008-2012 and 2013-2018. The dependent variables capturing varieties, Agg1, Agg2, and Agg3 (in logs), follow Kaplan and Menzio (2015) and are described in Section B.1. "Household Internet" is an instrumented variable from the first stage. The time-varying county controls are population (in logs) and income per capita (in logs). The time-invariant county controls are the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas. The variables are described in Section 2.2.1 and Appendix B. Robust standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.11: Household Internet Access and Product Varieties: Alternative Samples

	Log Products			Log Brands		
Panel A: Food and Health & Beauty Products						
Household Internet	0.873***	0.956***	0.417***	0.649***	0.707***	0.034
	(0.003)	(0.006)	(0.036)	(0.003)	(0.005)	(0.030)
R^2	0.159	0.218	-0.168	0.162	0.229	-0.001
Observations (1,000s)	2,683	2,677	2,496	2,683	2,677	2,496
Panel B: All Product Categories						
	Log Products			*Log Brands*		
Household Internet	0.874***	0.978***	0.384***	0.654***	0.729***	0.092***
	(0.003)	(0.005)	(0.030)	(0.002)	(0.004)	(0.024)
R^2	0.176	0.245	-0.137	0.175	0.249	-0.010
Observations (1,000s)	3,727	3,719	3,481	3,727	3,719	3,481
Time× Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Time× County Controls	No	Yes	Yes	No	Yes	Yes
County × Category FE	No	No	Yes	No	No	Yes

Note: The estimated regression coefficients from equation (2.2) with county-product category-time-level data for two alternative samples. Panel A uses a selected sample that includes Nielsen RMS product modules of food and health and beauty. Panel B includes all product modules in Nielsen RMS. Time is defined using 5-year periods: 2008-2012 and 2013-2018. "Household Internet" is an instrumented variable from the first stage. The time-varying county controls are population (in logs) and income per capita (in logs). The time-invariant county controls are the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas, as used in specification 2. The variables are described in Section 2.2.1 and Appendix B. Robust standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.12: Household Internet Access and Varieties: Unbalanced Sample

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(4)	(5)	(6)
Household Internet	0.626***	0.588***	0.634***	0.494***	0.463***	-0.486***
	(0.003)	(0.006)	(0.170)	(0.002)	(0.005)	(0.133)
Observations (1,000s)	2,333	2,328	2,176	2,333	2,328	2,176
Time $\times$ Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Time $\times$ County Controls	No	Yes	Yes	No	Yes	Yes
County $\times$ Category FE	No	No	Yes	No	No	Yes
1st stage F-stat	159,743	95,326	103	159,743	95,326	103

Note: The estimated regression coefficients from equation (2.2) with county-product category-time-level data for all stores in Nielsen RMS (as opposed to the balanced sample of stores used in the benchmark analysis). Time is defined using 5-year periods: 2008-2012 and 2013-2018. The dependent variables capturing varieties are either barcodes (in logs) or brands (in logs). "Household Internet" is an instrumented variable from the first stage. The time-varying county controls are population (in logs) and income per capita (in logs), used in specifications 2 and 3. The time-invariant county controls are: the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas, as used in specification 2. The variables are described in Section 2.2.1 and Appendix B. The 1st stage F-stat is the Kleibergen-Paap rk Wald F statistic. Robust standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table C.13: Household Internet and Varieties: Excluding Local Retail Chains

	Log Products			*Log Brands*		
	(1)	(2)	(3)	(1)	(2)	(3)
Household Internet	0.417***	0.363***	0.425***	0.235***	0.149***	0.042
	(0.005)	(0.009)	(0.038)	(0.004)	(0.006)	(0.032)
Observations (1,000s)	1,849	1,845	1,689	1,849	1,845	1,689
Time$\times$ Category FE	Yes	Yes	Yes	Yes	Yes	Yes
Time $\times$ County Controls	No	Yes	Yes	No	Yes	Yes
County $\times$ Category FE	No	No	Yes	No	No	Yes

Note: The estimated regression coefficients for equation (2.2) with county-product category-time-level data. Time is defined using 5-year periods: 2008-2012 and 2013-2018. The dependent variables are either barcodes or brands (in logs) in a county $\times$ time $\times$ category sold by multi-state retail chains that sell in the number of states in the top quartile of the distribution of the number of states the retails chains sell in. "Household Internet" is an instrumented variable from the first stage. The time-varying county controls are population (in logs) and income per capita (in logs), as used in specifications 2 and 3. The time-invariant county controls are the share of teenagers, share of young, share of seniors, share with college or higher degree, average population density per square foot, categorical variables for urban-rural status, and the share of households in urban areas, as used in specification 2. The variables are described in Section 2.2.1 and Appendix B. Robust standard errors are shown in parentheses. The ***, **, and * asterisks represent statistical significance at the 1%, 5%, and 10% levels, respectively.

D Consumer's Problem

Each consumer solves the problem

$$\max_{c,\{q(j)\}_0^M} \quad \{\theta \ln c + (1-\theta)\int\limits_0^M S(j)^\kappa \frac{q(j)^{1-\kappa}}{1-\kappa} dj\},$$

subject to

$$c + \int\limits_0^M p(j)q(j)dj = w + \pi \equiv y.$$

Let $\hat{y}$ be the inverse of the Lagrange multiplier attached to the consumer's budget constraint. Solving for the consumer's first-order conditions yields the following two conditions:

$$\frac{\theta}{c} = \frac{1}{\hat{y}}, \tag{D.1}$$

and

$$(1-\theta)S(j)^\kappa q(j)^{-\kappa} = \frac{p(j)}{\hat{y}}, \text{ for all} j. \tag{D.2}$$

Multiplying the last equation by the quantity demanded and integrating over all product lines delivers

$$(1-\theta)\int_0^M S(j)^\kappa q(j)^{1-\kappa} dj = \frac{1}{\hat{y}} \int_0^M p(j)q(j)dj. \tag{D.3}$$

Summing equations (D.1) and (D.3), and rearranging yields expression (3.3) for $\hat{y}$ or

$$\hat{y} = \frac{y}{\theta + (1-\theta)\int_0^M S(j)^\kappa q(j)^{1-\kappa} dj}.$$

This implies the solutions (3.2) and (3.4), or

$$c = \theta \hat{y}, \tag{D.4}$$

$$q(j) = S(j)\left[\frac{(1-\theta)\hat{y}}{p(j)}\right]^{1/\kappa}, \text{ for all } j. \tag{D.5}$$

Use the consumer's demand to retrieve the level of welfare as

$$W = \theta \ln(\theta \hat{y}) + \frac{\left[(1-\theta)\,\hat{y}^{\kappa-1}\right]^{1/\kappa}}{1-\kappa} \int_0^M S(j)\,p(j)^{(1-\kappa)/\kappa} dj.$$

The inverse of the Lagrange multiplier can in turn be expressed as

$$\hat{y} = \frac{y}{\theta + \left[(1-\theta)\,\hat{y}^{\kappa-1}\right]^{1/\kappa} \int_0^M S(j)\,p(j)^{(1-\kappa)/\kappa} dj}.$$

E Firm's Problem

The solution to the firm's maximization problem (3.10) 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 firm's marginal revenue net of unit operating costs to the marginal cost of digital advertising ,

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

Similarly, the intensity of traditional advertising satisfies

$$\underbrace{\left(p - w\Xi \frac{n^\eta}{\eta}\right)(1-a_d)\,q_t(p)}_{\substack{\text{Marginal revenue net of} \\ \text{marginal operating cost} \\ \text{related with traditional ads}}} = \underbrace{wBa_t^{\nu-1}}_{\substack{\text{Marginal cost of} \\ \text{traditional advertising}}}. \tag{E.2}$$

For the number of varieties the following must hold

$$\underbrace{\left(p - w\Xi \frac{n^\eta}{\eta}\right) a_d \frac{q_d(n,p)}{\sigma_d(n)} \frac{0.25\lambda}{n^2}}_{\substack{\text{Marginal revenue net of} \\ \text{marginal operating cost} \\ \text{related with varieties}}} = \underbrace{w\Xi n^{\eta-1}\left[a_d q_d(n,p) + a_t(1-a_d)\,q_t(p)\right]}_{\substack{\text{Marginal operating cost} \\ \text{of additional varieties}}}. \tag{E.3}$$

Last, the first-order condition for output price is

$$\underbrace{[(\kappa-1)/\kappa][a_d q_d(n,p)+a_t(1-a_d)q_t(p)]}_{\text{Marginal revenue}}=\underbrace{w\Xi(1/\kappa)(1/p)[a_d q_d(n,p)+a_t(1-a_d)q_t(p)]}_{\text{Marginal cost}}.$$

This first-order condition can be rewritten to obtain an equation where prices are given by a markup over marginal operating cost:

$$p=\underbrace{\left(\frac{1/\kappa}{1/\kappa-1}\right)}_{\text{Markup}}\quad\underbrace{w\Xi\frac{n^{\eta}}{\eta}}_{\text{Marginal operating cost}}. \tag{E.4}$$

Rearranging (E.4) results in

$$p-w\Xi\frac{n^{\eta}}{\eta}=(\frac{\kappa}{1-\kappa})w\Xi\frac{n^{\eta}}{\eta}. \tag{E.5}$$

The following propositions, which rely heavily on the firm's first-order conditions, are presented now.

Proposition 3. *(Neutrality of process innovation on advertising and targeting) Let Ξ rise to $\xi\Xi$. Then, the demands for specialized varieties resulting from digital and traditional ads, q_d and q_t, fall by a factor of ξ; i.e., q_d declines to q_d/ξ and q_t shrinks to q_t/ξ. The time price of varieties increases from p/w to $\xi p/w$. The variables a_d, a_t, and n remain constant.*

Proof. Suppose the proposition is true. Will the new equilibrium be satisfied at the conjectured solutions? It is immediate from (E.4) that p/w shifts to $\xi p/w$, when n is fixed. By using (E.5), the first-order conditions for a_d and a_t can be rewritten as

$$(\frac{\kappa}{1-\kappa})\frac{n^{\eta}}{\eta}\Xi[q_d-a_tq_t]=Aa_d^{\zeta-1}\text{ and }(\frac{\kappa}{1-\kappa})\frac{n^{\eta}}{\eta}\Xi\,(1-a_d)\,q_t=Ba_t^{\nu-1}.$$

If q_d falls to q_d/ξ and q_t drops to q_d/ξ when Ξ moves up to $\xi\Xi$, then a_d and a_t will remain constant. Using (E.5) in (E.3) yields

$$\left(\frac{1}{1-\kappa}\right)a_d\frac{q_d}{\sigma_d(n)}\frac{0.25\lambda}{n^2}=\frac{\eta}{n}\left[a_dq_d+a_t\,(1-a_d)\,q_t\right].$$

When the q's fall by the same conjectured factor, there is no need for n to change when a_d and a_t are constant. Observe that neither N or w enter directly into the above equations. Hence, a_d, a_t, and n will not be affected by changes in these variables provided that the q's change as postulated.

Now, from (3.4) it is easy to see that a requirement is that $\hat{y}/p$ grows to $(\hat{y}/p)/\xi^{\kappa}$. If so, then the argument is complete. Represent the initial value of $\hat{y}/p$ by ι. Therefore, the new value of $\hat{y}/p$ is ι/ξ^{κ}. Then, to find the new value for N and w, solve the two equations below:

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

and

$$\{\frac{w+(1-\alpha)x(\alpha x/w)^{\alpha/(1-\alpha)}}{\theta+(1-\theta)^{1/\kappa}\,(i/\xi^{\kappa})^{(1-\kappa)/\kappa}N[a_d+a_t(1-a_d)]}\}/[\left(\frac{1}{1-\kappa}\right)\xi\Xi\frac{n^{\eta}}{\eta}]=\iota/\xi^{\kappa},$$

where a_d, a_t, and n remain at their original values. The first equation is the labor market clearing condition. The left hand side of the second one provides a formula for $\hat{y}/p$ based on (3.14) and (E.4). The situation is shown in Figure E.1. □

.

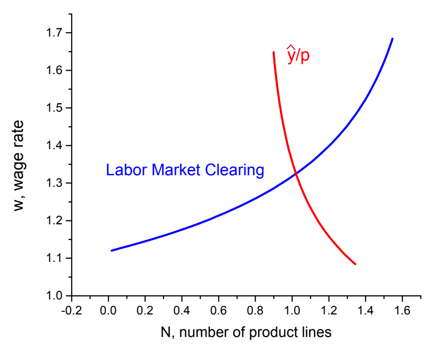

Figure E.1: The labor-market-clearing and $\hat{y}/p$ loci in (w, N) space.

Note: The plot uses values from the 2015 calibrated equilibrium. The lines cross at the 2015 solutions for $N$and w; specifically, $N = 1.17$ and $w = 1.38$. It is easy to deduce theoretically that the loci have the properties shown.

Proposition 4. *(Neutrality of technological progress in the generic goods sector on advertising, the number of product lines, and varieties per product line) Suppose that due to technological progress in the generic goods sector, x rises to ξx. Then, z increases to ξz, for $z = c, p, w, \hat{y}$. The variables a_d, a_t, n, M, N, and the q's all remain constant.*

Proof. Assume that the above statement is true. The argument proceeds using the guess and verify technique. From the first-order condition (3.11), the amount of labor hired in the generic sector does not change. Consequently, generic goods output will rise from o_g to ξo_g. Equation (3.2) guarantees that generic consumption will also grow from c to ξc, if $\hat{y}$ grows as conjectured. Also, profits step up from π to $\xi\pi$ – see (3.17). Therefore, y changes to ξy, because both w and π climb by a factor of ξ. Formula (3.6) implies then that the $q's$ will not change because p and $\hat{y}$ have shifted by the same proportion and n is constant. Now, in the firm's problem (3.10), replace p, w,and $\hat{y}$ by ξp, ξw, and $\xi \hat{y}$. It is immediate that the objective function just scales up by ξ so that the choice of a_d, a_t, and n remains the same. This implies that the new solution for prices must be ξp. Also note that the zero-profit condition (3.16) still holds. Thus, N remains fixed. Additionally, the labor-market-clearing condition (3.15) is unchanged at the new conjectured equilibrium. Hence, the conjectured hike in wages is correct. Finally, from (3.12), (3.13), and (3.14), it is clear that $\hat{y}$ moves up to $\xi \hat{y}$. □

F Calibration

Values need to be assigned to the following sets of parameters: preferences, $\{\theta, \chi, \lambda, \kappa\}$; specialized firms' advertising and cost parameters, $\{\Xi_{1995}, \Xi_{2015}, A_{1995}, A_{2015}, B, \eta, \zeta, \nu, \phi_{1995}, \phi_{2015}\}$; and generic firms' technology parameters $\{\alpha, x_{1995}, x_{2015}\}$. Some of these parameters are set using a priori information, others are backed out using a theory-based identification strategy, and the rest are estimated using a minimum distance criterion function that relates moments from the data with the model. These are discussed below.

F.1 Parameters Set Using A Priori Information

The first step in the calibration is to use a priori information to set values for the inverse of the price elasticity of demand for specialized products, κ, and for the generic output elasticity with respect to labor, α. These two parameters relate to firms' markups in each sector. A markup of 1.2445 is chosen for both sectors. This is in line with the average markup reported in Marto (2023) for the 1995-2015 period. Values for α and κ are chosen to match the average markup in the generic and specialized sectors in 1995 and 2015. This implies that

$$\texttt{Markup} = \left(\frac{1/\kappa}{1/\kappa - 1}\right) = \frac{1}{\alpha} = 1.2445.$$

F.2 Parameter Values Calibrated Based on an Exact Fit–Inner Loop

The second step employs a theory-based identification strategy. In this step the parameter vector $\rho_f \equiv (\Xi_{1995}, \Xi_{2015}, A_{1995}, A_{2015}, B, \eta, \phi_{1995}, \phi_{2015}, x_{1995}, x_{2015})$ is chosen. These 10 parameters are selected to match exactly 7 data targets pertaining to the years 1995 and 2005, and 3 restrictions on the values of the model's variables in 1995. The seven data targets are:

1. The advertising-to-GDP ratio in the United States in 1995 and 2015. This implies that

$$\frac{\texttt{Ad Cost}}{\texttt{GDP}} = \frac{\int_0^N w\left[Aa_d(j)^\zeta/\zeta + Ba_t(j)^\nu/\nu\right]dj}{o_g + \int_0^N p(j)o_s(j)dj} = \{0.0220, 0.0220\},$$

where o_g and $o_s(j)$ are the outputs of firms in the generic and specialized sectors.

2. The ratio of digital-to-traditional advertising spending in 1995 and 2015,

$$\frac{\texttt{Digital Ad Cost}}{\texttt{Traditonal Ad Cost}} = \frac{\nu A \int_0^N a_d(j)^\zeta dj}{\zeta B \int_0^N a_t(j)^\nu dj} = \{0.0230, 0.9655\}.$$

3. The growth in the total number of varieties between 1995 and 2015:

$$\Delta n = \frac{n_{2015}}{n_{1995}} - 1 = 1.1453.$$

4. The growth in the number of product lines between 1995 and 2015,

$$\Delta N = \frac{N_{2015}}{N_{1995}} - 1 = 0.1700.$$

5. The growth in income between 1995 and 2015,

$$\Delta y = \frac{y_{2015}}{y_{1995}} - 1 = 0.3573.$$

The 3 restrictions on the model's variables for 1995 are

$$w_{1995} = N_{1995} = n_{1995} = 1.0.$$

F.3 Parameter Values Calibrated to Minimize the Model's Prediction Error

In the third step, the preference parameters, θ, χ, and λ, as well as the advertising cost exponents, ζ and ν, are estimated to find the best fit between the following targets and their model counterparts. For convenience, this step assumes that ζ=ν. The second step represents an inner loop or a constraint on the best fit minimization problem.

1. The size of the specialized-products sector in 1995 and 2015, where for each period

$$\texttt{Size Specialized Goods Sector} = \frac{\int_0^N p(j)o_s(j)dj}{o_g + \int_0^N p(j)o_s(j)dj}.$$

2. The elasticity of sales with respect to total advertising,[37]

$$\texttt{Sales Elasticity} = \frac{\int_0^N w\left[Aa_{d_t}(j)^\zeta/\zeta + Ba_{t_t}(j)^\nu/\nu\right]dj}{\int_0^N p(j)o_s(j)dj} \times \frac{\partial \int_0^N p(j)o_s(j)dj}{\partial \int_0^N w\left[Aa_d(j)^\zeta/\zeta + Ba_t(j)^\nu/\nu\right]dj}.$$

The target is 0.2. The comparable moment in the model corresponds to the average elasticity between 1995 and 2015.

3. The elasticity of the number of total varieties with respect to digital intensity,[38]

$$\texttt{Total Varieties Elasticity} = \frac{\int_0^N a_d(j)dj}{\int_0^N n(j)dj} \times \frac{\partial \int_0^N n(j)dj}{\partial \int_0^N a_d(j)dj}.$$

This elasticity is estimated from the data to be 0.84, the average of columns 3 and 6 of Panel B in Table C.11 in Appendix C. The comparable moment in the model corresponds to the average elasticity between 1995 and 2015.

[37] The derivative term is computed using variation in A according to $\{\partial \int_0^N p(j)o_s(j)dj/\partial A\}/\{\partial \int_0^N w[Aa_d(j)^\zeta/\zeta + Ba_t(j)^\nu/\nu]dj/\partial A\}$.

[38] The derivative term is computed using variation in A according to $[\partial \int_0^N n(j)dj/\partial A]/[\partial \int_0^N a_d(j)dj/\partial A]$.