Appendix: New Economic Forces Behind the Value Distribution of Innovation

This version: 7/23/2025

A. Radio Stations Data Processing

An observation in the Berry-Waldfogel (1993) data is a station/city pair. There are 5476 of these. A station can occur in multiple cities. There are 3836 unique stations, appearing in from 1 to 18 cities each. According to Arbitron, there are 257 cities, each a local radio market. These cities range from Cheyenne, WY, to NY, NY, indicating that the largest local market has approximately 374 times more listeners than the smallest.

Table A1. Tabulation of the number of cities in which individual stations appear.

	Freq.	Percent	Cum.
1	2965	77.29	77.29
2	538	14.03	91.32
3	170	4.43	95.75
4	74	1.93	97.68
5	24	0.63	98.31
6	21	0.55	98.85
7	20	0.52	99.37
8	7	0.18	99.56
9	5	0.13	99.69
10	6	0.16	99.84
11	1	0.03	99.87
12	2	0.05	99.92
13	1	0.03	99.95
16	1	0.03	99.97
18	1	0.03	100.00
Total	3836	100.00	

As Table A1 shows, seventy percent of stations appear in one city, and 12 appear in 10 or more. We sum listeners over the 1 to 18 local markets in which they appear.

A "station" can be AM, FM, or AM/FM. Some AM/FM stations broadcast in the same format. Duncan (1993) provides an outline. We treat those as matched to 2013 if there is a match in either brand. Other AM/FM stations have a different format for each band. We treat them as matched if a band/format match exists for either half.

While a call sign is a satisfactory unique identifier for a cross-sectional analysis (the original purpose in Berry-Waldfogel), this is far less true for identity over time. The FCC relicenses call signs; some are no longer used at other stations. We treat a call sign in 2013 as associated with the same product in 1993 if 1) it is close to one of the cities where the 2013 call sign was competing, and 2) it is in the same band. Most "exits" occur because the call sign is not in use anywhere in 2013, but some arise because it is in use but has migrated hundreds of miles or (less frequently) switched bands.

Some chains do not permit local radio websites. Based on the 2013 lists of their members, we exclude CBS, K-Love, and iHeart stations. Other chains imposed this rule later, and we do not exclude them. Chain-like services, such as Tunein or Streema, offer access to local radio but do not block station-specific websites. We do not exclude those. We also exclude five stations for which Arbitron records zero audience.

We match the 1993 Arbitron stations to the 2013 stations listed in Wikipedia. The idea is to assign turmoil regressors to each 1993 station. It might have exited or changed its product radically over the past 20 years. Accordingly, we look for a 2013 station that shares the 1993 station's callsign, band (AM/FM), and location, and then compare its format (e.g., Top 40, News/Talk).

The match is complicated by several factors. We describe our solutions here. First, a call sign is not a station's unique identifier over time. When stations exit, their call signs can later appear in an entirely unrelated business in a distant state. Second, the Arbitron data used by Berry-Waldfogel locates stations in all the cities they serve, while the Wikipedia data reports the location of a station's (main) tower. Third, the band is complicated by station product differentiation practices. The Arbitron data include stations that are AM, that are FM, that are AM/FM broadcasting the same content on both bands, and that have different program products on both bands. Fourth, while the Arbitron data are categorized into consistent format categories, the Wikipedia entries are not, reporting many more formats. Fifth, some stations, particularly in smaller markets, are categorized as being in more than one format or as being "Full Service," which is another name for operating in multiple formats at different times of day. Sixth, the Arbitron data do not include stations' broadcast frequencies.

Most of these complications arise because the Arbitron data are designed to support advertisers and are thus market-oriented, whereas the Wikipedia data are crowdsourced and listener-oriented. Some attractive potential solutions are blocked because FCC data are a snapshot of licenses and do not make finding a 20-year history feasible. A positive fact is that there were far more Wikipedia stations in 2013 than Arbitron stations, more than 15,000.

We first look for a match on location, band, and call sign for each Arbitron station. For stations with multiple cities, we look in all the cities for the location. For multiple-band stations, we look for both bands. A match on location is a station whose 2013 tower is in a one-degree circle around the center of any of the 1993 cities. This is to catch stations whose tower is, for example, on a nearby hilltop. We hand-check all stations that match on call sign, but not the other two variables, drawing on station histories in Wikipedia, which often reflect call-sign changes. This check does not lead to any further matches.

For stations matched on those three criteria, we check whether there is also a match on band and format. Band is simple. Formats are more complex. According to the 1993 Arbitron data, they fall into 44 unique categories, some of which are blends, such as "G/B." Formats in the 2013 Wikipedia data are uncategorized, often merely reflecting station self-descriptions, with thousands of unique values. We define four levels of format match certainty. Same, Nearly same, Nearly different, and different.

Results from estimation using Alexa for Radio Stations.

We replicated the results in Table 2 using ranks in Alexa rather than Quantcast as the dependent variable. Alexa ranks the top 1 million websites on a worldwide basis, while Quantcast is US-focused. Alexa rankings are based on the behavior of users who install an Alexa toolbar in their browser and are calculated based on daily unique visitors plus average (over users) pageviews. The mean and standard deviation of ln(Alexa rank) conditional on it being observed are 13.73 and .353.

Table A2. Upper-Constrained Tobit ln(Rank) in 2013 (Alexa Rank)

X	Coefficient	S.E.	Coeff/S.E.	Coeff	S.E.	Coeff/S.E.
LnListeners93	-0.783	0.069	-11.29	-0.776	0.069	-11.28
Format Same	-0.136	0.118	-1.15	0.210	0.241	0.87
Format Nearly Same				0.161	0.390	0.41
Format Different				0.705	0.249	2.83
D_Music (Format)	0.056	0.136	0.41	0.091	0.140	0.65
D_AM (Band)	-0.417	0.152	-2.74	-0.593	0.156	-3.81
C	18.596	0.413	44.97	18.394	0.463	39.72

Notes: See notes to Table 2.

B. Novel Value Creation: Case Studies

Several examples of significant value creation became visible during the deployment of CCTs and the growth of internet superstars. These co-inventors created new services, operating models, and, ultimately, new markets and industries. In this subsection, we offer a few examples to illustrate the factors at work and motivate open questions.

Amazon. After its founding in 1994, Jeff Bezos located his entrepreneurial effort in Seattle in anticipation of improved logistics with the existing book supply chains. In addition to the favorable tax treatment of his business, Bezos sought to collaborate with a significant book wholesaler, Ingram, whose warehouse was located nearby (Stone, 2013), and capitalize on the technical labor market developed as a byproduct of Microsoft's growth. The young firm sought a thick labor market of technically fluent people. With Microsoft already there, Seattle was on that long list.[1] Seattle was also located in a not-very-populous state, reducing sales tax exposure.

In its first few years, Amazon had to solve complex technical and commercial problems associated with supporting electronic commerce and confront numerous logistical issues around the processes for fulfilling demand. As a result, it aspired to invent an online bookstore. This endeavor drew on existing markets for applications to support online commerce (e.g., online credit and payment processing), and it co-invented many aspects of this service, such as listing, search, and recommendation engines for the items it carried.[2]

As CCTs deployed after the millennium, Amazon increased the scale of its activities in the US and worldwide, improving the quality of customer engagement with its search and recommendation tools. It experimented with numerous design changes in the pages of its websites. Amazon also broadened its product lines and co-invented new commercial models to complement the breadth (e.g., Prime). It also became a vital media seller, introducing new inventive devices in the form of e-readers and new operational models for new services, such as selling e-books and videos on a large scale.[3]

Google. In 1998, Google began an effort to commercialize a solution to a complex technical problem in the search for web pages. It started with a model to license to portals. After the millennium, it co-invented many commercial elements of the consumer-oriented large-scale search business. The site responded to consumers' questions with organic listings ranked by anticipated relevance and ads of relevance to users ranked by advertisers' willingness to bid for clicks from users. Scaling this activity was the principal challenge faced by the firm in its early years. The deployment of CCTs coincided with the scaling of Google's service, and Google began growing a large team of technically skilled labor to improve many aspects of this service.[4]

---

[1] Microsoft was in Seattle for that most arbitrary reason: the founder's family.
[2] Eyewitnesses describe many questions and crises the organization had to confront, and the many inventive efforts and co-inventive activities. See, e.g., Bezos, 2021; Rossman, 2021; Bryar and Carr, 2021.
[3] Rossman, 2021, and Bryar and Carr, 2021.
[4] Greenstein, 2016, chapter 13, describes and analyzes the building of Google's dominant consumer-oriented search engine. It involved a combination of commercial aspirations, frontier computer science, competitive imitation, extraordinary intuition, and persistent exploratory experimentation of new designs. It

Developing a targeted ad business became Google's focus after the millennium. The largest advertising suppliers in the US were not located nearby and did not offer high industry-specific joint location attractions to the young firm. The early growth challenges lay more in building scale in a quality-weighted pay-per-click position auction for targeted ads that ran next to user query search results. After they succeeded in building this auction next to their search results, they faced challenges in educating potential advertisers to use the novel processes associated with bidding on auctions and arranging campaigns.

Google extended its services in numerous ways. Its founders and employees co-invented methods to assemble campaigns around the hundreds of millions of search terms. Google also altered the user experience in its search engine by adding visual details and news summaries, thereby increasing the relevance of ads. This included adjustments to address the gaming of the search algorithm. Additionally, after acquiring YouTube, Google began introducing new services. This led to the co-invention of processes for large-scale recognition of copyright holdings and incremental compensation against ad revenue, accommodating different copyright systems globally. Other notable co-inventions include scalable methods for segmenting the geographic source of queries.[5]

Facebook. Founded in 2004, Facebook developed processes for storing and recalling each user's social graph. After growing with university users, the founder, Mark Zuckerberg, moved the firm to Silicon Valley to hire labor with technical and commercial skills to help scale the business. Initially, addressing the technical challenges of scaling the business was more predominant in that location choice and access to technical talent dominated. The commercialization issues quickly became salient.[6] In 2008, Sheryl Sandberg moved from Google to become Facebook's COO, and the firm began accelerating the scaling of advertising. This involved inventing an automated ad auction that resembled Google's scale and precision but differed in the profiles of users targeted by advertisers.[7]

The social graph supported a new form of ad targeting and motivated experiments with many services. The graph contributed to the development of personalized news feeds on a large scale. The chart and newsfeed went through several iterations as the firm co-invented processes for identifying the intensity of preferences (e.g., "likes"), categories of selections (e.g., group discussions), and procedures for complementary activities (e.g.,

---

was neither solely a technical invention nor exclusively an inventive commercial business model, and it would be a mischaracterization to regard it in such narrow terms.

[5] Schmidt and Rosenberg, 2017 describe many of these efforts at length.

[6] See Baloun, 2007, and Kirkpatrick, 2011, for inside accounts of many of the earliest challenges.

[7] Like Google, Facebook was not the only firm experimenting with targeted advertising built around a social graph. For example, Twitter tested similarly, albeit with less financial success. See Levy, 2020.

Zynga's games). In addition, Facebook had to co-invent a new display format to provide a satisfying experience within the constraints of the most common mobile networks.

More visual forms of social networks also became popular, a trend that Facebook led and followed, depending on the dimension. In April 2012, it bought Instagram, which was founded in October 2010 and had become a central place for social media around pictures. Facebook retained the brand name, and the two platforms operated independently during the period covered by our data.

Entertainment and the Arts. Inventive and co-inventive activities focused on many market activities based around copyright-based markets, such as music, writing, visual arts, and video entertainment. As discussed below, Apple pioneered one opportunity in music, as its iTunes online store complemented its devices. Another opportunity emerged in short-form videos organized around user-generated content.[8] YouTube, based in Silicon Valley, became a focal platform with a large US audience. After its initiatives lost market share in competitive events, Google bought YouTube in October 2006[9] and subsequently invested in a sizeable multi-year effort to adapt its advertising technologies to the site. By 2013, YouTube had grown beyond its roots into a major music venue. General news reports indicated Google had spent considerable sums on the co-invention costs. By the time of our data collection, in 2013, it was turning the corner on developing a profitable service.[10]

Netflix pioneered the opportunity to stream long-form videos. Founded in 1997, Netflix first invented a better model for a video rental store. It began with a mail-in, mail-out model, then expanded its business by transitioning to an online streaming delivery model. The new model heavily leveraged broadband deployment, as the firm pioneered high-volume video delivery for mass market consumers. Large-scale streaming requires more co-invention in contracts with carriers to deliver high volumes of data and in contracting arrangements with content holders. Netflix also made visible progress in the co-invention of targeting recommendations for additional entertainment to users. Challenging practices in existing production also co-invented new processes for building video entertainment from historical viewing data.[11]

---

[8] See e.g., https://www.theringer.com/2016/10/10/16042354/google-youtube-acquisition-10-years-tech-deals-69fdbe1c8a06, accessed July, 2023.

[9] This purchase occurred slightly over a year after their beta service launched in May 2005. https://www.nytimes.com/2006/10/09/business/09cnd-deal.html, accessed July 2023.

[10] See e.g., https://www.theringer.com/2016/10/10/16042354/google-youtube-acquisition-10-years-tech-deals-69fdbe1c8a06, accessed July, 2023.

[11] Netflix was one of many organizations to co-invent in media and entertainment, but many existing organizations resisted its novel approach to value creation. See Smith and Telang, 2016, and Hastings and Meyers, 2020.

Disney is a notable and special case of novel co-invention in entertainment. Headquartered in the Los Angeles area, Disney-owned properties in television (i.e., ABC, ESPN) operated many theme parks (e.g., Disneyland) and boat cruises across the globe, and held a valuable library of films and content (e.g., Snow White). After the millennium, it acquired several movie studios (e.g., Pixar, Marvel, Lucasfilm), adding additional storytelling capabilities to its existing Studio.[12] This restructuring gave it an extensive portfolio, redeployed across many new services.

Disney restructured its online activities to capitalize on the overlap between its television, video, movie, and family vacation and entertainment offerings within its amusement parks. Some steps targeted potential customers with specific needs (e.g., family entertainment) or geographic preferences (e.g., Florida, California). For example, the firm undertook experiments in personalization using tracking technologies at entertainment parks, which served over one hundred and twenty million visitors worldwide in 2012.[13] Disney co-invented an online presence, principally through its portal, Go, that complemented its entertainment offerings.

Banking. As CCTs were deployed, many banks experienced adjustment costs when moving to online banking. The largest banks, such as Bank of America, adopted a strategy to reduce costs associated with branch personnel by providing users with convenience through self-service ATMs. The earliest movements toward online banking continued this trend, increasing the convenience of home computer users and gradually accumulating improvements over time. The services were inherently digital, so some co-invention involved straightforward and incremental changes in displaying information and developing secure login and identity verification processes. Additionally, online channels enabled transactions outside working hours, albeit at the expense of some services, such as deposits and cash dispensing.[14] Initially, the movement to mobile banking accelerated customer convenience by providing users with access anytime, from anywhere, as long as they had their phone. [15]

---

[12] These efforts began after 2005 when Robert Iger became CEO. Iger, 2019.

[13] Disney experimented with co-invented ID bracelets and park tickets to track usage. In addition, it co-invented policies to alter costs, induce changes in use within the properties, induce usage at non-peak moments, and reduce queues at peak times. Those identities also informed online services. Uhl et al., 2014; Iger, 2019.

[14] Online customers increased transactions, making them potentially more expensive, but with the benefit of higher retention later and possibly higher profitability. See Campbell and Frei, 2009; Xue et al., 2011.

[15] Customers experienced online banking services as a unified whole when the online links delivered a picture of their financial status and options. Mobile banking also started to yield changes in the demographics of participation, with growth in underserved communities (Abhishek and Li, 2019; Buchholz et al., 2014; Broeders and Khanna, 2015).

It would be inaccurate to call all the changes incremental. New services emerged gradually and accumulated over time, following considerable testing due to heightened concerns about security and privacy in these services. These services included co-inventing transaction processes, such as low-balance and fraud alerts, check recognition, loan approval, credit card issuance, and payments. By 2013, more than half of US smartphone users were using their smartphones for banking services, including transactions, interactive account access, and payment solutions. The effects on user behavior were subtle and depended on branch availability and preexisting customer usage. By 2013, some banks had made only incremental changes, while others had gone far beyond.

Retailing. Many retailers with a nationwide presence responded to the deployment of CCTs and the increasing prominence of Amazon's competitive pressure by developing omnichannel distribution strategies. Online actions lifted engagement by personalizing attention with unique selections and sales offerings based on user histories. Omnichannel strategies accelerated with the deployment of CCTs. Many prominent retailers – e.g., Barnes and Noble, Walmart, Safeway, Target, Walgreens, CVS, Staples, and Home Depot – supported an online channel to complement activities at existing brick-and-mortar outlets. Bell et al., 2014. Brynjolfsson et al, 2013.

However, that came with challenges in coordinating online and offline activities.[10] Many retailers implemented a hybrid approach, where physical stores could act as "showrooms" to support online sales. Physical stores also helped sales from online channels with programs to " buy online, pick up at the store" or " research online and purchase offline." The latter mattered for products that require " fit and feel." Online channels could also support users who bought online for home delivery but returned to the store.

All these combinations could involve co-inventing order fulfillment and inventory management systems that coordinate channels to avoid stock-outs and tailor sales to available inventories. [11.] That required changes to operational processes to support the low-friction customer experience. That involved changes to assortment planning, inventory availability and salvaging, daily planning for sales and weather, and reaction to stockouts and order fulfillment, both in the immediate and medium term (Gallino and Moreno, 2019).

Some applications successfully built engagement with services such as "availability checking" at retail stores, triggered by online or mobile searches; "couponing" on a mobile device or website, distinct from in-store; "informing" customers about product types with reviews; and “soliciting” reviews after in-store searches. However, developing these functions into an integrated and seamless service for existing users required considerable co-invention at each firm. By 2013, some retailers had made only incremental changes, while others had gone far beyond.

That conclusion aligns with another characteristic of the environment: the constant fluidity of co-inventive activity. Co-invention generally proceeds in a series of steps, some novel, some incremental. We classify a firm as novel if it has made a novel step by 2013.

Automated refills in retail pharmacies require a complex co-invention effort involving modifications to an operating model. This included installing automated notifications, reminders, and notices with personalized links to medical prescriptions. In a retail setting, personalization occurs on two levels: in the refill order and through personal interaction. There is a presumption that the system did not eliminate face-to-face interaction at the point of sale, where the patient interacts with the pharmacist during in-person pickup. In many cases, these interactions were quite valuable for the retail outlet, and there was no desire to eliminate them.

After the millennium, many pharmacies used automated systems and manual processes to send reminders for medication refills. Investments in electronically personalized automated reminders for refills were typically part of a broader strategy to automate operations, an investment that involved both hardware and software upgrades. The most advanced retail pharmacies implemented the first upgrades in the 1980s, and many did so in the 1990s. As the web diffused, many advanced systems were installed after 2000, and the emergence of phone apps further improved those systems.

The ideal workflow at a retail pharmacy with automated refills involves many activities. A prescription must be entered into the system, either transferred from a doctor’s office or by the pharmacist. In most systems, alerts about refills are not sent without verification of insurance and stock availability. A calendar alert system triggers a computerized reminder in the patient database, which, in turn, generates a message – either via text, call, email, or app notification. In some systems, the patient must confirm a refill. Either automatically (or after patient confirmation), pharmacists must prepare and store the medication for pickup. Again, patients are notified when the prescription is ready using the same systems as the reminder. Alerts are triggered, and follow-up messages are sent if the prescription is not picked up.

To achieve this ideal workflow, pharmacies install a PMS (Pharmacy Management System), which tracks numerous items, including patient information, medication name, dosage, refill eligibility, expiration date, patient history, and insurance status. These can be (and typically are today) integrated with other electronic medical record systems from doctors' offices and the networks of local hospitals. To generate automated refills, the PMS must be integrated with a notification system that sends reminders via text, email, and app notifications. Most of these notification systems have a web and/or app that enables patients to acknowledge receiving the notification and respond in some circumstances.

Pharmacists in a retail setting must adjust their workflow in numerous ways after these notification systems and automated refills are installed. However, the precise adjustments depend on how far the pharmacy has progressed in installing electronic systems as part of a general program to reduce paperwork. The most significant adjustments can include: a) less calling of patients and doctors to fix communication errors and more monitoring of alerts; b) more follow-up on alerts about inventory; c) less time doing manual entry of essential information and more time spent on manual entry correcting input errors, particularly insurance errors and incorrect patient information; d) reduction in phone inquiries and increase in web/app-based inquiries; e) managing alert fatigue.

These investments brought about changes in service for the user and in the routines for the pharmacists. Pharmacists used to conduct inventory checks before patients came in for refills, and then phone to arrange pickups. They used to conduct an insurance check upon the patient's arrival. They used to do both checks for a new script after the patient came in. After the change, the process is much smoother for the patient. Some of those checks are moved to the system and automated. The per-patient face-to-face time can be much shorter, but only if the order is packaged and nothing goes wrong. That means pharmacists must stay on top of "alerts" that prefigure the patient visit. The patient arrives expecting everything to be ready since they got the notice that says, "Come get your refill/script." Doing damage control with the patient when something happens – like insurance doesn't cover it or inventory has run out –is a more interesting part of the job. Fixes require training. Pharmacists must become familiar with newly installed information technology, where and how errors arise, and how to fix issues.

C. Selection into our main sample

Our commercial datasets are both selected to include larger products. For the websites, we start with the 2,500 most-used websites over the year. A comparison with estimates of total web use suggests that we have over 80% of web usage time. For apps, our data source features a monthly sample of 5,000 users for iPhones and 5,000 users for Android phones. An app is reported as active in a month if it has six or more users.

Our web usage dataset records the time the consumer's browser is pointed at a URL[16]. Some of these URLs are not consumer-facing websites. These include router addresses, similar local links, and background sites providing services, such as JSON services, to consumer websites. This issue does not apply to mobile apps. We exclude these URLs.

We excluded non-US commercializers. We identified firms outside the US through two mechanisms. The first was the location of the firm's web server, as in another country, queried using standard Internet utilities. (The location of the firm's web server within a country is not a reliable indicator of what city it is in, and we do not use it for assigning firms to cities.) If the web server is not found, for example, for firms that exited after 2013, or if the web server is in the USA, we use the location search methods described above. The largest excluded properties are Pornhub.com (a Canadian website), Spotify (a Swedish-based firm), and Candy Crush Saga (a Swedish/Maltese mobile game).

We exclude illegitimate products, which fall into four main categories. We used security sites to identify many malware and scam sites. Others in those two categories, as well as pornography and file-sharing sites, were identified by hand in the course of our search for firm identity. Most illegitimate products are websites rather than apps, as our definition of illegitimate is close to the iTunes Store's criterion to exclude apps. The Android store had, in 2013, a somewhat less stringent criterion, though it is tightening over time.

Table A3 displays the number of products and total usage for several classes related to our exclusions, as well as for the included products.

Table A3: Product Exclusions.

Reason	**Number**	**Total Usage (minutes 10^6)**
Not a product[a]	94	13439.5
Not in the US	667	125093.4
Not Legit[b]	473	74682.0
Included in Analysis	2426	5520964.6

Notes
a URLs that are not websites, such as a home router address.
b URLs that point to scams, spyware, pornography (spyware again), online gambling, piracy, related file sharing, and sales of illicit goods and services.

[16] For additional details on online time use characteristics, see Boik et al. (2019).

The ordering of rows matters in Table A3. The Not in US exclusion and the Illegitimate exclusion have a substantial overlap, as sketchy sites serving the US are overwhelmingly located in other countries. Thus, the larger exclusions for Not in the US arise largely because we impose this edit first. The total exclusions are just under 4% of usage.

Just over 2% of product usage is of the excluded non-US products. This means US consumers mostly use US sites. We do not interpret this to mean that the scope of commercialization is largely within the country. A study of most other Western countries would show significant imports from the United States.

D. Classification of Incremental and Novel

We classify the regime based on information about the product and preexisting services. Conceptually, a product is in the novel regime if it offers a service unavailable from the co-inventing firm before it used CCTs. Otherwise, it is in the incremental regime.

This definition contains both easy and challenging elements. We classify the products of young firms that used CCTs from when they were founded in the novel regime. For firms founded before 1994, there is both fact-gathering and judgment. We compare the services offered through the CCTs to pre-existing services. For example, we classify a radio station that sends the same audio through an app as incremental. Also, incremental is a bank that moves existing ATM services (except cash back) to a website. Yet, we do not classify a reminder from CVS to a customer to refill their prescription as incremental. That is a new service. Those examples illustrate that our definition is based on co-invention, not invention, since the new website may need help to write, keep it secure, etc.

Novel innovation from existing firms is complex because they started with refined business processes aimed at specific customer segments before deploying CCTs. While adopting CCTs and co-inventing new services might enhance operations, it also risks hurting value delivery. On the other hand, many established firms already employ a workforce familiar with frontier IT. This could give such firms advantages in conceptualizing new uses and addressing opportunities earlier. Examples of established firm investments with CCTs will illustrate these differences.

Above, we reviewed the experiences in three areas with established firms that operate businesses with features quite different from those supported by Microsoft and Apple, such as Disney, banking, and retailing. New services emerged notably at Disney, which leveraged the reach and engagement of CCTs. For example, it created services that tied online and entertainment experiences together. It also emerged from retailers that used the mobile internet to enable customers to personalize the management of purchases. For example, CVS and Walgreens co-invented personalized order fulfillment that raised customer engagement. This included notifications, reminders, and notices with links to medical prescriptions. These services drew on IT-intensive components, such as databases of past purchases and additional databases of past searches, combined and directed to prompt new consumer searches. We conclude that co-invention at established firms contains one crucial difference with entrepreneurial firms: established firms face a trade-off when building new services, which might undermine the value of existing ongoing operations.

We also consider the possibility of random classification error on our part. To address this concern, we employ the misclassification model developed by Aigner (1983). We undertook a bootstrapping exercise, where we remade the tables of this section thousands of times by randomly re-assigning products in each regime at some rate. We increase the misclassification rate until the qualitative results of this section no longer appear in the average reclassified table. This exercise requires unrealistic classification error rates for our inferences to be wrong. See Appendix L.

Table A4 reflects not just the high returns of the FAANG[17] firms near the top of the distribution, but it is a more general phenomenon. After all, there are 77 (38) incremental products and 165 (82) novel products in the top 90 (95) percentiles. This can only result from the actions of one or two firms if those firms are multiproduct firms with numerous high-value products from numerous novel co-inventive efforts. That is different from what this data shows.

Table A4: Quantiles of the Usage Distribution for Incremental and Novel Products

Percentile	Incremental	Novel
1%	0.90	0.88
5%	3.31	5.81
10%	6.71	18.85
25%	35.91	119.17
50%	156.89	226.34
75%	323.87	643.78
90%	771.10	2101.03
95%	1373.12	5780.01
99%	4125.00	41610.22

[17] FAANG stands for Facebook, Apple, Amazon, Netflix, and Google.

A characteristic of novel co-invention, in general, is uncertainty about returns. This manifests as a large upper tail for valuable GPTs, such as the CCTs. The 20 most-used products are listed in Table A5: Top 20 Products by **Usage**.

Table A5: Top 20 Products by Usage

Product	Firm	Product Classification	Usage
Facebook (Mobile App)	**Facebook**	**Targeted Advertising**	8.23e+05
Facebook.com	**Facebook**	**Targeted Advertising**	5.82e+05
Youtube.com	**YouTube**	**Information Products (Media)**	3.92e+05
Google.com	**Google**	**Targeted Advertising**	3.89e+05
Apple iTunes (Mobile App)	**Apple**	**Retail Trade**	2.71e+05
Pandora Radio (Mobile	**Pandora Media**	**Information Products (Media)**	2.57e+05
Yahoo.com	**Yahoo!**	**Information Products (Other)**	1.50e+05
Instagram (Mobile App)	**Instagram**	**Targeted Advertising**	1.21e+05
YouTube (Mobile App)	**Google**	**Targeted Advertising**	96257.630
Gmail (Mobile App)	**Google**	**Targeted Advertising**	76458.402
Tumblr.com	**Tumblr**	**Targeted Advertising**	66451.892
Words With Friends Free	**Zynga**	**Information Products (Games)**	61220.948
Twitter (Mobile App)	**Twitter**	**Targeted Advertising**	61117.388
Apple App Suite (Mobile	**Apple**	**Retail Trade**	52525.292
Msn.com	**Microsoft**	**Information Products (Other)**	45974.598
Google Maps (Mobile App)	**Google**	**Targeted Advertising**	41610.218
Android Email (Mobile	**Google**	**Targeted Advertising**	39147.420
Aol.com	**AOL**	**Information Products (Other)**	38393.195
Google Play (Mobile App)	**Google**	**Targeted Advertising**	32997.336
Netflix (Mobile App)	**Netflix**	**Information Products (Media)**	32460.974

The familiar products in Table A5 illustrate some of the important aspects and definitions of our data. Facebook's top two products are its mobile app and its (very similar) webpage. Google, a larger firm at this juncture by advertising revenue than Facebook, appears through multiple products. Facebook is winning here in terms of reach and engagement. Google, whose customers often say what product they are trying to buy, has a more lucrative targeted advertising monetization model.[18] While such distinctions

[18] YouTube and Instagram are listed as separate organizations, though they were acquired by Google and Facebook before 2013. Where possible, the organization in our data is recorded as of the moment of co-invention – pre-merger, in these cases. In addition, as earlier noted, the fifth-largest product by usage, Apple's music store, is classified in retail trade. Product classification refers to the categorization of products, not organizations.

receive considerable attention from business analysts, the similarities outweigh those differences for our purposes. Both firms have more in common at the top of the value distribution for novel co-invention.

The table shows some of the many new firms. It includes firms founded in 1994 or later, except for Apple and Microsoft, which were founded in the PC era. Long-established firms are largely absent. We note that there is little ambiguity about the regime definition here. These products are novel.

Table A5 also illustrates the ambiguity of (some) classification boundaries. For example, the boundary between advertising and media products is inherently arbitrary. Still, there are differences between products like YouTube or Pandora Radio, which are novel forms of music and video distribution, compared with Facebook and Google, which have new means of attracting consumers whose attention can be sold to advertisers. The boundary between the Information Products (Other) category and Media or Targeted advertising products is less arbitrary. The most significant “other” products are portals. The category also includes information products from IMDB, Verizon Wireless, etc. Yet, the considerable firm-level assets of many incremental co-inventors (and of other potential co-inventors who do not appear in our data) do not appear to have given those firms clear advantages – instead, the reverse.

This ambiguity relates to a fundamental fact about novel co-invention. By definition, it provides new services at the firm level. Sometimes, those services are also new to the economy. They are often related, as here, by being two sides of a platform, e.g., advertising and content, or by departing from existing product markets such as advertising-supported media in many different ways. In any case, the new services upset the status quo.

Those were firms founded to take advantage of the business opportunities offered by CCTs. Their business processes are built " greenfield " in this new technological era. This provides an obvious source of novel co-invention. These entrepreneurial firms treated the CCTs as enabling the invention of novel business processes and sometimes whole new product classes or industries. Not listed in this table, but shown in the statistics of Table 8, are the many newly founded firms that aspired to those lofty goals but did not reach them.

E. Firm Identity Search

Conceptually, we are searching for the firm that co-invented the product at the time of co-invention. We seek to exclude contract developers, viewing the buyer, not the seller, of custom software as the co-inventor. We aim to “unmerge” mergers that occurred between the time of commercialization and 2013. Each of these tasks comes up at different points in our search. We start from a product name, either the domain name for a website or an app’s name. For apps, we also frequently obtain the “developer’s website” and other information from the Android and iPhone app stores.

We have a “Preferred Firm Name” supplied by ComScore, our data source for a small subset of products. That name is the product's owner in 2013, so we still need to search for earlier co-inventors, if any. For apps, we frequently have incomplete information about the supplying firm from the app stores. The Play Store for Android has a developer field, while the Apple App Store has three fields, which might be the name of the developer, the firm, or the “ artist.” There is also a numeric “ developer ID.”   Developers of iPhone apps seem to treat the three alphameric fields as interchangeable. In both stores, we know from earlier work that the names tend to be filled in when the product is first submitted and rarely changed later.

Step one is to read the “ about us” page of websites or developer websites for apps, if they exist, in the Internet Archive. For websites, we do this for the oldest “ about us” page we can find in the Archive, since that is likely to reveal the owner when the website was first commercialized. This won’t always work for developer websites since mergers often move the billing and customer support functions to a new website.   So, we focused on the 2013 versions of the developer website to know at least the owner at the time of our data.

The rest of the steps are a series of structured web searches. We search for the product's name plus “ founded” or “ introduced,” which often gives us the time of commercialization and sometimes gives us the name of a person involved at the beginning, a news story about the introduction of the product, or the founding of the firm supplying it.   If this search turns out to be too broad to be useful, we restrict it to TechCrunch and PRNewswire. Suppose this search turns out to be too restrictive to be useful. In that case, we drop “founded” and/or ”introduced” and see if there is information about the firm owning the product, typically trying “founder” or “developer” before searching just with the product name.

If we have a time of commercialization, we restrict the following searches to the year following that. If these lead to a firm name, we take that as the commercializer.

For the still-unfound firms, we continue to search, now often using one or more (possibly human) names in TechCrunch and CapIQ. A particular problem arises for commercializers like churches and government agencies. These are not firms, but the website (or developer’s website, for apps) typically names the parent organization, leaving only a few long web searches.

Some entrepreneurial firms are not in either CapIQ or TechCrunch (if they have never considered venture capital, for example). We find many of these in open web searches, often in the contemporary trade press and sometimes in Wikipedia. Wikipedia is more likely to have an entry on a firm founder than on a small firm, and we often search based on a product name and one or more human names. Others we find starting on LinkedIn, where employees or contractors frequently list their work on an app or website development project as part of their resume. When they include the product name, that gives us a bridge to the firm name.

There are some products for which we cannot find a commercializing firm. These fall into two categories. Some products have gone out of business and have no entry in TechCrunch, nor are affiliated personnel discoverable by web search (in LinkedIn, TechCrunch, Wikipedia, or elsewhere). There are products – Satoshi Nakamoto products – where the commercializing entrepreneur has taken extraordinary measures to avoid discovery. Many of these are fan fiction sites, likely seeking anonymity in an era of strong copyright. One is a math teaching program where we know the entrepreneur is a former teacher, but we cannot learn her location or last name. We searched diligently and, in several cases, turned up discussions among site users on Reddit or in a site's discussion boards who could not identify the entrepreneur. We exclude these 25 products since we cannot assign firm or region variables.

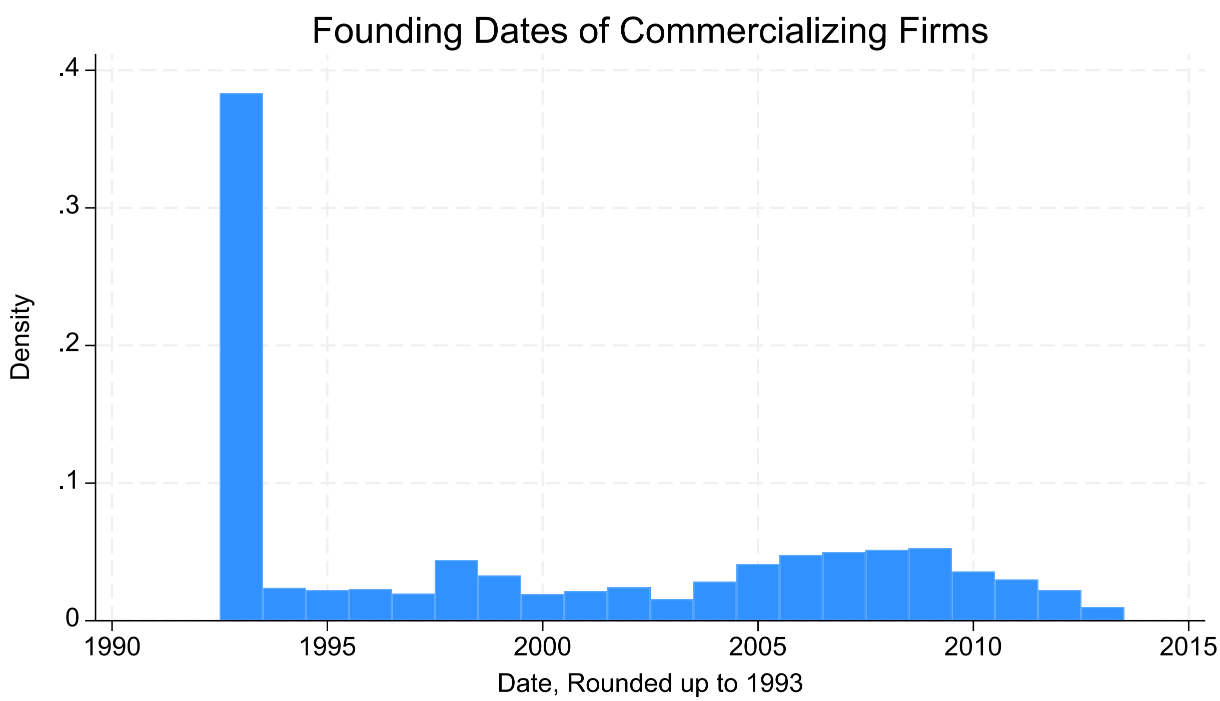
Some approaches did not work well at all. We attempted more automated mechanisms for finding parent firms and their location, and report to future scholars how badly they went. The ownership of domain names is semi-public information. The “semi” arises because our hard-to-find firms, and not only do they use shell corporations to own domain names, but our out-of-business firms have often transferred domain names to brokers. We attempted to use standard IP tools to locate the server since we have an IP address for all the websites and most apps (on the developer’s website). This worked at national boundaries but had at least two complex problems within the US locations. First, many servers are located away from the commercializing firm, for example, located in Ashburn, Va. Second, many reported IP locations are highly ambiguous, such as being at the center of the United States (under a lake) or at the center of a region (under a bowling alley, in one instance).

F. Founding date of commercializing firm

Finding a firm in TechCrunch or CapIQ yields a founding date. Failing that, we find a local or trade press news story that interviews a founder about the firm’s founding. We find others in firm Wikipedia entries. We assign a founding date from public sources for other entities not in those sources. Since our data processing and statistical efforts do not distinguish among founding dates before 1993, we often conclude the search for old "firms." We assign, e.g., the State of Florida a founding date some centuries ago.

Our searches attempt to assign the founding date of the original commercialized in a case where there has been a merger. For example, YouTube and Instagram have founding dates that are different from those of Google and Facebook. We found CapIQ's merger history too clumsy to help here. Our " founder" and " founded"-based searches for firm identity typically led us to the original commercialization date.

In Appendix Figure A1, we show the histogram of founding dates, where foundings before 1993 are rounded up to 1993.



G. Some details on location assignment

Our goal in location is to place the commercializing activities associated with the product inside one of the 558 statistical areas of the US or to become certain it lies outside them.

Most locations are the headquarters of an established firm, typically found in CapIQ, or the sole location of a new firm, typically found in CrunchBase. In some cases, as described above, we did not find the firm in those sources and used other search strategies to find the firm, looping back to those data sources for location. A few locations require more adjustment. As with the founding date, sometimes a parent company's HQ location has been assigned to a recent acquisition. Relatedly, a parent company's HQ location is sometimes assigned to a subdivision with a distinct identity in another location. In those instances, we are looking for the locus of the firm's assets complementary to the co-invention. For young firms, sometimes there has been a move to be closer to complementary assets at the regional level, such as the move of the very young Facebook to Palo Alto. In those cases, we use the "to" location as our goal is to find the regional complementary assets.

We discuss the very few firms outside the 558 regions below.

There is very little co-invention with sufficient usage to make it into our sample in low-density parts of the United States. First, of the 558 primary statistical areas (CSA, MSA, and Micropolitan Statistical Areas) of the US, 433 do not have any commercializers in our dataset. Second, seven of our 2426 products were commercialized in micropolitan statistical areas that are not part of a CSA or MSA. Third, five commercializing firms are outside the 558 primary statistical areas. We discuss these 12 firms in detail.

The forecast that this new opportunity would enable widespread rural entrepreneurship (Cairncross, 1997) is clearly incorrect. By examining the twelve firms and their products, we can learn more about the forces restricting rural entrepreneurship, specifically the forces limiting rural entrepreneurship that seek a broader-than-local market.

We now go through the 12 firms, grouping similar stories.

Three of the twelve firms, two ISPs, and a custom furniture and cabinet store, are in vacation/seasonal home areas. Their customers include people living in the areas year-round and seasonal residents. Connecting to customers when they are not in town offers an immediate benefit from improved communications media. The ISPs, for example, have seasonal pricing plans for their customers and enable, through simple e-commerce, turning the service on or off. Similarly, the cabinetry store enables customers to shop for vacation house furnishings from their primary residences in the city. While this is a connection to remote customers, it is, at the same time, a connection to existing local customers. The fixed costs of having one of these businesses can be remarkably low. The ISPs, for example, are typically run by technically fluent entrepreneurs who offer ISP service and PC and networking installation and repair. These entrepreneurs themselves prefer to live in the vacation spot.

Two school districts, each with multiple campuses, are located in sprawling districts with students spread out over many small towns. They provide routine school information to widely scattered but still local students.

The next two firms are less local. One is a consulting firm with a widely dispersed workforce and a rural HQ (where the founder wants to live). The Cabela's retail chain serves hunters and other primarily rural outdoor enthusiasts. It was founded as a mail-order store in rural Nebraska. The chain kept its HQ in Sidney, NE, while opening stores there and in distant locations. The online Cabela's is an improvement over the mail-order business, creating a multichannel retailer that can serve rural customers throughout the country.

The final three firms on this list undertook novel co-invention and served nationwide markets from rural locations.

The mobile and website advertising agency Adhitzads has a rural location and aspires to become a national business in the three-sided market of content providers, advertisers, and users. The firm is four orders of magnitude smaller than the significant national firms in this business. It has, however, survived until today, employs a few people, and has various Bitcoin sites as its largest content providers and a partially overlapping set of Bitcoin market participants as its primary advertisers.

The Zippo lighter company exemplifies radical innovation by an existing business. It is located in the town where its founder and his descendants (it is still a family firm) have resided since the 1930s. With the decline of smoking, Zippo was not a growth business. They retained a mobile development firm to make a virtual lighter, which makes a smartphone look like a lighter, though there is no flame. This was a very attractive mass-market proposition that peaked at 15 million users and was profitable. Like physical Zippos, virtual Zippos can be customized, have a low cost (in the virtual case), and offer a high revenue marketing opportunity. Zippo was able to leverage its very widely known brand name to gain value from the new opportunity.

Finally, the word game maker Blue Ox Family Games is located in a very rural part of Maine. Its most successful product has been adopted to help 4th graders learn more vocabulary. This is another company founded by a rural IT professional, in this case, a self-taught programmer who bought a TRS-80 as a child. After a stint at a local business that failed, he started an IT consulting business, which he then transitioned into the app business. Marketing ideas come from the founder, who tries the games out on friends and family; there is no need to co-locate with customers or markets because the games are written for people with the founder's tastes. (Although he and his friends do not seem to have unusual tastes. One of the authors can testify that at least one app is a reasonably engaging word game.) Employees are scattered around rural Maine. This is, in short, an example of the rural entrepreneurship forecast's cliché firm. It is, alas, also the lone example.

None of these small-town and rural firms needed to gain marketing knowledge about markets or customers outside of rural America. They indeed had access to technology that allowed them to take advantage of the new opportunity. In short, the bottlenecks to large-scale rural entrepreneurship in this area lie in co-invention costs, not invention costs.

Google and Facebook chose their location to be near other "tech" firms before pivoting to targeted advertising as their primary revenue source. Each faced numerous commercial issues with implementing operations to generate revenue at scale. Amazon's decision to locate in Seattle combined both technical and commercial motives. By 2013, however, we were observing more than these leading firms. Hundreds of others are collocated in these clusters. A cluster can thrive in many places, but the arbitrariness of the initial experience should not take away from the importance of the positive feedback that emerges after a cluster begins.

Table A6 provides a count of products, showing where the clusters arose and where they did not arise. This table displays an unweighted count of the locations of properties, which are color-coded according to the top decline in population (dark blue), the next decile in population (light blue), and the third decline in population (black).

Table A6: Sample Distribution of Location of Products

Location (CSA, CBSA, or Rural Town)	Freq.	Location (CSA, CBSA, or Rural Town)	Freq.
San Jose-San Francisco-Oakland, CA	595	**Pittsburgh-New Castle-Weirton, PA-OH-WV**	12
New York-Newark, NY-NJ-CT-PA	422	**St. Louis-St. Charles-Farmington, MO-IL**	12
Los Angeles-Long Beach, CA	209	**Grand Rapids-Kentwood-Muskegon, MI**	11
Seattle-Tacoma, WA	126	**Portland-Vancouver-Salem, OR-WA**	11
Washington-Baltimore-Arlington, DC-MD	100	**Sacramento-Roseville, CA**	11
Boston-Worcester-Providence, MA-RI-NH	86	**Charlotte-Concord, NC-SC**	10
Atlanta–Athens-Clarke County–Sandy Sp	68	**Raleigh-Durham-Cary, NC**	10
Dallas-Fort Worth, TX-OK	65	**Richmond, VA**	10
Chicago-Naperville, IL-IN-WI	60	**Tampa-St. Petersburg Clearwater, FL**	10
Philadelphia-Reading-Camden, PA-NJ-DE	46	**San Antonio-New Braunfels-Pearsall, TX**	9

Austin-Round Rock, TX	44	**Indianapolis-Carmel-Muncie, IN**	8
Denver-Aurora, CO	31	**Knoxville-Morristown-Sevierville, TN**	7
Minneapolis-St. Paul, MN-WI	28	**Nashville-Davidson–Murfreesboro, TN**	7
Salt Lake City-Provo-Orem, UT	27	**Boise City-Mntn Home-Ontario, ID-OR**	6
Miami-Port St. Lucie-Fort Lauderdale,	26	**Buffalo-Cheektowaga-Olean, NY**	6
San Diego-Carlsbad, CA	25	**Fayetteville-Springdale-Rogers, AR-MO**	6
Hartford-East Hartford, CT	20	**Louisville/Jefferson County–Elizabeth**	6
Columbus-Marion-Zanesville, OH	19	**New Orleans-Metairie-Hammond, LA-MS**	6
Kansas City-Overland Park-Kansas City	19	**Omaha-Council Bluffs-Fremont, NE-IA**	6
Phoenix-Mesa, AZ	18	**Des Moines-Ames-West Des Moines, IA**	5
Orlando-Lakeland-Deltona, FL	16	**Memphis-Forrest City, TN-MS-AR**	5
Cincinnati-Wilmington-Maysville, OH-K	14	**Milwaukee-Racine-Waukesha, WI**	5
Las Vegas-Henderson, NV	13	**Springfield, MA**	5
Cleveland-Akron-Canton, OH	12	**Little Rock-North Little Rock, AR**	4
Detroit-Warren-Ann Arbor, MI	12		
Houston-The Woodlands, TX	12	**79 Places with 1, 2, or 3 Products**	125
... continued in the next column			

Source: Authors' calculations.

The correlation between employment in the region and products in table 5 is .733. This is calculated over the subset of US places with any products. Most small places have no products and are thus excluded. We provide an analysis of this in Section 7, especially Table 13.

H. Multiproduct Firms

Multiproduct firms and the unit of observation. Many firms in the sample supply multiple products. Sometimes, a firm's products fall into different classifications, especially for firms launching new product lines in response to the CCT opportunity. We resist the convention – commonly used in industry reporting databases – to undertake firm-level product classification. Apple, for example, classifies itself as a manufacturing firm. We keep that designation for the Apple.com website, which sells and supports Apple hardware, but we place the iTunes Store in retail trade.

This observation has practical implications for analysis. The most sensible observation unit for most analyses is usually the product, which does not confuse product classes. However, that choice raises questions about the prevalence of multiproduct firms. We find 1657 distinct firms that commercialized 2426 products. Table A7 shows that almost 80% of those firms have only one product with sufficient prominence to be included in the dataset, and another 229 firms have two products.

A few firms (named in the rightmost column) have a dozen or more products. These include some game firms (which tend to have one app per game), media firms such as ESPN and Disney (which tend to have one app and website per show), and firms that have transformed industries. Apple, AOL, and Google are multiproduct firms (at least on the consumer-facing side of their platform supply). Additionally, Verizon has substantial distribution advantages in mobile, which has led it to introduce modestly successful imitations of many leading apps. In summary, no simple pattern explains the behavior of all multi-product firms.

Table A7. Number of Products Per Firm

N products	Freq	Percent	Firm
1	1315	79.36	
2	229	13.82	
3	55	3.32	
4	25	1.51	
5	5	0.30	
6	7	0.42	
7	3	0.18	
8	2	0.12	
9	4	0.24	
10	2	0.12	
12	1	0.06	**Verizon**
14	2	0.12	**AOL, Apple**
17	1	0.06	**Yahoo!**
18	1	0.06	**ESPN**
19	1	0.06	**Microsoft**
22	1	0.06	**Zynga**
27	1	0.06	**Disney**
30	1	0.06	**Electronic Arts**
55	1	0.06	**Google**
Total	1657	100.00	

In Table A8, we further explore the behavior of multi-product firms with exactly two products. Two-thirds (153) of these firms have one website and one app in our data. In the most typical situation, the website was first developed in the 1990s, and the app was created after the spread of smartphones. Looking further, while most multi-product firms (66.8%) supply both a mobile app and a website, a substantial number offer two products using the same CCT ecosystem, either exclusively as apps or websites.

Table A8 Two Product Firms

Breakdown	N	Percent
Both apps	40	17.5

Split	153	66.8
Both web	36	15.7
Total	229	100.0

I. Product type classification

Table A9 lists the new classes and illustrates them with the largest (by usage) product in each one. The table is sorted in declining order of usage of the largest product in the class.

Table A9: Leading Products and their Classification

Property	Classification
Facebook (App)	Targeted Advertising
Youtube.com	Information Products (Media)
Apple iTunes (App)	Retail Trade
Yahoo.com	Information Products (Other)
Words With Friends Free (App)	Information Products (Games)
Apple.com	Manufacturing
Calorie Counter & Diet Tracker by MyFitnessPal (App)	Health Care and Wellness Information
Bank of America (App)	Finance and Insurance
Ecollege.com	Educational Services
Starbucks (App)	Accommodation and Food Services
Bible (App)	Other Services (except Public Administration)
Answers.com	Travel Agents Removed: rest of Administrative and Support and Waste Management and Remediation Service
Puffin Web Browser Free	Professional, Scientific, and Technical Services (less Advertising)
REALTOR.com Real Estate Search	Real Estate, Rental, and Leasing
Expedia.com	Travel Sales (not by travel companies)
Ca.gov	Public Administration
Ticketmaster.com	Arts, Entertainment, and Recreation
Fedex.com	Transportation and Warehousing
Wonatrading.com	Wholesale Trade
Westchesterhealth.com	Health Care and Health Finance
Teambizwiz.com	Utilities and Construction

The new product classification has three main elements: 1) young firms tend to classify themselves in NAICS 51, whatever they do, 2) products in our sample come from an odd corner of a 2-digit industry, and 3) some multiproduct firms self-classify in a way unrelated to their CCT-based product.

Table A10: Crosstab of NAICS and new product classification

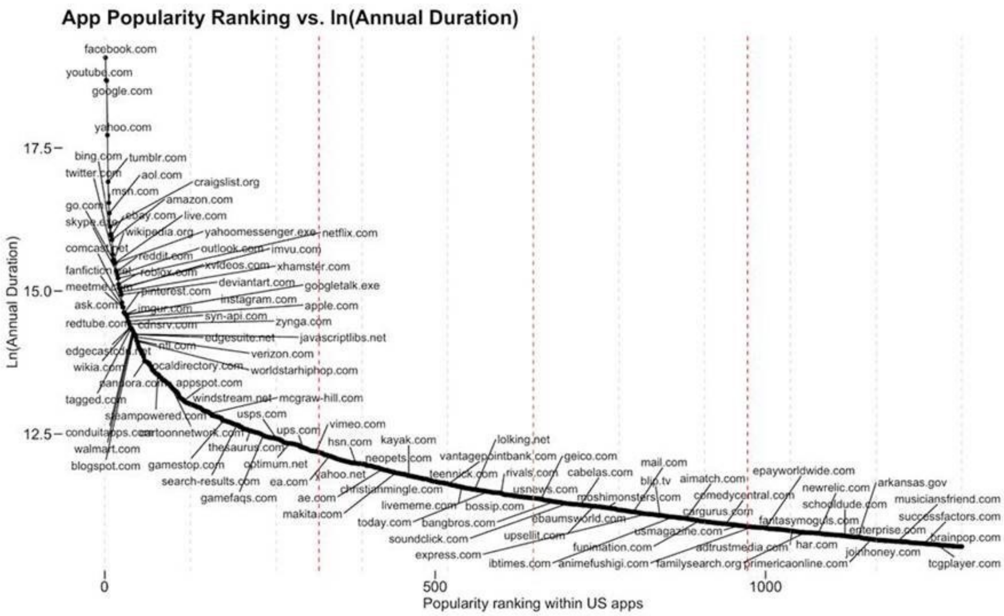
	2 digit NAICS													
Product Classification	31-33	42	44-45	48	51	52	53	54	56	61	62	71	72	Total
Accommodation and Food Services	0	0	0	0	0	0	0	0	0	0	0	0	16	16
Arts, Entertainment, and Recreation	0	0	0	0	0	0	0	0	0	0	0	17	0	17
Educational Services	0	0	0	0	0	0	0	0	0	117	0	0	0	117
Finance and Insurance	0	0	0	0	0	93	0	0	1	0	0	0	0	94
Health Care and Health Finance	0	0	0	0	1	0	0	0	0	0	7	0	0	8
Health Care and Wellness Information	0	0	0	0	12	0	0	0	0	0	38	0	0	50
Information Products (Games)	4	1	8	0	330	0	0	2	0	2	0	7	1	355
Information Products (Media)	1	0	5	2	525	1	0	2	0	0	0	17	0	553
Information Products (Other)	1	0	0	0	533	0	0	1	0	0	0	0	0	535
Manufacturing	22	0	0	0	0	0	0	0	0	0	0	0	0	22
Professional, Scientific, and Technical Services (less Advertising)	0	0	0	0	0	0	0	36	0	0	0	0	0	36
Real Estate and Rental and Leasing	0	0	0	0	0	0	29	0	0	0	0	0	0	29
Retail Trade	11	0	195	0	25	0	0	0	0	0	0	0	0	231
Targeted Advertising	0	0	1	0	146	0	1	78	0	3	0	1	0	230
Travel Agents Removed: rest of NAICS 56	0	0	0	0	0	0	0	0	6	0	0	0	0	6
Travel Sales (not by travel companies)	0	0	0	0	1	0	0	0	21	0	0	0	0	29
Total	39	8	209	24	1,573	94	30	119	28	122	45	42	17	

Not shown in Table A10 because all in one product class:

NAICS 22, 1 product in Utilities and Construction
NAICS 23, 2 products in Utilities and Construction
NAICS 47, 7 products in Travel Sales (not by Transportation Companies)
NAICS 81, 12 products in Other Services (except Public Administration)
NAICS 92, 57 products in Public Administration

J. Usage and Rank

Figure A1: Rank versus Usage, Ordered by Rank.



Many studies assume a descriptive inverse-proportional relationship between usage and rank. According to our website data, as shown in Figure A1, the relationship holds better outside the top 100. Similar results are documented for App data in Bresnahan-Davis-Yin (2015). Following a similar logic to that paper, we estimate the descriptive relationship outside the top-ranked websites.

We estimate the relationship between usage and rank for sites ranked between 2000 and 10000. The estimates for the incremental websites are ln(Y) = 20.14 – 1.22*ln(rank). For prediction purposes, that implies an estimate of b in $\ln(rank_{2013}) = a + b*\ln(audience_{1993})$ is 0.78.

This has one additional implication. We found that incremental co-invention accounts for only 6% of total value—i.e., the value of returns from investment in novel co-invention exceeds the value of the investment in incremental co-invention by a large amount. However, that is only *observed* incremental success. There is likely more *unobserved* success in incremental co-invention. Could the value of these unobserved investments bring the value of incremental co-invention investments to a higher order of magnitude? These estimates suggest not.

These estimates provide a bound on the unobserved gains from spreading incremental co-invention and a sense of the order of magnitude of the unobserved usage. Specifically, it is possible to calculate the usage assuming that the functional form remains the same and that adoption is not limited. By integrating under the function, we can calculate the ratio between the observed usage in our dataset (for web users) and the maximum unobserved usage. The ratio is 2.39.[19]

In other words, the unobserved value cannot exceed 2.4 times the observed value (and is likely less). With the *observed* value being 6% of the total, adding twice its value would still result in less than 18%, which is still a small fraction.

To summarize, because the ratio of novel to incremental value is so large, the long tail likely does not contain additional unobserved value of sufficient magnitude to change the conclusion.

K. Local Media and Firm Location

National firms own many local news outlets. Every local media outlet is placed in its local region, even if the co-invention may have come from the national firm, not the local subsidiary.

We examine all local media in our main dataset, predicting usage by city size. There is a noisy but proportional relationship, which encourages us to utilize further city-level measures for all industries, including those lacking product-level pre-CCT demand measures. We estimate an association between the value created by all products in our dataset and the assets associated with each product's location.

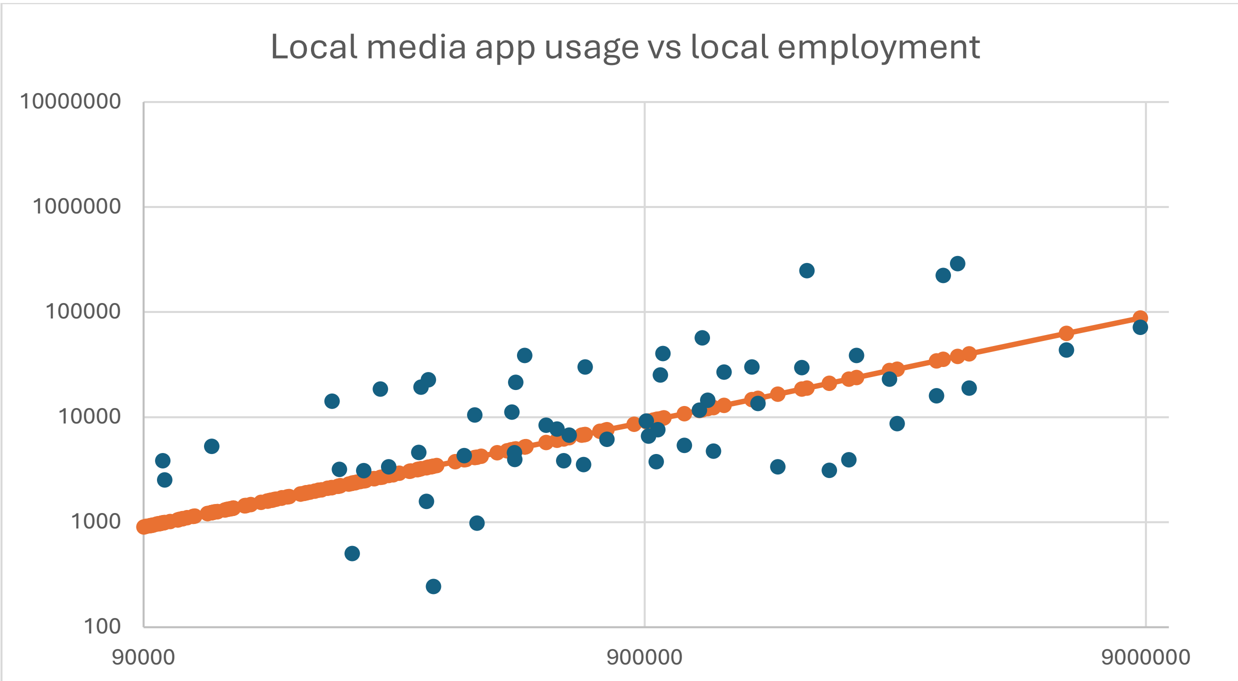
These regressions estimate the relationships between co-invention outcomes and measures of local business assets. We test whether the local external effects among commercializers are more critical for novel co-invention. Relatedly, we test whether existing firm assets, using the former connections, are more essential for incremental co-invention.

The first evidence in our main dataset is the 146 local media observations. We plot their CCT usage (Web or App) against the size of their city, measured by total employment

[19] More precisely, this is a comparison of the value from incremental adoption for rank 500 to 2500, which we observe. It is compared to the value of adoption for 2500 to infinity. All the additional value beyond 2500 is no more than 2.39 times the value of what we observe. If the adoption stops at a rank of one million, the ratio of observed to unobserved is 1.72. If the adoption stops at a rank of ten million, the ratio is approximately 2.

there. To show scale, we add a line where usage expands proportionally to city size. While there is variation, a local media product's usage is approximately proportional to city size.

Figure A2: Local Media versus Local Employment



Source: authors' data and US Census. The "city" of a local media product refers to the location in which it is published, not necessarily the location of its innovator.

Thus, for local media, we observe the implication of incremental co-invention discussed in Section 2. The usage of CCTs is approximately proportional to the existing market size (as shown by the red line, which simulates proportionality).

Figure A2 is limited because local media tend to be too small to be included in our main database, where we only consider the largest local media products. Additionally, our model suggests proportionality to existing customers, whereas the Figure examines local market size. To sharpen the analysis, we turn to an examination of local radio co-invention and local and formerly local newspaper co-invention, where we can observe prior audience size.

L. The Misclassification Model of Aigner (1983) and an Alternative Regime Classification

There is judgment in our empirical regime definition. We have two strategies for checking our judgment. First, we examine the results of an alternative, more stringent definition of novel co-invention, which requires the new service to change the scale or scope of the firm significantly. For example, Verizon Wireless has the V Cast Music app (the studio that wrote it was Microsoft). Although it is novel in our base definition, it is used only by a trivial portion of Verizon phone customers. We reclassify it, along with other similar apps, as "incremental" in our stringent definition.

Appendix Table A11

Novel Incremental Strict Definition Stats of Usage

Summary statistics: mean, sd, count by(R2)

R2	Mean	SD	N
Incremental	395.9	1490.34	839.00
Novel (strict)	3269.57	31088.69	1587.00

The second approach is to use Aigner's model of misclassification of a dummy regressor, like our regime classification (Aigner 1973). The model permits us to assess our findings that the mean and standard deviation of usage are larger in the novel regime. We use it directly: how poor our regime classification must be, relative to the true intentions of the firms we study, for those findings to be incorrect.

The strong assumption of the model is that misclassification is independent across observations. The probability of misclassification may vary, however, depending on the true regime. For our purposes, it is easiest to write the probabilities of misclassification conditional on the observed regime. (These are determined, via Bayes Law, from the probabilities conditional on true regime.)

Let v be the probability a product was misclassified, conditional on it being observed in the Incremental regime, and let η be the probability it was misclassified, conditional on it being observed in the Novel regime. We assume that $v+\eta<1$, so that there is *some* information about the true regime in the observed regime. We then perform a simple sampling exercise in which $v=\eta$ and takes on values from .01 (almost no classification error) to .4 (almost no information in observed classification).

For each value, we randomly draw 10,000 samples. If the observed regime for a product is N, in each sample, it is N with probability $1-\eta$ and I with probability η, and similarly for products observed in regime I. In each sample, we calculate the mean and standard deviation of usage for Incremental and Novel products. We then average these across all

10,000 samples and report the result in Appendix Table A12. Rows of this table are values of v=η. The first two columns show the degree of misclassification and correct classification in the row. The rest of the columns show the mean and standard deviation of sampled usage by regime, and their ratios.

Appendix Table A12

ν, η	1- ν, 1-η	μ_I	σ_I	μ_N	σ_N	μ_N / μ_I	σ_N/ σ_I
0	1	409	1547	3154	30494	7.711491	19.7117
0.01	0.99	466	2614	3141	30409	6.740343	11.63313
0.05	0.95	691	6477	3084	30023	4.463097	4.635325
0.1	0.9	932	10010	3018	29607	3.238197	2.957742
0.15	0.85	1153	12869	2946	29151	2.555074	2.265211
0.2	0.8	1364	15457	2862	28563	2.09824	1.847901
0.25	0.75	1552	17526	2778	27932	1.789948	1.593746
0.3	0.7	1709	19035	2699	27455	1.579286	1.442343
0.35	0.65	1875	20776	2598	26597	1.3856	1.280179
0.4	0.6	2015	21968	2501	25925	1.241191	1.180126

The first row assumes no misclassification and thus reports the observed differences across regimes. These are substantial, with the Novel regime having a 7.7 times larger mean usage and a 19.7 times larger standard deviation of usage.

The degree of misclassification needs to be substantial for the regime differences to be small. For misclassification probabilities of up to 0.2, either the mean or standard deviation of the implied true distribution in the Novel Regime is more than double that of the Incremental Regime.

Appendix References

Abhishek, V., Li, B., 2019. "Omnichannel Customer Behavior in Retail Banking." In (Eds) Gallino, S, and A Moreno, *Operations in an Omnichannel World,* Springer, Cham.

Baloun, K., 2007. *Inside Facebook: Life, Work, and Visions of Greatness*. Trafford Publishing.

Berry, S., and J. Waldfogel, 1993, "Free Entry and Social Inefficiency in Radio Broadcasting," *Rand Journal of Economics,* 30(3), 397-420.

Bezos, J., 2021, *Invent and Wander, The Collected Writings*, Harvard Business Review Press.

Boik, A., S. Greenstein, J. Prince, 2019"The Persistence of Broadband User Behavior: Implications for Universal Service and Competition Policy." *Telecommunications Policy*. 43 (8).

Bresnahan, T.F., Davis, J.P., Yin, P.-L., 2015. Economic Value Creation in Mobile Applications, in Jaffe, A., Jones, B. (Eds.), *The Changing Frontier: Rethinking Science and Innovation Policy*. The University of Chicago Press.

Broeders, H., & Khanna, S. 2015. Strategic choices for banks in the digital age. New York, NY: McKinsey & Company

Bryar, C., and B. Carr, 2021, *Working Backwards, Insights, Stories, and Secrets from Inside Amazon*, Saint Martin's Press; New York.

Buchholz, D., M. B. Gross, A. A. Lopez-Fernandini, M. D. Schmeiser, 2014. "Consumers and Mobile Financial Services 2014," *Reports and Studies*, 1098, Board of Governors of the Federal Reserve System (U.S.).

Cairncross, F. 1997, *Death of Distance. How the Communications Revolution will Change our Lives.* Cambridge, MA; Harvard Business Review Press.

Campbell, D., and F. Frei. 2009, "Cost Structure, Customer Profitability, and Retention Implications of Self-Service Distribution Channels: Evidence from Customer Behavior in an Online Banking Channel," *Management Science,* 56, 1.

Duncan J. H., 1993, *American Radio,* Various reports. Indianapolis. Duncan's American Radio Inc.

Gallino, S, and A Moreno. 2019. *Operations in an Omnichannel World,* Springer, Cham.

Greenstein, S., 2016. *How the Internet Became Commercial. Innovation, Privatization, and the Birth of a New Network.* Princeton Press. Princeton, NJ.

Hastings, R., and E. Meyers. 2020. *No Rules Rules: Netflix and the Culture of Reinvention.* Random House; New York.

Iger, R. 2019. *The Ride of a Lifetime: Lessons Learned from 15 Years as CEO of the Walt Disney Company,* Penguin Random House, New York.

Kirkpatrick, D. 2011. *The Facebook Effect. The Inside Story of the Company that is Connecting the World*. Simon and Schuster; New York.

Levy, S. 2020. *Facebook. The Inside Story*. Penguin Publishing Group; New York.

Rossman, J., 2021. The Amazon Way: Amazon's 14 Leadership Principles. Clyde Hill Publishing; Seattle and Washington, D.C.

Schmidt, E., and Rosenberg, J. 2017. *How Google Works*. Grand Central Publishing; New York.

Smith, M.D., and R. Telang, 2016. *Streaming, Sharing, Stealing: Big Data and the Future of Entertainment.* MIT Press; Cambridge, MA.

Stone, B., 2013. *The everything store: Jeff Bezos and the age of Amazon.*

Uhl, A., L. A. Gollenia. 2014, *Data Enterprise Transformation, A Business-Driven Approach to Leveraging IT.* Routledge.

Xue, M., L. Hitt, P. Y. Chen, 2011, "Determinants and Outcomes of Internet Banking Adoption." *Management Science,* 57, 2.