ICT Behavior at the Periphery: Exploring the Social Effect of the Digital Divide through Interest in Video Streaming

Erik W. Hanson and Justin C. LoTurco

Under the supervision of

Dr. Michelle P. Connolly

Dr. Grace Kim

Honors thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

Duke University
Durham, North Carolina
2019

Acknowledgements

We would like to thank our thesis advisors, Dr. Michelle Connolly and Dr. Grace Kim, for their guidance, patience and wisdom. Both advisors helped us to narrow the scope of our interest to a single topic, and provided us with extensive tools and resources to assist our research. We are also grateful to our Spectrum Lab peers, as well as our Honors Thesis Seminar classmates, for listening to our developments and offering valuable feedback.

Abstract

We investigate the factors that influence changes in consumer behavior with regard to video streaming. We focus our analysis on the effect of bandwidth impairment to explore a potential consequence of the digital divide. To measure the change in relative popularity of video streaming services, we use Google Trends data as a proxy. We then investigate whether broadband speed improvements in rural vs. urban regions affect the proxy differently. We find that increasing the broadband speeds in rural regions appears to stimulate greater interest in video streaming than equivalent speed increases in urban regions.

JEL Classifications: C33; J11; L96

Keywords: Digital divide, video streaming, Google Trends

"Since my first day as Chairman of the FCC, my number one priority has been closing the digital divide and bringing the benefits of the Internet age to all Americans" - FCC Chairman Ajit Pai[1]

I. Introduction

The digital divide between urban and rural regions within the United States becomes apparent when access to high-speed fixed broadband is compared. Under the FCC's existing speed benchmarks of 25 Mbps for downloads and 3 Mbps for uploads, only 65% of Americans living in rural regions have access to high-speed fixed services; compared to 97% in urban regions.[2] Past studies have attempted to prove the economic, political and health disparities that result from bandwidth impairment (Crandall & Jackson, 2001; Rains, 2008; Greenstein & McDevitt, 2011; Miner, 2015). Through our research, we hope to explore whether or not the consequences stemming from the digital divide extend to social effects, specifically with regard to video streaming services.

According to Netflix, streaming a movie requires bandwidth speeds of somewhere between .7 Mbps to 5.3 Mbps. Increasing the quality of the audio or video will increase the level of bandwidth needed. Netflix recommends 25 Mbps for optimal quality of streaming.[3] This infers that 35% of rural Americans are incapable of experiencing optimal quality when streaming Netflix, with some regions having not enough broadband to stream videos whatsoever. With the rural US having significantly less access to broadband than its urban counterpart, and with bandwidth-intensive video streaming service content becoming increasingly prevalent in American culture, we set out to test if broadband speed improvements in rural cities have a greater impact on video streaming activity than in urban cities.

Google search frequency data relating to each video platform are used as a proxy for user activity because these streaming services do not publicize the relevant data needed. We select three states — Alabama, Alaska and Tennessee — to scrape the google search data for based on a set of predetermined criteria. For each selected state within the US, Google search data pertaining to multiple cities, some rural and some urban, are collected across 6-month periods of time from 2009 to 2014. For identical time intervals, each city's population, income, age, rurality

[1] https://www.fcc.gov/about-fcc/fcc-initiatives/bridging-digital-divide-all-americans
[2] https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2018-broadband-deployment-report
[3] https://www.otelco.com/how-much-internet-speed-do-i-really-need/

level and broadband speeds are also collected. A panel data analysis is constructed with cities as the cross-sectional unit being compared longitudinally over the 12 semiannual periods.

Before the search frequency data is used, a regression is created to test whether population, income, age or level of rurality correlates with speed. This test allows us to determine what, if any, attributes of a city are significant in identifying regions with bandwidth-impairment. Should a variable be found significant, it could help the government in identifying and prioritizing regions for broadband subsidies and bridging the digital divide. The second regression uses panel data with the same spatial and temporal units to investigate how certain variables influence changes in video search queries. If we find that an equivalent percentage increase in broadband speed has a greater impact on video streaming interest in rural regions than in urban regions, it would suggest that demand for broadband is not being met in rural regions. If this result is significant, it could be said that the consequences of the digital divide in the US extend to the social effect of unmet demand for video streaming platforms. Unequal access to these entertainment sites would inhibit rural Americans from experiencing the same level of participation in the popular culture of the 21st century.

II. Related Literature

A. Measuring Broadband's Impact on Societal Advantages

Multiple studies have set out to test the legitimacy and magnitude of the anticipated benefits of broadband on communities. Interest in this topic was first galvanized when economic growth potential was first linked to broadband adoption. Crandall and Jackson (2001) were the first to estimate a demand function for high-speed broadband access, with consumer surplus being calculated in an effort to quantify the economic benefit of ubiquitous broadband adoption in the US. Greenstein and McDevitt (2011) improve upon this demand function by redefining consumer surplus for broadband adoption as measured by taking the difference between what broadband users actually paid and what they would have paid had dial-up continued and not been replaced by broadband, which proved to be more accurate.

In addition to economic benefits, access to broadband has been found to also provide political and health advantages. In testing whether greater broadband speeds also imply regions are at a political advantage, Miner (2015) examined whether regions in Malaysia with higher internet penetration experience higher voter turnout rates. At the time of the study, internet speeds in Malaysia were increasing rapidly due to an influx of municipal broadband investments, hence why the region was used as a case study. Using OLS estimates comparing internet penetration to voting behavior, Miner found that Malaysian regions with higher internet penetration resulted in higher turnouts and higher turnovers of the incumbent. Shifting focus to the potential for broadband to bring health benefits, Rains (2008) examined inequities in the adoption of broadband technology and used the health information seeking (CMIS) model to measure the potential benefit.[4] Using data from a population-based survey conducted by the National Cancer Institute in 2005, Rains found that those with a broadband connection were more likely to use the Internet for health-related information seeking and communication than those with a dial-up connection.

B. Google Trends as a Proxy

[4] The Comprehensive Model of Information Seeking, or CMIS, is a theoretical construct designed to predict how people will seek information. It was first developed by J. David Johnson and has been utilized by a variety of disciplines including Library and Information Science and Health Communication.

This paper also relies heavily on the potential for Google trends data to reveal the intent of an internet user to make an economic decision in the near future. In an IMF working paper titled "In Search of Information: Use of Google Trends' Data to Narrow Information Gaps for Low-income Developing Countries," authors Futoshi Narita and Rujun Yin provide a brief yet comprehensive overview of the use of Google's SVI in forecasting and now casting economic variables (Narita and Yin, 2018) Citing Google's 90% share of global users and the quality, availability, and timeliness of the information aggregated by online search engines, the authors demonstrate the effectiveness of Google search behavior in predicting a range of macroeconomic variables from car sales to the unemployment rate to tourist arrivals to home and oil prices. This paper operates under the assumption that the aggregation of Google searches which include the name of popular pure play OTT content providers such as Netflix within the query can be used as a reliable predictor of the searchers' intent to use the related online service. This paper treats this basic predictive functionality and the observed correlation between search term and real-world activity as the dependent variable, which is an unusual approach.

C. Behavioral Economic Theory

Behavioral economic theory stipulates that the consumption of leisure services necessitates the a priori use of some decision-making process. Several models assume a multi stage process within which consumers engage. Simon (1960) first centralized the decision-making process in three sequential stages of activities, including intelligence activity, design activity, and choice activity. Kollat, Engel, and Blackwell (1968) proposed the traditional five-stage model of the consumer buying process which stated that consumers move through the stages of need recognition, information search, evaluation of alternatives, purchase, and post-purchase behavior. Other papers provided alternative conceptualizations for consumer behavior from the vantage point of brands. The Nicosia Model, for example, identifies a similar flow of events through different stages except formulated on the basis of communication between brand and consumer (Sengupta & Nicosia, 1969). Sheth and Howard (1969) proposed a general theory of the buyer behavior of individuals over time in order to show how the degree of repetition in interaction with brands created a hierarchal model of consumer decision making processes including extensive problem solving, limited problem solving, and habitual response behavior.

Despite the fluid and chaotic nature of the online information gathering process, this paper assumes the basic proposition outlined by the Mintzberg model that a basic structure underlies these “unstructured” processes (Mintzberg, Raisinghani, & Theoret, 1976). This paper also accepts that consumer choice may be a function of multiple consumption values distributed across a multi stage process, but considers functional value to be the primary driver of choice and independent from social, conditional, emotional, and epistemic values, which are contextually and personality-driven and therefore cannot be easily measured (Sheth, Newman, & Gross, 1991). The focus of this paper is the differential contribution bandwidth speed to the salient functional, utilitarian, and physical attributes of OTT video streaming platforms.

III. Data

3.1 Dependent Variable

This paper uses Google Trends data to estimate consumers' intent to engage with over-the-top (OTT) entertainment video services. Large data sets of search engine queries have been shown to contain signals representative of real-life patterns and previous literature has employed publicly available historical search query data to forecast social and economic trends with significantly superior accuracy, granularity, and timeliness than traditional methods. The reason why these tools have proved to be such effective prediction mechanisms isn't because they measure the actual phenomena in question, but rather because search engines play a critical intermediary role in connecting users with what they want when they know that they want it. Nanoecomic data is an immediate and powerful predictor of transactions. Moreover, the overwhelming majority of users include only a few queries per search while 76% do not go beyond their first and only query, indicating that a search engine's focus on keyword relevancy enables it to direct the majority of internet traffic to their final search destination with minimal friction (Ching 2006).

3.1.2 Google Trends Basics

To measure the change in consumer behavior with regard to bandwidth-intensive applications, Google search queries pertaining to video streaming domains were explored. The OTT services that consume the most bandwidth in the United States were selected: Netflix, YouTube, Amazon Prime Video, and Hulu.[5] Because these companies do not publicize regional user activity data, Google search queries relating to attempts to access their webpages were used as a proxy. There are a few other considerations we should keep in mind. The percentage of incoming traffic for any online service is distributed across a range of channels, including direct, email, referrals, social, organic search, paid search, and display ads. For popular online multimedia applications, traffic is heavily concentrated in the direct-to-website category (approximately 70%), while organic search is a distant second (approximately ~ 20% for platforms included). Google Trends pulls unbiased samples from organic searches input into the Google search engine. If "Netflix" was the keyword, the data would not account for any

[5] https://www.businessinsider.com/which-services-use-the-most-bandwidth-2015-12

searches that did not include Netflix in the actual query, even if the user clicked on the Netflix website on the results page. It would account for all searches that include Netflix, even if the user did not click on the Netflix website on the results page.

Our choice of using Google Trends data works for several reasons. Organic search is only responsible for a minority of the incoming traffic for the platforms included in our query, but it is the optimal means of identifying an increase in interest that is specifically correlated with an increase in broadband speed. Traffic generated by email, referrals from other websites, social media platforms, paid search, and display ads is automatically influenced by the medium of transmission. Additionally, while the direct category accounts for the range of different devices through which online multimedia applications can be accessed, including desktop and laptop computers, video game consoles, smartphone and tablet apps, smart TVs, and others, we are specifically focused on interest in use, not actual use. That is, an increase in use from one period to the next is likely strongly correlated with an increase in consumption on a user-by-user basis. Direct-to-website traffic implies that a company's website was previously saved in an individual's internet browser or that its app was previously downloaded on an alternative device. Given that we are unable to access data on the number of unique visitors and downloads accruing to a platform sorted by a geographical region as granular as the city, organic search is an appropriate proxy for newfound interest.

One limitation of obtaining data from Google Trends is that the results are pulled from only a sample of the total search volume. To prevent search queries for different streaming services being pooled from different sample sets, all four domain names were grouped together into a single search query. Grouping multiple keywords together returns the proportion of search terms containing at least one of the service names over the total number searches from one sample set. For example, inputting the phrase "Netflix + YouTube + Hulu" will pick up any Google search that contains either name. Because the Google Trends data proves that Amazon Prime Video is often colloquially referenced to as "Amazon Video" and "Prime Video," and inputting the entire phrase would require all three words to be present for a query to register a match, we extended the search query to be "Netflix + YouTube + Hulu + 'Amazon Video' + 'Prime Video'." The quotations around "Amazon Video" and "Prime Video" require that the search contains the name "Amazon" or "Prime" directly before the word "Video" to prevent unrelated queries like “Video of the Amazon River” from registering as false positives. Our selected query was further extended to avoid type 1 errors with the addition of “- who - what - where - when - why - how.” Using a subtraction symbol followed by a word will eliminate any searches containing these words from the set of matches. This way, searches related to questions regarding each OTT service will not be mistaken for attempts to find and access the site.

In constructing our query, we did not consider SNS’s and or services provided directly by telecom operators, instead focusing on ‘Pureplay’ offerings and selecting the most well-known brands in the space, namely Netflix, Hulu, Prime Video, and YouTube. There are a few obvious observations:

1) There is considerable variation in the business models and the nature of value exchanges within the business models of the companies included in our query (Verno, Teixeira, & Brochado, 2017). Netflix is the only standalone subscription-based video-on-demand services (SVOD). Hulu implements a hybrid model that combines SVOD with Ad-supported video on-demand (AVOD). YouTube’s audience is drawn through a wide-range of free-to-air content and monetized through Ad-supported video on-demand (AVOD) functionality, while Prime Video is a bundled product offered to Amazon Prime subscribers but is open to the general public as well through a transactional video-on-demand (TVOD) model.
2) Consumer demand for digital video has increased dramatically across the board, but distribution, promotion, and consumptions is becoming increasingly influenced by social video. Major SNS’s like Facebook have scaled their social media platforms and mobile-based social video apps to become massive video distribution hubs, although the inability to separate users’ intentions for visiting a social media website negates the rationale for including these companies in our analysis.

3.1.3 Google Search Level Construction

The normalized frequency for a given search query q in a city i during a 6-month time period t is:

$$\widetilde{queryshare}_{t,i} = \frac{\#\ searches\ for\ query\ in\ city\ i\ during\ period\ t}{\#\ of\ total\ searches\ in\ city\ i\ during\ period\ t}$$

When extracting city-level data points from a sub-region, such as a U.S. state, Google Trends scales the normalized frequencies of each city from 0 to 100, with each point divided by the highest $\tilde{z}_{q,t}$ within the selected time period, which is assigned a value of 100. The equation for relative popularity is denoted below, where the i → N subscript denotes that the $\widetilde{queryshare}$ value in the denominator will only select city i if the other N – 1 cities in the state have lower normalized frequencies.

$$\widetilde{querypop}_{t,i} = \frac{\widetilde{queryshare}_{t,i}}{\max(\widetilde{queryshare}_{t,i \to N})} \text{X } 100$$

The proxy we used for consumer interest in OTT video applications was as follows:

netflix + hulu + youtube

+ "prime video" + "amazon video"

- why - how - what - where - when - who

This search term asks Google Trends to return data for searches containing one or more of the words "Netflix" OR "Hulu" OR "Youtube" including any variation such as "watch Youtube," either of the phrases "prime video" OR "amazon video" which can also be modified around their edges but must be arranged in the proper two-word sequence, and excludes all searches which fit the prior criteria but include one or more of the words "why" OR "how" OR "what" OR "where" OR "when" OR "who." Substituting search term s in for query.

$$\widetilde{termshare}_{t,i} = \frac{\#\ of\ searches\ implied\ by\ search\ term\ s\ in\ city\ i\ during\ period\ t}{\#\ of\ total\ searches\ in\ city\ i\ during\ period\ t}$$

$$\widetilde{termpop}_{t,i} = \frac{\widetilde{termshare}_{t,i}}{\max(\widetilde{termshare}_{t,i})} \text{X } 100$$

There are numerous problems associated with using the relative popularity values returned by Google Tends for our search term. First and foremost, $\widetilde{termpop}_{t,i}$ is dependent on the $\widetilde{termshare}_{t,i}$ values exhibited by other cities in the area, as Google Trends scales the normalized frequencies of all cities in an area through its establishment of an upper bound and changes in our dependent variable over time become dependent on changes observed in other cities, thereby making it infeasible to analyze time series data longitudinally. Previous authors have suggested using a benchmark term to deal with this effect and enable comparison of the numbers provided by Google Trends across different time periods, as the entry of a 2nd or 3rd search term inverts the analysis so that the number returned by Google Trends is no longer scaled relative to other geographic locations but rather the other search term. In accordance with previous literature, we used "Google Website" as our benchmark term and evaluate our proxy's search frequency level as a multiple of the benchmark "Google Website" at the city level. This is called the ratio for city i during period t, denoted:

$$combine_{t,i} = \#\ of\ searches\ for\ (s\ and\ b)\ in\ city\ i\ during\ period\ t$$

$$search\%_{t,i} = \frac{\#\ of\ searches\ for\ search\ term\ s\ in\ city\ i\ during\ period\ t}{combine_{t,i}}$$

$$bench\%_{t,i} = \frac{\#\ of\ searches\ for\ benchmark\ b\ in\ city\ i\ during\ period\ t}{combine_{t,i}}$$

$$ratio_{t,i} = \frac{search_{t,i}}{bench_{t,i}}$$

While previous authors stopped here, our paper differs in two significant ways.

1) Many of the other studies using Google Trends only analyzed data across a single geographic unit. A few compared search results across different geographies, but none attempted to both compare data across cities and normalize that data on a temporal scale. This is significant because the city is the only geographic unit for which Google Trends does not provide "interest over time" data.
2) Previous authors that used a benchmark term did so in order to compare search volumes for *different* terms across time periods. Our analysis only involves one search term which can be leveraged in an unconventional way to complete the normalization process across time. In doing so, we hope to provide a basic model for future papers attempting to study statistics across time and between cities for a single encompassing search term.

We input our benchmark term into one of the three remaining comparison boxes without clicking on the suggested entry from the drop down box or removing the existing entries used in equation (1) so as to not resample the data we've already collected. We edit the filter of the added benchmark term within its individual box, assigning it the subsequent time periods, t + 1, but keeping the same benchmark term used previously within time period t.

$$bench\%_{it} = \frac{benchfreq_{it}}{combine_{it} + benchfreq_{i(t+1)}}$$

We can use this relationship to solve for the increase in search for the benchmark term between periods and adjust for the change in our dependent variable accordingly.

3.2 Independent Variables

3.2.1 Internet Speed

National Broadband Map and FCC Form 477 Fixed Broadband Deployment semiannual data were initially considered for analysis. On these forms, ISPs are required to report maximum advertised upload and downloads speeds are for all census blocks where they provide at least one connection to end-user premises. Fitting block-level data into an empirical model would necessitate the haphazard summation of thousands of data points, even though the maximized advertised download speeds appearing in Form 477 art not informed by penetration rates, the quantity of services, or the general distribution of reported speeds across census blocks. Moreover, the biannual data from FCC Form 477 Fixed Broadband Deployment collected across December 2014 to December 2016 provide speeds in disaggregated values while the biannual data from the National Broadband Map collected across June 2010 to June 2014 provide speeds in speed bands. The risks of overgeneralization and inconsistency within our key variable were traded in for real-world speed data.

This paper uses separate monthly data on broadband upload and download speeds initially collected by Ookla, the global leader in fixed broadband and mobile network testing applications. Although Ookla sells its data and insights to enterprises, we managed to gain access through readily available data from "The Evolution of U.S. Spectrum Values Over Time", which previously downloaded the dataset and use broadband speeds as a proxy for existing communications infrastructure across markets (Connolly, Zaman, Roark, & Trivedi's, 2018). Ookla's broadband performance tests are initiated by users looking to assess the speeds of their individual connections on any device in a given time and place, meaning the aggregated data should reflect real-world internet speeds. This is optimal as it removes any potential gap that exists between advertised speeds and actual performance. The broadband speed data include multiple observations over time for most locations starting in 2008 and ending in 2015. We restrict our matches to 2009 through 2014 as we would be unable to complete the necessary calculations for our speed variable in 2008 and 2015 without data for 2007 and 2016.

Missing cities were omitted from our analysis, although cities and time periods with missing monthly data were not unless 0 observations were recorded for the city and 6-month period in question. Monthly speed observations (month m, observation x) were matched by city, i, and averaged across time unit, t, to form our independent variable for speed, denoted below:

$$Speednom_{it} = \sum_{m=1}^{6} x_i, \quad Speednom_{it} = \sum_{m=7}^{12} x_i \ \ldots$$

$$speed = \ln\left(speednom_{it}\right)$$

3.2.2 Rurality Index

Our six-level categorization of cities along a continuum of urbanization level uses the Urban-Rural Classification Scheme first developed by the NCHS in 2001 (Ingram and Franco, 2014). The NCHS urban-rural scheme was based on the 1990 Office of Management and Budget standards for defining metropolitan statistical areas using 1990 census data. It was updated in 2006 in order to account for micropolitan statistical areas delineations and then again in 2010 to reflect 2010 census data. Within this scheme, counties located in the largest metropolitan areas (MSA's or Metropolitan Statistical Areas with population of 1 million are more) are further disaggregated into two groups, the large central metro, which includes counties that contain all or part of the area's principal city, and the large fringe metro, which includes the surrounding counties. Countries contained in an MSA of a population of 250,000 to 999,999 are considered a medium metro, while those contained in an MSA of a population of less than 250,000 are considered small metro. Areas with populations of less than 250,000 are divided amongst two nonmetropolitan categories. First, there are counties contained in micropolitan statistical areas, labeled micropolitan, which consist of one or more urban clusters or smaller urban areas with 2,500 to 49,9999 inhabitants. Then, there are counties which did not quality as metropolitan or micropolitan, labeled noncore.

While the classification scheme makes distinctions based on population statistics, the categorization of a county is not at all based on the population of that county, but the population of the larger region that it is contained with – metropolitan or micropolitan. Because we are taking the second order of this analysis by additionally associating the city with the county it is contained within, this is extremely relevant to our paper for two reasons. First, noncore counties and noncore cities are not only not contained within urban areas and urban clusters, but they are strictly defined as not being within the general proximity of such areas, meaning they are essentially "off the map" and most likely rural. Second, the large central metro includes the inner city, meaning that network congestion in these areas may create a divide between advertised speed and real-world speed should be captured by our Ookla speed data. However, while slower speeds might be captured by speed test data, search engine activity is not nearly as affected by congestion as is high-quality video streaming. This points to the larger problem of bias being introduced when peak bandwidth utilization induces buffering for bandwidth-intensive applications in an area such as video streaming, but the affect on functionality is not captured by Google Trends data which reflects the intent to use over a 6-month period, but does not capture the how intent to use at a certain hour in the day accumulates over a 6-month period.

When deciding which U.S. states (or sub-regions) to include in our analysis, we had to account for the fact that Google Trends systematically excludes data corresponding to low search volumes. This technical adjustment ensures that urban cities with large populations and thus high absolute search volumes will almost always show up as an extractable data point, regardless of the level of geographic granularity specified by the filter. Search results from rural areas with low absolute volumes, on the other hand, have a much lower probability of being returned by Google Trends, even if the search term in question is relatively popular in that area. Because our analysis depends on the inclusion of data points along the entire continuum of the rurality index, states were selected on the basis of the likelihood that the underlying distribution of cities within the state and the population densities of those cities would, in turn, increase the probability that Google Trends returned non-core and micropolitan data points for that state. In order to do this, the five U.S. states with the lowest standard deviation of population densities at the city-level were selected from an aggregated list of all U.S. cities, and three of those states – Alabama, Arkansas, and Tennessee – were used in our analysis.

The rationale behind this sorting methodology is as follows. Google Trends selection method seems to be used in order to account for the statistical problems imposed by unequal sample sizes. Equal-sized groups maximize statistical power, meaning the "relative popularity" statistic for a city in a state is more likely to be returned by Google Trends if it is able to compare and thus scale the normalized frequency number associated with that city against other cities with approximately equal-sized sets of search query data to draw from. A low standard deviation of population density amongst all cities contained within a state suggests that the underlying population of that state is not first overwhelmingly concentrated in several metropolitan areas, and then dispersed amongst numerous rural areas. Instead, a low standard deviation implies that the underlying population of people and accordingly search volumes are equally distributed, therefore enabling comparison between numerous regions within that state, even if they are of smaller size.

3.2.3 Demographics of Geographical Regions

This paper first aims to determine what other factors, aside from broadband speeds, might also influence consumer behavior with regard to video streaming that we can feasibly include within our analysis. The primary resource we used to find relevant variables pertaining to American cities across different time periods was the United States Census Bureau. Through the Bureau's American Community Survey (ACS) and Populations Estimates Program, we were able to acquired annual data regarding population, income and age at the city-level. For the purpose of alignment with the dependent variable, data for these variables were converted into 6-month time periods by taking the average between each pair of consecutive years from 2009 to 2015, and assigning each average as the value for the second time period in the earlier of the two years.

A rise in video streaming activity might be due, in part, to an increase in the overall population, but because Google Trends is expressed as a percentage of all search queries, fluctuations in population should already be accounted for. We decided to still include population within the regression in case it plays an unexpected role in influencing video search i.e. a city experiences a drastic increase in population in which overcrowding leads to network congestion. Despite the inclusion of population, we don't expect this variable to have any significant effect. Contrarily, median household income can easily change consumer behavior with regard to any product. Although YouTube offers free access, the other four streaming platforms included in our search query require a monthly subscription. Therefore, a rise in median household income would make the video streaming services affordable to a larger percentage of the city population. This would theoretically lead to subscriber growth among these services, thus increasing the relative frequency with which these platforms are searched for.

As indicated by a YouGov survey of 2000 subscribers, Netflix users tend to be young, with just under 50% under the age of 35.[6] Conversely, a report from Moffett Nathanson Research found that 45 to 54 year-olds are the fastest growing demographic of users on Netflix. Additionally, YouTube's most recent release of in-house statistics indicates the the 35+ and 55+ age groups are the fastest growing YouTube demographics.[7] For these reasons, we would expect younger cities to have a higher percentage of video search interest, but would also expect older cities to have a greater rate of change. To more easily differentiate between young cities and old cities, we categorized age into four dummy variables: *youngest*, *young*, *old*, and *oldest*. Parameters for the buckets were determined by calculating the first through third quartile of all median age observations. The variable ranges are the following:

age1 → age < 33.4
age2 → 33.4 ≤ age < 36.5
age3 → 36.5 ≤ age < 39.6
age4 → age ≥ 39.6

---

[6] http://www.businessofapps.com/data/netflix-statistics/
[7] https://www.omnicoreagency.com/youtube-statistics/

For the aforementioned reasons, and given that other studies testing for changes in consumer behavior at the city-level also test across rurality, population, income and age, these were the secondary explanatory variables selected for our regression. We also sought out data pertaining to the rising interest in streaming services and planned to benchmark it against population growth to omit bias from rising popularity of streaming services in general, but no publicized data capable of acting as a reasonable proxy exists for this purpose.

IV. Empirical Methodology

The following regressions used in our analysis use panel data to endow both spatial and temporal dimensions. Our panel's spatial dimension pertains to the cross-sectional unit that is the city level, while the temporal dimension is the 6-month time period (January 2009 – December 2014).

The first concept tested was the possibility that one of our explanatory variables plays a role in contributing to the digital divide. This was done through regressing level of rurality, population and income all on speed under a random-effect regression. The level of rurality is determined by population density and proximity to larger cities. The results from this regression can be seen in Table 1, and summary statistics regarding the data used to build the regression can be found in Table 5.

$$Speed_{it} = \beta_0 + \beta_1 \cdot noncore_i + \beta_2 \cdot micropolitan_i + \beta_3 \cdot small\_metro_i + \beta_4 \cdot medium\_metro_i + \beta_5 \cdot large\_fringe_i + \beta_6 \cdot large\_central_i + \beta_7 \cdot population_{it} + \beta_8 \cdot income_{it}$$

The second concept tested was the effect of changes in speed on changes in google search frequencies. This was done through another random-effect regression using our panel data. The logarithmic transformation of search and speed were used to test for the percentage change of each variable. The dependent variable was measured as a lead variable in order to capture the effect of the explanatory variables on the search frequencies for the following time period. The dependent variable was also regressed with a lead of two time periods and with no lead at all to determine when the captured effect is greatest. To test how speed affects search frequency differently in cities with differing ages or opposing levels of rurality, interaction variables were constructed. Because dummy variables were used to categorize levels of age and rurality, the *oldest* age group and the *large fringe metro* county-type were omitted under the regression due to multicollinearity. The equation below represents the dependent variable with a lead of one period. The results for this regression are shown in Table 2. The results from the final model containing every explanatory variable regressed with no lead and a 1-year lead can be found in Table 4. The summary statistics for the data used to build the regression can be found in Table 6.

$$
\begin{aligned}
\ln(search)_{i(t+1)} \\
&= \beta_0 + \beta_1 \cdot \ln(speed_{it}) + \beta_2 \cdot income_{it} + \beta_3 \cdot population_{it} + \beta_4 \\
&\cdot youngest_{it} + \beta_5 \cdot young_{it} + \beta_6 \cdot old_{it} + \beta_7 \cdot oldest_{it} \\
&+ \beta_8(\ln(speed_{it}) \cdot youngest_{it}) + \beta_9(\ln(speed_{it}) \cdot young_{it}) \\
&+ \beta_{10}(\ln(speed_{it}) \cdot old_{it}) + \beta_{11}(\ln(speed_{it}) \cdot oldest_{it}) + \beta_{12} \cdot noncore_i \\
&+ \beta_{13} \cdot micropolitan_i + \beta_{14} \cdot small\_metro_i + \beta_{15} \cdot medium\_metro_i + \beta_{16} \\
&\cdot large\_fringe_i + \beta_{17} \cdot large\_central_i + \beta_{18}(\ln(speed_{it}) \cdot noncore_i) \\
&+ \beta_{19}(\ln(speed_{it}) \cdot micropolitan_i) + \beta_{20}(\ln(speed_{it}) \cdot small\_metro_i) \\
&+ \beta_{21}(\ln(speed_{it}) \cdot medium\_metro_i) + \beta_{22}(\ln(speed_{it}) \cdot large\_fringe_i) \\
&+ \beta_{23}(\cdot \ln(speed_{it}) \cdot large\_central_i)
\end{aligned}
$$

V. Results

Table 1: Regression results for speed equations

	MODEL 1	MODEL 2	MODEL 3
	b/se	b/se	b/se
NONCORE	-1.8432303	-1.8566073	-2.1003604
	(1.56941)	(1.57287)	(1.62829)
SMALL_METRO	1.3392794	1.493975	0.2350096
	(1.77031)	(1.77674)	(1.85854)
MEDIUM_METRO	1.3183249	1.625793	-1.1816072
	(1.55836)	(1.59612)	(1.75126)
LARGE_FRINGE_METRO	2.2781747	2.3767946	-2.2445804
	(1.60473)	(1.60946)	(1.91620)
LARGE_CENTRAL_METRO	-2.3500833	-2.3074604	-5.1543760*
	(2.45479)	(2.46060)	(2.61624)
MICROPOLITAN	0	0	0
	(omitted)	(omitted)	(omitted)
POPULATION		-0.0000095	-0.0000094
		(0.00001)	(0.00001)
INCOME			0.0001680***
			(0.00003)
CONSTANT	14.6932125***	14.8396662***	9.3637303***
	(1.12141)	(1.12738)	(1.62391)
OBSERVATIONS	1283	1278	1278
NUMBER OF LOCATIONS	195	194	194
WITHIN	0	0.0025	0.0697
BETWEEN	0.0443	0.0434	0.0442
OVERALL	0.0338	0.0373	0.0389
STANDARD ERRORS IN PARENTHESES			
*** P<0.05, ** P<0.01, *** P<0.001**			

Table 1 portrays the results for the random-effect regression testing the effects of rurality, population and income on speed. Model 3 represents the effect of all three explanatory variables under the same regression. With an overall adjusted R-squared value of 3.89%, the three independent variables chosen explain very little of the variance in speed. However, our results are consistent with the notion that qualities of the city itself do not play a role in determining the broadband speed. Rather than speed being determined by variables describing the city and its consumers, it now seems more likely that the speed is more greatly influenced by the supply-side, with variables pertaining to Internet Service Providers potentially explaining the variance that is unaccounted for. It is also possible that the level of rurality would have a greater statistical significance had the classifications been made at the city-level rather than the county-level.

Despite the regression's low adjusted R-squared value, two variables in the model produced statistically significant relationships. Because the income variable is positive and significant, our results are consistent with the notion that a $10,000 increase in a city's median household income would increase broadband speeds by 1.68 Mbps holding all else constant. The *large central metro* dummy variable for rurality indicates that a city within a large central metro is, on average, 5.15 Mbps slower than a city within a micropolitan (the omitted dummy variable). Given that *large central metro* is our most urban classification, the resulting coefficient is counterintuitive to our initial hypothesis that increased rurality would be correlated with higher broadband speeds. To understand the reason behind this, the average broadband speed for each level of rurality was calculated in Table 3. Of the six levels of rurality, the *large central metro* variable has the lowest average broadband speed. The variable also has significantly less instances than the other five classifications, which could be the reason behind the outlier. In addition to average speed and count, the standard deviation for speed, and the average and standard deviation for income, were calculated for each category. Given that urban counties often have a large percentage of impoverished regions, and given that the positive relationship between income and speed is statistically significant, the relatively high standard deviation of income for *large central metro* might help to explain the surprisingly low average broadband speed.

Table 2: Regression results for %Δsearch equations

	MODEL 1	MODEL 2	MODEL 3	MODEL 4
	b/se	b/se	b/se	b/se
LN(SPEED)	0.223614***	0.215398***	0.150298***	0.204274**
	(0.01736)	(0.01751)	(0.03190)	(0.06100)
INCOME		0.0000016*	0.0000017*	0.0000019*
		-	-	-
POPULATION		0.0000001	0	0
		0	-	-
AGE1			0.2831167	0.2383036
			(0.11667)	(0.11988)
AGE2			0.2183138	0.2278402
			(0.11946)	(0.12329)
AGE3			0.2069179	0.1447019
			(0.11399)	(0.11605)
AGE4			0	0
			(omitted)	(omitted)
LN(SPEED)*AGE1			-0.1135133*	-0.0987967*
			(0.04561)	(0.04654)
LN(SPEED)*AGE2			-0.0981681*	-0.1021049*
			(0.04628)	(0.04785)
LN(SPEED)*AGE3			-0.071724	-0.0505999
			(0.04466)	(0.04543)
LN(SPEED)*AGE4			0	0
			(omitted)	(omitted)
NONCORE				-0.4017110*
				(0.16775)
MICROPOLITAN				-0.0407078*
				(0.17898)
SMALL METRO				0.1316558
				(0.19468)
MEDIUM METRO				-0.2019069
				(0.16135)
LARGE CENTRAL METRO				0.1138049
				(0.26639)
LARGE FRINGE METRO				0
				(omitted)

LN(SPEED)*NONCORE				0.1427503***
				(0.06318)
LN(SPEED)*MICROPOLITAN				0.0106451*
				(0.06612)
LN(SPEED)*SMALL_METRO				-0.0563146
				(0.07146)
LN(SPEED)*MEDIUM_METRO				0.0619124
				(0.05939)
LN(SPEED)*LARGE_FRINGE				0
				(omitted)
LN(SPEED)*LARGE_CENTRAL				-0.0435355
				(0.10755)
CONSTANT	0.4647808***	0.5274688***	0.4969687***	0.5713407***
	(0.04518)	(0.05203)	(0.08462)	(0.17114)
OBSERVATIONS	1055	1055	1055	1055
NUMBER OF LOCATIONS	182	182	182	182
R-SQUARED WITHIN	0.2484	0.2464	0.2516	0.2834
R-SQUARED BETWEEN	0.0015	0.0083	0.0099	0.0534
R-SQUARED OVERALL	0.0882	0.1033	0.1076	0.1469
STANDARD ERRORS IN PARENTHESES				
*** P<0.05, ** P<0.01, *** P<0.001**				

The second regression attempts to model the factors that influence changing interests in video streaming services across cities and time. It was regressed three times with different lags: one with the dependent variable having a lead of one year (two time periods), a lead of six months (a single time period), and no lead at all. The six-month lead was used for its superior R-squared value. The regression was found to have an overall adjusted R-squared value of 14.69%, rendering the fitted regression line again unreliable. Although the equation is insufficient in accurately predicting the change in search frequencies, the results suggest that some explanatory variables contribute to the changing consumer behavior measured by the search frequency proxy. First, the percentage change in speed has a positive and statistically significant relationship with the percentage change in search frequencies. Interpreting the coefficient for this effect, a 1% increase in the broadband speed would lead to a 20.43% increase in relative video search popularity holding all else constant.

The second regression also suggests that for every $10,000 increase in a city's median household income, relative search popularity will increase 1.9%. Considering four out of five of the video streaming services included in our search query require a monthly subscription, it is logical that an increase in the affordability of these video services would correlate with more interest. The age and speed interaction variables containing the bottom two quartiles of age also produced statistically significant p-values. The coefficients are relative to the *oldest* dummy variable representing the fourth quartile, since this variable was omitted under the regression. The results imply that a 1% increase in speed for a city with a median age under 36.5 is correlated with an approximate 10% decrease in search frequency relative to the oldest age category, suggesting that an increase in broadband speeds is more likely to galvanize video streaming interest from older people than younger people.

The other variables that produced an effect on relative search interest are the two most rural classifications, noncore and micropolitan. The two variables are statistically significant in both their isolated and interaction terms. The relationship terms are somewhat counterintuitive at first, as the isolated dummy determines a negative relation whereas the interaction term is positive. Focusing just on noncore, the results indicate that a noncore city is correlated with experiencing a 40% drop in search popularity when compared to large fringe metros, the omitted dummy. We did not expect the isolated dummy to have any effect on change in search, however, this coefficient might be large due to the fact that cities in noncore counties contain less population density, thus are more prone to small sample bias when obtaining data for search popularity. That being said, this bias is less likely affecting the interaction term considering its two additional levels of statistical significance. The results for this variable suggest that a 1% increase in speed that takes place in a noncore region is correlated with a 14.3% increase in search interest when compared to changes in speeds within large fringe metros, the omitted dummy variable. The same relationship can be drawn for cities in micropolitan counties, but with a 1.1% increase in search. This implies that for an equivalent percentage increase in broadband speed, individuals living in rural regions are significantly more likely to increase engagement with video streaming services than those in urban regions.

In an effort to further explore the relative effect of this interaction variable, Table 7 was created to illustrate the average absolute value of percentage change in speed and search for all panel data instances for each dummy variable. The noncore dummy clearly has the largest ratio of average percent change in speed to average percent change in search, and the interaction variable is relative to the *large fringe metro* dummy, which clearly has the lowest ratio of average percent change in speed to average percent change in search. This could potentially imply that the interaction term composed of *ln(speed)* and *noncore* is underestimating the true effect of speed on search in these rural areas.

VI. Discussion

Given our results are consistent with the notion that broadband speed improvements have a greater positive effect on video streaming activity in rural areas than in urban areas, it can be said that the disadvantages generated by the digital divide extend to social consequences, with specific regard to the entertainment culture created of OTT platforms and their original content. Currently, through a fund known as the Lifeline program, broadband subsidies are given to low-income households in an effort aimed at bridging the digital divide. However, at least with respect to social consequences, our results infer that level of rurality, rather than income, would be a more effective determinant as to which cities are elected to receive the broadband subsidy. Future studies should build on our analysis by testing whether rurality is more effective than income in measuring the economic, political, and health effects of bandwidth impairment. In the event that our results hold true across all bandwidth impairment detriments, the FCC should rethink current criteria with which Lifeline program cities are chosen.

Another disadvantage of the digital divide we hoped to explore was the delocalization of content. Uploading content is upwards of 8 times more bandwidth-intensive than downloading content, thus inadequate broadband capacity can result in large consumptions of non-local content, and the inhibition of local businesses keeping up with the digital age. The inability to efficiently allocate bandwidth to "uploaders" not only poses an existential threat to communities on the periphery unable to document and share local content on the global stage, but also damages institutions who cannot respond to the erosion of their spatial-socio economic ties with local residents by digitizing their services. We hoped to use domain creation as a proxy for local content production, but Google Trends queries relating to this were to scarce at the city-level to produce enough cities for a reliable panel regression. Future studies should search for an effective proxy for localization of content. If found, this proxy could be used to determine whether providing improved broadband speeds within these rural regions would actually provide the economic benefit for "uploaders" that we would expect; by allowing them to bring their businesses to the digital world.

VII. Appendix

Table 3: Summary statistics for levels of rurality

RURALITY	AVG SPEED	COUNT	STD OF SPEED	AVG INCOME	STD OF INCOME
LARGE CENTRAL METRO	11.83	67	5.486	47274	17381.1
LARGE FRINGE METRO	17.03	272	6.59	57441	25449.5
MEDIUM METRO	15.21	374	10.25	47300	13522.5
SMALL METRO	15.55	197	6.41	37741	9257.7
MICROPOLITAN	14.15	297	6.91	32484	5136.1
NONCORE	12.36	247	8.35	30861	7185.4

Table 4: %Δsearch regression with no lead and 1-year lead

	MODEL 4 W/ NO LAG	MODEL 4 W/ 1-YEAR LAG
	b/se	b/se
LN(SPEED)	0.1971362**	0.1635177*
	(0.05385)	(0.07130)
INCOME	0.0000016	0.0000021*
	-	-
POPULATION	-0.0000001	0.0000003
	-	-
AGE1	0.2006967	0.1690087
	(0.11648)	(0.13094)
AGE2	0.1294707	0.0099542
	(0.11691)	(0.13893)
AGE3	0.1641368	0.0499267
	(0.10765)	(0.13549)
AGE4	0	0
	(omitted)	(omitted)
LN(SPEED)*AGE1	-0.0877387*	-0.0691422
	(0.04382)	(0.05174)
LN(SPEED)*AGE2	-0.0683972	-0.0097971
	(0.04422)	(0.05449)
LN(SPEED)*AGE3	-0.0697902	-0.0223893

	(0.04131)	(0.05386)
LN(SPEED)*AGE4	0	0
	(omitted)	(omitted)
NONCORE	-0.4170609*	-0.2759019*
	(0.15474)	(0.18785)
MICROPOLITAN	-0.1059411*	0.0430504*
	(0.16419)	(0.20208)
SMALL METRO	-0.2311968	0.3873722
	(0.17690)	(0.22467)
MEDIUM METRO	-0.2847729	-0.1203566
	(0.14748)	(0.18332)
LARGE CENTRAL METRO	0.1248555	0.0866551
	(0.24952)	(0.27819)
LARGE FRINGE METRO	0	0
	(omitted)	(omitted)
LN(SPEED)*NONCORE	0.1453670**	0.0924239*
	(0.05575)	(0.07236)
LN(SPEED)*MICROPOLITAN	0.0439891	-0.0318818
	(0.05806)	(0.07639)
LN(SPEED)*SMALL_METRO	0.0836668	-0.1328285
	(0.06198)	(0.08543)
LN(SPEED)*MEDIUM_METRO	0.1025389*	0.0405256
	(0.05182)	(0.06972)
LN(SPEED)*LARGE_CENTRAL	-0.0465284	-0.02239
	(0.09750)	(0.11288)
LN(SPEED)*LARGE_FRINGE	0	0
	(omitted)	(omitted)
CONSTANT	0.7712728***	0.4993851***
	(0.15974)	(0.19479)
OBSERVATIONS	1231	869
NUMBER OF LOCATIONS	186	172
R-SQUARED WITHIN	0.2493	0.1764
R-SQUARED BETWEEN	0.012	0.031
R-SQUARED OVERALL	0.1224	0.1145
STANDARD ERRORS IN PARENTHESES		
*** P<0.05, ** P<0.01, *** P<0.001**		

Table 5: Summary statistics for regression on speed

VARIABLE	OBS	MEAN	STD. DEV.	MIN	MAX
SPEED	1,283	15.09853	8.24032	1.488633	98.28233
NONCORE	1,283	0.1683554	0.3743275	0	1
SMALL METRO	1,283	0.1410756	0.3482353	0	1
MEDIUM METRO	1,283	0.2416212	0.4282328	0	1
LARGE FRINGE	1,283	0.1886204	0.3913593	0	1
LARGE CENTRAL	1,283	0.0514419	0.2209835	0	1
MICROPOLITAN	1,283	0.2088854	0.4066709	0	1
POPULATION	1,278	36599.46	60798.2	63	648051
INCOME	1,282	43668.68	17475.41	21076.5	140114

Table 6: Summary statistics for regression on %Δsearch

VARIABLE	OBS	MEAN	STD. DEV.	MIN	MAX
LN(SPEED)$_{T+1}$	1,055	0.3730709	0.2801199	-0.447312	1.265666
LN(SPEED)	1,283	2.564845	0.5782439	0.3978585	4.587844
INCOME	1,282	43668.68	17475.41	21076.5	140114
POPULATION	1,278	36599.46	60798.2	63	648051
AGE1	1,283	0.2478566	0.4319365	0	1
AGE2	1,283	0.2322681	0.4224438	0	1
AGE3	1,283	0.2579891	0.4376986	0	1
AGE4	1,283	0.2618862	0.4398325	0	1
LN(SPEED)*AGE1	1,283	0.612331	1.106761	0	3.591683
LN(SPEED)*AGE2	1,283	0.6012565	1.127219	0	3.67159
LN(SPEED)*AGE3	1,283	0.6709053	1.171066	0	4.08182
LN(SPEED)*AGE4	1,283	0.6803525	1.183321	0	4.587844
NONCORE	1,283	0.1683554	0.3743275	0	1
MICROPOLITAN	1,283	0.2088854	0.4066709	0	1
SMALL METRO	1,283	0.1410756	0.3482353	0	1
MEDIUM METRO	1,283	0.2416212	0.4282328	0	1
LARGE CENTRAL	1,283	0.0514419	0.2209835	0	1
LARGE FRINGE	1,283	0.1886204	0.3913593	0	1
LN(SPEED)*NONCORE	1,283	0.3943354	0.926804	0	3.733808
LN(SPEED)*MICROP.	1,283	0.5288375	1.057333	0	3.635032
LN(SPEED)*SMALL.	1,283	0.3752757	0.9416366	0	3.446545
LN(SPEED)*MEDIUM.	1,283	0.6214353	1.141512	0	4.587844
LN(SPEED)*LARGE_C.	1,283	0.1210786	0.5333759	0	3.293734

LN(SPEED)*LARGE_F.	1,283	0.5238827	1.102052	0	3.67159

Table 7: Variations in absolute values of search and speed for levels of rurality

	AVG |% CHANGE IN SEARCH|	**AVG |% CHANGE IN SPEED|**
LARGE CENTRAL METRO	11.06%	14.83%
LARGE FRINGE METRO	19.95%	14.12%
MEDIUM METRO	12.08%	18.69%
SMALL METRO	13.08%	17.35%
MICROPOLITAN	17.17%	19.42%
NONCORE	14.98%	26.26%

References

Chiang, K. (2006). Clicking instead of walking: Consumers searching for information in the electronic marketplace. *Bulletin of the American Society for Information Science and Technology, 32*(2), 9-11. doi:10.1002/bult.1720320205

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record, 88*, 2-9. doi:10.1111/j.1475-4932.2012.00809.x

Connolly, M. P., Ss, N., Zaman, A., Roark, C., & Trivedi, A. (2018). The Evolution of U.S. Spectrum Values Over Time. *SSRN Electronic Journal*. doi:10.2139/ssrn.2928332

Crandall, R. W., & Jackson, C. L. (2001). *The $500 Billion Opportunity: The Potential Economic Benefit of Widespread Diffusion of Broadband Internet Access* (Rep.). Criterion Economics, L.L.C.

Data Access and Dissemination Systems (DADS). (2010, October 05). American FactFinder. Retrieved April 8, 2019, from https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

Governo, F., Teixeira, A. A., & Brochado, A. M. (2017). Social Multimedia Computing: An Emerging Area of Research and Business for Films. *Journal of Creative Communications, 12*(1), 31-47. doi:10.1177/0973258616688967

Greenstein, S., & Mcdevitt, R. C. (2011). The broadband bonus: Estimating broadband Internet's economic value. *Telecommunications Policy, 35*(7), 617-632. doi:10.1016/j.telpol.2011.05.001

Ingram, D. D., & Franco, S. J. (2014, April). *NCHS Urban-Rural Classification Scheme for Counties*. Retrieved April 8, 2019, from U.S. Department of Health and Human Services.

Kollat, D. T., Engel, J. F., & Blackwell, R. D. (1970). Current Problems in Consumer Behavior Research. *Journal of Marketing Research, 7*(3), 327. doi:10.2307/3150290

Miner, L. (2015). The unintended consequences of internet diffusion: Evidence from Malaysia. *Journal of Public Economics, 132*, 66-78. doi:10.1016/j.jpubeco.2015.10.002

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The Structure of "Unstructured" Decision Processes. *Administrative Science Quarterly, 21*(2), 246. doi:10.2307/2392045

Narita, F., & Yin, R. (2018). In Search of Information: Use of Google Trends' Data to Narrow Information Gaps for Low-Income Developing Countries. *IMF Working Papers, 18*(286), 1. doi:10.5089/9781484390177.001

Punj, G. (2012). Consumer Decision Making on the Web: A Theoretical Analysis and Research Guidelines. *Psychology & Marketing, 29*(10), 791-803. doi:10.1002/mar.20564

Rains, S. A. (2008). Health at High Speed. *Communication Research, 35*(3), 283-297. doi:10.1177/0093650208315958

Sengupta, S. S., & Nicosia, F. M. (1969). Consumer Decision Processes: Marketing and Advertising Implications. *Econometrica, 37*(1), 166. doi:10.2307/1909226

Sheth, J. N., Newman, B. I., & Gross, B. L. (1991). Why we buy what we buy: A theory of consumption values. *Journal of Business Research, 22*(2), 159-170. doi:10.1016/0148-2963(91)90050-8

Stankevich, A. (2017). Explaining the Consumer Decision-Making Process: Critical Literature Review. *Journal Of International Business Research And Marketing, 2*(6), 7-14. doi:10.18775/jibrm.1849-8558.2015.26.3001