

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems

School of Information Systems

11-2016

Content sampling, household informedness, and the consumption of digital information goods

Ai Phuong HOANG

Singapore Management University, aphoang.2013@phdis.smu.edu.sg

Robert J. KAUFFMAN

Singapore Management University, rkauffman@smu.edu.sg

DOI: <https://doi.org/10.1080/07421222.2018.1451958>

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research



Part of the [Databases and Information Systems Commons](#)

Citation

HOANG, Ai Phuong and KAUFFMAN, Robert J.. Content sampling, household informedness, and the consumption of digital information goods. (2016). *Journal of Management Information Systems*. 35, (2), 575-609. Research Collection School Of Information Systems.

Available at: https://ink.library.smu.edu.sg/sis_research/4137

This Journal Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Published in Journal of Management Information Systems,
November 2016, Pages 575–609
DOI 10.1080/07421222.2018.1451958

Content Sampling, Household Informedness, and the Consumption of Digital Information Goods

AI-PHUONG HOANG AND ROBERT J. KAUFFMAN

AI-PHUONG HOANG (aphoang.2013@phdis.smu.edu.sg; corresponding author) is a Ph.D. candidate in the Interdisciplinary Doctoral Programme in Information Systems and Marketing at Singapore Management University. Her research interests involve consumer behavior, social network and marketing strategy for digital information goods, and methodology innovations for causal inference with consumers' digital trace data. Her research work appears in *Electronic Commerce Research and Applications* and in several leading IS conferences and workshops. From 2015 to 2016, she was a visiting Ph.D. student at the Heinz College of Information and Public Policy, Carnegie Mellon University.

ROBERT J. KAUFFMAN (rkauffman@smu.edu.sg) is a Professor of Information Systems at the School of Information Systems, Singapore Management University. He also serves as Associate Dean (Faculty). He was a Distinguished Visiting Fellow at the Center for Digital Strategies at the Tuck School of Business at Dartmouth. He has served on the faculty of the business schools of New York University, University of Rochester, University of Minnesota, and Arizona State University. His graduate degrees are from Cornell and Carnegie Mellon. His interdisciplinary research spans strategy, information, information technology, economics, marketing and consumer behaviour, the fintech revolution, environmental sustainability, and computational social science and data analytics. He is a frequent keynote speaker on these subjects. His research has appeared in *Management Science*, *Information Systems Research*, *Journal of Management Information Systems*, *MIS Quarterly*, *Organization Science*, and the *Review of Economics and Statistics*, among others.

ABSTRACT: Technology and media are delivering content that is transforming society. Providers must compete for consumer attention to sell their digital information goods effectively. This is challenging, since there is a high level of uncertainty associated with the consumption of such goods. Service providers often use free programming to share product information. We examine the effectiveness of content sampling strategy used for on-demand series dramas, a unique class of entertainment goods. The data were extracted from a large set of household video-on-demand (VoD) viewing records and combined with external data sources. We extended a propensity score matching (PSM) approach to handle censored data, which permitted us to explore the main causal relationships. Relevant theories in the marketing and information systems disciplines informed our research on consumer involvement and informedness for decision making under uncertainty, the consumption of information goods, and seller strategies for digital content. The results show that content

sampling stimulates higher demand for series dramas, but in a more nuanced way than was expected. Samples of the series reveal quality information to consumers and allow them to assess preference fit directly. As a result, they become more informed about their purchase decisions. Also, households seem to be willing to pay more to be better informed, and informed households tend to purchase more. This suggests that content providers should invest in strategies that help consumers to understand the preference fit of information goods.

KEY WORDS AND PHRASES: content sampling, data censoring, digital entertainment, household informedness, information asymmetry, information goods, online content, preference fit, propensity score matching, video-on-demand.

Disruptive technologies, such as digital content-streaming platforms, have boosted the production and consumption of entertainment content. Economies of scale now allow digital entertainment service providers to market and sell information goods directly to consumers on an on-demand, anytime, anywhere basis. Among the different types of content that are offered on-demand, video-on-demand (VoD) services are a key source of revenue for digital entertainment firms [47]. At the industry level, a consulting firm [60] has estimated that the VoD market of US\$47.25 billion in 2015 will grow to almost US\$75 billion by 2020, representing a compound annual growth rate (CAGR) of 9.63 percent.

In the past decade, TV series have experienced a great upswing in consumer interest. Because of this surge in market demand, all of the TV studios, including industry incumbents such as ABC and CBS, and content distributors such as Netflix, Hulu, and Amazon, are competing in the race for the next “big show.” They have invested heavily in original shows despite a high failure rate in the production stage, since the rewards for a successful hit come in so many different forms: more viewers, higher ad revenue, and most important perhaps, a competitive edge in sustaining the customer base [62].¹ For example, the Hulu TV original series, *The Handmaid’s Tale*, recently won eight Emmy awards. This success for the content distributor signals a whole new era for original on-demand content [66], and a growing global market.

Despite a recent audience report from Nielsen that reveals that Americans spend almost eleven hours each day staring at the screen and consuming media [41], content providers are struggling to market and sell their programming due to the high level of consumer uncertainty associated with the consumption of this class of products. A TV program’s quality is known only after it has been watched, and imperfect information about its content typically decreases a consumer’s willingness to pay [13, 14]. In addition, entertainment products are horizontally differentiated; their value relies heavily on the subjective evaluation of consumers. With a large amount of content available, it is hard for consumers to choose what they are likely to enjoy, or what fits them best. Across different industries, various forms of

sampling strategies have been used to communicate product information for experience goods to consumers. Readers of the *New York Times*, for example, can access up to ten articles each month, representing a metered model in the newspaper industry [35]. In addition, software companies provide the most basic version of their software free of charge or an extended version for free during a trial period [65]. Online music distributors, such as Apple and Spotify, also make it possible for listeners to sample all of their songs—but for only 30 seconds each [67]. Production companies, meanwhile, have been making trailers and sneak peeks of shows they produce too. And firms also employ sampling strategies at the service level, such as Netflix’s one-month basic membership trial.

The wide implementation of sampling-based strategy for digital goods has much to do with the one-time fixed cost of content digitization and the associated cheap cost of distribution. The impact of such strategies is more profound though. The interdisciplinary literature on sampling strategies for information and experience goods has often focused on online music and software [12, 21]. Such studies have investigated the determinants of consumer decision making and examined the consumption of these household purchases. We extend this literature with empirical evidence for the impact of sampled content on purchases of on-demand series dramas, a unique class of entertainment products. In this context, consumers are able to evaluate fit related to their preferences through the sampling of a series.

The theories we use are drawn from different streams of literature. The first deals with the specific characteristics of experience goods that create a high level of uncertainty [79]. We look at the impact of sampling strategy for physical goods [29, 58], and the implications for experience goods. The second stream focuses on how sampling influences consumer buying behavior under uncertainty [36, 55, 59]. We examine issues related to consumer viewing behavior [57]. To our knowledge, this research is the first to provide empirical support for the effectiveness of sampling strategies related to the purchase of VoD series dramas, a niche product that consists of a video bundle with multiple episodes. Previously, Markopoulos [54] examined sampling and video game purchases with a smaller, less granular data set, as Clemons et al. [15] later did for music sampling purchases, but not in the depth that we have.

We address two questions: (1) What are the impacts of different forms of content samples on a household’s VoD series purchases? and (2) How do a household’s choices of standard content and customized, add-on content affect its VoD series purchases? We also discuss the role of data analytics in effective implementation of sampling-based strategies for the marketing of digital information goods.

To answer these research questions, we designed a study to learn about the aggregate behavior related to free sampling and series purchases with an emphasis on the household level as our unit of analysis. We addressed causality and potential threats to the robustness of our main findings with additional econometric procedures. We used a blend of data analytics methods to establish evidence for causality. Our analysis work benefited from access to millions of TV viewing records, including those involving VoD content, across hundreds of thousands of households,

and multiple sources of data on series dramas. The period of observation for VoD viewing records was limited though—just one month.

Without access to additional data or the ability to construct a set of formal field experiments within the operations of the sponsor of this research, we implemented an innovative approach using propensity score matching (PSM). It uses iterative replacement methods to pair data across censored and noncensored data groups based on discoverable sequences over time, and patterns of observable past activities by the subjects—households, in our case. This allowed us to make inferences related to unobservable viewing records outside the study period, which caused data censoring. The overall approach enabled us to make causal arguments about the impact of free samples, on the basis of our extensive data analysis. The findings contribute to theory and practice by highlighting the importance of an effective sampling-based strategy in marketing digital information goods, while offering new managerial knowledge about how to offer effective sampling to consumers.

Theoretical Background

We now turn to the relevant streams of literature: (1) product uncertainty associated with the consumption of digital information goods; (2) selling strategy for digital information goods; and (3) consumer viewing and purchase behavior for digital information goods.

Uncertainties Associated with the Consumption of Digital Information Goods

Product uncertainty is viewed as an important construct in marketing and information systems (IS) research, as it directly affects consumers' willingness to pay for goods and services [1, 72]. Hong and Pavlou [40] distinguished between uncertainty about product quality and uncertainty about product fit with a consumer's taste. The product may not be in the promised condition [69], or the vendors may fail to communicate product information to consumers [22, 31], hence uncertainty about quality is normal. Fit uncertainty refers to the degree to which consumers are unable to assess whether a product's attributes match their preferences [40]. Imperfect information concerning quality and fit creates higher perceived transaction costs and tends to diminish a consumer's willingness to pay [51].

In another stream of research, Nelson [63] separated experience goods from search goods: the quality of search goods can be determined simply by inspection before purchase, whereas the quality of experience goods is realized only after use. Thus, the assessment of digital information goods, such as music, books, or movies, must involve personal experience [43, 56]. In fact, the actual source of quality is the experience itself, in which product fit plays a critical role [46]. A study on the craft beer industry has shown that firms with highly differentiated products experience higher revenue growth when consumers become more informed [13]. They are often

willing to pay more when the match between product characteristics and their preferences is improved. Different types and levels of informedness can also influence consumer choices [49]; for instance, elimination of product fit uncertainty for a digital experience good can increase the number of purchases and consumer loyalty [56]. In platforms on which entertainment is marketed and sold at the product level, the effects of consumer informedness about products and their fit become more pronounced.

Sales Strategy for Digital Information Goods

As streaming media have become affordable, and demand for content has increased, firms have had to adjust their strategies to be more effective with the selling of digital information goods. Online reviews and word of mouth are good sources of information on digital goods for consumers. Moretti [61] showed that social learning and peer effects have positive impacts on the consumption of movies. Nevertheless, it is hard to describe the characteristics of an experience good, especially when consumer tastes vary significantly [56]. A TV program is better from a consumer's perspective if it fits his or her viewing preferences. Signaling quality and content is achievable, while communicating fit is more complicated.

Previous studies have focused on selling strategies for digital information goods, and the market context and environment in which they are offered. Bhattacharjee et al. [6] looked at online music sales in the presence of online piracy, and showed that effective pricing options, search tools, and licensing structures are leading strategies to mitigate the related revenue losses for the music labels and artists. The search process for digital information goods is different from that for physical goods. Each product is unique and has its own characteristics, so consumers need to repeat the search process for every purchase. As a result, the associated search cost will vary greatly and be proportional to the number of options available. As part of the transaction cost, search costs can influence consumer purchase decisions [8, 42].

Product sampling lowers the search cost by effectively communicating product quality to consumers. Thus, it is a key promotional tool to stimulate sales for many products [58]. A sample is a portion of a product given to consumers to try for free before making a purchase decision. Consumers like to receive free goods. Thus, free samples can influence their behavior at the point of purchase, encouraging unplanned purchases and active switching to promoted brands [36, 70]. For retailers of physical products, sampling yields a higher purchase conversion rate and return on investment than other direct advertising [28, 29]. Nevertheless, it has mainly been used to enhance the effectiveness of traditional marketing only; the implementation of a sampling strategy is expensive, and the market reach is limited [58].

Considerable attention has also been given to sampling strategies for information and experience goods. Information goods are characterized by large sunk costs for development, and negligible costs of reproduction and distribution [79]. Digital content can be digitally broadcast, streamed, and stored at a relatively low cost.

Niculescu and Wu [65] explored the economics of free under perpetual licensing for two software business models. With a feature-limited freemium, consumers gain free access to a basic version of the software but have to pay for premium versions, while under uniform seeding, firms offer a full product for free to part of the market. Halbheer [35] studied the profitability of ad-supported content sampling for newspapers. In the entertainment sector, offering teasers or previews for movies and TV shows has become an industry norm; yet the implications are overlooked in the literature.

The execution of sampling strategies for digital content is not that straightforward though. Firms need to consider how individuals value the same product differently, reflecting customer heterogeneity, to design an appropriate strategy. For software products, the rate of learning by users determines the effectiveness of time-locked trials [21]. Using data analytics though, firms can help buyers find their nearly “perfect” product fit. Netflix, for example, shows different trailers of the same series to different market segments, so it is able to figure out their viewing preferences [10]. It may take longer for some consumers to reach a decision; yet offering lengthy samples is not desirable for most providers [37]. Free content may interfere with the market’s consumption of programming, and free content on the Internet decreases consumer willingness to pay for content in other channels [5].

Consumer Viewing and Purchase Behavior for Digital Information Goods

Research on consumer behavior has examined different aspects of TV viewing activity. Rubin [76] looked at the interaction between viewing patterns and motivation and identified two viewer types: one watches TV out of habit to pass time; the second seeks information and watches TV to learn. Viewing activity is recognized as a gratification-seeking process, in which viewers search for and watch the content that matches their preferences [53]. Viewers may also modify their viewing preferences, a variety-seeking behavior [57]. Variety-seekers respond positively to new programs, and new means of delivery across different platforms, such as their desktops, tablets, and phones.

Recently though, researchers have begun to focus more on specific types of programming, TV shows and series dramas. This has been due to the emergence of advanced content-streaming technology. A survey conducted by Harris-Netflix has shown that most viewers admitted to binge-watching [64]; they get hooked and watch multiple episodes of a series in one sitting [39]. Theoretical perspectives from multiple disciplines are helpful to explain this behavior. For example, connectedness, the relationship between a viewer and the characters, intensifies as the viewer spends more time watching a show [78], and not having closure on how a story ends may cause dissatisfaction and regret [4, 32]. The most prominent consideration is instant gratification, the desire to fulfill a need without delay [2]. If the content triggers a viewer’s interest, he or she will feel the impulse to purchase the show. On-demand

services make it easier for consumers to have access to extensive TV content, which influences consumption.

Personal experience with the viewing content is necessary, similar to other experience goods. A majority of viewers may agree on certain attractive features of a show, but it is unlikely that all will enjoy watching it. A successful movie is not necessarily suitable for every member of its audience. Given a choice, consumers want to learn as much as possible about products by experiencing their content, rather than by gathering information from secondary sources [59]. Overall, this is a trade-off between effort and accuracy; consumers always gather risk-diminishing information when there is uncertainty. They often choose options that are satisfactory, but suboptimal if decision costs were zero [36].

Development of Hypotheses

This research was made possible through a partnership with a large digital entertainment firm in Singapore. There are varied kinds of programming from a number of content clusters (also called genres), such as news and children's programs, and entertainment and educational shows. Customers can specify the clusters of content as well as premium channels to be included in their subscription packages. Most channels are available in high-definition format also. Monthly subscription fees reflect the number, type, and quality of channels accessible to households.

The service provider also delivered a wide selection of movies and series dramas on demand, on top of a household's TV subscription. VoD services can be expensive though: a series with multiple episodes can cost from \$3 to \$60 in the market we studied. For each VoD series purchased, a household obtains immediate access over a preset period—depending on the number of shows in the series. The service provider offers households the first episode of series dramas to watch for free before they make a purchase. We next develop hypotheses on content sampling, the purchase of VoD series, and the effects of subscriptions at the household level, based on different theoretical perspectives.²

Free Sampling and Consumer Purchases

Information acquisition is known to be a costly and time-consuming though valuable process [20]. Initially, households will be uncertain about the quality of a series and whether it fits their preferences. They actively seek fit-related, risk-diminishing product information before making purchase decisions, especially when there may be financial consequences [18]. Though they can learn about a TV series through various means—online and offline, such as through online reviews or viewership ratings—they will explore and update their evaluations of different series through the free episode samples. Samples give households direct and easy access to quality and preference fit information for a series,³ thus reducing the associated search cost. Free samples can reduce uncertainty too, given that a household obtains direct experience with the content of one episode [55]. Thus, we offer:

Hypothesis 1: (Household's Content Sampling) A household's free sampling of a series has a positive effect on its likelihood to purchase that series.

Even when a household identifies a series that its viewers will like, it is possible that the household members will sample a few other series to rule out the available alternatives. By sampling this way, they will be more informed in the decision to buy the VoD series. This greater involvement will likely lead to more than one purchase. First, the household members are more likely to find other acceptable entertainment goods that meet their preferences. Second, sampling also provides a way for a household to broaden its consumption. For instance, a household that normally prefers the comedy-related genre may sample a crime-related drama and find it interesting. Such variety-seeking behavior [57] may result in multiple purchases across different genres. And because the first episodes of all series are offered for free, the perceived search cost for a household is minimal. So by increasing the household's involvement, free sampling ought to increase VoD series demand in the household.⁴ We assert:

Hypothesis 2: (Household's Purchase Decision Involvement). A household's involvement in its purchase decisions via content previews increases the number of drama series that it purchases.

Paid Sampling and Consumer Purchases

Households are likely to purchase the series that satisfy them based on their experience with free samples. This does not imply that a one-episode free sample is effective for all series though. Such a sample may not be sufficient for households to evaluate fit, as it is rarely the pilot episode that gets consumers hooked on a series [45].⁵ After a household watches a first free episode of a series, they can purchase subsequent episodes of that series separately at the typical stated price, around \$1 or \$2 each, or purchase the whole series at a discount. The price of a series is fixed, regardless of how many episodes the household has already purchased. Hence, the best solution *ex post* is not the same as the best solution *ex ante*. The best option for those who like a series is to purchase it shortly after free sampling. If the household is still hesitant about buying the series, its members can also seek additional information from outside sources for further evaluation. This alternative option is not desirable though. For different series, the search costs involved can vary greatly, and yet the household will still not be able to evaluate fit.

Any episode purchased before the household has purchased the whole series is considered to be a paid sample, as the household pays to additionally sample the series. Purchasing a paid sample is preferable in this case, because continuing to watch the series is the most effective way to reduce uncertainty concerning fit; this is especially true after the household has already previewed the first episode. In addition, the consumer decision making process involves a trade-off between effort and accuracy [36], so households should be willing to pay more for direct fit over indirect fit information. As a household purchases more paid samples, it will become

more informed about whether the content is suitable, and this should increase the number of series purchases. Hence, we posit:

Hypothesis 3: (Household's Informedness About Fit). A household's informedness about the fit of any drama series increases the number of drama series that it purchases.

Pay TV and TV services represent a good source of revenue for service providers. It is useful to look at the interaction between the consumption of new service innovations, such as VoD series, and the consumption of existing services, especially when both are subject to time and budget constraints [3]. For instance, Liebowitz and Zentner [52] showed the impact of Internet consumption as a substitute for television viewing. While the household's overall subscription package reveals its demand and preferences for TV viewing, the next two hypotheses examine a more nuanced relationship between the household's choices of content and its purchases of VoDs.

Standard Content Choices and Consumer Purchases

A household's TV subscription usually includes a selected number of standard content clusters.⁶ A cluster includes multiple channels that are similar in nature. For example, consider the News cluster, which includes local, regional, and international news channels. The number of standard content clusters approximates how many channels the household has access to, as well as its monthly payment. Consequently, households with a variety of channels to choose from will be less interested in VoD content, especially because a VoD series is typically longer than other programming: a 20-episode drama, at 45 minutes per episode, takes about 15 hours to finish. A subtler implication is that even if a household likes the content of the series after the free episode, it is still less likely to purchase the series, due to time and budget constraints. The marginal utility from the consumption of a VoD series is likely to diminish. So the variety of choices in a household's subscription appears to interfere with its series purchases:

Hypothesis 4: (Standard Content Choice). The greater the number of choices of standard content in a household's TV subscription, the lower the number of series it purchases.

Customized Add-On Content Choices and Consumer Purchases

Households can also customize their viewing experience beyond standard content clusters by: (1) adding specific programs and niche channels; (2) adding more channels in the same content cluster; or (3) upgrading their subscribed channels to higher screen resolutions. These requests reveal a household's expected level of utility from TV viewing, and they reflect utility for additional paid content that goes beyond what is available in a typical household TV services subscription. The

members of a household are likely to experience different levels of utility, and not all of them will agree on the same programming content. For example, households with fewer members or those who do not have time for TV viewing are likely to be content with the basic channels; and yet households with small children may benefit from special educational programming. If TV viewing is the main form of entertainment for the household, then acquiring access to a more diversified set of channels beyond the basic subscription services is appropriate. Households with a higher level of utility are more likely to try out VoD services, and likely will have higher willingness to pay for more suitable content. Adding on more customized, paid services gives households more control over the content they watch, in the same manner that they were able to customize their packages when they initiated their service subscriptions. Thus, we assert:

Hypothesis 5: (Customized, Add-On Content Choices). The more customized, add-on choices a household's TV subscription service offers, the higher the number of series it purchases.

The household's choices for standard content versus its own customized, add-on choices have different impacts on its demand for VoD series purchases, as the service provider used different pricing structures for the standard clusters and the add-on channels.

Research Setting and Data

We first present our research setting and the data extraction approach that allowed us to gather information from various sources and handle the limitations that accompany it. Then, we analyze the data sets to discover the underlying causal relationships.

Research Setting and Data Extraction Approach

The VoD and household-related data were collected through smartcards that are used in digital set-top boxes for digital cable TV and satellite entertainment systems. Smartcards store a household's information, the channels to which it subscribed, and all of the viewing records the smartcard captured. The technology does not identify which individual members watched the programming though. The voluminous data that we use pertain to household-level VoD viewing activities for one month between September 30 and October 30, 2011, and include more than 17 million viewing sessions. A viewing session for a TV program occurs when a household starts watching, and ends when it switches to another channel or turns off the TV. There are three categories of VoD sessions: (1) free-sample sessions include the viewing of first episodes of a series; (2) paid-sample sessions involve the viewing of purchased episodes; and (3) series-purchase sessions record the viewing of purchased series.⁷ There were no holidays, promotions, or special events during this

period that might have influenced household viewing activities in ways that created anomalies in the data or household-level biases, which would have made our use of it problematic.

The large amount of set-top box data represents only one month of household viewing for the provider's market though. An important aspect of empirical research with consumer and household data at scale is to obtain as deep an understanding of behavior as the data will allow [11]. Thus, we used multiple data sources to bring together the household information, series characteristics, and VoD activities for this study. A problem arises when there are many observations at the level of the primary unit of analysis, but an incomplete set of variables across all the time periods or stratification.⁸ This forced us to make choices on how to construct a workable research design to support the overall research inquiry, while still yielding useful insights.

We implemented a data extraction approach, feature selection, to maximize the number of observations available for empirical testing. Feature selection refers to a process of strategically selecting a subset of variables that are relevant to address each research objective. We analyzed all VoD sessions for 14,596 different holds. This set of anonymized households is called the *Households with VoDs Only Data Set*. We used it to explore the sequences and patterns of household VoD consumption. Nevertheless, it was not possible to link the full household-level information to the viewing-related variables that would have supported an ideal research design at the household level for the series-drama sampling the households did. We could only match 8,939 households with their subscription information. We call this set of anonymized households as the *Households with VoDs and Subscription Information Data Set*, and used it to test our hypotheses related to household VoD activities (see [Figure 1](#)). Both data sets are representative of the entire customer population.

Preliminary Analysis of Households' TV Viewing and VoD Activities

In the *Households with VoDs Only Data Set*, the anonymized households viewed 28,214 free samples for the first episodes of the various series, and 10,164 paid samples of other episodes. There were 1,140 series purchased, which yielded a conversion rate for free samples to series purchases of 4.04 percent. A closer look at the volume of household sampling and purchasing activities throughout the weeks revealed an interesting pattern (see [Figure 2](#)).

We observed similar sampling and purchasing patterns. A surge of free-sample activity on Fridays was followed by a high number of paid samples and series purchases on Saturdays. These patterns provide visual evidence for the positive relationship between sampling and purchasing, and suggest that the anonymized households searched for shows so they could watch them during the weekend. All activities slowed down during the weekdays though; the households did not have as much time during the week for TV viewing. The gap between the number of free

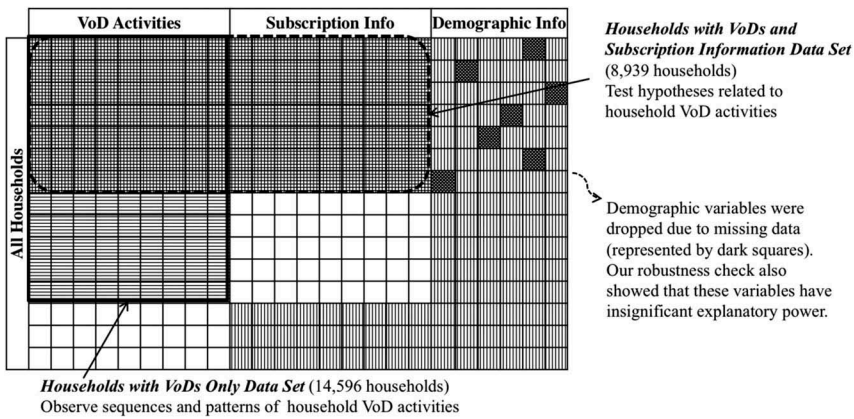


Figure 1. Approach Used to Extract Data for This Study

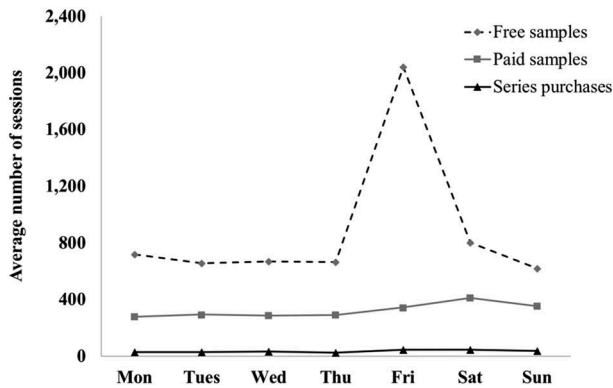


Figure 2. Average Number of Samples by Type and Series Purchased, by Day of the Week

samples and series purchases points to room for service providers to improve the conversion rate for VoD content.

We also looked more closely at the *Households with VoDs and Subscription Information Data Set* to examine the underlying relationships. The conversion rate for free samples to series purchases of these households is 5.02 percent. The descriptive statistics and correlation matrix for the variables in this data set are reported (see Table 1). *#SeriesPurchases* is the number of series that household *j* purchased in the study period, *#FreeSamples* is the number of one-episode free samples it watched, and *#PaidSamples* represents the number of episodes it bought. *ContentClusters* captures the number of groups of standard content to which household *j* subscribed. *PremiumChannels* refers to the number of customized channels selected when the service contract was signed. Together, they represent a household

j 's subscription package. Households with many content clusters were more likely to have more premium channels, so the correlation was 57.6 percent.

Other considerations in the anonymized households' VoD purchases are the nature of the service offerings and the characteristics of the series themselves. Factors such as ads, price, and rental time are likely to influence household purchase decisions. In our context, the service provider advertised all series dramas under "VoD Services," thus there were no advertisement effects for individual series. Higher-quality and more popular series from particular markets or genres may receive more attention from viewers, and thus they were more sampled and purchased. For example, romantic Korean dramas have attracted audiences worldwide in recent years. Consequently, we may overestimate the effect of free samples on a subset of popular dramas. Due to data scarcity, however, we cannot incorporate these factors into the main models, so we conducted a series-drama-level analysis separately. We also extracted outside quality information on the series, such as viewership, ratings, and award nominations to assess the impact of sampling versus outside quality information on series sales.

Research Methodology

We next present the explanatory empirical approach we used for causal inference in this study. (See [Figure 3](#) for an overview of the data analytics procedures.) To test the hypotheses on the overall effectiveness of sampling strategy on the consumption of series dramas, we used different count data models that can handle aggregated data at the household level over a one-month study period. We also implemented propensity score matching (PSM) to reduce selection bias due to household heterogeneity, and address the endogeneity issue, by using a suitable instrumental variable for a household's free samples. To test for a direct relationship between a household's free sample of a series and its likelihood of purchase for that series, we needed to handle the issue of left- and right-data censoring in our data set. Finally, we also implemented an identification strategy using heterogeneity across the VoD series.

Empirical Testing Procedures

Count data models

The variable of interest is the count value of VoD *#SeriesPurchases* for each household. This value is censored at 0, if a household did not purchase any series; censoring makes ordinary least squares (OLS) estimates inconsistent [33].

We captured the relationship between the number of *#SeriesPurchases* and other variables via this function: $\#SeriesPurchases = f(\#FreeSamples, \#PaidSamples, ContentClusters, PremiumChannels)$ for each household j ,⁹ and estimated:

Table 1. Descriptive Statistics and Correlation Matrix for the Households with VoDs and Subscription Information Data Set (8,939 households)

Variables	Descriptive Statistics							Correlation Matrix				
	Mean	SD	Min	Median	Max	1	2	3	4	5		
1. #SeriesPurchases _j	0.103	0.410	0	0	7	1.000						
2. #PaidSamples _j	0.950	3.727	0	0	93	0.195	1.000					
3. #FreeSamples _j	2.048	2.172	0	1	29	0.162	0.063	1.000				
4. ContentClusters _j	3.909	1.280	0	3	19	0.084	0.061	-0.031	1.000			
5. PremiumChannels _j	3.201	3.046	0	2	25	0.111	0.096	-0.029	0.576	1.000		

Notes: *j* denotes individual households; the least correlated variables are #FreeSamples_j and PremiumChannels_j (-2.9 percent), and the most correlated ones are ContentClusters_j and PremiumChannels_j (57.6 percent).

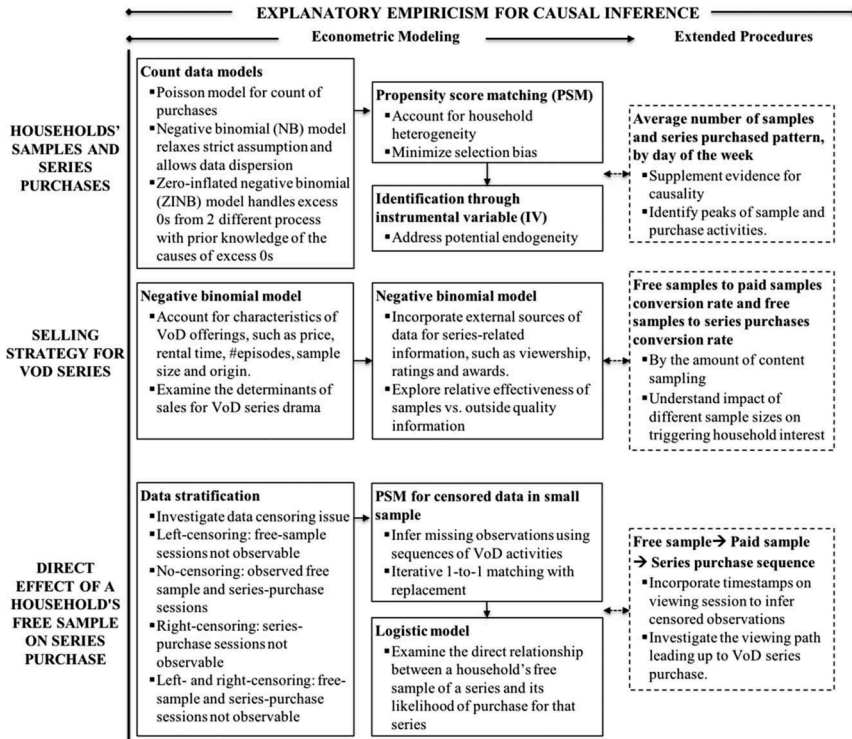


Figure 3. Overview of the Data Analytics Procedures in This Study

$$\begin{aligned} \#SeriesPurchases_j = & \beta_0 + \beta_1 \#FreeSamples_j + \beta_2 \#PaidSamples_j \\ & + \beta_3 ContentClusters_j + \beta_4 PremiumChannels_j + \varepsilon \end{aligned}$$

Since most households did not make many purchases and the maximum was just seven series, we assessed various count data models that are appropriate to handle these characteristics. Count models restrict the dependent variable to nonnegative integer values, and account for the mean and variance of the distribution used to characterize the dependent variable [9].

Poisson regression model

The most well-known of the discrete regression models for count data is the Poisson model, which takes the form of: $y_j \sim Poisson(\theta_j)$ for $j = 1, \dots, N$ and all $y_j > 0$; $\theta_j = \exp\left(\sum_j \beta_j x_j\right)$ and all $\theta_j > 0$; and finally $y_j \sim Poisson(\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n))$. Using the Poisson distribution, the events are estimated as independent of one another, without any restrictions on the independent variables. It constrains the conditional mean and variance of the dependent variable to be the

same though, which is not appropriate for our research. So, we use this model as an estimation baseline only.

Negative binomial (NB) model.

We observed a sparse dependent variable matrix, which is common in purchase conversion research, as the majority of households made no purchases or few purchases. In our data, this was a larger proportion than what we would see for a normal distribution (see Table 2).

Overdispersion occurs when the conditional variance of the dependent variable exceeds the conditional mean. As a result, the standard errors of the parameter estimates from the model will be underestimated [38], and the estimated values of the parameters will be greater than would be predicted based on the use of the Poisson distribution for the observed event counts. We checked for overdispersion by calculating the overdispersion ratio, which is more or less than 1 if there is overdispersion or underdispersion, respectively. Negative binomial regression generalizes the Poisson model and handles this issue. It has an extra parameter, α , to model the degree of overdispersion: the larger α is, the greater the amount of overdispersion in the data. The confidence intervals for the negative binomial model are also narrower compared to those of a Poisson regression model.

Zero-inflated negative binomial (ZINB) model

In addition to overdispersion, our data sets exhibited more zeros for no purchase decisions than those that the Poisson model can handle. The Poisson model also assumes that the zeros and nonzeros come from the same data-generating process [16]; this is not true in our setting though. The class of zero-inflated models relaxes this assumption [34], by modeling the response variable as a mixture of the Bernoulli and Poisson distributions.¹⁰

A household's zero-purchase decision may result from different processes. For example, if a household does not have money or time to consume the whole series,

Table 2. Conversion Rates of Free Sample for Households

Conversion Rate	<i>Households with VoDs and Subscription Information Data Set, %</i>
Free samples to paid samples only	8.09
Free samples to series purchases only	3.44
Free samples to paid samples and series purchases	3.56

Notes: The conversion rate of free samples to purchases at the household level is the number of the household's purchases, divided by the number of its free samples. All household data were anonymized.

they will not purchase regardless of whether they watched the free previews. And if the household purchases a VoD series, then its decision making process will have been a function of perceived quality and fit, in keeping with their unitary or aggregate preferences. This is a count process model, where the count is influenced by other variables. Based on our observation of the anonymized households' TV viewing activities, the consumption of on-demand content is bounded by several constraints. Thus, we modeled the expected count of *SeriesPurchases* as the result of a combination of two processes:

$$\begin{aligned} E(\#SeriesPurchases_j = k) &= Pr(HouseholdWithConstraints) \cdot 0 \\ &+ Pr(HouseholdWithoutConstraints) \cdot E \#SeriesPurchases_j \\ &= k \cdot Pr(HouseholdWithoutConstraints) \end{aligned}$$

To account for this, we chose the ZINB regression model, which has a logit model part and a negative binomial count data model part. The logit part models the probability of excess 0s independently; the probability of $\#SeriesPurchases = 0$, due to the fact that a household's purchases are bounded by some constraints. The covariate, *ContentClusters*, reveals some of these constraints for household j . The two parts do not need to use the same predictors, and the estimated parameters do not need to be the same either. Since y_j below represents $\#SeriesPurchases$, the number of series purchased by household j , the probability density function is:

$$\Pr(Y_j = y_j) = \begin{cases} \Phi + (1 - \Phi)(1 + k\mu_j)^{-k-1} y_j = 0 \\ (1 - \Phi) \frac{\Gamma(y_j + k - 1)}{y_j! \Gamma(k - 1)} \frac{(k\mu_j)^{y_j}}{(1 + k\mu_j)^{y_j + k - 1}} y_j > 0 \end{cases},$$

with $E(y) = \mu_j (1 - \Phi)$; and $Var(Y_j) = \mu_j (1 - \Phi) (1 + k\mu_j + \phi\mu_j)$, where μ_j and ϕ depend on the covariates. Here, ϕ is the density function governing the binary process such that $0 \leq \phi < 1$, and the dispersion parameter $k \geq 0$ is a scalar [48]. When ϕ or k is greater than 0, there is overdispersion. When $\phi = 0$, the equation reduces to a negative binomial, and for $k = 0$, it becomes a zero-inflated Poisson (ZIP) model.

Propensity Score Matching (PSM) to Address Selection Bias

Causal inference using observational studies has been a central pillar of many disciplines [23]. A causal effect is a comparison between the potential outcome of a treatment group and a control group, averaged over a population [77]. Without a randomized assignment, bias may arise due to systematic differences between the groups. In our business context, the households that watched free-sample episodes may be different from those that did not. The differences between these households produce bias in our estimations. Matching methods have been used effectively to address this problem [74, 75]; they involve the pairing of treated and controlled observations that are similar in some observable characteristics.

In our *Households with VoD and Subscription Information Data Set*, we identified 586 households without any free-sample sessions. We used the PSM approach and found comparable matches for these anonymized households, based on two sets of covariates that are likely to have influenced the households' decisions to sample VoD content. The treatment is the household's exposure to VoD sampling, and the outcome is *SeriesPurchases*. The first set of covariates consists of *LoyalCustomer*, *ValueCustomer*, *EarlyAdopter*, and *TechOptimist*, representing four different household relationships with the service provider. *TechOptimist* represents the households that typically respond promptly to new products and services, and *EarlyAdopter* represents the households that were first to subscribe to new offerings. *LoyalCustomer* refers to households that were observed to use multiple services from the provider, and *ValueCustomer* refers to those with high-value contracts with the provider.

The second set of covariates includes demographic variables such as *AgeBand*, *Ethnicity*, *HouseholdSegment*, *Housing*, and *Region*. The category of variables, *HouseholdSegment*, captures the diversity of the customer base, which may reflect the differences in viewing preference. *Housing* offers a way to control for household size and income, as larger and wealthier families tend to live in larger residences. We weighted the differences between the covariates for the households that were observed to have sampled VoD content and those that did not, in order to establish statistical equivalence between the treatment and control groups [50, 68]. This matching method yielded 1,655 households with free samples and 394 households without free samples.

Instrumental Variable (IV) Analysis for a Household's Free Episode Samples

Another issue in our econometric models is whether the variable, *#FreeSamples*, is exogenous. We handled this endogeneity issue by finding a suitable instrumental variable (IV) for a household's free samples. A suitable IV should be exogenously related to that household's tendency to sample VoD series, but not affect its VoD series purchases. We noticed that, at the time of the research, the service provider offered an interactive home entertainment service to households on a monthly subscription, on-demand basis. Anonymized households that subscribed to this service were able to access an extensive library of songs in various languages to watch or sing along with. It was offered on the same platform as the VoD series. Every time a household used this service, it was exposed to a variety of VoD series. Thus, households that used the service frequently were more likely to sample VoD series. Yet we did not expect to see a direct relationship between a household's usage of this service and its series purchases.

Propensity Score Matching (PSM) to Handle Data Censoring in a Small Sample

In marketing, medical epidemiology, and employment research, data censoring has been a common challenge since historical data for consumers, patients and employment are rarely available in complete form.¹¹ In addition, personally identifiable information on consumers must be masked due to privacy regulations. In this research, we encountered left- and right-data censoring for free- and paid-video sampling, as well as subsequent purchases, during the one-month time window. Thus, the number of observations in the noncensored data category is relatively small. This small set is also infeasible for empirical testing to gauge the effect of a household's free samples on its likelihood to purchase that series, as each free-sample session corresponds to a purchase session. Common computational and resampling approaches, such as the partial deletion, multiple imputation, and bootstrapping methods, are not suitable to handle this issue [26, 27].

Censored data create a roadblock for establishing a solid foundation for causal inference. We view this as an opportunity for a methodological advance, however. We propose a household behaviour matching method that requires the recognition of patterns and the adherence to a particular kind of ordering, or sequence in all of the data, to match them so that censored records for some household behaviour can be recovered. Our method extends the PSM and data imputation approaches to match and impute the values of the censored records from outside the observation window of the study based on a probabilistic model [19, 30, 71]. This is an advance for identifying causal links, by the completeness of the household-level data for causal inference.

Results

We offer the main empirical results from our econometric models, followed by analytical procedures to address concerns that a reader may raise, and we discuss the robustness of our identification strategy.

Household's Samples and Purchases of VoD Series

The estimation results obtained from count data models support the positive relationship between a household's samples and the number of VoD series it purchased. To strengthen this relationship, we include procedures to address selection bias and endogeneity issues arising from heterogeneity across different households and different VoD series.

Count data models results

Table 3 shows the results of the Poisson, negative binomial, and zero-inflated negative binomial models, at the household level of analysis.

We report the estimates of the ZINB model as our main results. The coefficients for *#FreeSamples*, *#PaidSamples*, and *PremiumChannels* are positive and significant. The coefficient for *ContentClusters* is negative as we expected, but not significant. The marginal effects of *#FreeSamples*, *#PaidSamples*, and *PremiumChannels* are 1.198 ($= e^{0.181}$), 1.095 ($= e^{0.091}$), and 1.092 ($= e^{0.088}$), respectively. The exponential values of the coefficients represent the incidence rate ratio, which is the relative risk of something occurring versus not occurring [25]. We further leveraged them to interpret the estimation results in terms of their statistical confidence intervals. If a household were to watch one free sample more, for example, its corresponding incidence rate ratio would be expected to increase by a factor of 1.198. Thus, households with an additional free sample will purchase dramas 19.8 percent more of the time, supporting the Household’s Purchase Decision Involvement Hypothesis (H2). Likewise, an additional paid sample caused a 9.4 percent increase in the number of series purchased, aligning with the Household’s Informedness About Fit Hypothesis (H3). An additional premium channel predisposed a household to have a 9.2 percent increase in the number of series purchased, which supports the Customized, Add-On Content Choices Hypothesis (H5) (see Table 4).

We did not find statistical support for the Standard Content Choice Hypothesis (H4) though. Interestingly, the results also reveal that the log odds of the excess zeros decreased by 1.505 for each content cluster that a household subscribed to. This implies that no-purchase decisions were less likely due to time and budget constraints.

Robustness of the ZINB model estimates

Our choice of the ZINB model is appropriate for this data set. The overdispersion ratio of 1.289 from the Poisson model suggests overdispersion estimation bias. The NB model, with an extra parameter that estimates the degree of overdispersion, produced coefficients that are slightly larger than those of the Poisson model ($0.178 > 0.135$, $0.094 > 0.041$, $0.090 > 0.080$, and $0.085 > 0.078$). Thus, we observed an improvement of the NB model, as expected, over the Poisson model.¹² Next, the ZINB model deals with the excess zeros for no-purchase decisions in the data set, by modeling “true zeros” and “inflated zeros” separately. The impact of free samples was stronger compared to the results from the prior models. The ZINB model fit the data better than the null intercept-only model in a statistically significant way.¹³ We used a closeness test to check whether the two models were indistinguishable [80]. Based on a Vuong test statistic of 1.75 ($p < 0.1$), we rejected the null hypothesis that the two models were equally close to the true data-generating process.

Table 3. Poisson, Negative Binomial (NB), and Zero-Inflated Negative Binomial (ZINB) Results

Variables	Poisson		Negative Binomial		Zero-inflated Negative Binomial			
	Coef.	SE	Coef.	SE	Count Data Part	Logit Part		
					Coef.	SE		
<i>Intercept</i>	-3.320***	(0.102)	-3.609***	(0.134)	-3.029***	(0.259)	3.670	(2.250)
<i>#FreeSamples_j</i>	0.135***	(0.008)	0.178***	(0.013)	0.181***	(0.016)		
<i>#PaidSamples_j</i>	0.041***	(0.003)	0.094***	(0.006)	0.091***	(0.009)		
<i>ContentClusters_j</i>	0.080***	(0.027)	0.090***	(0.035)	-0.009	(0.049)	-1.505*	(0.805)
<i>PremiumChannels_j</i>	0.078***	(0.011)	0.085***	(0.014)	0.088***	(0.015)		
<i>Ln(θ)</i>					-0.949***	(0.150)		

Notes: Model: Poisson; 8,939 obs.; dep. var.: #SeriesPurchases. Null dev.: 4,901.6, 8,939 d.f.; resid. dev.: 4,388.3, 8,939 d.f., pseudo R^2 : 0.080, AIC: 5,906.3. Model: Negative binomial; 8,939 obs.; dep. var.: #SeriesPurchases. Null dev.: 3,058.6; 8,939 d.f.; resid. dev.: 2,629.4, 8,939 d.f., pseudo R^2 : 0.067, AIC: 5,514. $\theta = 0.306$; degree of dispersion: $\alpha = 1/\theta = 3.27$. Model: ZINB; 8,939 obs.; dep. var.: #SeriesPurchases. Pseudo R^2 : 0.069, AIC: 5,507. $\theta = 0.387$. Signif.: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4. Incidence Rate Ratios for Coefficients from ZINB Model and Their Confidence Intervals

Variables	Coef.	Confidence interval	
		2.5%	97.5%
<i>Intercept</i>	0.048	0.029	0.080
<i>#FreeSamples_j</i>	1.198	1.162	1.236
<i>#PaidSamples_j</i>	1.094	0.075	1.115
<i>ContentClusters_j</i>	0.991	0.900	1.092
<i>PremiumChannels_j</i>	1.092	1.061	1.124

Note: The confidence intervals of 2.5 percent and 97.5 percent are the lower and upper bounds, respectively, of the 95 percent confidence intervals for coefficients.

ZINB model results after the PSM approach

The imbalance in the covariates may have affected the outcome of our results. Households that sampled free episodes are different from those that did not sample them, which influenced their series purchase decisions. We employed the PSM approach to match the households with and without free samples. Table 5 shows the ZINB model results after the PSM approach was applied. These coefficients align with our main results, which provides additional support for the impact of content sampling on the consumption of VoD series (see Table 5).

Two-stage least-squares (2SLS) IV results

We used the number of household-level home entertainment sessions as an IV for a household's free samples. We removed all duplicate sessions in the same day. We also conducted an endogeneity test on the 479 households that subscribed to home

Table 5. ZINB Model Results After the PSM Approach Was Applied

Variables	Count Data Part				Logit Part			
	Coef.	SE	z-Val.	<i>p</i> (> z)	Coef.	SE	z-Val.	<i>p</i> (> z)
<i>Intercept</i>	-1.945***	0.509	-3.823	< 0.001	2.497*	1.409	1.772	0.076
<i>#FreeSamples_j</i>	0.156***	0.025	6.355	< 0.001				
<i>#PaidSamples_j</i>	0.072***	0.014	5.107	< 0.001				
<i>ContentClusters_j</i>	-0.121	0.093	-1.303	0.193	-0.911*	0.487	-1.873	0.061
<i>PremiumChannels_j</i>	0.075**	0.029	2.569	0.010				
Ln(θ)	-0.149	0.441	-0.337	0.736				

Notes. ZINB model; 2,049 obs.; dep. var.: *#SeriesPurchases*. Pseudo R^2 : 0.055, AIC: 1,589, $\theta = 0.862$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 6. Linear Model Estimation Results with an Instrumental Variable (IV)

Variables	Linear Model without IV				2nd-stage estimates with IV				
	Coef.	SE	z-Val.	$p (> z)$	Variables	Coef.	SE	z-Val.	$p (> z)$
<i>Intercept</i>	-0.058	0.090	-0.647	0.518	<i>Intercept</i>	0.078	0.211	0.369	0.713
<i>#FreeSamples_j</i>	0.038***	0.009	4.073	< 0.001	<i>1stStageErrors</i>	-0.017	0.077	-0.219	0.827
<i>#PaidSamples_j</i>	0.036***	0.006	5.591	< 0.001	<i>#PaidSamples_j</i>	0.041***	0.010	4.253	< 0.001
<i>ContentClusters_j</i>	0.030	0.025	1.215	0.225	<i>ContentClusters_j</i>	0.025	0.027	0.943	0.346
<i>PremiumChannels_j</i>	0.000	0.009	0.045	0.964	<i>PremiumChannels_j</i>	0.002	0.010	0.247	0.805

Notes: Model: Linear without IV, estimated with OLS; 474 obs.; dep. var.: *#SeriesPurchases*; resid. SE = 0.571; 474 d.f.; R^2 : 0.112; adj. R^2 : 0.104; F -stat: 14.92 on 4 on 4 vars. and 474 d.f.; $p < 0.01$. Model: Linear with IV, estimated with 2SLS. 474 obs.; dep. var.: *#SeriesPurchases*; resid. SE = 0.581; 474 d.f.; R^2 : 0.081; adj. R^2 : 0.073; F -stat: 10.42 on 4 and 4 vars. and 474 d.f.; $p = < 0.01$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

entertainment services. The estimation results for the OLS and 2SLS models are reported in Table 6, suggesting that even if the *#FreeSamples* variable is considered to be endogenous, the results are still in alignment with our earlier findings (see Table 6). The Hausman IV test result ($\chi^2 = 0.511$) for endogeneity shows that *#FreeSamples* can be treated as exogenous, however.

Tables 3 and 4 offer empirical evidence of the positive impact of samples on the number of VoD series that households purchased. Those that viewed more free samples and paid samples ended up purchasing more VoD series. These results align with our main hypotheses: households that are more involved in the purchase decision, and more informed about the fit of VoD series dramas with their aggregate preferences will likely purchase more. In addition, we also wanted to see if the households' TV subscriptions influenced additional VoD purchases. As we expected, households that purchased more customized, add-on TV viewing content options beyond their basic TV subscriptions were more likely to purchase VoDs series. These findings remained robust after we addressed the issues of heterogeneity and endogeneity. More important, our results offer the service provider a directional reading on causality between content sampling and on-demand purchases, after all other covariates were accounted for.

Household's Free Samples and Likelihood of Purchase for VoD Series

Extended PSM for censored data in a small data set

The noncensored data contained fewer observations than were desirable for empirical testing. At the household level, there were only 193 observations for which we had a full reading of free-sample, paid-sample, and series-purchase activities, out of 30,006 observations in total. Thus, it was infeasible with this small a sample size to gauge the extent of a causal relationship between a household's decision to watch a free sample and then make a series purchase. So, we matched censored household-level data to noncensored household-level data, based on their sequences of activities. Then, we inferred a behavior in the censored data using the 90th percentile of the distribution for the viewing patterns associated with all noncensored data from that sequence of activities. The use of the 90th percentile of the distribution is appropriate based on the data. As more time goes by after watching a sample, the households were less likely to make a purchase, making the use of anything more than the 90th percentile unnecessary. And yet, using anything less than the 90th percentile would discard the households that needed more time to make their decision. As a result, we recovered 862 left- and 10,848 right-censored household-level data that were likely to have occurred just outside the study period.

Table 7. Logit Model Results

Variables	Coef.	SE	z-Val.	p ($> z $)
<i>Intercept</i>	-1.713***	0.114	-14.970	< 0.001
<i>FreeSample_j</i> (0/1)	1.366***	0.104	13.092	< 0.001
<i>#FreeSamples_j</i>	-0.007*	0.004	-1.927	0.054
<i>#PaidSamples_j</i>	-1.017***	0.003	-5.632	< 0.001
<i>ContentClusters_j</i>	0.010	0.014	0.729	0.466
<i>PremiumChannels_j</i>	0.002	0.006	0.401	0.688

Notes: Model: logit; 19,815 obs.; dep. var.: *SeriesPurchase* (0/1). Null dev.: 26,712; 19,814 d.f.; resid. dev.: 26,422; 19,809 d.f., pseudo R^2 : 0.011, AIC: 26,434. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Logit model results

We used a logit model to estimate the effect of whether a household samples a series on the likelihood of its purchase of that series. The binary dependent variable in this model represents whether a household j purchased a particular series i . Beyond all the independent variables that are used in the count data models above, we added a binary independent variable, *FreeSample_j* (0/1), to show whether the household j had watched the free episode of series i . This model tests for the direct effect of a series' free sample on the likelihood of a household's purchase of that series. The results from this model strengthened our findings above (see Table 7).

The coefficient of *FreeSample* is positive and significant; so a household that sampled a series was more likely to purchase that series. This supports the direct relationship between a household's sampling and purchase for each series, which is our Household's Content Sampling Hypothesis (H1). Overall, a free sample of a series directly influenced a household's purchase decision of that series. It also positively influenced the household's decision to purchase other VoD series.

Robustness Check Analysis for the Empirical Research Design

The main objective of this research has been to extend our understanding of entertainment content-service providers' sampling-based strategy in the context of digital information goods. Causal inference with observational data still remained a challenge though we were able to access more than 17 million digital traces of anonymized households' viewing activities. This entertainment service provider and the data set did not permit us to conduct a full test to infer causality in the manner we wished, since we had no control over the business setting. So we took a divide-and-conquer approach to understand more deeply the causality relationship between content sampling and purchases in a scientific manner. First, the count data models were useful for understanding this data set, and allowed us to reach a general conclusion: over the one-month study period, the more samples a household watched, the more series dramas it purchased. We conducted a matching procedure to address selection bias due to

household heterogeneity. And we addressed potential endogeneity with a Hausman test and a suitable instrument to increase our ability to claim the presence of a causal relationship.

An intriguing question remains: Did the households really purchase the same series that they had sampled? To address this question, we repurposed PSM to impute censored observations for a smaller data set, but still one that was entirely representative of our study's setting overall. This innovation provided us enough complete sequences of data to analyze the direct impact of sampling on series purchases. We also accounted for series drama heterogeneity, and examined the relative effectiveness of content sampling versus outside quality information. Our findings indicate that the impact of series samples on purchases remained significant. Households were likely to purchase series dramas that fit their viewing preferences and expectations, rather than those that they perceived as being of generally good quality.

Discussion and Limitations

Our findings suggest that there is not just an association but also a causal link between episode samples and series purchases. A household's free sample increases its likelihood to purchase the series. This suggests that sample content signals both vertical and horizontal differentiation on objective features. In addition, free-episode samples are effective in increasing the purchase conversion rate not only because they were made available to the customers; the customers actually watched the content to evaluate its fit related to their preferences. An additional free sample for a household caused a 19.8 percent increase in the number of series it purchased. This indicates that for entertainment goods, customers also searched for and evaluated different alternatives before making a purchase. Watching free-episode samples is a faster and cheaper way for them to gain experiential knowledge. Thus, this action had a positive impact on series purchases in our study.

An important finding from this research is that an additional paid-sample episode led to a 9.4 percent increase in the number of series a household purchased. This seems counterintuitive because purchasing individual episodes of a series will increase the transaction cost of buying the remaining content of that series. Yet this result aligns with our overarching theory in this research: customers are willing to pay to be more well-informed about the content they like to watch, and informed households will end up purchasing more series dramas. Several aspects of this research deserve further discussion, especially in terms of the business insights that they have to offer. Next, we discuss the implications for service providers for their use of sampling-based strategies.

Implications for Service Providers

Omni-platform consumption and binge-watching of digital content have become the new norms. Analytics with big data on consumers' digital traces also play a salient role in guiding business strategic planning. Our research contributes to the understanding of content sampling as a strategic marketing tool. It also raises important questions regarding more effective implementation of sampling-based strategy: (1) Is there an appropriate amount of content sampling that stimulates series purchases by households? Would it be easier to convince household's viewers to purchase a cheaper, shorter series after a single free episode? (2) Can a service provider influence consumer conversion rates for different types of TV series? If the service provider has limited screen space to advertise free-TV series episodes, should it promote a cheaper, shorter series or a longer, more expensive one?

To the best of our knowledge, none of the prior studies addressed the issue of how much free content is enough in the context of series dramas, largely due to the other authors' limited access to data; thus, the most important problems have remained unsolved. We attempted to provide a sneak peek of some answers in this study. Across the households, in many cases, it was evident that one free-sample episode for a series was not enough for a purchase to occur. Service providers gain an additional stream of revenue from paid-sample episodes, however, it is not a desirable approach for everyone involved. Paid samples impose additional transaction costs for households, making VoD content more expensive. For example, even if a household sampled Episode 1 for free and then purchased Episodes 2 and 3 of a 10-episode drama, it still would have had to pay a fixed price for the seven remaining episodes. This may dissuade households from purchasing the series, creating a potential opportunity for the provider that would be missed.

The diverse nature of the series dramas in the data sets allowed us to examine the effect of the amount of sampling on household purchases, when the provider offered one free-sample episode for each series. The number of episodes in a series is a proxy for its price: the longer the series, the more expensive it is, and vice versa. For longer series in episodes terms, the conversion rate for paid samples was also high, while the conversion rate for series purchases was low. This suggests that a small portion of the sample content was not effective to stimulate series purchases, as households ended up purchasing many paid samples for additional viewing. So service providers, as a result, may wish to offer more episodes as free samples for longer series dramas. When one episode represents around 5–6 percent of the episode length of a drama series, the conversion rate of the free samples to series purchases was at its maximum, suggesting that this amount may be sufficient to spark a household's interest in a series. Service providers apparently will not benefit from simply increasing the number of free episodes, as our results suggest that the conversion rates for purchases quickly diminished for short series with a larger percentage of free content (see [Figure 4](#)).

There are many possible explanations for this. When a household has watched a substantial portion of a drama series via free samples, the remaining portion will

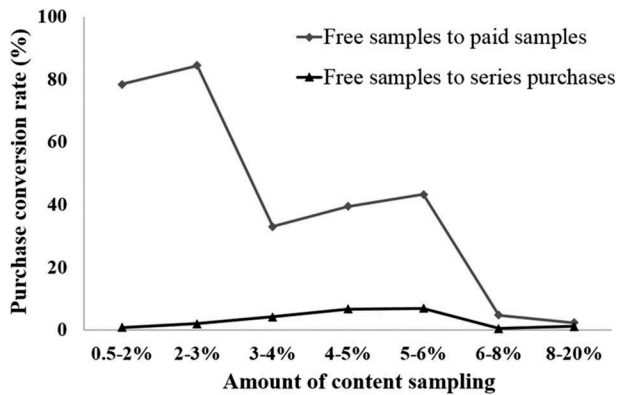


Figure 4. Conversion Rates by Amount of Content Sampled. Series dramas were sorted and aggregated based on their length in episode terms. So a one-episode preview for a 20-episode series is 5.0 percent, for a 30-episode series it is 3.3 percent, and for just 6 episodes it is 16.7 percent. The x-axis values represent the average conversion rate of all dramas within a given range of the amount of content sampling in percentages.

have become relatively more expensive, and a series purchase may be less attractive. Though our results only provide a glimpse into what really happened, the practical implications are important. Service providers should consider customizing their offerings of free samples and paid samples for different series dramas. An appropriate amount of free sampling is the amount that sparks a household's interest in a drama. An even more direct strategy is to offer a decreasing price scheme for the remainder of the series, encouraging more sampling and purchasing.

Another concern worth mentioning is that online piracy has taken a new form via illegal streaming services. It was estimated that there were over 141 billion visits across 200 million devices to the 14,000 largest piracy sites [7]. According to the same source, music and TV series are at the top of all illegally streamed content; streaming websites made up 73.7 percent of 78.5 billion visits to access pirated TV content in 2015. Offering content on an on-demand basis via legal streaming services has not been sufficient though: the rise of music streaming services has not killed music piracy [24]. This poses a major challenge and, at the same time, presents a new opportunity for content producers and service providers. Firms must leverage new technology and proprietary data for understanding consumer behavior more deeply to improve their market offerings, and to do so in a way that consumers cannot benefit from when they obtain programming from other illegal streaming sources.

Research Design Issues

Even with an innovative research methodology coupled with a strong theoretical foundation across different disciplines, the limited coverage of our one-month of

observational data hindered causal testing. This led us to modify our objective slightly, to making statistical inferences about important relationships that come very close to true causality, and at the same time, provide managerially important results. We formulated empirical testing models that worked well with the available data to make reasonable inferences about causality, based on an appropriate theoretical background. The different count data models that we used, with one improving on another, addressed the specific characteristics of set-top box viewing data. In addition, the key variables that we selected for these models relate directly to the VoD business. Next, we adapted the PSM approach to handle selection bias. We also conducted a Hausman test and used an instrumental variable estimation to address endogeneity. Finally, our use of PSM to impute censored household-level records for the data set allowed us to achieve more convincing empirical test results.

Our research is unique in that we studied a specific area of digital goods, on-demand series dramas, very closely. Thus, our results may not be generalizable to other types of digital entertainment products. In addition, the study was done in Singapore, so it would be interesting to conduct a similar study in other markets, such as the United States, where TV series play a major role in media consumption. An extension of this work also should consider a nonunitary model of the household to account for the differences among households whose average consumption preferences are similar, but whose individual members express different preferences [73]. How much free content is appropriate to make available for sampling remains a question for researchers and managers alike, and opens up new empirical research opportunities. We call for future studies that explore new marketing strategies for digital information goods, and studies that assess causality more thoroughly, by building on our method.

Conclusion

This research provides an empirical validation for the common wisdom that information goods are experience goods too, and giving the consumer a glimpse of the experience will be the most effective way to stimulate more purchases. Series dramas represent a major source of revenue for digital entertainment service providers, and the market for VoD drama series is unique for the application of sampling strategies to the consumption of digital information goods. This research is the first to provide empirical support for how episode sampling works in the context of VoD drama series purchases. A free-sample episode of a series has a beneficial effect by reducing a household's fit uncertainty for that series.

Even when a household's members know what they want to watch, they may need to sample other dramas to rule out any alternatives. Thus, a free sample of a series serves as a point of comparison for other series. Households with more customized content in their TV services are more likely to purchase VoD content, yet the number of content clusters that a household subscribes to apparently interferes with its VoD purchase intention. In addition, recognizing that a one-episode free sample will have different implications for dramas with various lengths in episode terms permitted us

to gain insights on the appropriate amount of sampling that needs to be supported. Although households were willing to acquire paid samples to ensure that a series fit their tastes, service providers should offer free samples more strategically than on a common market-wide basis.

We emphasize the main message that personal experience—Experience me!—is more influential than second-hand information for digital information goods sales to household consumers. With this in mind, service providers should invest more in marketing strategies that provide useful information about the fit of their digital goods with household preferences, since such strategies will help firms to reduce their marketing costs and increase sales and revenue performance in the long run. Another possibility is a decision support system that offers specific recommendations based on household viewing pattern matches on the households' TV screens. This will allow like-minded viewers to share their comments about their choices of VoD series with others. Equally important, digital entertainment service providers should implement incentive schemes that encourage viewers to watch more episodes and eventually make purchases, instead of looking for alternate sources of entertainment.

Acknowledgments: This research was originally supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the Infocomm Development Authority. An earlier version was: A.P. Hoang and R.J. Kauffman, "Experience Me! The Impact of Content Sampling Strategies on the Marketing of Digital Entertainment Goods," in R. Sprague and T. Bui (eds.), *Proceedings of the 49th Hawaii International Conference on Systems Science* (Kauai, HI: IEEE Computer Society Press, Washington, DC, 2016), where it was nominated for the Best Research Paper Award in the Organizational Systems Track. In addition, some aspects of the empirical methodology related to censored data and propensity score matching were presented at the 2017 Statistical Challenges in Electronic Commerce Workshop in Ho Chi Minh City, Vietnam, in July 2017. We are grateful to the late Steve Fienberg for his support and feedback in seminars at Carnegie Mellon University. Vladimir Zwass and the anonymous referees helped to shape our intellectual contributions in this research. We also acknowledge Steve Miller, Kapil Tuli, Qian Tang, and other academic participants in seminars at the Living Analytics Research Centre (LARC) at Singapore Management University, for their thoughtful comments. Eric Clemons, Atanu Lahiri, Jennifer Zhang, Yabing Jiang, Avi Seidmann and the participants of the "Integrating Business Operations, Information Technologies, and Consumer Behavior Mini-Track," gave us many suggestions that have made this research stronger. Rong Zheng, Tuan Phan, Ting Li, and the participants of the 13th Statistical Challenges in E-Commerce Research (SCECR 2017) Workshop offered helpful suggestions, and Terence Saldanha and Zhoulun Li also gave us feedback. We are especially grateful for support and guidance from the anonymous reviewers. This research was conducted with the participation of a corporate sponsor under a binding nondisclosure agreement; thus, some details of the data and qualitative findings were disguised. The households and data used in this research were anonymized. Also, the identities of individual households and account holders cannot be traced back through our analysis. All errors and omissions are the responsibility of the authors.

NOTES

1. A series drama consists of 10, 20, 30, or more episodes. Most American TV series, packaged since the 1960s with 20 to 26 episodes a season, are in this format. The economic importance of paid TV series revenue streams has increased, while providers have been

fighting for profitability in the face of Internet delivery and digital convergence. Producing an original TV series requires a huge investment: about US\$2 million to shoot a half-hour pilot and about US\$5.5 million for an hour-long drama [62].

2. In this study, we consider a unitary model of the household in which the viewing time constraint, demand, and preferences of all household members are pooled [73]. The current technology in our setting did not permit tracking individual viewers.

3. Content sampling signals both horizontal and vertical differentiation on objective features of a series to consumers. If the content only signals vertical differentiation, then consumers just need to know such samples are available, and they do not actually need to watch any free-sample episodes.

4. There are some drawbacks to free content. A perception that free content is available may dissuade consumers from buying programs [44]. Also, unlimited access to free content makes other programs less attractive and decreases consumers' willingness to pay [5]. Further, some consumers may sample with no intention to purchase anything, though this is unlikely for a majority of them in the VoD setting for several reasons. Series dramas are unique, so a viewer's experience is not complete without seeing it all. So, after viewing the free sample of a series' first episode, viewers may feel connected and want to view the rest of the content [78]. Those that sample a portion of the series are more likely to purchase the remainder of it. In addition, since households will have many channels in their TV subscriptions, they are unlikely to watch a free sample episode of a series if they have no prior topical interest.

5. Netflix's method of releasing a series—in its entirety—has helped the company to understand customer viewing behavior for the different series it offers across various market segments. This is relevant to our context, since it shows that a one-episode free sample may not be sufficient for the viewers [45].

6. In our research context, the households decided on the number of standard content clusters in their TV subscriptions at the beginning of long-term service contracts.

7. Households often finish watching an episode across multiple viewing sessions, as each episode takes more than 30 minutes. So, if a household had three free-sample sessions for a series, we only admitted the earliest session to our data set based on its timestamp, and removed other duplicates. This was normally not permissible.

8. Meaningful stratification is sometimes difficult with in big data analytics research. Even though the researcher may have access to a lot of data, often it is surprisingly hard to develop research designs to support causal analysis, such as researcher-designed field experiments, and quasi-experimental designs that have "just right" conditions that can be leveraged to produce undeniably correct managerial insights.

9. In the different count data models that we used, we did not include any household demographic characteristics as control variables. Instead, we used them in our propensity score matching approach, so this would have been double-counting to add them as control variables also. These variables include the demographic segmentation of the household, such as the region of the residence, age band, and gender of the residents. Other specifics regarding the ethnicity of the anonymized households are not included or reported, due to our nondisclosure agreement with the research sponsor. In fact though, these variables did not add much explanatory capability for the dependent variable of interest.

10. Hurdle models also relax the assumption that the zeros and nonzeros in the data set come from the same data-generating process. They use a Bernoulli probability that governs the binary outcome for the count variable with a 0 or a positive count. Once the hurdle or threshold is crossed, and a positive number occurs, the conditional distribution is represented by a truncated-at-zero count data model. Since we had prior knowledge of the cause of the excess zeros, we chose to proceed with zero-inflated models.

11. In censored data, the total number of observations is known but full information is not available for some [17]. Left-censoring arises when the events of interest occurred before the study period; right-censoring refers to events that might or might not have occurred after the period of observation ended. Data without censoring are ideal for empirical testing.

12. We justify the use of the NB model by showing that the data are overdispersed. The Poisson model is nested in the NB model. It relaxes the assumption that the conditional variance is equal to the conditional mean. We use a likelihood ratio test to assess the null

hypotheses to see if this restriction is true: $\lambda = -2 \cdot (LL_{NB} - LL_{Poisson})$. We rejected the null hypothesis that it is appropriate in favor of the NB model, based on $\chi^2 = 394.29$. This exceeds 2.71 ($p < 0.001$), so overall the evidence suggested the data are overdispersed.

13. We show that the ZINB model fits the data better than the null intercept-only model does. The associated χ^2 value for the difference between the model-level log likelihoods, $\lambda = -2 (LL_{ZINB} - LL_{Null})$ is 408.64. So the ZINB model is preferred over the null intercept-only model.

REFERENCES

1. Ba, S., and Pavlou, P. A. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 26, 3 (September 2002), 243–268.
2. Baumeister, R.F., and Bushman, B.J. *Social Psychology and Human Nature*, 2nd ed. Boston: Cengage Learning, 2010.
3. Becker, G.S. A theory of the allocation of time. *Economic Journal*, 75, 299 (September 1965), 493–517.
4. Bell, D. E. Regret in decision making under uncertainty. *Operations Research*, 30, 5 (October 1982), 961–981.
5. Berger, B.; Matt, C.; Steininger, D.; and Hess, T. It is not just about competition with “free”: Differences between content formats in consumer preferences and willingness to pay. *Journal of Management Information Systems*, 32, 3, (July 2015), 105–128.
6. Bhattacharjee, S.; Gopal, R.; Lertwachara, K.; and Marsden, J.R. Consumer search and retailer strategies in the presence of online music sharing. *Journal of Management Information Systems*, 23, 2 (Summer 2006), 129–159.
7. BI Intelligence. Illegal streaming is dominating online piracy. *Business Insider*. (August 2016). <http://www.businessinsider.com/illegal-streaming-is-dominating-online-piracy-2016-8> (accessed on March 25, 2018).
8. Brynjolfsson, E.; Hu, Y.J.; and Simester, D. Goodbye Pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, 57, 8 (June 2011), 1373–1386.
9. Cameron, A.C., and Trivedi, P.K. *Regression Analysis of Count Data*, 2nd ed. Econometric Society Monograph no. 53. Cambridge: Cambridge University Press, 1998.
10. Carr, D. Giving viewers what they want. *New York Times*, February 24, 2013.
11. Chang, M.R.; Kauffman, R.J.; and Kwon, Y. Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, 63 (2014), 67–80.
12. Chellappa, R., and Shivendu, S. Managing piracy: Pricing and sampling strategies for digital experience goods in vertically segmented markets. *Information Systems Research*, 16, 4 (December 2005), 400–417.
13. Clemons, E.K.; Gao, G.; and Hitt, L. When online reviews meet hyper differentiation: A study of the craft beer industry. *Journal of Management Information Systems*, 23, 2 (Fall 2006), 149–171.
14. Clemons, E.K.; Gu, B.; and Spitler, R. Hyper-differentiation strategies: Delivering value, retaining profits. In R.H. Sprague Jr. (ed.), *Proceedings of the Thirty-Sixth Hawaii International Conference on System Science*. Los Alamitos, CA: IEEE Computer Society Press, 2003. 225b.
15. Clemons, E.K.; Spitler, R.; Gu, B.; and Markopoulos, P. Information, hyperdifferentiation, and delight: The value of being different. In S. Bradley and R. Austin (eds.), *The Broadband Explosion: Leading Thinkers on the Promise of a Truly Interactive World*. Boston: Harvard Business School Press, 2005, pp. 137–164.
16. Cragg, J.G. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 39, 5 (September 1971), 829–844.
17. David, F.N., and Johnson, N.L. Statistical treatment of censored data I: Fundamental formulae. *Biometrika*, 41, 1 (June 1954), 228–240.

18. De Matos, M.; Ferreira, P.; Smith, M.D.; and Telang, R. Culling the herd: Using real-world randomized experiments to measure social bias with known costly goods. *Management Science*, 62, 9 (February 2016), 2563–2580.
19. Dehejia, R.H., and Wahba, S. Propensity score-matching methods for non-experimental causal studies. *Review of Economics and Statistics*, 84, 1 (February 2002), 151–161.
20. Demski, J. *Information Analysis*. Reading, MA: Addison-Wesley, 1980.
21. Dey, D.; Lahiri, A.; and Liu, D. Consumer learning and time-locked trials of software products. *Journal of Management Information Systems*, 30, 2 (Fall 2013), 239–268.
22. Dimoka, A.; Hong, Y.; and Pavlou, P.A. On product uncertainty in online markets: Theory and evidence. *MIS Quarterly*, 36, 2 (June 2012), 395–426.
23. Ding, P.; VanderWeele, T.; and Robins, J. Instrumental variables as bias amplifiers with general outcome and confounding. *Biometrika*, 104, 2 (June 2017), 291–302.
24. Dunn, J. The rise of music streaming services hasn't killed music piracy. *Business Insider* (April 2017). <http://www.businessinsider.sg/music-piracy-streaming-chart-2017-4/> (accessed on March 25, 2018).
25. Dupont, W.D. Multiple Poisson regression. Chapter 9 in *Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data*. Cambridge: Cambridge University Press, 2002, pp. 295–318.
26. Efron, B. Censored data and the bootstrap. *Journal of American Statistical Association*, 76, 374 (June 1981), 312–319.
27. Efron, B., and Tibshirani, R. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1993.
28. Faugère, C., and Kumar, G.T. Designing free software samples: A game theoretic approach. *Information Technology and Management*, 8, 4 (September 2006), 263–278.
29. Freedman, A.M. Use of free product samples wins new favor as sales tool. *Wall Street Journal*, August 28, 1986.
30. Gemici, S.; Rojewski, J.W.; and In-Heok, L. Use of propensity score matching for training research with observational data. *International Journal of Training Research*, 10, 3 (June 2012), 219–232.
31. Ghose, A. Internet exchanges for used goods: An empirical analysis of trade patterns and adverse selection. *MIS Quarterly*, 33, 2 (June 2009), 163–291.
32. Gilovich, T., and Medvec, V.H. The experience of regret: What, when, and why. *Psychology Review*, 102, 2 (April 1995), 379–395.
33. Greene, W.H. *Econometric Analysis*. London: Pearson, 2012.
34. Gurmu, S., and Trivedi, P.K. Excess zeros in count models for recreational trips. *Journal of Business and Economic Statistics*, 14, 4 (October 1996), 469–477.
35. Halbherr, D.; Stahl, F.; Koenigsberg, O.; and Lehmann, D.R. Choosing a digital content strategy: How much should be free? *International Journal of Research in Marketing*, 31, 2 (June 2014), 196–206.
36. Haubl, G., and Trifts, V. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Management Science*, 19, 1 (February 2000), 4–21.
37. Heiman, A.; McWilliams, B.; Shen, Z.; and Zilberman, D. Learning and forgetting: Modeling optimal product sampling over time. *Management Science*, 47, 4 (April 2001), 532–546.
38. Hilbe, J.M. *Negative Binomial Regression*, 2nd ed. Cambridge: Cambridge University Press, 2011.
39. Holloway, K. The reasons you can't stop binge watching. *Alternet.com*, December 30, 2015.
40. Hong, Y., and Pavlou, P. A. Product fit uncertainty in online markets: Nature, effects, and antecedents. *Information Systems Research*, 25, 2 (April 2014), 328–344.
41. Howard, J. Americans devote more than 10 hours a day to screen time, and growing. *CNN.com*, July 29, 2016.
42. Johnson, E.J.; Moe, W.W.; Fader, P.S.; Bellman, S.; and Lohse, G.L. On the depth and dynamics of online search behavior. *Management Science*, 50, 3 (March 2004), 299–308.
43. Jones, R., and Mendelson, H. Information goods vs. industrial goods: Cost structure and competition. *Management Science*, 57, 1 (January 2011), 164–176.

44. Kamins, M.A.; Folkes, V.S.; and Fedorikhin, A. Promotional bundles and consumers' price judgments: When the best things in life are not free. *Journal of Consumer Research*, 36, 4 (November 2009), 660–670.
45. Kastranekes, J. Netflix knows the exact episode of a TV show that gets you hooked. *TheVerge.com*, September 23, 2015.
46. Kward, Y.; Chen, J.; and Raghunathan, S. Online product reviews: Implications for retailers and competing manufacturers. *Information Systems Research*, 25, 1 (March 2014), 93–110.
47. Lafayette, J. Threat becomes profit center as TV leverages technology. *BroadcastingCable.com*, January 6, 2014.
48. Lawal, B. Zero-inflated count regression models with applications to some examples. *Quality and Quantity*, 46, 1 (January 2012), 19–38.
49. Li, T.; Kauffman, R.J.; Van Heck, E.; Vervest, P.; and Dellaert, B.G.C. Consumer informedness and firm information strategy. *Information Systems Research*, 25, 2 (May 2014), 345–363.
50. Li, X. Could deal promotion improve merchants' online reputations? The moderating role of prior reviews. *Journal of Management Information Systems*, 33, 1 (June 2016), 171–201.
51. Liebeskind, J., and Rumelt, R.P. Markets for experience goods with performance uncertainty. *RAND Journal of Economics*, 20, 4 (Winter 1989), 601–621.
52. Liebowitz, S.J., and Zentner, A. Clash of the titans: Does Internet use reduce television viewing? *Review of Economics and Statistics*, 94, 1 (February 2012), 234–245.
53. Lin, C.A. Modeling the gratification-seeking process of TV viewing. *Human Communication Research*, 20, 2 (December 1993), 224–244.
54. Markopoulos, P.M. Product information dissemination in Internet markets and markets for product information. Ph.D. diss., University of Pennsylvania, Philadelphia, 2004.
55. Markopoulos, P.M., and Clemons, E.K. Reducing buyers' uncertainty about taste-related product attributes. *Journal of Management Information Systems*, 30, 2 (Fall 2013), 269–299.
56. Matt, C., and Hess, T. Product fit uncertainty and its effects on vendor choice: An experimental study. *Electronic Markets*, 26, 1 (February 2016), 83–93.
57. McAlister, L., and Pessemier, E. Variety seeking behavior: an interdisciplinary review. *Journal of Consumer Research*, 9, 3 (December 1982), 311–322.
58. McGuinness, D.; Gendall, P.; and Mathew, S. The effect of product sampling on product trial, purchase and conversion. *International Journal of Advertising*, 11, 1 (January 1992), 83–92.
59. Mehta, N.; Rajiv, S.; and Srinivasan, K. Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Science*, 22, 1 (Winter 2003), 58–84.
60. Mordor Intelligence. Global video on demand market by revenue model, platform, applications, industry verticals, geography and vendors: market shares, forecasts and trends (2015 -2020). *PR Newswire* (September 2015). <https://www.prnewswire.com/news-releases/global-video-on-demand-market-by-revenue-model-platform-applications-industry-verticals-geography-and-vendors—market-shares-forecasts-and-trends-2015—2020-300148230.html> (accessed on March 25, 2018).
61. Moretti, E. Social learning and peer effects in consumption: evidence from movie sales. *Review of Economic Studies*, 78, 1 (January 2011), 356–393.
62. Nathanson, J. The economics of a hit TV show. *Priceonomics.com*, October 17, 2013.
63. Nelson, P. Information and consumer behavior. *Journal of Political Economy*, 78, 2 (March–April 1970), 311–329.
64. *Newswire*. Netflix declares binge watching is the new normal: Study finds 73% of TV streamers feel good about it. December 13, 2013.
65. Niculescu, M., and Wu, D. J. Economics of free under perpetual licensing implications for the software industry. *Information Systems Research*, 25, 1, (March 2014), 173–199.
66. New York Times. The best and worst moments of the 2017 Emmys. *New York Times*, September 18, 2017.

67. O'Kane, S. Spotify unveils Touch Preview, a beautiful new music discovery tool. *TheVerge.com*, January 22, 2015.
68. Oestreicher-Singer, G., and Zalmanson, L. Content or community? A digital business strategy for content providers in the social age. *MIS Quarterly*, 37, 2 (June 2013), 591–616.
69. Pavlou, P.; Liang, H.; and Xue, Y. Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective. *MIS Quarterly*, 31, 1 (March 2007), 105–136.
70. Pinsker, J. The psychology behind Costco's free samples. *Atlantic*, October 2014, 1–6.
71. Pirracchio, R.; Resche-Rigon, M.; and Chevret S. Evaluation of the propensity score methods for estimating marginal odds ratios in case of small sample size. *BMC Medical Research Methodology*, 12, 1 (May 2012), 70–79.
72. Rao, A.R., and Monroe, K.B. Causes and consequences of price premiums. *Journal of Business*, 69, 4 (October 1996), 511–535.
73. Rode, A. Literature review: Non-unitary models of the household (theory and evidence). Working paper, University of California, Santa Barbara, 2011.
74. Rosenbaum, P., and Rubin, D. Constructing a control group using multivariate matched sampling methods that incorporate the propensity. *American Statistician*, 39, 1 (February 1985), 33–38.
75. Rosenbaum, P., and Rubin, D. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 1 (April 1983), 41–55.
76. Rubin, A.M. Television uses and gratifications: The interactions of viewing patterns and motivations. *Journal of Broadcasting*, 27, 1 (January 1983), 37–51.
77. Rubin, D. Matching to remove bias in observational studies. *Biometrics*, 29, 1 (March 1973), 159–183.
78. Russell, C.A.; Norman, A.T.; and Heckler, S.E. The consumption of television programming: Development and validation of the connectedness scale. *Journal of Consumer Research*, 31, 1 (June 2004), 150–161.
79. Shapiro, C., and Varian, H.R. *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business School Press, 1999.
80. Vuong, Q. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57 2 (March 1989), 307–333.