2013

Do Household Cable TV Viewing Patterns Demonstrate Efficiency and Concentration?

Rae CHANG
Singapore Management University, mrchang@smu.edu.sg

Pulak GHOSH
Indian Institute of Management

Gwangjae JUNG
Singapore Management University, gwangjaejung@smu.edu.sg

Robert J. KAUFFMAN
Singapore Management University, rkauffman@smu.edu.sg

Peiran ZHANG
Singapore Management University, przhang@smu.edu.sg

Do Household Cable TV Viewing Patterns Demonstrate Efficiency and Concentration?

Ray Chang[a], Pulak Ghosh[b], Gwangjae Jung[a], Robert J. Kauffman[a], Peiran Zhang[a,c]
[a]Singapore Mgmt. Univ.; [b]Indian Institute of Mgmt., Bangalore; [c]Fudan Univ.

Last revised: May 18, 2013

Abstract

Understanding consumer behavior for cable TV services consumption creates the basis for greater value for the stakeholders involved: consumers through higher viewership satisfaction, and providers through higher revenues per user. This research explores a new large data set on cable TV services subscriptions and viewing at the household level of analysis. We construct household viewership preference clusters, and then use econometric methods to assess the relative efficiency and concentration of channel viewership patterns. We estimate a system of limited dependent variable models for the different measures: one with a beta distribution for proportional dependent variables; and the other with a quasi-likelihood function-based regression that exploits the asymptotic requirements of the model for larger cuts of the data but does not employ a fully-specified distribution for the dependent variables. Our findings suggest that households' cable TV viewing behavior is affected by their channel subscription, genre preference, and available time to watch TV. Taking those factors into consideration will help service providers to understand viewer characteristics better and redesign their program offerings.

Keywords: Cable TV, consumers, data analytics, households, preferences, viewership

1. Introduction

In media industries such as cable TV services, understanding viewer preferences and predicting consumer behavior are precursors of profitability for the provider. Due to problems with household-level data collection and the diverse nature of consumer preferences, it has been hard for services providers to obtain meaningful information on what kinds of content are demanded and to what extent consumers want to watch it. Today though, two-way set-top boxes that deliver cable TV services to households have made it possible for providers to collect nearly complete viewership information: micro-level data on household viewership patterns can be tracked continuously, including information on what channel is being watched at a given time, and on the continuous stream of clicks that are made via the remote control handset. New methods, especially data analytics for business, consumer and social insights, have become available for use.

Cable TV services are *subscription-based information goods*, so understanding household subscription choices and viewership patterns will provide useful information as a basis for refining product and service designs. Prior research has focused on pricing ads (Wilbur 2008), and on customer retention and services churn (Niculescu et al. 2012). We explore household TV viewing behavior based on a *subscribe-and-view process*, which is typical in understanding consumer behavior in this context (Crawford and Yurukoglu 2012). Two problems are evident. One is to determine what viewership patterns and behavior households exhibit. Another is to identify how viewing behavior is shaped by the channels to which households subscribe, and their available viewing time and content preferences. We ask: To what extent do cable TV service subscribers make use of the full spectrum of contents of their channel subscriptions? Do we observe efficient and concentrated channel viewing patterns? How useful are patterns of household channel preferences, bundle choices, viewing time and household demographics in explaining the observed outcomes? The results we will report suggest that cable TV services providers may need to rethink how they design and offer television bundles and channel promotion campaigns.

2. Research Process, Models and Data

We will specify cable TV viewing models at the household level, considering factors such as channel subscriptions, viewing times, and observed channel viewing preferences. We consider two different dependent variables. One captures household-level TV channel viewing efficiency relative to the channel bundle selected, and the other assesses the extent of concentration of TV channels viewed. Our models estimate the marginal effects of different factors on the observed household-level outcomes. We next discuss our dependent, main effects, and control variables.

Dependent variables. To reflect different aspects of channel viewing behavior at the household level, we propose two metrics. One is *ViewingEfficiency*. It measures the proportion of available channels a household watches at least thirty minutes a month, to capture whether all subscribed services are necessary. The other is *ViewingConcentration*, which gauges the distribution of channel viewing time, reflects the extent of households' diversity seeking behavior. These variables are important to measure business performance in the TV business.

Main effects variables. Several factors affect viewing behavior. The number of *Subscribed Channels* captures the viewing constraint imposed on a household due to its channel bundle choice. The total *ViewingTime* measures how much time the members of a household spend watching TV during a month, and determines the viewing efficiency and concentration. Our field study analysis suggested the presence of a significant relationship between the differences in household preferences and their observed viewing behavior. For this, we applied the *k*-means algorithm with Euclidean distance similarity (Landau et al. 2011) on a *genre level clustering analysis*. Prior research has suggested classifying programming content into multiple genres to better capture common features within hundreds and thousands TV programs (Creeber et al. 2001). We segment TV channels into eight genres based on program content: *Movies, Children's, Drama, Documentaries, Lifestyle*, *Sports, News* and *Music*. For each observation, we calculated household viewing time spent on each genre, and divided by total household viewing time to yield a normalized eight-dimensional vector of values to prepare for clustering feature set. Dummy variables are used to indicate household *PreferenceClusters* after clustering.

Control variables. Control variables are appropriate: *#Rooms*, a proxy for the number of viewers in a household; and *SubscriberAge*, which may influence subscription choice and viewing time. Other demographics information including ethnic, gender or profession might also affect observed viewing behavior, but since our observation unit has been limited to household, which usually contains multiple individuals, more individual level demographic variables included in models will not generate meaningful implications.

Estimation models and methods. We propose two separate but related models with the two different dependent variables:

$$\begin{aligned} \mathit{ViewingEfficiency} = {} & f\,(\mathit{SubscribedChannels}, \mathit{ViewingTime}, \mathit{PreferenceClusters}, \\ & \mathit{\#Rooms}, \mathit{SubscriberAge}) + \varepsilon \qquad (1) \end{aligned}$$

$$\mathit{ViewingConcentration} = g\,(\mathit{SubscribedChannels}, \mathit{ViewingTime}, \mathit{PreferenceClusters},$$

$$\mathit{\#Rooms,\ SubscriberAge}) + \xi \tag{2}$$

The main effects and control variables' coefficients can be estimated for each household, with the error terms ε and ξ determined across the same set of households for the same time periods. The dependent variables' values are proportions (0% to 100%), and their variances are different. Their variance approaches 0 as the mean approaches 0 or 1 (Kieschnick and McCullough 2003). So limited dependent variable estimation models are appropriate.

We initially estimated the models in two different ways. One approach involves a fully specified distribution assumption for the dependent variables, the *beta distribution*, which is well known for estimating models with dependent variables that represent proportions. Our second estimation involves a *quasi-likelihood function-based approach* to obtain the model's parameters, $L_i(\boldsymbol{\beta})$. It only considers the moments of the distribution of the dependent variable, but the distribution is not fully specified. See the Estimation Models Appendix for additional details.

3. Data and Model Diagnostics

Data and *k*-means analysis. We randomly sampled data for 10,000 cable TV-subscribing households from a larger dataset with several hundred thousand Singaporean households for one month in 2011. The viewing data we collected cover thirty-second time-stamped intervals, with the household set-top box as the unit of analysis. After controlling for the dwelling type of the households, our final data set contained 4,720 records. Using randomly-sampled data from the larger data not only supports the effectiveness of data analyses for the whole population, but also provides a way to check for robustness. For *PreferenceClusters*, we applied *k*-means clustering analysis on the identified channel genre viewing times for the households. We determined that nine clusters are appropriate: eight clusters with preferences for each of the eight genres; and one cluster with mixed preferences for all genres.

Model diagnostics. See Table 1 for descriptive statistics of the data.

Table 1. Descriptive Statistics

VARIABLES		N	MEAN	STD. ERR	MIN	MAX
ViewingEfficiency		4,728	0.195	0.112	0.017	0.674
ViewingConcentration		4,728	0.565	0.151	0.008	0.970
SubscribedChannels		4,728	64.388	18.032	23	145
ViewingTime (hours)		4,728	120.556	103.063	1.150	798.833
SubscriberAge	< 20	43	12.581	2.771	1	19
	20 - 40	2,221	32.590	4.403	20	39
	40 - 60	2,013	48.082	5.628	40	59
	> 60	445	80.243	21.672	60	111
#Rooms		4,728	4.166	0.772	1	5
PreferenceClusters (with *Movies* as the base case category)	*Children's*	4,728	0.256	0.436	0	1
	Drama		0.178	0.383	0	1
	Documentary		0.118	0.322	0	1
	Lifestyle		0.051	0.219	0	1
	News		0.077	0.267	0	1
	Sports		0.042	0.200	0	1
	Music		0.019	0.135	0	1
	MixedGenre		0.127	0.333	0	1

We checked pair-wise correlations between the variables; most are lower than 0.30, except *ViewingTime* and *ViewingEfficiency* at 0.61, and *ViewingTime* and *ViewingConcentration* at 0.46. We also checked to for multicollinearity via variance inflation factors. The highest value is 2.22

for the *Children's* viewing cluster, so multicollinearity is not a problem (Kennedy 1998).

We currently are evaluating whether endogeneity is an issue, based on whether *ViewingTime* and the dependent variables exhibit simultaneity. We considered the following possible instrumental variables for *ViewingTime*:

- *#ChildrensChannels* measures how many children's channels a household subscribes to. This is a proxy for whether children are viewers in the household. A report published by Nielson (2009) shows that children two to five years old spend more than 32 hours a week on average in front of a TV screen. Children are important TV viewers and often represent most of a household's TV viewing time.
- *#AddOnChannels* measures how many add-on channels a household subscribes to. These channels are different from others included in basic bundles. Customers are not required by cable TV providers to subscribe to add-on channels. Instead, they can freely choose what they want to watch, but they will have to pay, and this may results in their spending more viewing.

We performed a Hausman test with the null hypothesis that our estimators for the *Viewing-Efficiency* and *ViewingConcentration* models are consistent, but could not reject it for either model. We are continuing to explore the empirical modeling issues here, and will report updated results at CSWIM 2013.

4. Results and Discussion

Table 2 shows our results for the *ViewingEfficiency* and *ViewingConcentration* models. *SubscribedChannels* is negative and significant in the *ViewingEfficiency* model. This negative relationship indicates that the number of channels viewed does not increase proportionally with the number of subscribed channels. Instead, households apparently focus on a limited number of channels with programs that match their interests, rather than use the entirety of their subscribed bundles. Some households may have low *ViewingEfficiency* too. This may be unrelated to whether they subscribe to fewer channels or discontinue their service accounts later.

Also, *ViewingTime* positively influences the two dependent variables, *ViewingEfficiency* and *ViewingConcentration*, in our models. This implies that viewers who spend more time watching TV have a higher propensity to watch a more diverse set of channels, whereas investing more time on their favorite channels than the others. From a business perspective, letting viewers explore more channels makes it more likely that they will upgrade their existing subscriptions. Cable TV service providers should encourage customers to watch more different channels when they are expected to have more available time, say, during weekends or holidays. Free channels and promotions can be designed based on these findings.

The empirical results further show how the program genre preferences affect viewing behavior. One interesting finding is that the households that belong to the *Documentaries* and *Lifestyle* preference clusters seem to have higher channel viewing efficiencies than those in the *Sports* and *News* preference clusters. Households exhibiting focused viewing of *Children's* channels have more concentrated viewing behavior, which indicates the important role children play in TV viewing. The different effects of preferences offer a new perspective on channel bundle design. Bundles with more channel genres will encourage customers to explore more channels.

Table 2. Estimation Results

Models / Variables		*ViewingEfficiency* Model: Fully-Specified Beta Distribution for Dep. Var. Coef. (Std. Err.)		*ViewingEfficiency* Model: Not Fully-Specified Distribution for Dep. Var. Coef. (Std. Err.)		*ViewingConcentration* Model: Fully-Specified Beta Distribution for Dep. Var Coef. (Std. Err.)		*ViewingConcentration* Model: Not Fully-Specified Distribution for Dep. Var. Coef. (Std. Err.)	
Constant		-1.497***	(.061)	-1.179***	(.064)	0.466***	(.064)	0.402***	(.062)
SubscribedChannels		-0.006***	(.000)	-0.006***	(.000)	-0.001	(.000)	0.000	(.000)
ViewingTime		0.395***	(.008)	0.398***	(.010)	0.302***	(.009)	0.304***	(.009)
SubscriberAge		0.048***	(.012)	0.013	(.013)	-0.009	(.013)	-0.002	(.012)
#Rooms		0.029***	(.010)	-0.013	(.011)	-0.040***	(.011)	-0.024**	(.011)
PreferenceClusters Dummy Variables									
2	*Children's*	0.072***	(.028)	0.031	(.030)	0.103***	(.029)	0.070**	(.027)
3	*Drama*	0.119***	(.029)	0.061**	(.030)	0.019	(.030)	-0.014	(.028)
4	*Documentary*	0.229***	(.033)	0.141***	(.038)	-0.056*	(.034)	-0.026	(.034)
5	*Lifestyle*	0.178***	(.041)	0.149***	(.044)	0.050	(.042)	0.025	(.040)
6	*News*	0.011	(.037)	0.017	(.043)	0.044	(.037)	-0.031	(.035)
7	*Sports*	0.063	(.046)	-0.057	(.050)	0.116**	(.047)	0.036	(.049)
8	*Music*	0.072	(.064)	-0.072	(.073)	-0.039	(.063)	0.057	(.073)
9	*MixedGenre*	0.444***	(.030)	0.463***	(.033)	-0.126***	(.032)	-0.152***	(.030)
Model R^2		42.6%		39.8%		21.2%		25.1%	
AIC		8.84		10.56		10.17		9.87	
BIC		92.81		94.54		94.15		93.85	
Observations		4,720		4,720		4,720		4,720	

Note: Each model has a limited dependent variable estimated at the household level. They were estimated in two ways for comparison: (1) with a beta distribution for the dependent variable, and (2) with a quasi-likelihood-based function for the independent variables, and without a fully-specified distribution for the dependent variable. Signif.: *** $p < .01$, ** $p < .05$, * $p < .10$. The *Movies* cluster (Cluster 1) is the base case for the *PreferenceClusters* dummy variables.

5. Conclusion

The main purpose of this research was to identify TV viewing patterns in terms of efficiency and concentration. We used the data on household TV viewing patterns, and implemented two complementary types of estimation models. Our results show that subscribed channels, viewing time, and viewer preferences affect the efficiency and concentration of viewing patterns. They also suggest that more personalized marketing strategies may be on more effective promotion and channel bundle redesign.

To further develop this research, we are exploring how to implement more detailed measurements for the viewing patterns we observed. We plan to expand our data sets to include time-based viewing patterns and find out how TV viewing patterns change by time. We are also exploring in greater depth how subscription constraints may affect viewing behavior, and how efficiently customers use the channel bundles. We are further assessing the link between household viewing patterns and household cable TV bundle subscription changes too.

References

Crawford, G.S. and Yurukoglu, A. "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review* (102:2), 2012, pp. 643-685.

Creeber, G., Miller, T., and Tulloch, J. (eds.). *The Television Genre Book*. British Film Institute, London, UK, 2001.

Kennedy, P. 1998. *A Guide to Econometrics*, 4th ed. MIT Press, Cambridge, MA, 205–217.

Kieschnick, R., and McCullough, B.D. “Regression Analysis of Variates Observed on (0, 1): Percentages, Proportions and Fractions,” *Statistical Modelling* (3:3), 2003, pp. 193-213.

Landau, S., Leese, M., and Stahl, D. *Cluster Analysis*, Wiley, New York, NY, 2011.

Niculescu, M., Shin, H., and Whang, S. “Underlying Consumer Heterogeneity in Markets for Subscription-Based IT Services with Network Effects,” *Information Systems Research* (23:4), 2012, pp. 1322-1341.

Nielson. "TV Viewing among Kids at an Eight-Year High." October 26, 2009. Available at www.nielsen.com/us/en/newswire/2009/tv-viewing-among-kids-at-an-eight-year-high.html.

Papke, L., and Wooldridge, J. “Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates,” *Journal of Applied Econometrics* (11:6), 1996, pp. 619–632.

Wilbur, K. C. “A Two-Sided, Empirical Model of Television Advertising and Viewing Markets,” *Marketing Science* (27:3), 2008, pp. 356-378.

Appendix. Estimation Models

For the *beta distribution estimation model*, we assume a two-parameter distribution $Beta(p, q)$, with $f(y) = 1 / Beta(p, q)\, y^{p-1} (1 - y)^{q-1}$, and a $0 \leq y \leq 1$ dependent variable (Kieschnick and McCullough 2003). With this distribution, the conditional expectation of the dependent variable is given by $E(y_i|\mathbf{x_i}) = u_i = h(\eta_i) = 1 / (1 + \exp(-\eta_i)) = 1 / (1 + \exp(-\mathbf{x_i}'\boldsymbol{\beta}))$, where $\eta_i = g(u_i) = \ln(u_i / (1 - u_i)) = \mathbf{x_i}'\boldsymbol{\beta}$. By relating this to the two parameters of the beta distribution, p and q, we can obtain $q(\mathbf{x_i}') = p \exp(-\mathbf{x_i}'\boldsymbol{\beta})$. Next, let $\mu = p / (p + q)$ and $\phi = p + q$, with $p = \mu\phi$ and $q = (1 - \mu)\,\phi$. Substituting these into the definition of the beta distribution $f(y)$ gives the conditional distribution of the beta-distributed variable that represents the dependent variables for efficiency and concentration of household channel consumption in each regression. Then we can apply maximum likelihood estimation to obtain coefficient estimates. For the *quasi-likelihood function-based estimation model*, Papke and Wooldridge (1996) proposed a log-likelihood model, $L_i(\boldsymbol{\beta}) = y_i \ln [G(\mathbf{x_i}, \boldsymbol{\beta})] + (1 - y_i) \ln [1 - G(\mathbf{x_i}, \boldsymbol{\beta})]$ for continuous proportions with the logistic function, $G(\mathbf{x_i}, \boldsymbol{\beta})$ and $0 < G(\mathbf{x_i}, \boldsymbol{\beta}) < 1$. This is what we implement.

Acknowledgments. This research was supported by the Singapore National Research Foundation, under the International Research Centre @ Singapore Funding Initiative, administered by the Interactive Digital Media Program Office (IDMPO). Rob Kauffman also thanks the Lee Kuan Yew Faculty Fellowship for Research Excellence for generous support.