

Predictors of Customer's Subscriptions to Movie and Sport Packages

Yingge Qu
Mississippi State University

William W. Hill
Mississippi State University

Subscription-based television services providers, which operate in a highly competitive industry with high saturation, are finding it exceedingly difficult to grow their customer base. The ability to identify customers with a higher possibility of subscribing to the premium plans, is undoubtedly critical to increasing future sales. In that vein, we seek to identify predictors to premium packages selection behavior by jointly modeling customer subscriptions of premium movie, premium sport and basic sport packages. We show that, the customer subscriptions of different TV packages can be highly correlated. Moreover, our model profiles likely subscribers to premium movie, premium sport and basic sport packages using both subscriber demographic and lifestyle information.

INTRODUCTION

The subscription-based television service operates in a highly competitive business environment. In recent years, with the emergence of streaming TV service from the Internet, the competition has become even more severe. The television service market has transformed from Cable TV owning nearly the entire viewing market to multiple platform providers (satellite, Internet, etc.) of multi-channel subscriptions options available to each household (The Internet & Television Association, 2017). Facing such roaring competition, the television service providers are finding it exceedingly difficult to grow their customer base. Instead, they are eager for effective strategies to help them increase welfare from their existing customers. To achieve this substantive goal, service providers must understand customer subscription behavior so that they can design better marketing strategies to incite customer's subscriptions. A key area of profitability in subscription-based television services comes from the premium movie or premium sports packages. More than 40 percent of TV subscribers spending is attributed to sport programs, and revenues are expected to exceed \$20 billion by 2020 (James, 2016). According to Statista (2018), the subscription revenue of HBO, the oldest movie and TV series program, increased from 4.23 billion in 2013 to 5.5 billion U.S. dollars in 2017.

The goal of this paper is to identify the customers with a higher likelihood of subscribing to a specific TV package, especially premium offerings. Using the dataset from a major television service provider in U.S. market, we aim to address the following questions: how do subscriber demographics and lifestyle information shape TV subscriptions? Are customer subscription behaviors, e.g. the subscription of different TV packages, correlated? How do movie package subscribers differ from sport subscribers, and

how do premium package subscribers differ from basic subscribers? We recruit a multivariate probit model to capture the customer subscriptions to the three TV packages of premium movie, premium sport and basic sport. We found that the correlation of customer subscriptions between different TV packages are statistically significant. We also determined that our model can simultaneously estimate the probabilities of subscriptions to premium movie, premium sport, and basic sport packages using the demographic and lifestyle information that can be obtained easily from external market research companies.

LITERATURE REVIEW

Our study falls in the research stream of pay TV subscription market which typically focuses on three areas: the comparison between bundle vs. á la carte service, the competition in the market, and the exploration of customer's subscriptions behaviors. In U.S. market, majority of the TV services are provided in bundle. Past research suggests the bundling services benefit the firm, because firms can strategically design the bundle to reduce the heterogeneity in customer preference (Crawford, 2008), induce the price discrimination (McAfee et al., 1989; Bakos and Brynjolfsson, 1999) and extract the consumer surplus (Crawford and Cullen, 2007). The debate has long existed on the issue of whether or not customers are better by purchasing preferred channels singly (e.g. á la carte service) or in a bundle. For example, the Federal Communication Commission (FCC) arrived at opposing results on whether or not á la carte would drive customer's spending on TV service (FCC, 2004; FCC, 2006). By empirically analyzing the bundling effects in the TV subscription market, Crawford et al. addressed that the cost change of providing TV service under á la carte is the key to determining whether or not customers can be beneficial (Crawford and Cullen, 2007; Crawford and Yurukoglu, 2012).

When looking at the competition in the TV subscription market, studies mainly focus on the two major players: cable service and direct broadcast satellite (DBS). Findings typically center on what factors influence customer's choice between the two service providers and consequences of the competition. Wise and Duwadi (2005) addressed that both customer's choice between the two providers can be related to firm attributes, customer demographic, price changes in the basic cable service, and switching costs. With more competition in the market, customers can gain welfare from both services because the penetration of DBS both elevates the quality (Goolsbee and Petrin, 2004) and reduces the price of the cable service (Savage and Wirth, 2005).

Researchers explore customer's TV subscriptions behaviors from two aspects. One is the overall demand in the market. For example, Campmajó (2007) addressed that the level of competition in the market and the service contract features (such as whether high quality or premium program is included in the contract) determines the penetration of pay TV platform. Karikari et. al (2003) empirically identified that, in U.S. market, the penetration of DBS service can be influenced by the regulation of basic cable services, competition of local exchange telephone carriers and the upgrades of cable providers. Uri (2005 & 2006) found that, not only price, but also the overall market size, service features, quality factor influences the demand of both DBS and TV cable services. The other research stream is the customer-level subscription behaviors. LaRose and Atkin (1988) showed that, customer satisfaction, demographics and service cost affect the customer's intentions to disconnecting a cable service. Burez and Van den Poel (2007) developed an analytical CRM model to help elevate the firm's profits through reducing the customer churn. With the emergence of online streaming, more recent studies focus on modeling how consumers choose between paid subscription and online streaming (Prince and Greenstein, 2017).

Our research focuses on exploring the customer subscription behavior in the pay TV market but takes an unconventional approach compared to existing studies. Specifically, we take a closer view of customer-level subscriptions and evaluate the predictors of customer interest in subscribing to specific TV packages (sport, premium, and basic packages). To the best of our knowledge, our research is the first to quantify the potential drivers of customer subscriptions between multiple TV packages. Wang et. al (2005) presented a survey study with a similar approach in online paid subscription. In their study, they found that the consumers' willingness to pay for online content depends on the importance of the content,

quality of the online service, and consumer’s usage rate. Our study is different from Wang’s study in both the study context and the empirical method of analyzation. We use the customer actual subscription data to understand their choices of specific TV packages.

DATA DESCRIPTION

The research sample (n=100,000) used includes customer information from a leading subscription TV company. Specifically, the dataset consists of customer subscription TV package data, and corresponding demographic and lifestyle information for each of the 100,000 subscribing customers. Following an assessment of the data file, three customer subscription TV packages were identified—Premium movie, Premium sport, and Basic sport. These subscription TV packages categories are coded and defined as follows.

- premium movie subscribers (premium movie package = 1); purchasers of the premium movie package
- premium sport subscribers (premium sport package =1); purchasers of the premium-sport-package
- basic sport subscribers (basic sport package = 1); purchasers of the basic sport package

These three dichotomous variables will be used as dependent variables to demonstrate the customer TV subscription behaviors.

The analysis examines subscriber demographic and lifestyle information as predictor variables. The demographic variables assessed were subscriber gender, marital status, age, income, child presence, and household age range. The lifestyle “interest” variables examined were music, gardening, hunting, fitness, Internet, and home video, etc. A detailed explanation of both the demographic and lifestyle variables is shown in Table 1. Note that the age, age range, and income variables are continuous, while the remaining predictor variables are binary.

MODEL SETUP

A multivariate probit model is utilized for both testing the correlation between the customer subscriptions of the three TV packages (e.g. the three dependent variables) and estimating the empirical significance of the predictor variables (e.g. the subscriber demographic and lifestyle information). The multivariate probit model is well-known for quantifying the correlated dichotomous dependent variables (Ashford and Sowden, 1970; Amemiya, 1974; Song and Lee, 2005). In our case, the customer subscriptions of the three TV packages are likely to be correlated, thus, we require one framework to jointly model the three dependent variables: premium movie, premium sport and basic sport packages.

$$\begin{aligned}
 Y_1^* &= \beta_1' X_1 + \varepsilon_1, & Y_1 &= 1(Y_1^* > 0) \\
 Y_2^* &= \beta_2' X_2 + \varepsilon_2, & Y_2 &= 1(Y_2^* > 0) \\
 Y_3^* &= \beta_3' X_3 + \varepsilon_3, & Y_3 &= 1(Y_3^* > 0)
 \end{aligned}
 \quad
 \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \sim N_3 \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}
 \tag{1}$$

where:

$Y_1, Y_2,$ and Y_3 represent the three customer subscriptions TV packages (e.g. three dependent variables).

$\varepsilon_1, \varepsilon_2$ and ε_3 are the random terms that capture the unobserved effects. We allow the random terms to be correlated with each other to quantify the correlation of customer subscriptions between the three TV packages.

ρ is the tetrachoric Correlation for measuring the correlation between each pairs of the binary dependent variables. Consequently, the ρ_{12} indicates the correlation between Y_1 and Y_2 ; ρ_{13} indicates the correlation between Y_1 and Y_3 , and ρ_{23} indicates the correlation between Y_2 and Y_3 .

X_1, X_2, X_3 are the predictor variables in the model, which include both the subscriber demographic and lifestyle variables. We rely on the maximum likelihood method for model estimation.

TABLE 1
DESCRIPTION OF THE CUSTOMER'S DEMOGRAPHIC AND LIFESTYLE VARIABLES

Demographic Variables	Characteristics	Explanation
Age	Continuous	The TV subscriber's age
Age Range	Continuous	The age range (oldest-youngest) of the subscriber's household
Income	Continuous	The subscriber's income
Gender	Binary, Male=1	The subscriber's gender
Marital	Binary, Marriage=1	The subscriber's marital status
Children Presence	Binary, With Children=1	Whether the subscriber has children at home
Lifestyle Variables	Characteristics	Explanation
Music	Binary (interest=1)	Whether interest in stereos/records/tapes/CDs
Internet	Binary (Use =1)	whether uses internet at home
Video Game	Binary (Use =1)	Whether uses/owning computer video games at home
Gardening	Binary (Interest=1)	Whether interest in gardening or plants
Hunting	Binary (Interest=1)	Whether interest in hunting/shooting/fishing
Travel	Binary (interest=1)	Whether Interest in Travel
auto interest	Binary (interest=1)	Whether interest in automotive related activities and magazines
Health	Binary (interest=1)	Whether interest in fitness/exercise
Upscale	Binary (Has=1)	Whether has credit cards issued by upscale retail store

RESULTS AND DISCUSSION

SAS 9.4 software was used to perform all the statistical analysis including the multivariate probit model. We used descriptive statistics of the major variables in the model to illustrate the model findings.

Dependent Variables

As we noted previously, there are three dichotomous dependent variables describing customer subscription behavior in this study—premium movie, premium sport, and basic sport packages. In order to determine if a reasonable quantity of observations (e.g. subscribers) exist in all three customer subscription TV packages, the frequency and percentage of both subscribers and non-subscribers were assessed, and are presented in Table 2. In our data, we observed 44838 customers (44.8%) subscribing to the premium movie package, 11061 customers (11.1%) purchasing the premium sport package, and 5911 customers (5.9%) adopting the basic sport package (Table 2). Therefore, a reasonable quantity of observations (e.g. subscribers) was evident in all three TV customer subscription TV packages.

TABLE 2
FREQUENCY TABLE OF SUBSCRIBERS ON MAJOR TV CHANNELS

	Frequency	Percent
Premium Movie Package		
0	55162	55.16%
1	44838	44.84%
Premium Sport Package		
0	88939	88.94%
1	11061	11.06%
Basic Sport Package		
0	94089	94.09%
1	5911	5.91%

Next, we create a cross-tabulation table between each pairs of the dependent variables to demonstrate the potential correlations between the customer subscription TV packages (Table 3). A Chi-square test was used to detect the significance of the correlation. Shown in the result, we observed 10148 customers who subscribe to both premium packages (e.g. premium movie and premium sport). This group of customers represents about 22.6% of the total premium movie subscribers (e.g. 10148/44838=22.6%) and 91.7% of the total premium sport subscribers (e.g. 10148/11061=91.74%). The large amount of multi-packages subscribers indicates that customer subscriptions between premium movie and premium sport packages are highly correlated. Additionally, the p-value of <0.0001 in the chi-square test further confirms the significance of the correlation between the premium movie and premium sport subscription.

We were also interested in exploring if the customer premium package subscription is correlated with basic package subscription. We found that there are 2883 customers who subscribe to both premium sport and basic sport packages. This number represents 26.1% (e.g. 2883/11061=26.1%) of the total premium sport subscribers and 48.8 % (e.g. 2883/5911=48.8%) of the total basic sport subscribers (Table 3). We also observed 4029 customers who subscribed to both the premium movie and the basic sport packages (Table 3). Both the frequency of multi-packages subscribers and the Chi-square tests indicate that basic sport subscription is significantly correlated with both the premium sport and the premium movie subscriptions.

As shown in the descriptive statistics, the three dependent variables are significantly correlated. Therefore, an independent model (e.g. separately modeled each dependent variable) can result in a biased estimation result. For this reason, a joint modeling approach was required to simultaneously evaluate the customer subscription on the three TV packages (e.g. premium movie, premium sport and basic sport).

Predictor Variables: Continuous

The subscriber demographic information includes both the continuous variables of age, age range, and income, and the binary variables of gender, marital status, and child presence. We present the descriptive statistics for the continuous demographic variables in Table 4. We observed that, on average, the TV subscribers in our study are about 43 years old and have an annual income of 34,753.08 US dollars (Table 4). Their household has an average age range of 7.5 years which means that the age difference between the oldest and youngest member in the household is on average 7.5 years.

In Table 4, you find the results of the two-sample t-test analysis between the subscribers (e.g. Y=1) and non-subscribers (e.g. Y=0) for each of the three TV packages. Since we are interested in the relationship between the customer TV subscription behavior and the predictor variables, the t-test serves as a preliminary model-free evidence of such relationship. We find that, for both premium movie and premium sport packages, the means of all three predictor variables are significantly different between the subscribers and non-subscribers (Table 4). Further, we identify a positive mean-difference for both age range and income and a negative mean-difference for age (Table 4). The result suggests that all three predictor variables are correlated with the subscriptions of the two premium TV packages. For both premium movie and premium sport packages, the subscribers tend to have younger age, wider household age range, and higher income than the non-subscribers. When we examined the basic sport subscription, necessitated by the result of the t-test, we found the significant predictor variables are more likely to be both income and age.

**TABLE 3
CROSS TABULATION TABLE BETWEEN DEPENDENT VARIABLES**

Premium Movie Package	Premium Sport Package		χ^2 test
	0	1	P-value
0	54249	913	<0.0001
1	34690	10148	
Premium Sport Package	Basic Sport Package		χ^2 test
	0	1	P-value
0	85911	3028	<0.0001
1	8178	2883	
Premium Movie Package	Basic Sport Package		χ^2 test
	0	1	P-value
0	53280	1882	<0.0001
1	40809	4029	

TABLE 4
DESCRIPTIVE STATISTICS FOR CONTINUOUS DEMOGRAPHIC VARIABLES

Predictor Variable	Total Sample (N=10,000)		t-test					
			Premium Movie Package		Premium Sport Package		Basic Sport Package	
	Mean	Std Dev	Mean Diff	P-Value	Mean Diff	P-Value	Mean Diff	P-Value
Age range	7.542	12.227	0.259	0.0009	0.325	0.0083	-0.070	0.6690
Income	34,753.08	19,125.79	947.80	<.0001	1,915.20	<.0001	2,679.50	<0.0001
Age	43.174	14.138	-1.876	<.0001	-2.101	<.0001	-2.535	<0.0001

*Mean Diff = $\bar{X}|(Y = 1) - \bar{X}|(Y = 0)$

Predictor Variables: Dichotomous

The dichotomous predictor variables include the demographic variables of gender, marital status, and child presence, along with the list of lifestyle variables. Since both the dependent variables and predictor variables are dichotomous, we use the cross-tabulation table and the associated chi-square test to show the potential relationship between them (Table 5). The frequency of the demographic variables shows that 58.7% of the subscribers are female, 60.0% of subscribers are unmarried, and 55.7% have children in the household (Table 5). All three demographic variables are relatively balanced in sample size between the two binary groups for model estimation.

TABLE 5
CROSS-TABULATION TABLE BETWEEN BINARY
PREDICTOR AND DEPENDENT VARIABLES

Predictor Variables	Total Sample (N=100,000)		Premium Movie Package		Premium Sport Package		Basic Sport Package		χ^2 test	
	0	1	χ^2 test	0	1	0	1			
	Demographic Variables									
Gender	0	58662	32838	25824	<0.0001	52952	5710	55888	2774	<0.0001
	1	41338	22324	19014		35987	5351	38201	3137	
Marital	0	60089	33477	26612	<0.0001	53704	6385	56770	3319	<0.0001
	1	39911	21685	18226		35235	4676	37319	2592	
Children Presence	0	44313	25317	18996	<0.0001	39460	4853	41565	2748	0.0005
	1	55687	29845	25842		49479	6208	52524	3163	
Lifestyle Variables										
Music	0	79931	44222	35709	0.0383	71105	8826	75291	4640	0.0046
	1	20069	10940	9129		17834	2235	18798	1271	
Internet	0	73328	40929	32399	<0.0001	65451	7877	69298	4030	<0.0001
	1	26672	14233	12439		23488	3184	24791	1881	
Video Game	0	95441	52952	42489	<0.0001	85002	10439	89832	5609	0.0366
	1	4559	2210	2349		3937	622	4257	302	
Gardening	0	84367	46345	38022	0.0007	74897	9470	79332	5035	0.0759
	1	15633	8817	6816		14042	1591	14757	876	
Hunting	0	88426	48909	39517	0.009	78677	9749	83150	5276	0.0394
	1	11574	6253	5321		10262	1312	10939	635	
Travel	0	73264	40548	32716	0.054	65273	7991	69066	4198	<0.0001
	1	26736	14614	12122		23666	3070	25023	1713	
auto interest	0	88377	48836	39541	0.0899	78580	9797	83126	5251	0.258
	1	11623	6326	5297		10359	1264	10963	660	
Health	0	72327	40006	32321	0.1214	64277	8050	68129	4198	0.0206
	1	27673	15156	12517		24662	3011	25960	1713	
Upscale	0	85020	46896	38124	0.9612	75694	9326	80131	4889	<0.0001
	1	14980	8266	6714		16245	1735	13958	1022	

The lifestyle variables are relatively unbalanced in sample size between the two binary groups (Table 5). Of all the lifestyle variables, travel and internet have the highest number of interest observations. We observed about 26.7% of the customers are interested in travel and 26.7% of them use the Internet at home (Table 5). The least preferable lifestyle is video game use and hunting, which subscribers show interest at 4.6% and 11.6% levels, respectively (Table 5). The unbalanced samples of the lifestyle variables can result in less observations in the cross-tabulation table with dependent variables. For example, within the group of customers who are interested in video game ($n=4559$), we observed 2349, 622, and 302 customers who subscribe to premium movie, premium sports and basic sports, respectively. Although the lifestyle sample is less balanced, we still have at least hundreds of observations in each cell of the cross-tabulation table for each lifestyle variable, which is adequate for model estimation purpose.

Next, we examined the potential relationship between the predictor and dependent variables. The chi-square tests suggest that all the demographic variables have significant correlation with the subscriptions of premium movie and basic sport packages (Table 5). For premium sport package, both gender and marital status variables are significantly correlated with the customer subscription, but the “child presence” variable does not show a strong correlation.

The relationship between the subscriber lifestyle and their TV subscriptions is more diversified. For the premium movie package, the variables of Internet, video game, gardening, and hunting all show a strong correlation ($p\text{-value} < 0.01$ in Table 5), the variables of music, travel and auto interest show a moderate correlation ($p\text{-value} < 0.1$ in Table 5), and the variables of health and upscale show no significant correlation ($p\text{-value} > 0.1$ in Table 5). The subscription of premium sport package is significantly correlated with internet, video game, gardening, travel and upscale, but not correlated with music, hunting, auto interest and health. Additionally, the basic sport subscription appears to have moderate to strong correlation with all lifestyle variables except for auto interest (Table 5). This result suggests that the customer lifestyle can play different roles when s/he chooses different TV packages. For example, the music interest variable shows to have a greater influence on premium movie subscription than premium sport. Some lifestyle interests, such as Internet and video game use, may impact the customer subscription choice behavior for all three TV packages.

In summary, the descriptive analyses provide a preliminary evidence on the two substantive aspects. First, the customer’s choices of TV packages are highly correlated. Second, both the customers demographic and lifestyle can significantly influence their TV subscription behaviors. Moreover, we observed reasonable distribution of the continuous predictor variables, and adequate observations in the categories of binary predictor variables for model parameter estimation. Our result also demonstrates that, the effects of predictor variables, especially lifestyle, on the customer subscriptions of TV packages, can be diversified. To more accurately quantify the impacts of both demographic and lifestyle variables on the customer subscription behaviors, we need the joint modeling approach to evaluate both predictor and dependent variables in one framework.

Multivariate Probit Model Estimation Results

The multivariate probit model estimation result is presented in Table 6. To make the comparison between the three TV packages more straightforward, we only show the parameter estimation sign and significance for all predictor variables. The complete estimation outcomes including the estimates, standard errors, t-value and p-value are given in Appendix A.

We can evaluate the parameter estimation outcomes from two perspectives. First, we compare the customer subscriptions between the premium and basic packages. Since both premium movie and premium sport are premium TV packages, we can view both of them as a category of premium packages, and study the difference of customer subscription between the premium packages (e.g. both premium movie and premium sport) and the basic packages (e.g. basic sport). Next, we can compare the customer subscriptions between the movie and sport packages. Similarly, we combine the premium sport and basic sport into the category of sport packages and evaluate the difference between the movie (e.g. premium movie package) and the sport (e.g. both premium and basic sport packages) subscriptions.

From the parameter estimation result (Table 6), we observed four different types of predictor variables. The first type of predictor variables, which include gender, marital, age, income, Internet and gardening, appear to have a consistent effect on the subscription choice behavior for all three TV packages. The parameter estimations of gender, marital, income and internet variables are significantly positive and those of the age and gardening variables are significantly negative in the results of all three TV packages. This suggests that, if a customer is male (gender =1), married (marital=1), has higher income and uses the Internet at home, then he is more likely to subscribe all three TV packages. On the other hand, if the customer is older (age increase) and/or likes gardening, s/he may be less interested in subscribing to any TV packages.

TABLE 6
MULTIVARIATE PROBIT MODEL PARAMETER ESTIMATION

	Premium Movie Package (Y1)	Premium Sport Package (Y2)	Basic Sport Package (Y3)
Parameter	Parameter Sign and Significance		
Intercept	N.S	- (***)	- (***)
Demographic Predictor Variables			
¹ Gender	+ (***)	+ (***)	+ (***)
¹ Marital	+ (*)	+ (***)	+ (***)
³ Children Presence	+ (***)	- (***)	- (***)
¹ Age	- (***)	- (***)	- (***)
² Age Range	+ (***)	+ (***)	N.S
¹ Income	+ (**)	+ (***)	+ (***)
Lifestyle Predictor Variables			
Music	- (*)	N.S	+ (*)
¹ Internet	+ (***)	+ (***)	+ (***)
² Video Game	+ (***)	+ (***)	N.S
¹ Gardening	- (***)	- (***)	- (***)
² Hunting	+ (***)	+ (**)	- (***)
³ Travel	N.S	+ (**)	+ (**)
³ Auto Interest	N.S	- (**)	- (***)
Health	N.S	- (*)	N.S
³ Upscale	N.S	+ (**)	+ (***)
Tetrachoric Correlation			
	Estimate	SE	P-value
ρ_{12}	0.692	0.0046	<.0001
ρ_{13}	0.294	0.0077	<.0001
ρ_{23}	0.573	0.0066	<.0001

*--significant at 10% level; **--significant at 5% level; ***--significant at 1% level

1: The predictor variable has consistent impacts on all three TV subscription

2: The predictor variable has different impacts between Premium and Basic TV subscription

3: The predictor variable has different impacts between Movie and Sport subscription.

The second type of predictor variables, which include age range, video games, and hunting, show different impacts between premium and basic packages subscription. We found that all three variables have a significant positive influence on the subscription of premium packages. This means that, if the subscriber has a wider age range household structure, uses video game at home, and likes hunting, s/he may prefer subscribing the premium packages. On the other hand, for basic package subscription, both age range and video game use are not significant, while hunting is negatively significant.

Child presence, travel, auto interest and upscale belong to the third group of predictor variables whose influences on subscription are different between movie and sport packages. As suggested by the estimation results (Table 6), when a subscriber has children, s/he is more likely to subscribe movie package (e.g. the estimation is positive) but less likely to choose sport package (e.g. the estimation is negative). When we look at the three lifestyle variables of travel, auto interest, and upscale, we found that, none are significantly related to the movie package, but all are significantly related to sport packages. Specifically, a customer will be more likely to subscribe the sport packages if s/he is interested in travel and/or upscale stores, but less interested in sport packages if s/he likes automotive activities.

There also exists predictor variables, such as music and health, which show distinct estimation results with each of the three TV packages. For example, the estimation of music is negative in the premium movie, is non-significant for the premium sport, and is positive in basic sport. The estimation of health is non-significant in both premium movie and basic sport, but negative in premium sport.

Another important estimation is the Tetrachoric Correlation, which is used to determine the correlation between the three dependent variables. The estimation results show that the correlation is 0.693 (e.g. P-value < 0.0001) between the two premium packages (e.g. premium movie and premium sport), and is 0.573 (e.g. P-value < 0.0001) between the two sport packages (e.g. premium sport and basic sport). This result confirms that the customer's subscription behaviors are highly correlated between TV packages within the same category (e.g. category of premium packages or category of sport packages). We also noticed that the correlation between premium movie and basic sport is statistically significant (e.g. $\rho_{13}=0.294$, P-value<0.001). This suggests that, the customer's subscription behaviors can also be highly correlated between different categories of TV packages (e.g. movie vs. sport or premium vs. basic).

To evaluate the model prediction power, we computed the cross-tabulation table between observed and predicted subscriptions for all three TV packages (Table 7). From the result, we found that, the model can achieve more than 87% overall hit-rate for all three TV packages. Specifically, the model can simultaneously identify 40615 out of 44838 (90.6%) subscribers for premium movie package, 10021 out of 11601 (86.4%) subscribers for premium sport package and 4981 out of 5911 (84.3%) subscribers for basic sport package. In the non-subscriber group, the model can reach at least 84% accuracy for each of the three TV packages. This hit-rate outcomes are much better than the 50 percent "by chance" criterion, thus confirming the model's prediction power.

TABLE 7
CROSS-TABULATION TABLE BETWEEN OBSERVED AND PREDICTED TV SUBSCRIPTIONS

Predicted Subscription		Observed Subscription					
		Premium Movie Package		Premium Sport Package		Basic Sport Package	
		0	1	0	1	0	1
Count	0	48363	4223	79296	1580	82684	930
	1	6799	40615	9643	10021	11405	4981
Percentage	0	48.4%	4.2%	79.3%	1.6%	82.7%	0.9%
	1	6.8%	40.6%	9.6%	10.0%	11.4%	5.0%

CONCLUSION

Operating in a highly competitive market with high saturation, the subscription-based television service providers need to understand the customer's subscription behaviors such that they can design better marketing actions to incite the customer's purchase intention. To help the service provider achieve this substantive goal, our study attempted to profile likely subscribers to premium movie, premium sport and basic sport packages using both subscriber demographic and lifestyle information.

A multivariate probit model was proposed for jointly evaluating the customer subscriptions of the three TV packages: premium movie, premium sport, and basic sport. We found a significant correlation between the different TV packages. The high correlation necessitated simultaneously investigating the customer subscriptions of multiple TV packages rather than evaluating each TV package separately. The model estimation results show that, the customer subscription behaviors, e.g. the subscriptions of TV packages, are strongly correlated with the subscriber's demographic and lifestyle information. The influence of the demographic and lifestyle variables on the customer TV subscriptions choices is diverse. For example, customers who like hunting prefer premium TV packages, but not the basic packages. Customers with children are more likely to subscribe to movie packages than sports packages. The service provider should pay attention to the varied demographic and lifestyle influences to avoid making the wrong target decisions.

An extension of this study would consider examination of the effects of other types of predictor variables on the customer TV subscription behaviors, such as price, promotion, and advertising influences. Price is always a determinant factor shaping a customer purchase decision. Therefore, it would make sense to add covariates in this area to improve both the model application and prediction power. A further extension of this study would be to expand the dataset from cross-sectional to panel data, such that we could also evaluate the dynamic features of the customer subscription behaviors.

REFERENCES

- Amemiya, T. (1974). Bivariate probit analysis: Minimum chi-square methods. *Journal of the American Statistical Association*, 69(348), 940-944.
- Ashford, J. R., & Sowden, R. R. (1970). Multi-variate probit analysis. *Biometrics*, 535-546.
- Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2), 277-288.
- Bakos, Y., & Brynjolfsson, E. (1999). Bundling information goods: Pricing, profits, and efficiency. *Management science*, 45(12), 1613-1630.
- Campmajó, L. D. (2009). Competition between TV Platforms. *Documentos de trabajo (XREAP)*, (9), 1.
- Crawford, G. S., & Cullen, J. (2007). Bundling, product choice and efficiency: Should cable television networks be offered á la carte? *Information Economics and Policy*, 19(2007), 379-404
- Crawford, G. S. (2008). The discriminatory incentives to bundle in the cable television industry. *Quantitative Marketing and economics*, 6(1), 41-78.
- Crawford, G. S., & Yurukoglu, A. (2012). The welfare effects of bundling in multichannel television markets. *American Economic Review*, 102(2), 643-85.
- Federal Communications Commission. (2004). Report on the packaging and sale of video programming services to the public. *Washington, DC*.
- Federal Communication Commission. (2006). Further report on the packaging and sale of video programming to the public. *Washington, DC*.
- Goolsbee, A., & Petrin, A. (2004). The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica*, 72(2), 351-381.
- James, Meg. (2016). The rise of sports TV costs and why your cable bill keeps going up. Retrieved from <http://www.latimes.com/business/hollywood/la-fi-ct-sports-channels-20161128-story.html>

- Karikari, J. A., Brown, S. M., & Abramowitz, A. D. (2003). Subscriptions for direct broadcast satellite and cable television in the US: an empirical analysis. *Information Economics and Policy*, 15(1), 1-15.
- LaRose, R., & Atkin, D. (1988). Satisfaction, demographic, and media environment predictors of cable subscription. *Journal of Broadcasting & Electronic Media*, 32(4), 403-413.
- McAfee, R. P., McMillan, J., & Whinston, M. D. (1989). Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics*, 104(2), 371-383.
- Prince, J., & Greenstein, S. (2017). Measuring Consumer Preferences for Video Content Provision via Cord-Cutting Behavior. *Journal of Economics & Management Strategy*, 26(2), 293-317.
- Savage, S. J., & Wirth, M. (2005). Price, programming and potential competition in US cable television markets. *Journal of Regulatory Economics*, 27(1), 25-46.
- Song, X. Y., & Lee, S. Y. (2005). A multivariate probit latent variable model for analyzing dichotomous responses. *Statistica Sinica*, 645-664.
- Statista. (2018). HBO's subscription revenue from 2011 to 2017 (in billion U.S. dollars). Retrieved from <https://www.statista.com/statistics/329295/hbo-subscription-revenue/>
- The Internet & Television Association. (2017). America's competitive TV and internet markets. Retrieved from <https://www.ncta.com/whats-new/americas-competitive-tv-and-internet-markets>
- Uri, N. D. (2005). Estimating the demand for cable service in the US in the presence of measurement error in the data. *Technological Forecasting and Social Change*, 72(5), 585-607.
- Uri, N. D. (2006). The market for subscription television service in the United States. *The Engineering Economist*, 51(3), 205-236.
- Wang, C. L., Zhang, Y., Ye, L. R., & Nguyen, D. D. (2005). Subscription to fee-based online services: What makes consumer pay for online content?. *Journal of Electronic Commerce Research*, 6(4), 304.
- Wise, A., & Duwadi, K. (2005). Competition between cable television and direct broadcast satellite – It is more complicated than you think. *FCC Media Bureau Staff Research Paper No. MB 2005-1; FCC International Bureau Working Paper No. 1B-3*.

APPENDIX A

MULTIVARIATE PROBIT MODEL PARAMETER ESTIMATION RESULT

Parameter Estimates				
Parameter	Estimate	Std. Error	t-Value	P-Value
Premium Movie Package				
Intercept	0.026	0.018	1.47	0.1425
Gender	0.054	0.008	6.65	<.0001
Marital	0.017	0.009	1.89	0.0586
Children Presence	0.032	0.009	3.65	0.0003
Age	-0.006	0.000	-18.16	<.0001
Age Range	0.002	0.000	4.91	<.0001
Income	0.568	0.222	2.55	0.0107
Music	-0.023	0.014	-1.66	0.0974
Internet	0.047	0.011	4.25	<.0001
Video Game	0.136	0.021	6.53	<.0001
Gardening	-0.071	0.014	-4.95	<.0001
Hunting	0.043	0.016	2.73	0.0062
Premium Sport Package				
Intercept	-1.121	0.023	-48.03	<.0001
Gender	0.167	0.011	15.68	<.0001
Marital	0.036	0.012	3.02	0.0025
Children Presence	-0.050	0.011	-4.39	<.0001
Age	-0.007	0.000	-15.69	<.0001
Age Range	0.002	0.000	3.91	<.0001
Income	1.713	0.290	5.9	<.0001
Internet	0.059	0.014	4.12	<.0001
Video Game	0.111	0.026	4.2	<.0001
Gardening	-0.092	0.019	-4.86	<.0001
Hunting	0.051	0.020	2.51	0.0121
Travel	0.031	0.014	2.2	0.0277
Auto Interest	-0.044	0.020	-2.22	0.0262
Health	-0.028	0.016	-1.75	0.0809
Upscale	0.035	0.015	2.32	0.0203
Basic Sport Package				
Intercept	-1.396	0.029	-48.9	<.0001
Gender	0.239	0.013	18.24	<.0001
Marital	0.050	0.015	3.45	0.0006
Children Presence	-0.109	0.014	-7.92	<.0001
Age	-0.008	0.001	-16.1	<.0001
Income	1.706	0.358	4.77	<.0001
Music	0.041	0.023	1.81	0.0701
Internet	0.128	0.017	7.44	<.0001

APPENDIX A (Continued)

MULTIVARIATE PROBIT MODEL PARAMETER ESTIMATION RESULT

Parameter Estimates				
Parameter	Estimate	Std. Error	t-Value	P-Value
Basic Sport Package				
Gardening	-0.069	0.023	-3.02	0.0025
Hunting	-0.082	0.025	-3.24	0.0012
Travel	0.036	0.017	2.11	0.0346
Auto Interest	-0.072	0.024	-2.98	0.0029
Upscale	0.073	0.018	4.01	<.0001
Tetrachoric Correlation				
ρ_{12}	0.692	0.005	150.12	<.0001
ρ_{13}	0.294	0.008	37.96	<.0001
ρ_{23}	0.573	0.007	87.33	<.0001