Matt Helbig
7 December 2018
ECON 176

Marketing Analysis - Uber

Task Description:

The goal of this project is to determine the viability of a fleet of self-driving cars for Uber. In order to determine viability, consumer preference related to self-driving cars will be measured by administering a survey. The survey will collect income and age, as well as ask three valuation questions and eight discrete choice questions. The three valuation questions will measure how the responder values price and ride duration, as well as measure interest in riding in a self-driving vehicle. The eight discrete choice questions will present the responder with an option of multiple ride types, in which the responder will choose between UberX, Uber Pool, UberX Auto (self-driving option), and UberPool Auto (self-driving option). There will be varying prices and ride durations for each of the questions presented to the survey responders. In addition, there will be one question in which the ride is free of charge. Multiple regressions will be conducted using the data collected from the survey responders. The focus of the regressions will primarily be on consumer willingness-to-pay for self-driving cars, but other coefficients worthy of note will be analyzed if they present themselves. The data will also be subsetted by consumer characteristics in order to identify potential differences between market segments. The valuation questions will be analyzed to assess the robustness of the discrete choice data, as ideally the values indicated by the consumers will match the results from the survey data. Finally, an action plan will be recommended to Uber based on the results and analysis of the regressions, with the focus being primarily on the plan for self-driving cars.

Summary Statistics:

VARIABLES	(1) mean	(2) sd	(3) min	(4) max
price	14.32	8.855	6.330	48.67
eta	20.25	12.90	7	57
income	32,333	37,684	0	160,000
age	22.09	1.636	20	27
questionb1	4.519	0.855	1	5
questionb2	4.370	0.909	1	5
questionb3	3.370	1.252	1	5

Our variables of interest are price (Measured in dollars), eta (Ride duration, measured in minutes), income (Measured in dollars), age (Measured in years), questionb1 (Response to the statement: "I take into consideration price when booking a ride with Uber."), questionb2 (Response to the statement: "I take into consideration ride duration when booking a ride with Uber."), and questionb3 (Response to the statement: "I would be interested in Uber creating rides with autonomous vehicles." Price has a mean of $14.32, and ranges from $6.33 to $48.67. ETA has a mean of 20.25 minutes and ranges from 7 minutes to 57 minutes. From these, an average price-per-minute of $0.71 can be measured. Income has a mean of $32,333 per year, ranging from $0 to $160,000 per year. Age has a mean of 22.09 years, and ranges from 20-27. While this is not an exceptionally large range, subsets can still be created, grouping ages 20-23 into the low age range and ages 24-27 into the high age range. Questionb1, questionb2, and questionb3 range from 1 to 5. Questionb1's mean of 4.519 indicates that many people selected 5, meaning they took price highly into consideration. Questionb2's mean of 4.37 indicates again that many people selected high numbers, meaning they took ride duration highly into consideration. Questionb3's mean of 3.37 means that on average people were fairly lukewarm about being interested in rides with autonomous vehicles. This gives an early indication into the viability of pursuing a fleet of self-driving cars, as already the interest seems like it might be minimal.

Main Regression:

VARIABLES	(1) With Income	(2) Without Income
Uber Pool	0.193	-0.329*
	(0.212)	(0.174)
UberX Auto	-0.268	-0.544***
	(0.173)	(0.129)
UberPool Auto	-0.0915	-0.562***
	(0.197)	(0.159)
Price	-0.149***	-0.143***
	(0.0218)	(0.0214)
Duration	-0.110***	-0.106***
	(0.0190)	(0.0187)
Inc*Pool	-1.64e-05***	
	(3.91e-06)	
Inc*UberX Auto	-7.77e-06**	
	(3.40e-06)	
Inc*UberPool Auto	-1.47e-05***	
	(3.85e-06)	
Observations	1,944	1,944

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regression formulas:

$U_T = 0.193UberPool – 0.268UberXAuto – 0.09UberPoolAuto – 0.149Price - 0.110Duration - 1.64e\text{-}05Inc*Pool – 7.77e\text{-}06Inc*UberXAuto – 1.47e\text{-}05Inc*UberPoolAuto + \varepsilon_T$

$U_T = -0.329UberPool – 0.544UberXAuto – 0.562UberPoolAuto – 0.143Price – 0.106Duration + \varepsilon_T$

In each regression formula, Uber Pool, UberX Auto, and UberPool Auto will be dummy variables that equal one when the responder selected that option from among the choices presented to them. There were four total options, but UberX has been normalized and omitted to avoid collinearity. Price and Duration represent the added utility to the consumer of a one-unit increase in that variable. Inc*Pool, Inc*UberX Auto, and

Inc*UberPool Auto are interacted variables for the income provided by the survey responders. Income has been interacted with the choice variables to avoid collinearity. For each of the regression formulas presented, the coefficients on the interacted income variables are small enough to the point of being zero, so the primary focus of analysis will be on the regression formulas where income was excluded. A coefficient of zero indicates that the variable had virtually no impact on the value of the independent variable. When the regression without income is examined, it is apparent that each coefficient is statistically significant, with Uber Pool significant at the 90% confidence level and the others significant at the 99% confidence level. The values of the coefficients are all measured in comparison to the base of UberX, which has been normalized and excluded from the model. The negative coefficients on the Uber alternatives indicate that the consumers derive less utility from Pool, Auto, and Pool Auto than they do from the UberX choice. The negative coefficients on price and duration indicate that increases in these variables lead to a decrease in utility for the consumer, which makes sense intuitively. The willingness-to-pay for each variable is the coefficient divided by the coefficient on price of -0.143, all times negative one. The willingness-to-pay for Uber Pool is (-)(-0.329/-0.143) = -2.3, for Auto is -3.8, for Pool Auto is -3.93, and for Duration is -0.74. All have negative coefficients when measured against UberX, but the coefficients on the Auto options (representing self-driving cars) have greater negative values than the coefficient on Uber Pool, indicating that consumers are even less willing to pay for a self-driving alternative. From this regression, it is indicated that consumers are less willing to pay for a solo ride in a self-driving vehicle than a carpool ride in a human-driven car.

Age-Subsetted Regression:

VARIABLES	(1) Low Age With Income	(2) Low Age Without Income	(3) Old Age With Income	(4) Old Age Without Income
Uber Pool	-0.0398	-0.310	1.466**	-0.386
	(0.228)	(0.191)	(0.733)	(0.448)
UberX Auto	-0.355*	-0.602***	0.508	-0.369
	(0.189)	(0.147)	(0.518)	(0.277)
UberPool Auto	-0.380*	-0.539***	1.058	-0.940**
	(0.217)	(0.176)	(0.702)	(0.403)
Price	-0.196***	-0.195***	-0.112***	-0.0878***
	(0.0313)	(0.0311)	(0.0358)	(0.0289)
Duration	-0.120***	-0.119***	-0.170***	-0.152***
	(0.0230)	(0.0229)	(0.0492)	(0.0458)
Inc*Pool	-1.08e-05**		-3.01e-05***	
	(4.96e-06)		(1.04e-05)	
Inc*UberX Auto	-9.93e-06**		-1.15e-05*	
	(5.01e-06)		(5.97e-06)	
Inc*UberPool Auto	-6.08e-06		-3.35e-05***	
	(4.85e-06)		(1.10e-05)	
Observations	1,620	1,620	324	324

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Again, since the interacted income coefficients are essentially zero for both the subsetted options, the focus will lie primarily on the models where income has been normalized.

Regression formulas:

Low, Without: $U_T = -0.310UberPool - 0.602UberXAuto - 0.539UberPoolAuto - 0.195Price - 0.119Duration + \varepsilon_T$

Old, Without: $U_T = -0.386UberPool - 0.369UberXAuto - 0.940UberPoolAuto - 0.0878Price - 0.152Duration + \varepsilon_T$

Here, the data has been subsetted by age, creating a low age group that ranges from 20 to 23 and a high age group that ranges from 24-27. This will allow for differences between the two market segments to be analyzed, as variations in behaviors and preferences will become more readily apparent. This will create the potential for different action plans for each segment. For the low age subset, the coefficients on UberX Auto, UberPool Auto, Price, and Duration are all significant at the 99% confidence level. The willingness-to-pay is -3.08 for UberX Auto, -2.76 for UberPool Auto, and -0.61 for Duration. These coefficients are still largely negative, but they are closer to positive than the coefficients were in the main regression that covered all age groups, indicating that people in this low age range are slightly more willing to try self-driving cars. Similarly, the willingness-to-pay for Duration was larger for this young age group, which means they are more willing to take rides that take a little bit longer. What is worthy of note is that the willingness-to-pay for the self-driving carpool was higher than the willingness-to-pay for the self-driving solo ride, which could mean that younger people want someone else in the vehicle when the car is driving itself. This could be a marker for a lack of trust in the self-driving technology, and is worth keeping an eye on.

For the high age group with income normalized, only UberPool Auto, Price, and Duration are statistically significant, with UberPool Auto significant at the 95% level and Price and Duration significant at the 99% level. The willingness-to-pay for UberPool Auto is -10.44, and for duration it is -1.68. This value for UberPool auto is much lower than some of the other willingness-to-pay we've seen from other coefficients, indicating that people in the high age range are much less willing to pay for a self-driving carpool option. The willingness-to-pay for duration is also roughly twice as low as it was for the young age group, and 1.6 times as low as it was in the main regression. This indicates that people in the high age range want to get places quicker, so ride duration is certainly a potential area of focus when marketing Uber's overall product to people in a higher age bracket. Finally, the coefficient on the price variable was much closer to zero than it was in the previous regressions, which is a sign that people in the higher age group derive less utility from a lower price relative to their younger counterparts. People in this higher age group likely value other factors outside of price that may not have been captured in the survey, such as ride comfort or friendliness of the driver.

Price Valuation Regression:

VARIABLES	(1) Agree, With Income	(2) Agree, Without Income	(3) Do Not Agree, With Income	(4) Do Not Agree, Without Income
Uber Pool	0.0974	-0.283	2.990	-1.125
	(0.216)	(0.178)	(1.913)	(0.799)
UberX Auto	-0.357**	-0.516***	1.542	-1.018*
	(0.178)	(0.133)	(1.282)	(0.547)
UberPool Auto	-0.241	-0.519***	3.269*	-1.499*
	(0.203)	(0.163)	(1.804)	(0.786)
Price	-0.157***	-0.155***	-0.415**	-0.0667
	(0.0241)	(0.0239)	(0.206)	(0.0478)
Duration	-0.116***	-0.114***	-0.139	-0.0173
	(0.0199)	(0.0198)	(0.138)	(0.0752)
Inc*Pool	-1.28e-05***		-0.000116**	
	(4.19e-06)		(4.68e-05)	
Inc*UberX Auto	-4.87e-06		-4.04e-05**	
	(3.72e-06)		(1.95e-05)	
Inc*UberPool Auto	-9.05e-06**		-0.000477	
	(4.04e-06)		(0.0606)	
Observations	1,836	1,836	108	108

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regression formulas (without income):

Agree, Without: $U_T = -0.283UberPool - 0.516UberXAuto - 0.519UberPoolAuto - 0.155Price - 0.114Duration + \varepsilon_T$

Do Not Agree, Without: $U_T = -1.125UberPool - 1.018UberXAuto - 1.499UberPoolAuto - 0.0667Price - 0.0173Duration + \varepsilon_T$

The two subsetted groups are those who agreed that they took price into consideration when making their decisions (indicated by answering either Somewhat agree or Strongly agree) and those who did not agree that they took price into consideration when making their decisions (indicated by answering Neither agree nor disagree, Somewhat disagree, or Strongly disagree). For those who agreed, the coefficients that are statistically significant are UberX Auto, UberPool Auto, Price, and Duration. The willingness-to-pay for UberX Auto is -3.33, for UberPool Auto is -3.35, and for Duration is -0.74. The willingness-to-pay for the self-driving options are in line with the previous responses, but in order to test the value of this valuation question, it will have to been seen if price has a higher absolute value on its coefficient than it did in the main regression. As people agreed that they took price into consideration, it would make sense empirically that price played a bigger part in the utility of the consumer. As it turns out, the price coefficient was roughly the same as it was in the main regression, mostly due to the fact that virtually all responders agreed that price played a role in their decision.

For those that did not agree that price played a role in their decision, the only coefficients with significance were those on UberX Auto and UberPool Auto. The coefficients were much more negative than in previous models, indicating that for those who didn't take price into consideration, the perceived utility of a self-driving car was much lower. This could indicate that for those who are not focused on price, the focus instead is on their distaste for self-driving automobiles.

Duration Valuation Regression:

	(1) Agree With Income	(2) Agree Without Income	(3) Do Not Agree With Income	(4) Do Not Agree Without Income
VARIABLES				
Uber Pool	0.210	-0.204	-0.223	-1.582**
	(0.222)	(0.182)	(0.932)	(0.629)
UberX Auto	-0.324*	-0.451***	0.795	-1.466***
	(0.184)	(0.135)	(0.882)	(0.477)
UberPool Auto	-0.186	-0.469***	0.816	-1.490**
	(0.209)	(0.166)	(0.930)	(0.581)
Price	-0.149***	-0.147***	-0.489***	-0.124**
	(0.0235)	(0.0233)	(0.166)	(0.0555)
Duration	-0.114***	-0.112***	-0.203*	-0.0505
	(0.0199)	(0.0198)	(0.109)	(0.0630)
Inc*Pool	-1.36e-05***		-0.000101***	
	(4.25e-06)		(3.59e-05)	
Inc*UberX Auto	-3.68e-06		-0.000312*	
	(3.69e-06)		(0.000179)	
Inc*UberPool Auto	-8.96e-06**		-0.000280**	
	(4.09e-06)		(0.000132)	
Observations	1,764	1,764	180	180

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regression formulas (without income):

Agree, Without: $U_T = -0.204UberPool - 0.451UberXAuto - 0.469UberPoolAuto - 0.147Price - 0.112Duration + \varepsilon_T$

Do Not Agree, Without: $U_T = -1.582UberPool - 1.466UberXAuto - 1.490UberPoolAuto - 0.124Price - 0.0505Duration + \varepsilon_T$

The two subsetted groups are those who agreed that they took duration into consideration when making their decisions (indicated by answering either Somewhat agree or Strongly agree) and those who did not agree that they took duration into consideration when making their decisions (indicated by answering Neither agree nor disagree, Somewhat disagree, or Strongly disagree). For those that agreed, the coefficients on UberX Auto, UberPool Auto, Price, and Duration were statistically significant. The willingness-to-pay for UberX Auto is -3.06, for UberPool Auto is -3.19, and for Duration is -0.76. These coefficients are all in line with the main regression, again because of the fact that a majority of the responders are present in this group that agreed that duration was taken into consideration.

For those that did not agree, UberPool, UberX Auto, UberPool Auto, and Price were all statistically significant. The variable most worth measuring, duration, was much lower than normal but was not statistically significant. The willingness-to-pay for UberPool is -12.75, for UberX Auto is -11.82, and for UberPool Auto is -12.01. The coefficient on price is roughly the same as in previous regressions, so the willingness-to-pay has primarily to do with the utility from the alternatives to UberX. Those who are not focused on duration derive a great deal more utility from UberX as compared to the alternatives. The coefficients for the three alternatives are roughly the same, so this isn't a case where the consumers derive less utility from solely the self-driving options.

Autonomous Interest Regression:

	(1) Agree With Income	(2) Agree Without Income	(3) Do Not Agree With Income	(4) Do Not Agree Without Income
VARIABLES				
Uber Pool	-0.170	-0.485*	0.558*	-0.307
	(0.304)	(0.263)	(0.327)	(0.254)
UberX Auto	0.189	0.0532	-0.907***	-1.223***
	(0.221)	(0.179)	(0.303)	(0.207)
UberPool Auto	0.434*	0.155	-0.905***	-1.441***
	(0.252)	(0.215)	(0.345)	(0.257)
Price	-0.105***	-0.102***	-0.247***	-0.232***
	(0.0230)	(0.0227)	(0.0458)	(0.0436)
Duration	-0.0877***	-0.0862***	-0.160***	-0.148***
	(0.0236)	(0.0234)	(0.0346)	(0.0331)
Inc*Pool	-1.19e-05*		-2.39e-05***	
	(6.10e-06)		(5.75e-06)	
Inc*UberX Auto	-4.42e-06		-7.83e-06	
	(4.25e-06)		(5.77e-06)	
Inc*UberPool Auto	-1.03e-05**		-1.46e-05**	
	(4.85e-06)		(6.47e-06)	
Observations	936	936	1,008	1,008

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regression formulas (without income):

Agree, Without: $U_T = -0.485UberPool + 0.0532UberXAuto + 0.155UberPoolAuto - 0.102Price - 0.0862Duration + \varepsilon_T$

Do Not Agree, Without: $U_T = -0.307UberPool - 1.223UberXAuto - 1.441UberPoolAuto - 0.232Price - 0.148Duration + \varepsilon_T$

The two subsetted groups are those who agreed that they would be interested in trying an autonomous vehicle (indicated by answering either Somewhat agree or Strongly agree) and those who did not agree that they would be interested in trying an autonomous vehicle (indicated by answering Neither agree nor disagree, Somewhat disagree, or Strongly disagree). For those who agreed, the coefficients on UberPool, Price, and Duration were statistically significant. The willingness-to-pay for UberPool was -4.76 and the willingness-to-pay for Duration was -0.85. The variables of interest here were UberX Auto and UberPool Auto, and while the two are positive for the first time in any of the regressions, the coefficients are not statistically significant.

For those that did not agree, UberX Auto, UberPool Auto, Price, and Duration were statistically significant. The willingness-to-pay for UberX Auto is -5.27, and for UberPool Auto is -6.21. These values are much more negative than the coefficients in the main regression. The conclusion from the data is that those who were willing to try self-driving vehicles did not provide statistically significant results, while those who were not willing had a much lower willingness-to-pay than the willingness-to-pay for the responders as a whole. Roughly the same amount of people were interested in trying self-driving cars as were not interested, so it seems as though there isn't a ton of clamor to try out the new technology.

Free Regression:

VARIABLES	(1) With Income	(2) Without Income
UberPool	-1.335**	-1.520***
	(0.566)	(0.417)
UberX Auto	-0.206	-0.827***
	(0.414)	(0.320)
UberPool Auto	-4.514**	-3.466***
	(1.823)	(1.016)
o.price	-	-
o.eta	-	-
Inc*Pool	-5.29e-06	
	(1.17e-05)	
Inc*X Auto	-2.52e-05*	
	(1.29e-05)	
Inc*UberPool Auto	1.83e-05	
	(1.95e-05)	
Observations	216	216

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Regression Formula (without income):

$$U_T = -1.520UberPool - 0.827UberXAuto - 3.466UberPoolAuto + \varepsilon_T$$

In this question, consumers were asked which option they would pick if given the opportunity for a free ride. Despite there being four choices overall, there were two options each for price and eta (Duration), so there exists a problem of perfect multicollinearity. Because of this, price and eta have been omitted from this regression model. The coefficients on UberPool, UberX Auto, and UberPool Auto are all statistically significant at the 99% confidence level, and we can see from the values of the coefficients that consumers derive less utility from sharing a carpool than an UberX Auto ride. Intuitively, this makes sense, as carpooling is often seen as a cost-saving option, which would not be necessary when given a free ride. Notably, the coefficient on UberPool Auto is much lower than the coefficient on UberPool, which indicates that people are less willing to try the self-driving option even when given a free opportunity to do so.

Action Plan:

Overall and under each segment, willingness-to-pay for self-driving cars was negative when compared to the baseline, UberX. The telling statistic is that in the main regression the willingness-to-pay for the self-driving solo ride was less than the willingness-to-pay for the UberX Pool with a human driver. This gives an indication of where the attitude towards autonomous technology is at that consumers would rather ride with other people than ride in a self-driving vehicle by themselves. At this point, there simply doesn't seem to be much interest in self-driving cars, and there appears to be a negative attitude towards the technology overall as compared to the standard UberX option. From our age subset regression, there is some potential evidence that the younger age group does not trust the self-driving technology, as they derived more value from riding with others in a self-driving car than riding alone in a self-driving car.

What became apparent during the multiple regressions was that UberX provided the most utility to consumers regardless of age or attitudes towards price and ride duration. This overwhelming favorability indicates that efforts should be focused on the UberX class of rides. Focus should be on methods to increase the overall quality of the UberX ride, and investments should be made in ways to decrease ride price and ride duration. Discount days could be implemented to improve consumer satisfaction, as consumers by and large derived a lower utility from an increase in price. From that, it only makes sense that they would derive a higher utility from an increase in price. Revamping the Uber app's map function in an attempt to travel on faster routes would also be an avenue worth exploring, as consumers universally derived lower utility from longer ride duration. As the average cost-per-minute was $0.71, decreasing the necessary time of the ride would be a relatively cost-effective way to boost the brand's favorability among consumers.

Going forward, the recommendation for Uber is to drop all plans to pursue self-driving technology for the foreseeable future. It may be wise to revisit the technology down the road when interest increases and consumer sentiment towards the technology improves, but right now it would not be wise to invest considerable resources into this venture.