Working Paper: Discrete Choice Analysis on Household Preferences

Kristopher C. Toll

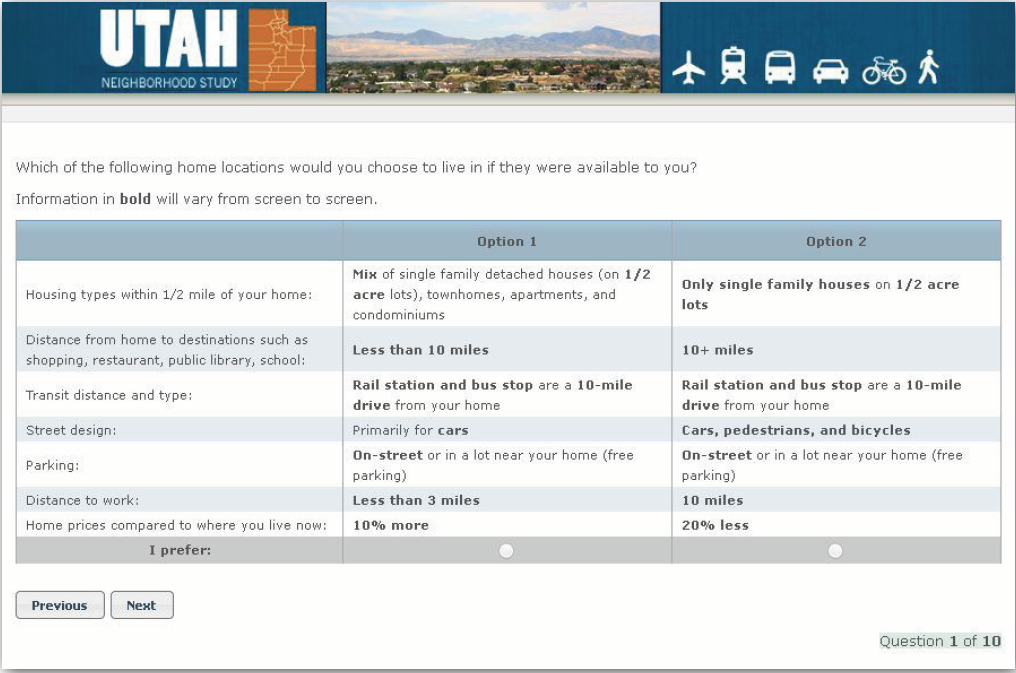
December 13, 2017

## Introduction

In 2012, the Wasatch Front Regional Council, the Mountainland Association of Governments, the Cache Metropolitan Planning Organization, the Dixie Metropolitan Planning Organization, the Utah Department of Transportation, and the Utah Transit Authority conducted a statewide survey on travel in Utah. The purpose of this study was to determine travel preferences, travel patterns and home location preferences. Utah as gone through much growth within the last decade. This has provided motivation for the aforementioned agencies to understand travel and growth patterns so as to better serve Utah Residents. This was done by administering several different surveys and asking Utah residents to keep a travel diary for one day. Data was collected on a household level, person level and trip level.

One of the surveys in the Utah Travel study asked resident to pick between two hypothetical homes. Each of these homes had differing attributes based off of type of neighborhood the home was located in, commute from destinations, distance from public transport, street design, parking, distance to work and home price compared to where the respondent currently lives. This stated preference survey was used to determine what kind of homes and locations Utah residents would prefer and how future development should be planned. An example of this question is located in figure 1.

Figure : Example of question from choice survey



One of the interesting attributes in the survey is about home prices. The attributes for this question describe the price of the other compared to where they live. These attributes are 20% less, 10% less, same, 10% more, and 20% more. If actual price data were obtained in this survey, it would be possible to extrapolate the marginal willingness to accept for each attribute.

## Method and Data Management

In order to determine marginal willingness to accept for each attribute, the same data for the Utah Travel Study will be used. Additionally, data from the Census Bureau will be used as a substitute for home prices. Specifically, the five year estimate of median home value for each census tract will be used. It will be adjusted to 2012 dollars. To make it comparable to the price attribute, the price difference will be calculated relative to the median value and substituted in for the home price attribute. For example, if the median value of the census track is $100,000 and if the attribute were 10% more, then -$10,000 will be put in place of the attribute 10% more. That figure will then be the comparative price difference.

It should be noted that this survey did include respondents who lived in mobile homes, multifamily homes, apartments and single family homes. Since the prices were calculated from a median value, only single family homes would be the most likely to have a comparable price to that value. As such, only single family homes will be considered in the model. Renters will not be used in the model as well. It is also speculated that homeowners how live in different neighborhoods will behave differently to prices and attributes. For this reason, separate models will be run on homes in rural areas and small towns, cities, and suburbs. This will limit the sample size to 287, 292, and 684 respectively for each model.

To estimate these parameters, a multi nominal logit model will be used. This model will determine the likelihood of choosing a given alternative based of off which attributes are present in the comparison of the two homes. Also, in an effort to limit the number of parameters and thus retain more degrees of freedom, each attribute will be refactored so that it is binary. This will also make it easier to interpret the results of the model. These new variable definitions are found in table 1 and 2. The variable chid is a used to keep track of which question is comparing which two alternatives.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Code: | Commute | Destinations | Homes | Streets | Transit | Parking on Driveway |
| 0 | Less than 5 miles | Less than 3 miles from shops, schools, doctors, ect. | (I need to redo this one) | For cars | Public transit is less than 5 miles away | Don’t park on drive way |
| 1 | More than 5 miles | More than 3 miles from shopes, schools, doctors, ect. | (I need to redo this one) | For cars, bikes and pedestrians | Public transit is more than 5 miles | Park on drive way |

|  |  |  |
| --- | --- | --- |
| Code: | Parking On-Street | Parking Off-Street |
| 0 | Don’t Park on street or in lot near home | Don’t Park off Street |
| 1 | Park on street or in lot near home | Park off street in a lot or garage |

## Results

Each coefficient from the model measures the likelihood of choosing an alternative given that an attribute is present in the comparison. For example, when commute changes from less than three miles to more than three miles, that alternative is less likely to be chosen by 0.41 or 41% for homeowner who live in the city. This particular estimate is significant at the 99% level.

For most homeowners, it seems that they have a preference for shorter commutes, closer distance to public transport, be closer to destinations such as shops, schools, et cetra and parking on the driveway. Homeowners also seem to have a preference to see their home value increase. For a $100,000 dollar increase, the likelihood of choosing that alternative will increase by 0.45 or 45% for home owners in the city.

### Model with City observations

## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)   
## 2:(intercept) -2.2140e-01 9.4724e-02 -2.3374 0.019421 \*   
## 2:commute -4.1033e-01 5.7768e-02 -7.1030 1.221e-12 \*\*\*  
## 2:destinations -4.3225e-01 5.7283e-02 -7.5460 4.485e-14 \*\*\*  
## 2:homes 4.2486e-01 6.4868e-02 6.5497 5.765e-11 \*\*\*  
## 2:streets 3.7690e-01 5.7792e-02 6.5217 6.953e-11 \*\*\*  
## 2:transit -1.8564e-01 5.7978e-02 -3.2019 0.001365 \*\*   
## 2:ParkingDriveway 1.1108e+00 6.4796e-02 17.1428 < 2.2e-16 \*\*\*  
## 2:ParkingOffStreet -8.4054e-02 8.4644e-02 -0.9930 0.320692   
## 2:NominalPrice 4.5198e-06 8.5065e-07 5.3134 1.076e-07 \*\*\*  
## 2:chid 1.3140e-08 5.2416e-06 0.0025 0.998000   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Log-Likelihood: -3564.8  
## McFadden R^2: 0.095817   
## Likelihood ratio test : chisq = 755.54 (p.value = < 2.22e-16)

### Model with Suburban Observations

## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)   
## 2:(intercept) -4.3529e-01 7.3797e-02 -5.8985 3.668e-09 \*\*\*  
## 2:commute -4.2272e-01 4.6049e-02 -9.1798 < 2.2e-16 \*\*\*  
## 2:destinations -3.5540e-01 4.5494e-02 -7.8121 5.551e-15 \*\*\*  
## 2:homes 7.4767e-01 5.1508e-02 14.5157 < 2.2e-16 \*\*\*  
## 2:streets 4.1995e-01 4.5895e-02 9.1502 < 2.2e-16 \*\*\*  
## 2:transit -1.7254e-01 4.6321e-02 -3.7248 0.0001954 \*\*\*  
## 2:ParkingDriveway 1.3490e+00 5.1641e-02 26.1224 < 2.2e-16 \*\*\*  
## 2:ParkingOffStreet -9.7793e-02 6.7099e-02 -1.4574 0.1449982   
## 2:NominalPrice 4.2205e-06 5.9039e-07 7.1487 8.762e-13 \*\*\*  
## 2:chid -6.0571e-06 4.1757e-06 -1.4505 0.1469067   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Log-Likelihood: -5712.8  
## McFadden R^2: 0.13315   
## Likelihood ratio test : chisq = 1754.9 (p.value = < 2.22e-16)

### Model with Small town/Rural Observations

## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)   
## 2:(intercept) -8.2937e-01 1.2515e-01 -6.6271 3.425e-11 \*\*\*  
## 2:commute -2.0943e-01 7.5390e-02 -2.7780 0.005469 \*\*   
## 2:destinations -3.4988e-02 7.4960e-02 -0.4668 0.640673   
## 2:homes 1.0021e+00 8.6145e-02 11.6323 < 2.2e-16 \*\*\*  
## 2:streets 4.1318e-01 7.5798e-02 5.4511 5.007e-08 \*\*\*  
## 2:transit -2.1895e-01 7.6050e-02 -2.8790 0.003989 \*\*   
## 2:ParkingDriveway 1.2512e+00 8.5776e-02 14.5873 < 2.2e-16 \*\*\*  
## 2:ParkingOffStreet 3.9022e-03 1.1011e-01 0.0354 0.971728   
## 2:NominalPrice 3.2150e-06 1.0634e-06 3.0233 0.002500 \*\*   
## 2:chid -5.4730e-06 7.0197e-06 -0.7797 0.435588   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Log-Likelihood: -2107.3  
## McFadden R^2: 0.12611   
## Likelihood ratio test : chisq = 608.22 (p.value = < 2.22e-16)

# Conclusions and notes

(I still have yet to examine marginal willingness to pay but all I should have to do is take the each coefficient and divide it by the Nominal Price coefficient)

(I made a mistake in refactoring the levels for the homes variable I missed 3 of them and I will have to go through and change that)

(I also need to do a goodness of fit test still and I have yet to figure that out for this particular model I have a manual, I just need to spend more time on it. However, the chisq test statistic is significant for all models so I imagine that they all work despite having g a low r^2 value)

(I do not think I needed three separate models as most of the preferences were the same for all and it may be better to have large sample size)

# References

Croissant, Y. (n.d.). *Estimation of multinomial logit models in R: The mlogit Packages.* Universite de la Reunion.

INC., R. S. (2013). *Utah Travel Study.* RSG INC.