Hannah Jun

02/12/2025

P2 Memo

Statement:

The following memo explains the usefulness and limitations of an alternative discrete choice models to predict household vehicle availability in comparison to the Boston Region's vehicle availability model. To deep dive into the code for each model please look https://github.com/urbanbyline/p2_hjun_2025.

Deep Dive:

The alternative developed vehicle availability model is based off inputs from the US National Housing Travel Survey (2017) and is a multinomial logistic regression model with a base category of sufficient. By contrast, the Boston Regional Travel Demand Model's vehicle availability model is based off inputs from the Massachusetts Travel Survey (2011). Keeping in mind the differences in data sources, the alternative model coefficients were selected and transformed to match as closely as possible with only two missing around intersection density and transit accessibility/highway accessibility as highlighted in red in Table 1. In addition, in the alternative model I added majority/minority renters in census block tract and home ownership/renting coefficients to better understand how homeownership influences vehicle availability as highlighted in green in Table 1.

In comparing the two models, there are some overarching similarities between the TDM vehicle availability model and the alternative vehicle availability model. In terms of similarities, the presence of children and workers decreases the utility of zero vehicles in comparison to sufficient vehicle households (Table 2 and Table 3). This tracks logically as there are more workers and children in a household would often prompt more need for a car to transport to school and places of employment. Another key similarity is around low-income households' utility for zero cars. What this suggests is as a household is low income their likelihood for zero car increases. In both models, low-income households have a positive coefficient for zero vehicle availability over sufficient car availability (Table 2 and Table 3). The TDM model does not provide all the coefficients in comparison to our multi-choice model so we cannot compare meaningfully if insufficient vehicle coefficients align for workers and children in household.

The alternative model illustrates the relationship between homeownership and vehicle availability (Table 2). Using two different coefficients, we can see the impact of household level home ownership ("homeownershipRENT") versus the impact of % home ownership on the census tract level ("HTHTNRNTminority_rent"). We see statistically significant positive relationships between renting on the household level and households choosing insufficient vehicles and zero vehicles. On the flip side, we see that if a minority of households rent in each census tract, the utility of insufficient vehicles and zero vehicles is negative and statistically significant. This relationship suggests that as there are more homeowners in a given census tract, more households will have as many cars as drivers in a household. We can hypothesize that homeownership highlights more financial independence to buy and maintain a vehicle which is why we see this correlation.

The alternative choice model provides several signals of its predictive power of vehicle availability. High level we see a McFadden R squared of 0.3035 (Table 3) and an AIC of a 37730.76 (Table 4). This contrasts with the model assignment's R squared of 0.272 and an AIC of 39049.78. While we can't directly compare the results, we can see that the alternative choice model has a higher R squared and a lower AIC which means that the model can explain more variation in the choice outcome and has better model fit even with more parameters/complexity. It is important to point out that while the accuracy of the confusion matrix 90%, the information rate is still very high at 0.8835 indicating that the model even without information would be accurate 88% of the time. The model overall predicts sufficient and zero well fairly well but has low success predicting insufficiency. This may be due to the fact that zero and sufficient cars are a clearer binary whereas personal situations define if households require as many cars as drivers.

Type of Model	**Boston Regional Travel Demand Model's vehicle availability**	**US National Housing Travel Survey**
	Workers in HH	Worker Count
	Children in HH	Number of Children
	CBD Dense Urban	Density
	Intersection Density	
	Low Income HH	Income Low/High
	Suburban or Rural	Urban or Rural
	Transit Accessibility/Highway Accessibility	
	Seniors in HH	Number of Seniors
	Number of Drivers above 2	Three Drivers, Extra Drivers
	High Income HH	Income Low/High
		Majority/Minority Renters in Census Block Tract
		Home Ownership/ Renting

Table 1: Comparing Coefficients from TDM and Alternative Model

Variable	**Value**
Zero Vehicle Coefficients	
ASC	-3.05
Workers in HH	-0.475
Children in HH	-0.371
CBD or Dense Urban	0.5
Intersection Density	1.2
Low Income HH	3.5
Suburban or Rural	-0.95
Transit Accessibility/Highway Accessibility	0.758
Insufficient Vehicle Coefficients	
ASC	-0.289
CBD or Dense Urban	0.768
Intersection Density	0.928
Low Income HH	0.6
Seniors in HH	0.315
Suburban or Rural	-0.537
Number of Drivers above 2	0.45
Presence of Drivers above 2	0.85

Sufficient Vehicle Coefficients	
ASC	1
All drivers are workers	0.955
High Income HH	0.292

Table 2: Vehicle Choice Availability Model Coefficients

```
Call:
mlogit(formula = choice ~ 0 | WRKCOUNT + n_child + n_seniors +
    n_extra_drivers + three_drivers + non_work_driver + income +
    HTHTNRNT + homeownership + URBRUR + density | 0, data = veh_dfidx_train,
    reflevel = "Suff.", method = "nr")

Frequencies of alternatives:choice
   Suff.   Insuff.     Zero
0.884761 0.068416 0.046823

nr method
9 iterations, 0h:0m:9s
g'(-H)^-1g = 9.73E-08
gradient close to zero

Coefficients :
                               Estimate Std. Error  z-value  Pr(>|z|)
(Intercept):Insuff.           -4.862130   0.097034 -50.1073 < 2.2e-16 ***
(Intercept):Zero              -0.834354   0.145251  -5.7442 9.235e-09 ***
WRKCOUNT:Insuff.               0.504977   0.032080  15.7410 < 2.2e-16 ***
WRKCOUNT:Zero                 -3.164911   0.073392 -43.1231 < 2.2e-16 ***
n_child:Insuff.                0.222462   0.018456  12.0537 < 2.2e-16 ***
n_child:Zero                  -0.177596   0.044921  -3.9536 7.699e-05 ***
n_seniors:Insuff.              0.488630   0.025530  19.1394 < 2.2e-16 ***
n_seniors:Zero                -0.147901   0.051755  -2.8577 0.0042670 **
n_extra_drivers:Insuff.        0.209448   0.059347   3.5292 0.0004168 ***
n_extra_drivers:Zero           1.060206   0.520041   2.0387 0.0414802 *
three_driversTRUE:Insuff.      0.838284   0.085184   9.8409 < 2.2e-16 ***
three_driversTRUE:Zero        -0.166476   0.711853  -0.2339 0.8150918
non_work_driver:Insuff.        1.345289   0.055417  24.2758 < 2.2e-16 ***
non_work_driver:Zero          -4.057743   0.077993 -52.0269 < 2.2e-16 ***
incomelow:Insuff.              0.431830   0.039500  10.9324 < 2.2e-16 ***
incomelow:Zero                 1.659053   0.060965  27.2130 < 2.2e-16 ***
incomehigh:Insuff.            -0.215460   0.050031  -4.3065 1.658e-05 ***
incomehigh:Zero                0.265067   0.124643   2.1266 0.0334526 *
HTHTNRNTminority_rent:Insuff. -0.345696   0.045681  -7.5676 3.797e-14 ***
HTHTNRNTminority_rent:Zero    -0.509607   0.062366  -8.1713 2.220e-16 ***
homeownershipRENT:Insuff.      0.768701   0.042909  17.9148 < 2.2e-16 ***
homeownershipRENT:Zero         1.428855   0.062634  22.8127 < 2.2e-16 ***
URBRURURBAN:Insuff.            0.338251   0.048017   7.0444 1.863e-12 ***
URBRURURBAN:Zero               0.297968   0.093016   3.2034 0.0013581 **
densityLow:Insuff.            -0.091636   0.042721  -2.1450 0.0319530 *
densityLow:Zero               -0.062573   0.066862  -0.9359 0.3493461
densityMedium:Insuff.          0.586150   0.062385   9.3958 < 2.2e-16 ***
densityMedium:Zero             1.301879   0.076170  17.0918 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -18837
McFadden R^2:  0.3035
Likelihood ratio test : chisq = 16416 (p.value = < 2.22e-16)
```

Table 3: Vehicle Choice Availability Model Coefficients

```
Confusion Matrix and Statistics

          Reference
Prediction Suff. Insuff.  Zero
   Suff.   54410    4127  1407
   Insuff.   203     185     8
   Zero      297       1  1511

Overall Statistics

               Accuracy : 0.9028
                 95% CI : (0.9004, 0.9051)
    No Information Rate : 0.8835
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3341

 Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

                     Class: Suff. Class: Insuff. Class: Zero
Sensitivity                0.9909       0.042894     0.51640
Specificity                0.2355       0.996352     0.99497
Pos Pred Value             0.9077       0.467172     0.83527
Neg Pred Value             0.7732       0.933153     0.97655
Prevalence                 0.8835       0.069398     0.04708
Detection Rate             0.8755       0.002977     0.02431
Detection Prevalence       0.9645       0.006372     0.02911
Balanced Accuracy          0.6132       0.519623     0.75569
[1] 37730.76
```

Table 4: Confusion Matrix and Statistics