Characterizing zero-vehicle households: A double-hurdle problem perspective

Divyakant Tahlyan† and Hani Mahmassani†

†Transportation Center, Northwestern University

Northwestern University Transportation Center

Northwestern University

Introduction

- About **8.6 %** households in the United States do not have a vehicle.
- While many households refrain from owning a vehicle due to **financial constraints** (and will likely buy a vehicle as income increases or vehicle cost decreases), there are many who do so due to **secondary reasons** like inability to drive or significant attitudinal preference for other modes.
- From household vehicle ownership modeling and forecasting perspective, three most popular frameworks include: un-ordered discrete choice models, ordered discrete choice models, and count data models.
- However, there is **limited attention** given to **zero-vehicle households** in the these frameworks, potentially leading to poor forecasting performance.
- We propose a **double-hurdle perspective** to vehicle-ownership modeling to account for **dual latent states** of no vehicle ownership. [1]
- We compare a **zero-inflated ordered probit** model [2] with a ordered probit model (without zero-inflation) for vehicle ownership modeling.
- Comparison is done using estimated parameters, **Vuong's closness test**, Akaike and Bayesian information criteria.
- The estimated model includes both socio-demographic and "soft" variables following **factor analysis**.

Zero-Inflated Ordered Probit Model

Let y be a random variable observed in terms of vehicles owned and takes discrete outcome 0 to J.

Variable	Description
y	Vehicles owned by a household
r	Binary variable, $r = 0$ for non-participants; $r = 1$ for participants
r^*	Latent variable representing propensity of participation in vehicle ownership decision process
x	Vector of exogenous variables
ε	Standard normally distributed random variable
$\Phi(\cdot)$	Cumulative distribution function of standard normal distribution
$\widetilde{y}^*$	Latent propensity function for ordered probit model
$\widetilde{y}$	Discrete random variable generated by ordered probit model
ψ_j	Estimable thresholds
$\mathcal{LL}$	Log-likelihood function

Tab. 1: Variable definitions

Propensity of participation in the vehicle ownership decision process:

$$r^* = x^{'}\beta + \varepsilon$$

Probability of participation in the vehicle ownership decision process:

$$Pr(r = 1|x) = Pr(r^* > 0|x) = \Phi(x^{'}\beta)$$

Propensity function of the ordered probit model:

$$\widetilde{y}^* = z^{'}\gamma + u$$

Relationship between latent propensity function $\widetilde{y}^*$ and $\widetilde{y}$:

$$\widetilde{y} = \begin{cases} 0 & \text{if} \quad \widetilde{y}^* \leq 0 \\ j & \text{if} \quad \psi_{j-1} < \widetilde{y}^* \leq \psi_j \quad \forall\, j \,\epsilon\, (1, ..., J-1) \\ J & \text{if} \quad \psi_{J-1} \leq \widetilde{y}^* \end{cases}$$

Ordered probit probabilities:

$$Pr(\widetilde{y}) = \begin{cases} Pr(\widetilde{y} = 0|z, r = 1) = \Phi(-z^{'}\gamma) \\ Pr(\widetilde{y} = j|z, r = 1) = \Phi(\psi_j - z^{'}\gamma) - \Phi(\psi_{j-1} - z^{'}\gamma) \quad \forall\, j \,\epsilon\, (1, ..., J-1) \\ Pr(\widetilde{y} = J|z, r = 1) = 1 - \Phi(\psi_{J-1} - z^{'}\gamma) \end{cases}$$

Full probabilities for y:

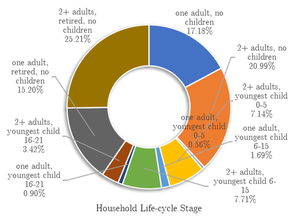
$$Pr(y) = \begin{cases} Pr(y = 0|z, x) = [1 - \Phi(x^{'}\beta)] + \Phi(x^{'}\beta)\Phi(-z^{'}\gamma) \\ Pr(y = j|z, x) = \Phi(x^{'}\beta)[\Phi(\psi_j - z^{'}\gamma) - \Phi(\psi_{j-1} - z^{'}\gamma)] \quad \forall\, j \,\epsilon\, (1, ..., J-1) \\ Pr(y = J|z, x) = \Phi(x^{'}\beta)[1 - \Phi(\psi_{J-1} - z^{'}\gamma)] \end{cases}$$

Log-likelihood function:

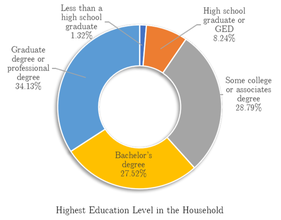
$$\mathcal{LL}(\theta) = \sum_{i=1}^{N}\sum_{j=0}^{J} h_{ij} \ln[Pr(y_i = j|x_i, z_i, \theta)]$$

Data and Factor Analysis

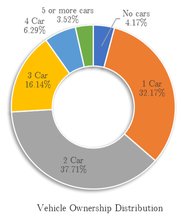
- 2017 National Household Travel Survey's (NHTS) **California add-on data** with information from 26,112 households.
- Available information included: vehicle ownership information; socio-demographics like household size, residential location, and life-cycle stage; travel behavior of all household members; attitudinal and life situation related information.
- Using attitudinal and life situation variables, two **latent factors** were identified: travel affected to due medical conditions; travel is a financial burden.



Household Life-cycle Stage



Highest Education Level in the Household



Vehicle Ownership Distribution

Fig. 1: Descriptive Statistics

Identified factor	Original Variables	Factor loading
Medical condition affects travel 25.2%*	Medical condition resulting a reduced day-to-day travel	0.863
	Medical condition resulting in asking others for rides	0.679
	Medical condition resulting in giving up driving	0.522
	Medical condition that makes it difficult to travel outside home	0.852
Travel is a financial burden 23.8%	Walk to reduce financial burden	0.853
	Bike to reduce financial burden	0.779
	Public transportation use to reduce final burden	0.701
	Price of gasoline affects travel	0.415
	Travel is a financial burden	0.370

* Percentage of variance explained by the factor

Tab. 2: Results for the factor analysis

Results

Comparison of Model fit measures:

- Since the two model structures are not nested, AIC, BIC and Vuong's closeness test used for comparison.
- $AIC = 2(k - \mathcal{LL})$ and $BIC = kln(n) - 2\mathcal{LL}$, where k is number of estimated parameters and n is number of observations.
- Test statistic in Vuong's closeness test written as:

$$\nu = \frac{\sqrt{N}(\frac{1}{N}\sum_{i=1}^{N} m_i)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_i - \overline{m})^2}}$$

where $m_i = log(\frac{f_1(y_i|x_i,z_i)}{f_2(y_i|x_i,z_i)})$, $f_h(y_i|x_i, z_i)$ is predicted probability using model h (=1 for OP and 2 for ZIOP) that y_i equals y, $\overline{m}$ is average value of m_i over all N. $\nu < -1.96$ favors model 2 (ZIOP).

- AIC and BIC values for the ordered probit model equal 51202.8 and 51397.2, respectively. For the zero-inflated ordered probit, AIC and BIC values equal 51141.2 and 51392.2, respectively. This suggests that the zero-inflated ordered probit model is the preferred model over the simple ordered probit model.
- The Vuong's test static equals -2.815, which is greater than the threshold -1.96 value favoring the zero-inflated model.

Results

Interpretation of estimated parameters

- Statistically significant parameters in the ordered model fall in the following categories:
  - Household socio-demographics including number of drivers and workers in the household, household income etc.
  - Life-cycle stage variables
  - Household's residential location characteristics
  - Factor variable that travel is affected due to medical issues.
- Variables in zero-inflation probit model includes income and both factor variables.

Variable	Ordered Probit Model		Zero-Inflated Ordered Probit Model	
	Parameter Estimate	t-statistic	Parameter Estimate	t-statistic
Ordered probit probability				
Constant	0.702	17.050	0.772	17.030
Household socio-demographics variables				
Number of drivers in the household	1.062	68.010	1.067	75.290
Number of workers in the household	0.120	11.550	0.120	11.440
Household income between $25,000 & $49,999 indicator	0.271	10.830	0.248	9.850
Household income between $50,000 & $99,999 indicator	0.466	19.550	0.447	18.880
Household income between $100,000 & $149,999 indicator	0.603	21.860	0.585	21.160
Household income more than or equal to $150,000	0.650	22.030	0.635	21.310
White ethnicity household indicator	0.041	2.200	0.039	2.050
Household with at least one individual with a college degree indicator	-0.140	-8.460	-0.142	-8.710
Life cycle stage variables				
At least two individuals in the household are related indicator	0.336	14.540	0.347	16.340
Single adult household with at least one child indicator	-0.222	-4.940	-0.224	-5.070
2+ adults household with youngest child 0-5 years in age indicator	-0.209	-7.290	-0.204	-6.140
2+ adults household with youngest child 6-15 years in age indicator	-0.108	-3.870	-0.106	-3.530
2+ adults household with youngest child 16-21 years in age indicator	-0.300	-7.270	-0.301	-7.070
Household's residential location characteristics				
Housing units per square mile (by 100) in census tract of household's home location	-0.005	-15.5	-0.005	-17.44
Workers per square mile (by 100) in the census tract of the household's home location	-0.002	-5.51	-0.002	-5.38
Home location in urban area indicator	-0.46	-19.99	-0.463	-21.85
Presence of rail in household's MSA	-0.095	-5.38	-0.095	-5.23
Home owned by the responding household indicator	0.481	25.51	0.475	24.58
Factor Variables				
Travel affected due to medical issues	-0.066	-8.1	-0.062	-8.05
Zero inflation probit probability				
Constant	—	—	2.871	11.77
Household income between $25,000 & $49,999 indicator	—	—	0.936	2.85
Household income between $50,000 & $99,999 indicator	—	—	0.991	3.07
Household income between $100,000 & $149,999 indicator	—	—	1.097	2.4
Household income more than or equal to $150,000	—	—	0.557	1.76
Travel affected due to medical issues	—	—	-0.114	-2
Travel is a financial burden	—	—	0.653	5.79
Thresholds				
1|2	2.323	188.6	2.381	74.56
2|3	3.905	352.44	3.967	116.98
3|4	4.827	368.03	4.89	138.62
4|5	5.517	296.88	5.581	151.03
Model fit measures				
Log-likelihood at convergence	-25577.442		-25539.623	
Number of estimated parameter	24		31	
Number of observations	24246		24246	
Log-likelihood for constants only model	-35154.751		-35154.751	
ρ_c^2	0.2724		0.2735	
AIC	51202.884		51141.246	
BIC	51397.188		51392.22	

— corresponding parameter not estimated

Tab. 3 Estimation results

Discussion

- Households that consider travel as a financial burden more likely to be in vehicle ownership state.
- Households with at least one member with a medical condition that affects travel less likely to be in vehicle ownership state.
- Presence of non-linear effect of income on probability of being in vehicle ownership state. However, the nature of non-linearity is different from as in the ordered probit probability model part.

References

[1] Cragg, J.G., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica: Journal of the Econometric Society, pp.829-844.

[2] Harris, M.N. and Zhao, X., 2007. A zero-inflated ordered probit model, with an application to modelling tobacco consumption. Journal of Econometrics, 141(2), pp.1073-1099.

[3] "Transportation Secure Data Center.", 2019. National Renewable Energy Laboratory. Accessed Jan. 19, 2021: www.nrel.gov/tsdc.

100th Annual Meeting of the Transportation Research Board (Paper ID: 21-04110)