**Online supplement to “Role of Childhood Context and Experience in Shaping Activity-Travel Choices in Adulthood”**

Kamryn Long, Denise Capasso da Silva, Felipe F. Dias, Sara Khoeini, Aarti C. Bhat, Ram M. Pendyala, Chandra R. Bhat (corresponding author)

## Characteristics of final sample

Table 1 of this online supplement presents the demographic characteristics of the final survey sample used in the paper as well as a comparison with corresponding characteristics of the U.S. adult population over 16 years of age for individual statistics (except for gender and race which are all age inclusive). The U.S. population characteristics were obtained from the 2016 American Community Survey (ACS). The cells that display a “---“ symbol represent variables for which data was not readily available in the 2016 ACS.

**Table 1 Sample and U.S. Sociodemographic Characteristics**

| Variable | Sample | | US |
|---|---|---|---|
| | Count | % | % |
| *Race* | | | |
| White (Non-hispanic) | 8,430 | 81.0 | 73.3 |
| White (Hispanic) | 475 | 4.6 | |
| Asian / Pacific Islander | 577 | 5.5 | 5.2 |
| Black / African American | 538 | 5.2 | 12.6 |
| Other (Primarily American Indian and Alaska native) | 384 | 3.7 | 8.9 |
| *Household Income* | | | |
| Low (Below $35,000) | 2,036 | 19.6 | 32.2 |
| Med ($35,000-$100,000) | 3,888 | 37.4 | 43.2 |
| High (Above $100,000) | 4,480 | 43.0 | 24.6 |
| *Generation/Age Group* | | | |
| Millennial (1980-1998) | 3,412 | 32.8 | 33.9 |
| Gen X (1965-1979) | 1,353 | 13.0 | 15.7 |
| Baby Boomer (1946-1964) | 4,220 | 40.6 | 32.4 |
| Silent (1925-1945) | 1,419 | 13.6 | 18.0 |
| *Residential Location* | | | |
| Urban | 2,510 | 24.1 | --- |
| Suburban | 5,426 | 52.2 | --- |
| Small Town | 1,389 | 13.4 | --- |
| Rural | 1,079 | 10.3 | --- |
| *Gender* | | | |
| Female | 5,575 | 53.6 | 49.2 |
| Male | 4,829 | 46.4 | 50.8 |
| *Individual Employment Status* | | | |
| Unemployed | 5,020 | 48.3 | 7.4 |
| Employed part or full | 5,384 | 51.7 | 92.6 |
| *Has driver’s license* | | | |
| No | 552 | 5.3 | --- |
| Yes | 9,852 | 94.7 | --- |

| Variable | Sample | | US |
|---|---|---|---|
| | Count | % | % |
| *Household Vehicle Ownership* | | | |
| 0 | 344 | 3.3 | 8.6 |
| 1 | 3,125 | 30.0 | 32.7 |
| 2 | 4,630 | 44.5 | 37.3 |
| 3+ | 2,305 | 22.2 | 21.5 |
| *Presence of Children in Household* | | | |
| No | 7,755 | 74.5 | 68.4 |
| Yes | 2,649 | 25.5 | 31.6 |
| *Multi-Worker Household* | | | |
| No | 6,636 | 63.8 | 64.1 |
| Yes | 3,768 | 36.2 | 35.9 |
| *Presence of Children in Household* | | | |
| No | 7,755 | 74.5 | 68.4 |
| Yes | 2,649 | 25.5 | 31.6 |
| *Household Size* | | | |
| 1 | 1,900 | 18.3 | 27.7 |
| 2 | 5,031 | 48.4 | 33.7 |
| 3 | 1,668 | 16.0 | 15.7 |
| 4 | 1,115 | 10.7 | 22.9 |
| 5 | 425 | 4.1 | |
| 6 | 180 | 1.7 | |
| 7+ | 85 | 0.8 | |
| *Number of Licensed Drivers in Household* | | | |
| 0 | 190 | 1.8 | --- |
| 1 | 2,495 | 24.0 | --- |
| 2 | 6,186 | 59.5 | --- |
| 3 | 1,079 | 10.4 | --- |
| 4 | 359 | 3.5 | --- |
| 5+ | 95 | 0.8 | --- |

As can be observed from Table 1, the sample is skewed towards individuals from the "White" race, higher income levels, higher vehicle ownership rates and unemployed individuals. Conversely, the proportions of individuals belonging to Baby Boomer and Millennial generations seem to be consistent with the general US population. This consistency is also seen in the proportion of multi-worker households as well as households with children in them. In general, the sample depicts socio-economic and demographic characteristics that would render it suitable for the type of analysis being undertaken in this study.

**The Model of Vehicle Ownership and Frequency of Transit Use**

In the following presentation, the index $q$ is used to denote respondents ($q$ = 1, 2,…, $Q$). For each respondent, let $c_q$ represent the index for the ordinal motorized vehicle ownership level, let $f_q$ represent the index for the ordinal frequency of transit use conditional on transit being available, and let $t_q$ be a binary variable taking the value of zero if transit is not available and one if transit is available. Let $m_q$ be the actual observed vehicle ownership level ($m_q$ can take values of 0, 1, 2,…, $M$; $M$=3 in this study, representing the vehicle ownership level of 3+ vehicles) and let $n_q$ be the actual observed transit use frequency ($n_q$ can take value of 0, 1, 2,…, $N$, where "0" represents "never used transit", "1" represents "less than twice per month" of transit use, and "2" represents "twice per month or more" of transit use; $N$=2 in this study). Then, for individual $q$, the following ordinal equation formulation may be written:

$$\begin{aligned} c_q^* &= \boldsymbol{\alpha}'\boldsymbol{x}_q + v_q,\ \delta_{m_q-1} < c_q^* < \delta_{m_q} \ \text{(because } c_q = m_q\text{)} \\ f_q^* &= \boldsymbol{\beta}'\boldsymbol{z}_q + \eta_q,\ \psi_{n_q-1} < g_q^* < \psi_{n_q} \ \text{(because } f_q = n_q\text{)} \end{aligned} \tag{1}$$

where $c_q^*$ and $f_q^*$ are the latent propensity measures associated with owning motorized vehicles and using transit, respectively; $\boldsymbol{x}_q$ and $\boldsymbol{z}_q$ are exogenous variable vectors (with no constant terms), including (1) demographic variables, (2) transportation-related variables, (3) all childhood experience variables, as well as (4) interactions of the many different types of variables; $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are corresponding coefficient vectors to be estimated; $v_q$ and $\eta_q$ are random error terms; the $\delta_m$ and $\psi_n$ terms represent thresholds that relate the latent propensity measures $c_q^*$ and $f_q^*$ to their observed counterparts $c_q$ and $f_q$ respectively, in the usual ordered-response fashion, i.e.,

$$(\delta_{-1} = -\infty, \delta_M = \infty; -\infty < \delta_0 < \delta_1 < \delta_2 < \ldots < \delta_{M-1} < \infty) \text{ and}$$
$$(\psi_{-1} = -\infty, \psi_N = \infty; -\infty < \psi_0 < \psi_1 < \psi_2 < \ldots < \psi_{N-1} < \infty).$$

The error terms $v_q$ and $\eta_q$ are assumed to be bivariate standard normally distributed with a correlation coefficient of $\rho$ (the use of standard normally distributed error terms $v_q$ and $\eta_q$ represents an innocuous normalization needed for econometric identification). The error terms $v_q$ are assumed to be independent and identically distributed (IID) across individuals $q$, and the error terms $\eta_q$ are also assumed to be IID across individuals $q$. For notational convenience, define $b_{m_q} = \delta_{m_q} - \boldsymbol{\alpha}'\boldsymbol{x}_q$ and $d_{n_q} = \psi_{n_q} - \boldsymbol{\beta}'\boldsymbol{z}_q$.[1] Also, let $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}, \delta_0, \delta_1, \ldots\delta_{M-1}, \psi_0, \psi_1, \ldots\psi_{N-1}, \rho)'$.

[1] The reader will note that a recursive endogenous effect of one observed ordinal variable on the propensity of the other observed ordinal variable is possible in the equation system of (1), although including both directions of observed effects leads to a logically inconsistent model structure because the probabilities of all the possible combinations of discrete observations will not sum to one (see Maddala, 1983, page 119 for a good discussion). Intuitively, the

With the preliminaries as above, the probability that the respondent *q*'s household has a vehicle ownership level of *m* and respondent *q*'s use of transit is characterized by level *n*, conditional on transit being available to respondent *q*, may be written as follows:

$$\begin{aligned}\text{Prob}[f_q = m, g_q = n] &= \text{Prob}[b_{m_q-1} < v_q < b_{m_q}, d_{n_q-1} < \eta_q < d_{n_q}] \\ &= \Phi_2[b_{m_q}, d_{n_q}, \rho] - \Phi_2[b_{m_q}, d_{n_q-1}, \rho] - \Phi_2[b_{m_q-1}, d_{n_q}, \rho] + \Phi_2[b_{m_q-1}, d_{n_q-1}, \rho]\end{aligned} \quad (2)$$

In the above expression, $\Phi_2[.,.,\rho]$ represents the bivariate cumulative standard normal distribution function. For respondents for whom transit is not available, the use of transit does not arise, and the only contribution of such individuals is to inform the vehicle ownership model:

$$\text{Prob}[f_q = m] = \text{Prob}[b_{m_q-1} < v_q < b_{m_q}] = \Phi[b_{m_q}] - \Phi_2[b_{m_q-1}] \quad (3)$$

The likelihood function to be maximized is then given by the following expression:

$$L(\boldsymbol{\theta}) = \prod_{q=1}^{Q} \left[ \text{Prob}[f_q = m, g_q = n] \times t_q + \text{Prob}[f_q = m] \times (1 - t_q) \right] \quad (4)$$

**Estimation of Size Effects**

Once estimated, the model can be used to quantify the relative contributions of each set of observed factors in explaining vehicle ownership and frequency of transit use (along with the remaining unexplained portion). The four factors are: (1) demographic variables (DEM), (2) transportation-related variables (TRAN), (3) direct childhood experience variables (DIRECT), and (4) parent childhood experience variables (PARENT). In the final model specification, almost all the interaction variable effects consisted of only demographic variable interactions, and so they are conveniently grouped with the DEM variables and not included separately.

To determine the size effects of each of the four categories of variables and the unobserved part of the propensity, it is appropriate to start from Equation (1). For presentation ease, focus on the size effects only for the motorized vehicle ownership propensity (because the procedure is identical for transit use frequency). Start by partitioning the vector $x_q$ into variables that correspond to the DEM, TRAN, DIRECT, and PARENT categories: $\boldsymbol{x}_q = (\boldsymbol{x}'_{q,\text{DEM}}, \boldsymbol{x}'_{q,\text{TRAN}}, \boldsymbol{x}'_{q,\text{DIRECT,}} \boldsymbol{x}'_{q,\text{PARENT}})'$, and correspondingly partition the vector $\boldsymbol{\alpha}$ into $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_{DEM}, \boldsymbol{\alpha}'_{TRAN}, \boldsymbol{\alpha}'_{DIRIECT}, \boldsymbol{\alpha}'_{PARENT})'$. Then:

$$c_q^* = \boldsymbol{\alpha}'_{DEM}\boldsymbol{x}_{q,DEM} + \boldsymbol{\alpha}'_{TRAN}\boldsymbol{x}_{q,TRAN} + \boldsymbol{\alpha}'_{DIRECT}\boldsymbol{x}_{q,DIRECT} + \boldsymbol{\alpha}'_{PARENT}\boldsymbol{x}_{q,PARENT} + v_q \text{ , and} \quad (5)$$

$$\text{Var}(c_q^*) = \text{Var}(\boldsymbol{\alpha}'_{DEM}\boldsymbol{x}_{q,DEM}) + \text{Var}(\boldsymbol{\alpha}'_{TRAN}\boldsymbol{x}_{q,TRAN}) + \text{Var}(\boldsymbol{\alpha}'_{DIRECT}\boldsymbol{x}_{q,DIRECT}) + \text{Var}(\boldsymbol{\alpha}'_{PARENT}\boldsymbol{x}_{q,PARENT}) + 1$$

The expression above results because Var($v_q$)=1 by normalization. From the above, and once the parameter vector $\boldsymbol{\alpha}$ is estimated, the total variance $\text{Var}(c_q^*)$ may be computed as $\text{Var}(\boldsymbol{\alpha}'\boldsymbol{x}_q) + 1$, where $\text{Var}(\boldsymbol{\alpha}'\boldsymbol{x}_q)$ itself is the variance of the value of $\boldsymbol{\alpha}'\boldsymbol{x}_q$ across respondents, and the size percentages of each of the contributions of the right-side elements of Equation (5) can then be

propensities are the precursors to the actual observed variables, and, when both the decisions are co-determined, it is impossible to have both observed variables structurally affect one another. In the current paper, models with each possible structural direction of impact were estimated, and the one that provided a better data fit was chosen. In this study, the best data fit was obtained with the effect going from vehicle ownership to transit use frequency (vehicle ownership is included as a transportation-related variable in the subsequent presentation, along with residential location choice). However, it is critical to note that, regardless of which directionality of structural effects comes out to be better (or even if both directions are not statistically significant), the system in (1) is a joint bundled system because of the correlation in unobserved factors affecting the underlying propensities.

obtained. It should be noted that one can also obtain the size percentages for different segments of the sample (such as Millennials versus non-Millennials) by focusing only on the respondents in each segment, and undertaking the same exercise as just described above for the entire sample. Doing so also recognizes the limited interactions we found between the PARENT variables and the DEM variables characterizing the generational divide.

## REFERENCES


American Community Survey. *American Fact Finder*. U.S. Census Bureau, 2016. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml.

Maddala, G.S. *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, UK, 1983.