<div align="center">

Online Supplement to

"Incorporating a Multiple Discrete-Continuous Outcome in the Generalized Heterogeneous Data Model: Application to Residential Self-Selection Effects Analysis in an Activity Time-use Behavior Model"

By Chandra R. Bhat (corresponding author), Sebastian Astroza, Aarti C. Bhat, and Kai Nagel

</div>

Descriptive Characteristics of the Sample

Table 1 provides descriptive statistics of the socioeconomic characteristics of the sample and that of the PSRC four-county region population as a whole from the 2010 Census. Of course, the comparison is not really appropriate (because we are unable to obtain, from the Census data, statistics solely on 1+-worker households with at least one person employed outside the home). But we provide the population statistics just for informational purposes.

According to the statistics provided in Table 1, a majority of the households are couple families (34.2%) or single person (27.8%), though there are also a sizeable number of nuclear families (20.2) and multi-adult households (the term "multi-adult is used here to represent more than two adults in the household; this category includes extended families and room-mates). The percentage of single parent families in our sample of 1+worker households is very low relative to the general population. The fractions of male and female adults within the household, when averaged across all sample households, are close to the 50% split observed in the population. Not surprisingly, the sample households in general are much more educated than the households in the population. This is also reflected in the high percentage of households with an annual income of over 75,000, though we do not have the income information for the PSRC region from the 2010 Census data. The distribution of the number of children shows a high percentage of childless households, consistent with the high percentage of single person, couple, and multi-adult households (though multi-adult households contribute to 3.3% of the 74.9% of childless households). The percentage of childless households is of the same order in the sample and the Census data, though the Census does not provide the breakdown by number of children for households with children. The fraction of adults by age in the sample, when averaged across all sample household, is highly loaded on the 35-54 year category relative to the entire population, with much smaller representation of individuals in the 65 years and beyond category. This is again not surprising given the focus on 1+-worker households in our sample. The work status

<div align="center">

1

</div>

distribution is not available from the Census, but the sample statistics on the fraction of adults in the household in each of four categories; full-time workers, part-time workers, self-employed workers, and non-workers; clearly indicates a high fraction of full-time workers, with the fraction of adults in the other three categories being of the same order and range from 0.066 to 0.128.

Computation of the Average Treatment Effects

For auto ownership, the measure is estimated as follows for each model:

$$\hat{\text{ATE}}_{ii'}^{auto} = \frac{1}{Q}\sum_{q=1}^{Q}\left(\sum_{g=0}^{\infty} g \cdot \left[P(\text{auto ownership} = g \mid a_{qi} = 1) - P(\text{auto ownership} = g \mid a_{qi'} = 1)\right]\right) \quad (19)$$

where a_{qi} is the dummy variable for the density category i for the household q. Although the summation in the equation above extends until infinity, we consider counts only up to $g=10$, which is the maximum vehicle ownership level observed in the data set. The standard error of the measure is computed using bootstraps from the sampling distributions of the estimated parameters.

For the activity time use variables (MDC variable), we focus only on the participation dimension here and compute the ATE measure for the out-of-home activity k ($k=1,2,\ldots,K–1$) as follows:

$$\hat{\text{ATE}}_{ii'}^{OH\,act.k} = \frac{1}{Q}\sum_{q=1}^{Q}\left(\left[P(t_{qk} > 0 \mid a_{qi} = 1) - P(t_{qk} > 0 \mid a_{qi'} = 1)\right]\right) \quad (20)$$

where t_{qk} is the time spent by individual q on the out-of-home activity k. To compute the probability that $t_{qk} > 0$, we drew, for each individual, 100 sets of 1000 realizations from a multivariate normal sampling distribution of estimated parameters and the distribution of the error terms involved. For each individual, each set, and each realization, we used the forecasting algorithm of Pinjari and Bhat (2014) to predict time allocations and, then, for each individual and each set, evaluated the share of the 1000 realizations that predicted $t_{qk} > 0$ for each of the two density categories involved (that is, i and i'). The treatment effect is then computed as in Equation (19) for each set, and the mean across all the 100 sets was computed as the final ATE effect and the standard deviation across the 100 sets was computed as the standard error estimate.

Reference

Pinjari, A.R., and Bhat, C.R. (2014). Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: application to residential energy consumption analysis. Technical paper, Department of Civil and Environmental Engineering. University of South Florida. Avaliable at:
http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Pinjari_Bhat_MDCEV_Forecasting_July21_2011.pdf.

Table 1. Sample Characteristics of Independent Variables

Socio demographic	Categories	Sample Distribution	Census Distribution
Family structure	Single person household	27.8%	31.0%
	Single parent family	1.6%	9.0%
	Couple family	34.2%	29.7%
	Nuclear family	20.2%	23.0%
	Multi-person household	16.2%	7.3%
Gender	Fraction of male adults in household (mean)	0.468	0.50
	Fraction of female adults in household (mean)	0.532	0.50
Educational attainment	Fraction of adults with High school or less in household (mean)	0.097	0.303
	Fraction of adults with Some college in household (mean)	0.233	0.327
	Fraction of adults with a Bachelor's degree in household (mean)	0.382	0.239
	Fraction of adults with Graduate degree in household (mean)	0.288	0.131
Household income	$0 to below $25,000	5.9%	NA
	$25,000 to below $35,000	7.0%	NA
	$35,000 to below $50,000	10.7%	NA
	$50,000 to below $75,000	18.3%	NA
	$75,000 and above	558.1%	NA
Number of children	No kids	74.9%	70.1%
	One kid	12.6%	NA
	Two kids	10.0%	NA
	Three or more kids	2.5%	NA
Age	Fraction of adults aged 18 to 34 in household (mean)	0.341	0.349
	Fraction of adults aged 35 to 54 in household (mean)	0.421	0.369
	Fraction of adults aged 55 to 64 in household (mean)	0.185	0.147
	Fraction of adults 65 years old or older in household (mean)	0.053	0.135
Adult work status	Fraction of full-time working adults in the household (mean)	0.700	NA
	Fraction of part-time working adults in the household (mean)	0.106	NA
	Fraction of self-employed working adults in the household (mean)	0.066	NA
	Fraction of non-working adults in the household (mean)	0.128	NA

*NA: Not available

Structural Equation System

Measurement Equation System

Effect coeff. vector (**d**)

Observed continuous outcomes (commute distance in empirical analysis)

Observed covariate vector (**x**)

Effect coeff. vector (**γ**)

Error vector (**η**)

Error vector (**ε**)

Unobserved threshold vector ($\widetilde{\boldsymbol{\psi}}$)

Observed covariate vector (**w**)

Effect coeff. vector ($\widetilde{\boldsymbol{d}}$)

Underlying latent variable vector ($\widetilde{\boldsymbol{y}}^*$) for ordinal variables

Observed ordinal outcomes (indicators in empirical analysis)

Endogenous effects (can be only in one direction)

Latent variable vector ($\boldsymbol{z}^*$)

Error vector ($\widetilde{\boldsymbol{\varepsilon}}$)

Effect coeff. vector ($\bar{\boldsymbol{\gamma}}$)

Unobserved threshold vector ($\breve{\boldsymbol{\psi}}$) including (**θ**) and (**φ**)

Endogenous effects (can be only in one direction)

Green lifestyle propensity

Effect coeff. vector ($\breve{\boldsymbol{d}}$)

Underlying latent variable vector ($\breve{\boldsymbol{y}}^*$) for count variables

Observed count outcomes (auto ownership in empirical analysis)

Luxury lifestyle propensity

Error vector ($\breve{\boldsymbol{\varepsilon}}$)

Effect coeff. vector ($\widetilde{\boldsymbol{b}}$)

Endogenous effects (can be only in one direction)

Effect coeff. vector ($\widetilde{\boldsymbol{\alpha}}$)

Effect coeff. vector ($\widetilde{\boldsymbol{\varpi}}$)

Utility vector **U** underlying nominal variable

Observed nominal outcome (residential location in empirical analysis)

Error vector ($\widetilde{\boldsymbol{\varsigma}}$)

Satiation parameter vector (**τ,α**)

Endogenous effects (can be only in one direction)

Effect coeff. vector (**δ**)

Utility vector $\widetilde{\boldsymbol{U}}$ underlying MDC variable

Observed MDC outcome (activity time use in empirical analysis)

Effect coeff. vector (**μ**)
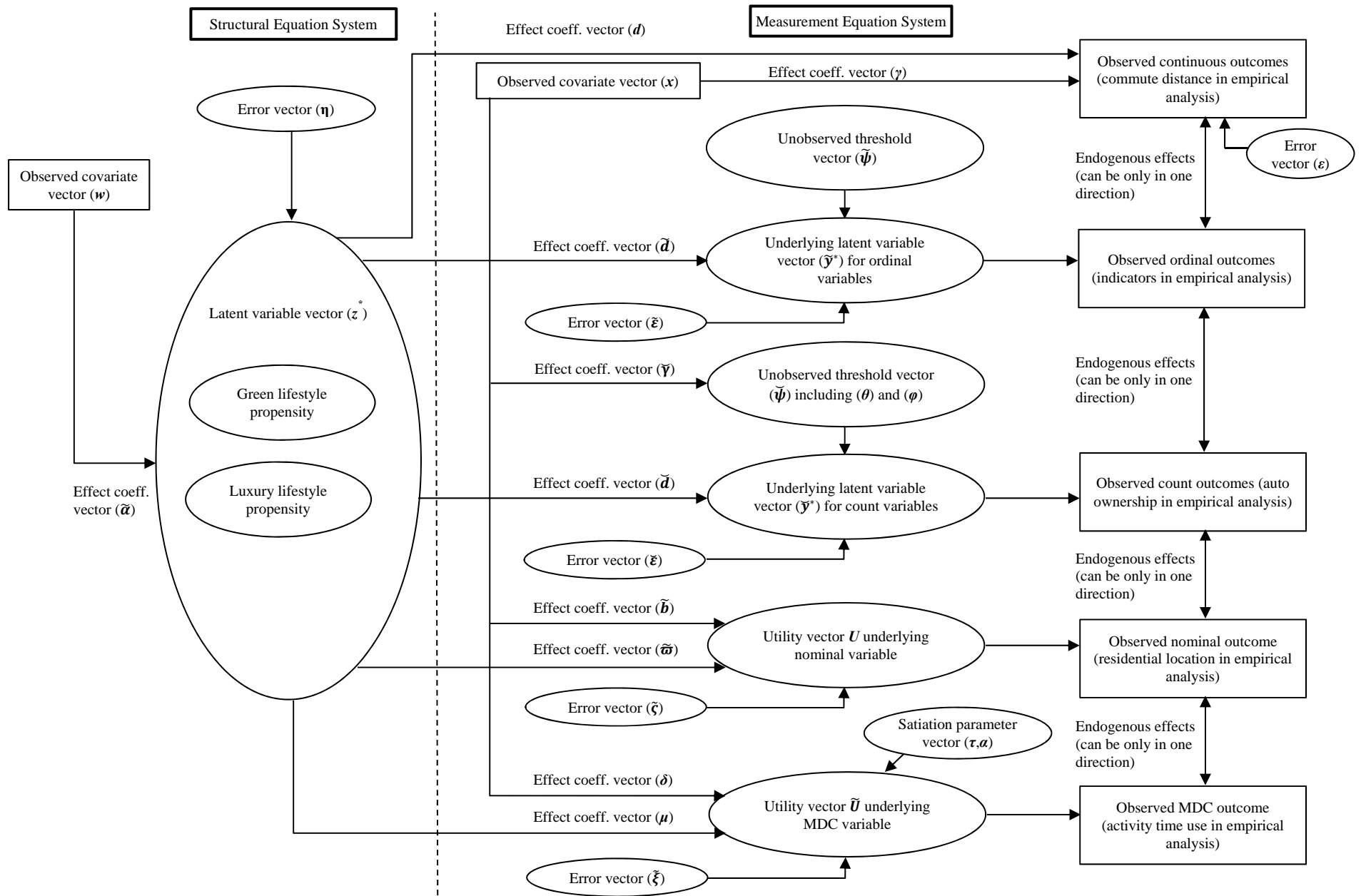
Error vector ($\widetilde{\boldsymbol{\xi}}$)

Figure A. Diagrammatic Representation of the Model System

5