**Appendices to**

**“A New Utility-Consistent Econometric Approach to Multivariate Count Data Modeling”**

by

Chandra R. Bhat\*

The University of Texas at Austin

Dept of Civil, Architectural and Environmental Engineering

301 E. Dean Keeton St. Stop C1761, Austin TX 78712

Phone: 512-471-4535, Fax: 512-475-8744

Email: [bhat@mail.utexas.edu](mailto:bhat@mail.utexas.edu)

Rajesh Paleti

Parsons Brinckerhoff

One Penn Plaza, Suite 200

New York, NY 10119

Phone: 512-751-5341

Email: [rajeshp@mail.utexas.edu](mailto:rajeshp@mail.utexas.edu)

Marisol Castro

The University of Texas at Austin

Dept of Civil, Architectural and Environmental Engineering

301 E. Dean Keeton St. Stop C1761, Austin TX 78712

Phone: 512-471-4535, Fax: 512-475-8744

Email: [m.castro@utexas.edu](mailto:m.castro@utexas.edu)

\*corresponding author

**Appendix A**

The notations used here will be the same as those used in the text. Before providing proofs for the theorems in the main text, we provide the following well established results for the multivariate normal distribution, collected together in a single Lemma (without proof).

*Lemma 1*

1) The multivariate normal density function and cumulative distribution function of dimension *R* are respectively given by  and , where .

2) Let  and  be normally distributed vectors of dimension  and , respectively. The corresponding mean vector and covariance matrix of  and  are  and . Defining ,  and , where  is the covariance matrix between  and , the conditional distribution of  given  is . Then,

.

In what follows, we present and discuss four theorems that are key to the proposal in this paper.

*Theorem 1*

The stochastic transformation of  aswhere  is a constant scalar parameter and  is a univariate normally distributed scalar , has a cumulative distribution function and density function as below:



Proof:



*Theorem 2 – Proposition (1)*

The probability function of  is given by  as follows:



where .

Proof:

The cumulative distribution function of is given by (see Tellambura, 2008):



The proof that the density function takes the form as given above can be shown by differentiating with respect to  and using the last result from Lemma 1.

*Theorem 2 – Proposition (2)*

The moment generating function of  is given by:

,

where , , , and .

Proof:





where .

The moment generating function of  is given by:











, where 







since  for all scalar *a,* vector***B*** and random variable  (see Marsaglia, 1963 and Gupta *et al.*, 2004).

Finally, .

**Theorem 2 – Proposition (3)**

This proposition can be proved through straightforward, but tedious, differentiation, and using the results of Lemma 1.

**References**

Gupta AK, González-Farías G, Domínguez-Monila JA, 2004. A multivariate skew normal distribution. *Journal of Multivariate Analysis* 89(1): 181-190.

Marsaglia G. 1963. Expressing the normal distribution with covariance matrix A + B in terms of one with covariance matrix A. *Biometrika* 50(3-4): 535–538.

Tellambura C. 2008. Bounds on the distribution of a sum of correlated lognormal random variables and their application. *IEEE Transactions on Communications* 56(8): 1241-1248.

**Appendix B: Consistency with Two-Stage Budgeting**

The proposed approach that combines a total count model with a model that allocates the count to different event types is analogous to the two stage budgeting procedure in utility-based consumer theory. The basic idea of two stage budgeting is to determine a budget for a specific group of commodities at a first stage (through the development of a scalar price index for the commodity group) in such a way that the first stage utility maximization can progress without the need to worry about allocations to particular commodities within the group. Once the budget is determined at the first stage, the allocation of the budget to individual commodities is pursued in a second stage. The approach makes use of the notion of weak separability of the direct utility function. Our presentation follows that of Hausman *et al.* (1995), except that there is a difficulty with the Hausman *et al.* formulation that makes it incompatible with two-stage budgeting, while our formulation is.

Consider a direct utility function in which a group of commodities is separable from the rest. The group of commodities corresponds to the one whose count is being modeled. So, it may correspond to recreational or grocery shopping trips (with the event type being alternative destinations), or to vehicle ownership level (with the event type being alternative body types), or, as in our empirical application, the number of out-of-home non-work episodes (with the event type being different time periods of the day). The notion of separability implies that the commodity group can be represented by a group utility function in the first stage of the two-stage budgeting process in which the overall budget allocation to the commodity group is being determined in the presence of other commodity groups. It also implies that the optimal allocation of the budget within the commodity group can be determined solely by the group utility function in a second stage, once the budget to the commodity group is determined in the first stage and the prices of individual commodities in the group are known (the reader is referred to Deaton and Muelbaurer, 1980, for a detailed description of the concepts of separability and two-stage budgeting; a comprehensive discussion is well beyond the scope of this paper).

An important issue in the two-stage budgeting is the question of how to determine the budget allocation to the commodity group in the first stage. While one can consider many different formulations, it would be particularly convenient if there were no need to explicitly consider the detailed vector information of the prices of all the individual commodities in the group in this first stage. The question then is whether one can use a group (scalar) price index for the commodity group at this first stage. Gorman (1959) studied this problem in his seminal research, and concluded that one can use a scalar price index if, in addition to the separability property of the overall utility function, this overall utility function in the first stage is additive in the group utility functions and the group indirect utility functions (corresponding to the group direct utility functions) follow what is now referred to as the Gorman Polar Form (GPF). We start with this group indirect utility function of the GPF form for the commodity group of interest. In the following presentation, we suppress the index *q* for the individual, and, as in Hausman *et al.* (1995), consider the group utility function to be homothetic. Then, we can write the group indirect utility function for the commodity group *D* as a function of the budget for the commodity group  and the vector of prices of the goods within the commodity group *D*:

 (B.1)

In the above GPF equation,  represents the group scalar price index. The functional form of must be homogenous of degree one. If this condition is satisfied, then information about the value of is adequate to determine the budget allocation to the commodity group in the first stage. That is, the entire commodity group can be viewed as a single commodity with price  in the first stage budgeting, which takes the form of maximizing a direct utility function that takes consumption in other goods and consumption in a single composite “good” representing the commodity group of interest as arguments (subject to the usual budget constraint). The number of units (the total count) of consumption in the commodity group becomes .

The second stage budgeting of the group budget  to individual commodities in the group can be obtained by applying Roy’s identity to the indirect utility function of Equation (B.1). Specifically, the conditional number of units of consumption of commodity *i* can be written as:

 (B.2)

where  is now the price of commodity *i* within group *D*. To view the above equation as the second stage of a two-stage budgeting procedure, there are two conditions that must satisfy: (1) it must be homogeneous of degree one (that is the requirement of the GPF), and (2)  (this allows the interpretation of  as the total units (or count) of consumption across all commodities in group *D*). Hausman *et al.* (1995) choose the expected consumer surplus (or accessibility) measure resulting from a multinomial logit model for . That is, they write  With this specification, we have , and therefore the second condition above on  is satisfied. However, the form used by Hausman *et al.* for  does not satisfy the first condition because of the presence of the log transformation. Specifically, Thus, as pointed out by Rouwendal and Boter (2009), Hausman *et al.*’s model specification is not consistent with a single utility maximization setting. Further, the use of any generalized extreme value (GEV) model for the second stage commodity choice is also not consistent with utility theory because the resulting expression for  is not homogeneous. Rouwendal and Boter (2009) comment that they have not been able to find an expression for  that satisfies both the conditions stated above. That is exactly where our proposed model comes in. To our knowledge, we are the first to propose a specification for  that satisfies both the required conditions discussed above for compatibility of the joint count-event type model with two-stage budgeting, while also allowing the probability of choice of commodity *i* to be a function of individual commodity prices (as they should be). In particular, as in Hausman *et al.*, we propose  except that we specify  to be multivariate normal (see previous section;  after suppressing the index *q* for individuals, where  plays the role of a generalized price vector for the set of individual commodities and is interchangeable with  in the theoretical model). This specification has not been considered in econometrics and utility theory in the past because the exact density function and moment generating functions for the maximum of multivariate normally distributed variables were not established until very recently. Specifically, it was not until the research of Arellano-Valle and Genton (2008) and Jamalizadeh and Balakrishnan (2009, 2010) that an exact density function and moment generating function was obtained for the maximum of arbitrarily dependent normally distributed random variables. These works show that the distribution of , when has a general multivariate normal distribution, is a mixture of unified univariate skew-normal distribution functions, and then use this mixture representation to derive the density and moment generating functions of  (in doing so, they invoke the density and moment generating functions of the unified univariate skew-normal distribution functions). In this paper, we derive, apparently for the first time, expressions for the density and the moment generating functions for  directly from first principles (rather than going through the circuitous route of using a mixture representation) and explicitly write out these expressions for (these are buried within the expressions for the general distribution of order statistics in Jamalizadeh and Balakrishnan, 2010). Also, we have not seen an expression for the first moment (or expected value) of  in the literature, which is important because that is the expression for  in our econometric model. We explicitly derive this expression from the moment generating function of . These results are collected below as Theorem 2.

*Theorem 2*

Let  be the sub-vector of  without the *i*th element, let  be the *i*th element of , let  be the sub-matrix of  without the *i*th row and the *i*th column, let  be the diagonal entry at the *i*th row and *i*th column of , and let  be the *i*th column of the matrix  minus the *i*th row element.

(1) Denote the probability density function of  by . Then:

 (B.3)

where .

(2) The moment generating function of  is given by:

, (B.4)

where , , , and .

(3) Let  be the vector  minus the *l*th row element,  the *l*th element of the vector ,  the *l*th element of the vector ,  the sub-matrix of  without the *l*th row and the *l*th column,  be the diagonal entry at the *l*th row and *l*th column of ,  be the *l*th column of the matrix  minus the *l*th row element, and the matrix .  (B.5)

With the expected value of  as above, we now present the following theorem that is crucial to the utility-consistent nature of our proposed model.

*Theorem 3*

as defined in Equation (B.5) is both homogeneous of degree one and satisfies the condition .

The fact that is homogeneous of degree one is proved by noting that  corresponds to the expected value of the maximum over random variables that are distributed . Then, by the application of Equation (B.5), we get 

The condition can be proved in many ways. The easiest is to first defineas an  matrix corresponding to an  identity matrix with an extra column of ‘’ values added as the  column. Then, statistically speaking, we can write:

. (B.6)

Then, by the first fundamental theorem of calculus,  and therefore .

Based on the results from Theorem 3, we have proved that setting  with  arising from a multinomial probit formulation for event type provides a theoretic underpinning to integrate the discrete choice model and a count data model into a single integrated utility maximizing framework. In particular, we can now write Equation (B.2) as:

 (B.7)

That is, the demand for commodity *i* is a product of the total count of the units of the commodity group consumed times the probability that commodity *i* is chosen. But everything above is predicated on using  from the MNP model in the count model. Without introducing this linkage, there is no way that prices of individual commodities enter into the total count model, and the resulting model is not utility-consistent. This linkage is precisely what we accomplish in Equation (10) in Section 2 of the paper, but with an important difference. In particular, we recognize that  has a distribution because of the presence of choice model errors. Thus, the precursor to the latent structure part of Equation (10), after reintroducing the index *q* for individuals, is as follows:

, (B.8)

Equation (10) is the net result.

**References**

Arellano-Valle RB, Genton MG. 2008. On the exact distribution of the maximum of absolutely continuous dependent random variables. *Statistics & Probability Letters* 78(1): 27-35.

Deaton A, Muellbauer J. (1980) Economics and Consumer Behavior, Cambridge University Press, Cambridge.

Gorman WM. (1959) Separable utility and aggregation, *Econometrics*, 27: 469-481.

Hausman JA. Leonard GK, McFadden D. 1995. A utility-consistent, combined discrete choice and count data model: Assessing recreational use losses due to natural resource damage. *Journal of Public Economics* 56(1): 1-30.

Jamalizadeh A, Balakrishnan N. 2009. Order statistics from trivariate normal and t-distributions in terms of generalized skew-normal and skew-t-distributions. *Journal of Statistical Planning and Inference* 139(11): 3799-3819.

Jamalizadeh, A., Balakrishnan, N., 2010. Distributions of order statistics and linear combinations of order statistics from an elliptical distribution as mixtures of unified skew-elliptical distributions. *Journal of Multivariate Analysis* 101(6): 1412-1427.

Rouwendal J, Boter J. 2009. Assessing the value of museums with a combined discrete choice/count data model. *Applied Economics* 41(11): 1417–1436.

**Appendix C: SAMPLE FORMATION PROCEDURES**

Several steps were involved in developing the sample used for the empirical analysis. First, only individuals over 18 years of age, and who participated in at least one work activity episode during the survey day on a weekday (Monday to Friday), were selected. Second, we eliminated individuals whose trip diary did not start or end at home. Third, records that contained incomplete information on individual, household, employment-related, and activity and travel characteristics of relevance to the current analysis were removed from the sample. Fourth, several consistency checks were performed and records with missing or inconsistent data were eliminated. The final estimation sample contained 2,113 person observations. Fifth the trip diaries of these 2,113 individuals were processed to obtain, for each individual, the total number of out-of-home non-work episodes undertaken during the survey day, along with the number of these episodes pursued during each of the five time-of-day blocks identified in Section 3.1. Finally, the accessibility measures by the fifteen different industry types were appended to each time-of-day block for each individual as follows. For the before-work (BW) block, the accessibility measures (by industry type) are based off the time the individual would have had to leave home if s/he went directly to work (computed as the individual’s work start time minus the estimated direct home-to-work commute time assuming auto mode of travel and an average speed of 30 mph). That is, the accessibility measures corresponding to the individual’s estimated departure time from home to work (assuming a direct home-to-work trip) and for the residential Census tract of the individual are designated as the home end accessibilities for the BW block. For the home-to-work commute (HWC) block, the accessibility measures are based off the individual’s work start time. For this block, we create two sets of accessibility measures, one for the home end (based on the Census tract of residence) and another for the work end (based on the Census tract of the individual’s workplace location). For the work-based (WB) block, the accessibility measures are based on the off-peak period for the work location Census tract. For the work-to-home commute (WHC) block, the accessibility measures are based off the individual’s work end time. For this block, we once again create both a home end set of accessibilities as well as a work end set of accessibilities. For the after home arrival from work (AH) block, the accessibilities are based off the time the individual would have arrived home if s/he went directly back home from work (computed as the individual’s work end time plus the estimated direct work-to-home commute time assuming auto mode of travel and an average speed of 30 mph). That is, the accessibility measures corresponding to the estimated arrival time back home and for the residential Census tract of the individual (assuming a direct work-to-home trip) are designated as the home end accessibilities. It is important to note that the accessibility measures, as discussed above, vary across the different time-of-day blocks for the same individual.

Table C.1 provides an unweighted summary of select individual, household, work-related and activity and travel characteristics of the final sample.

**Table C.1 Sample Characteristics**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | | **Share [%]** | **Variable** | | | | **Share [%]** | |
| **Individual characteristics** | |  | **Household characteristics** | | | |  | |
| *Race and ethnicity* | |  | *Household income [US$/year]* | | | |  | |
|  | Non-Hispanic Caucasian | 71.56 |  | Less than 80,000 | | | 46.66 | |
|  | Hispanic | 9.99 |  | 80,000 or more | | | 53.34 | |
|  | Non-Hispanic Asian | 9.37 | *Home location* | | | |  | |
|  | Non-Hispanic African-American | 4.45 |  | Urban cluster | | | 94.18 | |
|  | Non-Hispanic Other[[1]](#footnote-1) | 4.63 |  | Not in urban cluster | | | 5.82 | |
| *Gender* | |  | **Work-related characteristics** | | | |  | |
|  | Male | 52.25 | *Employment Industry* | | | |  | |
|  | Female | 47.75 |  | Professional, managerial or technical | | | 48.62 | |
| *Driver status* | |  |  | Sales or services | | | 23.32 | |
|  | Has driver’s license | 98.58 |  | Clerical or administrative support | | | 14.59 | |
|  | Does not have a driver’s license | 1.42 |  | Other[[2]](#footnote-2) | | | 13.47 | |
| *Highest education level* | |  | Is self-employed | | | | 9.51 | |
|  | At least some college education | 76.53 | Has flexible work start time | | | | 44.87 | |
|  | No college education | 23.47 | Has more than one job | | | | 9.13 | |
| *Past week primary activity* | |  | Has the option to work at home | | | | 13.06 | |
|  | Work | 94.18 | **Activity and travel characteristics** | | | |  | |
|  | Other activity | 5.82 | Survey day is Friday | | | | 17.79 | |
| *Shopped via internet in past month* | |  | Used public transportation on survey day | | | | 3.98 | |
|  | No | 57.31 | At least one walk trip in past week | | | | 63.98 | |
|  | Yes | 42.69 | At least one bike trip in past week | | | | 6.58 | |
| **Descriptive statistics** | | | | | | | | |
| **Variable** | | | **Mean** | | **Std. Dev.** | **Min.** | | **Max.** |
| **Individual characteristics** | | |  | |  |  | |  |
|  | Age [years] | | 46.67 | | 12.70 | 18.00 | | 86.00 |
| **Household characteristics** | | |  | |  |  | |  |
|  | Number of adults | | 2.40 | | 0.92 | 1.00 | | 7.00 |
|  | Number of non-adults | | 0.74 | | 1.05 | 0.00 | | 6.00 |
|  | Number of drivers | | 2.33 | | 0.92 | 0.00 | | 7.00 |
|  | Number of vehicles | | 2.59 | | 1.30 | 0.00 | | 12.00 |
|  | Number of workers | | 1.84 | | 0.82 | 1.00 | | 5.00 |
| **Work-related characteristics** | | |  | |  |  | |  |
|  | Distance to work [miles] | | 13.52 | | 12.56 | 0.11 | | 97.00 |
| **Dependent variable: Number of out-of-home non-work episodes** | | | | | | | | |
| **Time-of-day block** | | | **Mean** | | **Std. Dev.** | **Min.** | | **Max.** |
| Before-work (BW) | | | 0.12 | | 0.44 | 0.00 | | 6.00 |
| Home-to-work commute (HWC) | | | 0.20 | | 0.56 | 0.00 | | 11.00 |
| Work-based (WB) | | | 0.23 | | 0.47 | 0.00 | | 4.00 |
| Work-to-home commute (WHC) | | | 0.43 | | 0.83 | 0.00 | | 6.00 |
| After-home (AH) | | | 0.56 | | 1.12 | 0.00 | | 12.00 |
| Total non-work episodes | | | 1.54 | | 1.67 | 0.00 | | 13.00 |

**APPENDIX D: MODEL FIT ASSESSMENT USING PREDICTIVE MEASURES**

First, define   as an  matrix that corresponds to an  identity matrix with an extra column of ’s added as the  column. Following the notation in Equation (10) and immediately after, define . We can then write the probability that individual (consumer) *q* chooses alternative *i* at any choice occasion as:

|  |  |
| --- | --- |
| . | (D.1) |

Next, since this probability does not change across choice occasions, and the individual-specific preferences are already embedded in  (through the  vector), the multivariate probability of counts in each time-of-day block (*i.e*., event type), conditional on the total count level  ), takes the usual multinomial distribution form:

|  |  |
| --- | --- |
|  | (D.2) |

In our joint model of multivariate counts, the unconditional multivariate probability then takes the form indicated below (****, ****, ****):

|  |  |
| --- | --- |
|  | (D.3) |

with  as in Equation (13) after replacing (the actual observed total count for individual *q* in the estimation sample) with an arbitrary value . Using the properties of the multinomial distribution, the marginal probability of  counts for time-of-day block *i* is:

|  |  |
| --- | --- |
|  | (D.4) |

In the above expression, the upper bound of the summation is , though the probability values fade very rapidly beyond a  value of 10. For the purposes of this paper, we carry the summation up to 

With the above preliminaries, the model predictions can be used to evaluate data fit at both the disaggregate and aggregate levels, as well as for both the multivariate count distribution and the marginal count distribution. At the disaggregate level, we estimate the probability of the observed multivariate count outcome for each individual using Equation (B.3), and compute an average probability of correct prediction. Similarly, we also estimate the probability of the observed marginal count outcome separately for each time-of-day period using Equation (B.4), and compute an average probability of correct prediction. At the aggregate level, we design a heuristic diagnostic check of model fit by computing the predicted aggregate share of individuals for specific multivariate outcome cases (because it would be infeasible to provide this information for each possible multivariate outcome). In particular, we compute the aggregate share of consumers for each of six combinations. The first combination corresponds to no participation in any non-work episodes (which we will refer to as the “no participation” combination). The other five combinations correspond to participation in one or more episodes during a specific time-of-day block and no participation in any other time-of-day period (which we will refer to using such labels as the “BW participation only” combination or the “HWC participation only” combination). In addition to these aggregate shares of multivariate outcomes, we also compute the aggregate shares of the marginal outcomes of count values of 0, 1, 2, 3, and 4+ for each time-of-day period, as well as for the total count. As a yardstick to evaluate the performance of the joint model proposed here, we compare the predictions from the joint model with the independent model using the absolute percentage error (APE) statistic for each count value, and then compute a mean weighted APE value across the count values (of 0 1, 2, 3, and 4+) using the observed number for each count value as the weight for that count value.

The disaggregate-level data fit measures indicate an average probability of correct prediction of 13.9% for the multivariate counts and an average probability of correct prediction of 67.6% for the marginal counts. The corresponding values for the independent model are 13.6% and 65.0%, respectively, which are smaller in magnitude than those from the joint model. The aggregate fit measures are provided in Table D.1. The joint model provides a better (lower) APE value for all the multivariate outcomes in Table D.1 (see upper panel of the table), except for the WB participation only outcome. The APE values are sizeable for both the joint and independent values, but it should be noted that these predictions are for multivariate outcomes. Overall, the mean weighted APE value is about 12% higher for the independent model relative to the joint model. As expected, the APE values are lower for the marginal outcomes (see lower panel of Table D.1) than for the multivariate outcomes. The total count predictions from the joint model are much better than the total count predictions from the independent model. Also, the predictions for the other marginal counts are better from the joint model relative to the independent model (except for the WB block count). These results clearly show that the joint model proposed here outperforms the traditional independent model in the disaggregate level and aggregate level comparisons.

**Table D.1 Aggregate Data Fit Measures**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Aggregation Level** | **Combination Event** | | **Observed** | **Joint Model** | | **Independent Model** | |
| **Predicted** | **APE** | **Predicted** | **APE** |
| **Multivariate** | No participation |  | 676 | 669.7 | 0.9 | 656.1 | 2.9 |
| BW participation only | | 67 | 90.5 | 35.0 | 95.2 | 42.1 |
| HWC participation only | | 67 | 85.2 | 27.1 | 87.0 | 29.9 |
| WB participation only | | 168 | 63.0 | 62.5 | 67.4 | 59.9 |
| WHC participation only | | 230 | 153.6 | 33.2 | 137.0 | 40.4 |
| AH participation only | | 279 | 345.4 | 23.8 | 347.7 | 24.6 |
| **Overall mean weighted APE** | | | **19.9** | | **22.2** | |
| **Marginal** | Total count | 0 | 676 | 669.7 | 0.9 | 656.1 | 2.9 |
| 1 | 593 | 589.1 | 0.7 | 585.7 | 1.2 |
| 2 | 388 | 379.1 | 2.3 | 377.2 | 2.8 |
| 3 | 208 | 225.6 | 8.4 | 253.5 | 21.9 |
| 4+ | 248 | 249.5 | 0.6 | 240.5 | 3.0 |
| Weighted APE | | | 1.8 | | 4.3 | |
| BW block  count | 0 | 1926 | 1756.8 | 8.8 | 1745.3 | 9.4 |
| 1 | 147 | 295.8 | 101.2 | 305.7 | 108.0 |
| 2 | 24 | 47.8 | 99.1 | 50.8 | 111.5 |
| 3 | 13 | 7.6 | 41.5 | 8.1 | 37.7 |
| 4+ | 3 | 5.0 | 66.7 | 3.1 | 3.3 |
| Weighted APE | | | 16.5 | | 17.6 | |
| HWC block count | 0 | 1792 | 1739.1 | 3.0 | 1729.8 | 3.5 |
| 1 | 250 | 317.8 | 27.1 | 326.4 | 30.6 |
| 2 | 57 | 45.8 | 19.6 | 48.1 | 15.7 |
| 3 | 10 | 5.9 | 41.0 | 6.2 | 38.3 |
| 4+ | 4 | 4.4 | 10.0 | 2.5 | 37.5 |
| Weighted APE | | | 6.5 | | 7.3 | |
| WB block  count | 0 | 1660 | 1831.5 | 10.3 | 1809.9 | 9.0 |
| 1 | 421 | 244.8 | 41.9 | 261.2 | 38.0 |
| 2 | 29 | 29.2 | 0.7 | 35.0 | 20.7 |
| 3 | 2 | 3.6 | 77.5 | 4.7 | 133.2 |
| 4+ | 1 | 3.9 | 290.0 | 2.2 | 120.0 |
| Weighted APE | | | 16.7 | | 15.1 | |
| WHC block count | 0 | 1516 | 1560.2 | 2.9 | 1593.4 | 5.1 |
| 1 | 397 | 423.4 | 6.6 | 408.7 | 3.0 |
| 2 | 131 | 97.1 | 25.8 | 87.2 | 33.5 |
| 3 | 45 | 22.0 | 51.0 | 17.6 | 60.8 |
| 4+ | 24 | 10.3 | 57.1 | 6.1 | 74.6 |
| Weighted APE | | | 6.7 | | 8.4 | |
| AH block  count | 0 | 1465 | 1216.0 | 17.0 | 1201.7 | 18.0 |
| 1 | 394 | 576.9 | 46.4 | 589.8 | 49.7 |
| 2 | 90 | 214.3 | 138.1 | 221.7 | 146.3 |
| 3 | 103 | 71.9 | 30.2 | 71.6 | 30.5 |
| 4+ | 61 | 33.9 | 44.4 | 28.2 | 53.8 |
| Weighted APE | | | 29.1 | | 31.0 | |

**APPENDIX E: MODEL APPLICATION**

The joint model estimated in the paper can be used to examine the impact of changes in socio-demographic characteristics over time as well as the effects of policy actions that involve a change in the accessibility measures and work-related characteristics. In this paper, we demonstrate the application of this model by studying the effects of changes in three selected variables: distance to workplace, retail trade accessibility at the home location, and entertainment accessibility at the home location. These three variables are increased by 20% across all workers.

The impact on the frequency and organization of non-work activities is estimated by determining the percentage change in the expected number of non-work episodes (across all workers) for the entire day (*i.e.*, total count) and for each time-of-day block. To demonstrate the potentially misleading inferences from the independent model, we compute the percentage change as predicted by both the joint model as well as the independent model. The emphasis here is not on substantive empirical inferences as much as it is on demonstrating the differences in the inferences from the two models. Table E.1 provides the results.

Three observations may be made from Table E.1. First, in the independent model, a change in the retail trade and entertainment accessibility variables do not have any impact on the total count of non-work episodes over the entire day. This is, of course, because these variables appear only in the event discrete choice model and not the total count model (and the independent model does not have any link between the discrete choice model and the total count model). As indicated earlier in the paper, it is natural to expect that changes in the attributes impacting the attractiveness of alternatives in the choice model (retail trade and entertainment accessibilities in the specific case under discussion) will result not only in substitution among the counts of each discrete choice alternative, but also an overall change in the total count, as appropriately recognized by the joint model. Second, the positive effect (of an increase in the number of retail trade jobs) on the number of non-work tours during the BW, HWC, WHC, and AH periods is underestimated by the independent model, while the negative effect of the variable on WB non-work tours is overestimated by the independent model. Third, a similar result holds also for the influence of the number of entertainment jobs at the home location. Indeed, for this variable, the directionality of the effect on WHC non-work tours is itself different between the independent and joint models. These differences between the models highlight the potentially misinformed policy analyses that result from ignoring the linkage between the frequency of non-work episodes and their organization across time-of-day blocks.

**Table E.1 Aggregate Percentage Change in Expected Number of Non-Work Episodes**

|  |  |  |  |
| --- | --- | --- | --- |
| **Effect of 20% increase in …** | **Time-of-Day Block** | **Joint**  **Model** | **Independent Model** |
| Distance to work | All day | -2.51 | -2.30 |
| BW | -2.64 | -2.44 |
| HWC | -2.52 | -2.32 |
| WB | -2.30 | -2.08 |
| WHC | -2.46 | -2.27 |
| AH | -2.46 | -2.27 |
| Number of retail trade jobs at the home location | All day | 0.53 | 0.00 |
| BW | 0.75 | 0.31 |
| HWC | 0.85 | 0.40 |
| WB | -1.30 | -2.34 |
| WHC | 0.62 | 0.06 |
| AH | 0.69 | 0.31 |
| Number of entertainment jobs at the home location | All day | 3.44 | 0.00 |
| BW | 8.31 | 5.23 |
| HWC | 5.93 | 2.32 |
| WB | -2.41 | -6.94 |
| WHC | 1.52 | -1.75 |
| AH | 3.27 | 0.37 |

1. Non-Hispanic Other includes American Indian, Alaskan Native (1.23%), Native Hawaiian, or other Pacific Islander (0.52%), Multiracial (0.70%), and other (that is, specified in the survey capture itself as a catch all “other” category (2.18%) [↑](#footnote-ref-1)
2. This other category includes Manufacturing, construction, maintenance or farming (12.62%) and other (that is, specified in the survey capture itself as a catch all “other category”): 0.85% [↑](#footnote-ref-2)