Estimation of Choice Models for RELU Using Aggregate Data

Alex Anas and Debarshi Indra
Department of Economics
State university of New York at Bufalo

October 20, 2011

1. THE DATA

This section gives a detailed description of the preparation of the data set used in this study. The data set contains geographical, social and economic information for the year 2000 at the census tract level for six counties in the greater Los Angeles area. These counties are Imperial, Los Angeles, Orange, Riverside, San Bernardino and Ventura. These six counties together contain 3402 census tracts.

We put together the data set by combining information from CTPP2, CTPP3 and census bureau's SF3 file. CTPP stands for "Census Transportation Planning Package". CTPP provides various tabulations of census data. These tabulations are targeted at the transportation planning community as they offer a rich source of information on commute characteristics. CTPP2 provides journey to work data at the worker's place of work. CTPP3 also gives journey to work data but by the flow of workers between their residence and workplace for various geographical levels. For this study we use the flows at the census tract level. The census bureau's SF3 file augments the CTPP data by providing information on social and economic characteristics for the census tracts. In addition, it also provides data on certain geographical features of the tracts like its land area, latitude and longitude. All data have been aggregated from micro data to the tract level by the census bureau.

For this study we consider four modes of commuting which the census bureau calls **drove alone**, **2 or more person carpool**, **public transportation-excluding taxicab**, **all other excluding taxicab and worked at home**. We abbreviate these as **DA**, **CP**, **PT** and **RT** respectively. The reason for choosing this mode aggregation is that CTPP3 provides tabulation of worker flows between tracts by income groups for this mode of aggregation. The income groups for which this tabulation is done are, **all income**; **less than $30,000**; **$30,000 to $49,999**; **$50,000 to $74,999** and **greater than $75,000**. In table 1 $\boldsymbol{N_{DA}, N_{CP}, N_{PT}}$ and $\boldsymbol{N_{RT}}$ refer to the flow of *all* (household) workers (between a pair of census tracts) by mode of commute.

The CTPP3 gives aggregate commute times[1] for each of the four modes mentioned above. However, since we know the number of (total) workers commuting between two tracts for each mode, we can easily calculate the mean commute time[2] for each mode for each pair of census tracts. We label these mean commute times as $MTT_{DA}, MTT_{CP}, MTT_{PT}$ and MTT_{RT} respectively.

In summary, CTPP3 provides data on the flow of workers by mode of commute and income group and enables us to derive the average commute times for each mode between two tracts. The CTPP3 file acts as an anchor, to which information from CTPP2 and SF3 files are added. Recall that CTPP3 gives the flow of workers between their residence and workplace census tracts. Information about the workers' workplace

---

[1] This is the aggregate one way daily commute time for all workers traveling between two tracts.
[2] Mean commute time is simply aggregate commute time divided by the number of workers commuting.

tract is obtained from CTPP2. The data from CTPP2 relevant for this study is included in the census tabulation of “Industry by worker earnings in 1999”. This CTPP2 tabulation provides invaluable information about the number of workers employed in a census tract by industry[3] as well as the earnings distribution of workers within each industry. The CTPP2 tabulation was used to derive the average wage for each industry and an accessibility measure for the study area for each census tract.

$$W_{i,j} = \frac{1}{260 * 8 * 60}\left(\frac{1}{TW_{i,j}}\sum_{ear} ear_{mid} TW_{i,j,ear}\right)$$

In the above expression, $W_{i,j}$ refers to the average wage (per minute) in industry i, in census tract j. $TW_{i,j}$ denotes the total number of workers employed in that industry in the census tract while $TW_{i,j,ear}$ refers to the number of workers employed in the industry in earnings category ear. ear_{mid} refers to the midpoint of earnings category[4] ear.

Note that a census tract serves both as a commercial and a residential area. This means that information from CTPP2 can be combined with the residential side of workers census tract. This is precisely what is done with the accessibility measure created from the CTPP2 tabulation. The accessibility of residential census tract i is defined as,

$$ACCESS_{RET,i} = \sum_{j}\left(\frac{RW_j}{LAND_j} e^{-\theta d_{ij}}\right)$$

[3] The industry and earning categories are given in the appendix.

[4] As noted earlier, the earning categories are $0-$29,999; $30,000-$49,999;$50,000-$74,999 and $75,000 and above.

where RW_j and $LAND_j$ denote the total number of retail workers employed and the total land area in tract j respectively. d_{ij} is the (Great Circle) distance in miles between tract i and j, while θ is a positive constant to be calibrated.

The data set was completed by including socio economic information for each census tract from the census bureau's SF3 file. This data was added to the residential side of the CTPP3 data set. The census SF3 file was used to derive for each census tract an economic rent index (MHI, mean housing index) that captures the annual housing cost per dwelling unit, combining both rental and owner occupied housing units. The formula combines 12 times the monthly rent of rental occupied dwelling units with the annualized reported market value of owner-occupied units. The annualized market value is the Census average market value in the census tract multiplied by a capitalization rate.

Table 1a below gives some descriptive statistics for the variables by census tract that were included in the final data set.

VARIABLE	MEAN	MINIMUM	MAXIMUM	STANDARD DEVIATION
$ACCESS\_RET_i$	237736.86	5.94	460526.51	138045.02
$DIST$	7.9	0	184.80	8.46
MHI	49453.78	0	200001.00	24562.75
MTT_{CP}	28.53	1	350	24.82
MTT_{DA}	24.12	1	350	17.21
MTT_{PT}	50.12	1	350	37.84
MTT_{RT}	21.14	1	350	26.67
MNR	4.6	0	9.10	1.39
$MYSB$	1956	0	1999	135.09
N_{CP}	7.03	0	620	12.45
N_{DA}	32.59	0	2405	36.35
N_{PT}	1.88	0	385	6.97
N_{RT}	5.19	0	920	25.53
$RENT$	16700	1188	97260	11910.68
TWL	1034.24	0	31280	1842.97
$TWMe$	436.85	0	21635	985.88
$TWMi$	255.47	0	13070	647.60
TWH	193.05	0	14525	591.36
$WAGE_{TOT}$ ($/min)	0.23	0.08	0.44	0.04

Table 1a: DESCRIPTIVE STATISTICS

COUNTY	log(ACCESS_RET)	MHI
Imperial	7.246136	37534.41
LA	12.588998	49460.74
Orange	12.135663	61863.23
Riverside	10.280985	46415.03
San Bernardino	10.845686	46639.32
Ventura	10.925653	62471.00

TABLE 1b: THE ACCESSIBILITY TO RETAIL AND THE MHI BY COUNTY

Table 1b shows the value of the accessibility measure averaged for each county. As we can see LA is the most accessible county to live in, while Ventura and Orange are the most expensive in terms of the MHI.

2. ESTIMATION FOR THE RELU MODEL

We now report the results on the estimation o a worker's joint choice of workplace, residence and mode of commute which forms the foundation of "RELU", a computable general equilibrium model of regional economy and land use (see Anas and Liu, 2007). The RELU utility function to be used in the CGE model is not necessarily identical to the one estimated here, but the estimates here will form the basis for fixing the elasticity and other measures in the RELU CGE model. The estimation methodology implemented here follows that of Anas (1981) and Anas and Chu (1984).

Following the RELU model, space is discrete. For estimation purposes, the Greater Los Angeles area is divided into 3402 distinct zones where each zone corresponds to a census tract. The utility of a worker residing in a household from choosing zone i as workplace, zone j as residence and mode m for commute between i and j is given by U_{ijm}. We assume for convenience a linear-in-parameters utility function.

$$U_{ijm} = \delta X_i + \gamma Y_{ij} + \beta Z_{ijm} + \varepsilon_i + \varepsilon_{ij} + \varepsilon_{ijm} \quad i = Worlplace, j = Residence, m = Mode$$

U_{ijm} depends on observable attributes that vary by mode-residence location-job location, z_{ijm}, residence and job location Y_{ij}, and job location X_i. However, utility also depends on unobserved to the econometrician, factors ε that are random variables assumed to vary among the workers. Since utility is a random variable we can derive only the probability of a worker choosing (i, j, m) from a set of combinations of work zone, residence zone and mode of commute to work. This probability P_{ijm} is written as the product of conditional and marginal probabilities as shown above. $P_{m|ij}$ is the probability of choosing a mode, conditional on the choice of work zone and residence zone, $P_{j|i}$ is the probability of choosing residence zone conditional on the choice of a work zone, while P_i is the marginal probability of choosing a work zone.

$$P_{ijm} = P_{m|ij} P_{j|i} P_i$$

If the vector of unobserved utilities is jointly distributed Generalized Extreme Value (GEV) with a particular structure of the covariance matrix (see Train, 2003) then the conditional and marginal probabilities all become multinomial logit. This model is called the “Nested Logit Model (NLM)”. Under certain conditions the NLM has been shown to be consistent with utility maximization and the “Multinomial Logit Model (MNL)” is a special case of the NLM when the off diagonal terms in the covariance matrix are all zero.

$$P_{r|ij} = \frac{e^{\beta Z_{ijr}}}{\sum_{m|ij} e^{\beta Z_{ijm}}} \quad m = \text{ mode of transportation}$$

$P_{r|ij}$ is the conditional probability of mode choice given residence and job location. This conditional on the choice of i and j the mode choice probability is multinomial logit.

$$P_{s|i} = \frac{e^{\lambda I_{MODE_{is}} + \gamma Y_{is}}}{\sum_{j|i} e^{\lambda I_{MODE_{ij}} + \gamma Y_{ij}}} \quad j = resident\ census\ tract,$$

$$I\_MODE_{ij} = \ log\left(\sum_{m} e^{\beta Z_{ijm}}\right)$$

The conditional probability of residence zone choice $P_{s|i}$ also takes the multinomial logit form and is linked to $P_{r|ij}$ by I_MODE_{ij}. The term I_MODE_{ij} which appears as an explanatory variable in the utility function for residence choice is the log of the denominator of the conditional mode choice probability. This term is the expected utility from the mode choice decision once the worker chooses residence zone j conditional on workplace i. The coefficient λ is an approximate measure of the degree of correlation between the unobserved random utilities of the alternative commuting modes conditional on the choice of (i, j). In general λ could vary by (i, j), but for our study we will consider it to be constant across (i, j). Finally, the outer marginal probability of work zone choice $P_{s|i}$ also takes the Logit form and is linked to the conditional probability of residence zone choice by the log of the denominator of $P_{s|i}$.

$$P_t = \frac{e^{\mu I\_HOME_t + \delta X_t}}{\sum_i e^{\mu I\_HOME_i + \delta X_i}} \quad i = workplace\ census\ tract, \ \ I\_HOME_i = \ log\left(\sum_{j|i} e^{\lambda I\_MODE_{ij} + \gamma Y_{ij}}\right)$$

I_HOME_i is the utility the worker can expect from residence zone choice once the work zone is selected. The coefficient μ again is an approximate measure of the degree of correlation among the alternative residence zones conditional on the work zone choice.

For the NLM to be consistent with utility maximization the values of λ and μ must be between 0 and 1. A value of greater than one for any one of these coefficient is still consistent with utility maximization but only for certain range of values of the explanatory variables (see Train 2003). A value of less than 0 is inconsistent with utility maximization. If λ and μ are both equal to one, then the NLM is equivalent to the un-nested MNL (also known as the "simultaneous logit"), with no correlation across unobserved utilities.

3. ESTIMATION METHOD

The NLM mentioned in the previous section is estimated on aggregate data at the census tract level (see Anas (1981)) and Anas and Chu (1984) who also used the CTPP data of the 1970 Census. We follow a very similar procedure here.

For a particular (i, j, m) we observe the number of workers (living in households) who select that alternative, but we do not observe the attributes $\left(X_i, Y_{ij}, Z_{ijm}\right)$ of the alternative at the individual level. We observe the averages of the attributes by census tract which we label as $\left(\bar{X}_i, \bar{Y}_{ij}, \bar{Z}_{ijm}\right)$. Suppose we assume that,

$$X_i = \bar{X}_i + \xi_i, , Y_{ij} = \bar{Y}_{ij} + \xi_{ij}, where\ , Z_{ijm} = \bar{Z}_{ijm} + \xi_{ijm}$$

$$where\ E(\xi_i) = 0\ and\ E\left(\xi_{ij}\right) = 0\ and\ E\left(\xi_{ijm}\right) = 0.$$

In that case the utility function from the previous section becomes,

$$U_{ijm} = \delta\bar{X}_i + \gamma\bar{Y}_{ij} + \beta\bar{Z}_{ijm} + \varepsilon_i + \varepsilon_{ij} + \varepsilon_{ijm} + \delta\xi_i + \gamma\xi_{ij} + \beta\xi_{ijm},$$

$$U_{ijm} = \delta\bar{X}_i + \gamma\bar{Y}_{ij} + \beta\bar{Z}_{ijm} + \zeta_{ijm}.$$

If the vector ζ satisfies the assumptions mentioned in the previous section then the NLM with aggregate data is equivalent to NLM with individual data.

The NLM is estimated by a three step sequence. This yields a consistent but inefficient estimator of the parameters. The inefficiency arises because some information is lost when the parameters are not estimated jointly in a single step. However, it is well-known (see, for example, Anas and Chu (1984)) that one-step estimation can result in non-unique estimates since the one-step log-likelihood function is not concave in general.

LL_m, LL_j and LL_i are the log likelihood functions corresponding to mode choice, residence zone choice and workplace zone choice respectively.

$$LL_m = \sum_i \left[\sum_{j|i} \left[\left(\sum_{m|ij} (N_{ijm} \beta Z_{ijm}) \right) - N_{ij} log \left(\sum_{m|ij} \Delta_{ijm} e^{\beta Z_{ijm}} \right) \right] \right],$$

$$\Delta_{ijm} = 1 \; if N_{ijm} > 0 \; else \; \Delta_{ijm} = 0 \quad ;$$

$$N_{ij} = \sum_m N_{ijm} \;, \; N_{ijm} = Number \; of \; workers \; using \; mode \; m \; between \; i, j$$

In the expression for LL_m , Δ_{ijm} is a dummy variable that takes the value one to indicate that mode m is available between i and j. Δ_{ijm} equal to zero means that the mode is not available between i and j. In our study, mode m is assumed to be not available as a commute alternative between (i, j) if the observed number of workers using mode m between (i, j) is zero.

$$LL_j = \sum_i \left(\sum_{j|i} N_{ij} \left[\left(\lambda I\_MODE_{ij} + \gamma Y_{ij} \right) - log \left(\sum\nolimits_{j|i} e^{\lambda I\_MODE_{ij} + \gamma Y_{ij}} \right) \right] \right)$$

When estimating the residence zone choice probability model, the choice set of a worker observed to work in zone i includes his chosen residence census tract j and all other tracts from which other workers were observed to commute to census tract i. This method of restricting an agent's choice set is however valid. We know (see Train 2003) that in estimating a Logit model all the alternatives need not appear in the agent's choice set. Estimation on a subset of the choice set under certain conditions will yield consistent but inefficient estimators.

$$LL_i = \sum_i N_i \left[\left(\mu I_{HOME_i} + \delta X_i \right) - log \left(\sum_i e^{\mu I_{HOME_i} + \delta X_i} \right) \right]$$

$$N_i = \sum_{j|i} \sum_{m|ij} N_{mj}$$

Each worker's work zone choice set included all census tracts in the data set with positive employment.

4. ESTIMATION RESULTS

Selected results from the estimations are now presented. Table 2 below lists the variables used in the estimation and the definitions.

Variable	Source	Derived	Description
ACCESS_RET	CTPP2	YES	$\text{ACCESS}_{\text{RET}_i} = \sum_j \left(\frac{RE_j}{LAND_j}\right) e^{-0.0723 d_{ij}}, RE_j = RETAIL\ EMPLOYMENT\ IN\ CENSUS\ TRACT\ j,$ $LAND_j = LAND\ AREA\ (SQ\ MILES)\ OF\ CENSUS\ TRACT\ j, d_{ij} = GCD\ BETWEEN\ i\ AND\ J$
AGE	CENSUS SF3	NO	AGE = 2000-MEDIAN YEAR STRUCTURE BUILT
CP	NA	NA	CARPOOL SPECIFIC DUMMY
DA	NA	NA	DROVE ALONE SPECIFIC DUMMY
DIST	NA	NA	GREAT CIRCLE DISTANCE BETWEEN TWO CENSUS TRACTS. COMPUTED FROM THEIR LATITUDE AND LONGITUDE INFORMATION.
D_VT_RES	NA	NA	VENTURA COUNTY RESIDENCE DUMMY
D_LA_RES	NA	NA	LOS ANGELES COUNTY RESIDENCE DUMMY
D_OR_RES	NA	NA	ORANGE COUNTY RESIDENCE DUMMY
D_RV_RES	NA	NA	RIVERSIDE COUNTY RESIDENCE DUMMY
D_SB_RES	NA	NA	SAN BERNARDINO COUNTY RESIDENCE DUMMY
D_VT_WOK	NA	NA	VENTURA COUNTY WORKPLACE DUMMY
D_LA_WOK	NA	NA	LOS ANGELES COUNTY WORKPLACE DUMMY
D_OR_WOK	NA	NA	ORANGE COUNTY WORKPLACE DUMMY
D_RV_WOK	NA	NA	RIVERSIDE COUNTY WORKPLACE DUMMY
D_SB_WOK	NA	NA	SAN BERNARDINO COUNTY WORKPLACE DUMMY
D_RES_WOK	NA	NA	DUMMY VARIABLE TO INDICATE SAME RESIDENCE WORKPLACE COUNTY
HUM	SF3	NA	HOUSING UNITS WITH MORTGAGE

HUNM	SF3	NA	HOUSING UNITS WITHOUT MORTGAGE
INCOME_TOT	CTPP2	YES	WEIGHTED AVERAGE INCOME IN CENSUS TRACT OF EMPLOYMENT = $\frac{1}{TW_{i,j}} \sum_{ear} ear_{mid} TW_{i,j,ear}$
I_MODE	NA	NA	"log sum term" from MODE CHOICE NEST
I_HOME	NA	NA	"log sum term" from RESIDENCE CHOICE NEST
MCR	SF3	NA	MEDIAN CONTRACT RENT
MHI	CENSUS SF3	NO	MEDIAN HOUSEHOLD INCOME (1999) IN CENSUS TRACT OF RESIDENCE
MNR	CENSUS SF3	NO	MEDIAN NUMBER OF ROOMS IN HOUSING UNITS IN CENSUS TRACT
MTT_DA	CTPP3	YES	$MEAN\ TRAVEL\ TIME\ (DROVE\ ALONE) = \frac{TAB314X05}{TAB306X02}$
MTT_CP	CTPP3	YES	$MEAN\ TRAVEL\ TIME\ (CARPOOL) = \frac{\text{TAB314X09} + \text{TAB314X13}}{\text{TAB306X03} + \text{TAB306X04} + \text{TAB306X05} + \text{TAB306X06} + \text{TAB306X07}}$
MTT_PT	CTPP3	YES	$MEAN\ TRAVEL\ TIME\ (TRANSIT) = \frac{\text{TAB314X17} + \text{TAB314X21}}{\text{TAB306X08} + \text{TAB306X09} + \text{TAB306X10} + \text{TAB306X11} + \text{TAB306X12}}$
MTT_RT	CTPP3	YES	$MEAN\ TRAVEL\ TIME\ (OTHERS) = \frac{\text{TAB314X25}}{\text{TAB306X13} + \text{TAB306X14} + \text{TAB306X15} + \text{TAB306X16} + \text{TAB306X17}}$
PT	NA	NA	PUBLIC TRANSPORTATION SPECIFIC DUMMY
RT	NA	NA	"OTHER" MODE SPECIFIC DUMMY
RENT	CENSUS SF3	YES	$\frac{(((\text{HUM} + \text{HUNM}) * \text{SMV} * \mathbf{0.1}) + (\text{TROHU} * \mathbf{12} * \text{MCR}))}{(\text{HUM} + \text{HUNM} + \text{TROHU})}$
SMV	SF3	NA	SPECIFIED MEDIAN VALUE OF OWNER OCCUPIED HOUSING
TC	NA	YES	TOTAL ANNUAL OPPURTUNITY COST OF COMMUTING = WAGE($/min)*MTT(min/year)
THU	CENSUS SF3	YES	TOTAL HOUSING UNITS IN CENSUS TRACT = OCCUPIED + VACANT
TROHU	SF3	NA	TOTAL RENTER OCCUPIED HOUSING UNITS
TWL	CTPP2	YES	TOTAL WORKERS IN A CENSUS TRACT IN INCOME GROUP $0 TO $29,999
TWMe	CTPP2	YES	TOTAL WORKERS IN A CENSUS TRACT IN INCOME GROUP $30,000 TO $49,999
TWMi	CTPP2	YES	TOTAL WORKERS IN A CENSUS TRACT IN INCOME GROUP $50,000 TO $74,999
TWH	CTPP2	YES	TOTAL WORKERS IN A CENSUS TRACT IN INCOME GROUP $75,000 AND ABOVE
TW	CTPP2	NO	TOTAL WORKERS EMPLOYED IN EACH CENSUS TRACT
WAGE	NA	YES	$WAGE = \frac{INCOME\_TOT}{260 * 8 * 60}$

TABLE 2: VARIABLES USED IN ESTIMATION

Tables 3a and 3b present the estimation results of the three stages and includes the goodness of fit measure, rho-squared, which is cumulative over the stages. According to Train (2007), the goodness of fit measure is called the likelihood ratio index. It is one minus the ratio of the log-likelihood function evaluated at the estimated solution which maximizes likelihood and the log-likelihood evaluated when all parameters are set to zero.

$$1 - \frac{LL(\hat{\beta})}{LL(\hat{0})}$$

If the model fits perfectly then the likelihood ratio will be close to zero and the index will be equal to one. If the model fit is very badly then the ratio will be close to one and the index will be zero. But he stresses the point that this index has no obvious interpretation, unlike R^2 in regression analysis. An index of goodness that has such an interpretation is rho-squared:

$$\rho^2 = \frac{LL(\hat{\beta}) - LL(0)}{LL(*) - LL(0)}$$

$LL(\hat{\beta})$: the value of the log-likelihood when the choice probabilities are evaluated at the estimated coefficients which maximize log likelihood;

$LL(0)$: the value of the log-likelihood when all estimated coefficients are set to zero, so the probabilities are equal. This then is the value of log-likelihood under the null hypothesis;

$LL(*)$: the value of the log likelihood when the probabilities obtain the values given by the relative frequencies in the data. That is assuming the probabilities perfectly replicate the shares. This is also the highest possible value the log-likelihood could have achieved in estimation.

Note then that the denominator of the rho-squared is the total possible improvement over the null hypothesis (that the model explains nothing) and the numerator is the amount of improvement actually achieved by the estimated model. The rho squared above is then between zero and 1. The rho-squared can be calculated for each stage and also cumulatively for a particular stage and the stage lower than that.

	ALL INCOME	INCOME 1	INCOME 2	INCOME 3	INCOME 4
MODE CHOICE					
CP	-0.912 (-523.13)	-0.293 (-46.24)	-0.370 (-65.10)	-0.426 (-78.88)	-0.823 (-220.30)
PT	-0.797 (-217.74)	-0.098 (-9.05)	-0.356 (-26.99)	-0.415 (-28.10)	-0.829 (-68.49)
RT	-0.255 (-115.17)	0.185 (25.13)	-0.056 (-7.03)	-0.062 (-7.83)	0.085 (16.60)
MTT	-0.008 (-146.80)	-0.005 (-28.14)	-0.005 (-26.62)	-0.005 (-26.33)	-0.006 (-50.61)
Rho-squared (stage 1)	**0.38**	**0.10**	**0.13**	**0.16**	**0.36**
RESIDENCE LOCATION CHOICE					
I_MODE	0.674 (320.89)	0.964 (157.67)	1.02 (165.12)	1.09 (177.17)	0.97 (215.81)
Log(THU)	0.484 (324.66)	0.328 (62.35)	0.319 (67.69)	0.335 (76.51)	0.44 (146.27)
D_LA_RES	-0.214 (-66.65)	-0.069 (-4.75)	-0.005 (-0.47)	-0.078 (-9.31)	-0.158 (-30.32)
D_RES_WOK	0.126 (46.19)	-0.019 (-1.72)	0.088 (10.30)	0.069 (9.83)	0.073 (15.49)
Log(DIST) /Log(ACCESS_RET)	-5.558 (-556.06)	-2.575 (-69.99)	-2.389 (-72.73)	-2.186 (-68.41)	-3.842 (-146.06)
Log(AGE)	-0.144 (-93.06)	-0.148 (-21.52)	-0.124 (-21.93)	-0.116 (-24.91)	-0.146 (-55.88)
Log(RENT)	-0.187 (-52.41)	-0.085 (-6.82)	-0.413 (-32.83)	-0.475 (-41.86)	-0.009 (-1.47)
Log(MHI)	0.311 (65.87)	-0.363 (-27.56)	0.135 (8.76)	0.366 (24.52)	0.587 (56.55)
Log(MNR)	0.124 (25.11)	-0.093 (-6.12)	0.034 (2.27)	0.135 (9.09)	0.219 (21.09)
Rho-squared (stage 2)	0.53	0.53	0.45	0.42	0.46
JOB LOCATION CHOICE					
I_HOME	0.342 (219.36)	0.941 (242.45)	0.788 (186.81)	0.853 (206.72)	0.765 (290.48)
Log(TWL)	0.685 (316.94)	0.348 (45.79)	0.378 (55.97)	0.233 (39.28)	0.145 (36.10)
Log(TWMe)	0.038 (13.55)	-0.082 (-9.85)	0.122 (14.02)	0.052 (6.27)	-0.072 (-12.50)
Log(TWMi)	0.138 (62.04)	-0.036 (-5.79)	-0.020 (-3.14)	0.057 (9.01)	0.126 (30.22)
Log(TWH)	0.054 (41.80)	-0.044 (-11.36)	-0.109 (-28.62)	-0.047 (-13.15)	0.216 (73.31)
D_LA_WOK	-0.264 (-219.50)	-0.144 (-36.47)	-0.187 (-51.15)	-0.114 (-31.21)	-0.102 (-42.21)
Rho-squared (stage 3)	0.96	0.94	0.95	0.96	0.96

TABLE 3a: THREE-STAGE ESTIMATION RESULTS (Travel cost is entered as MTT)
(t-scores in parentheses)

	ALL INCOME	INCOME 1	INCOME 2	INCOME 3	INCOME 4
MODE CHOICE					
CP	-0.914 (-524.77)	-0.293 (-46.21)	-0.367 (-64.53)	-0.424 (-78.61)	-0.823 (-220.34)
PT	-0.785 (-211.85)	-0.088 (-8.03)	-0.321 (-23.94)	-0.386 (-25.79)	-0.812 (-67.53)
RT	-0.256 (-115.97)	0.184 (25.11)	-0.062 (-7.77)	-0.073 (-9.19)	0.064 (12.38)
Log(RENT+TC)	-1.204 (-144.36)	-1.128 (-28.14)	-0.724 (-30.54)	-0.569 (-29.08)	-0.799 (-55.51)
Rho-squared (stage 1)	0.38	0.10	0.13	0.16	0.36
RESIDENCE LOCATION CHOICE					
I_MODE	0.505 (273.74)	0.768 (137.34)	0.964 (172.46)	1.076 (193.09)	0.780 (97.80)
Log(THU)	0.491 (330.52)	0.338 (64.59)	0.322 (68.53)	0.337 (76.94)	0.469 (155.48)
D_LA_RES	-0.211 (-65.68)	-0.075 (-5.20)	-0.003 (-0.29)	-0.077 -9.00	-0.154 (-30.07)
D_RES_WOK	0.157 (57.06)	0.017 (1.51)	0.098 (11.49)	0.071 (10.08)	0.090 (19.68)
Log(DIST) /Log(ACCESS_RET)	-6.07 (-566.29)	-3.395 (-83.48)	-2.413 (-79.97)	-2.069 (-74.72)	-4.334 (-197.56)
Log(AGE)	-0.127 (-81.48)	-0.156 (-22.64)	-0.118 (-20.93)	-0.115 (-25.28)	-0.106 (-41.56)
Log(MHI)	0.603 (154.89)	0.050 (4.13)	0.240 (20.255)	0.334 (30.82)	1.115 (148.25)
Log(MNR)	0.224 (47.26)	0.203 (14.22)	0.076 (5.169)	0.127 (8.98)	0.344 (33.67)
Rho-squared (stage 2)	0.52	0.50	0.45	0.41	0.44
JOB LOCATION CHOICE					
I_HOME	0.333 (217.04)	0.892 (244.01)	0.776 (186.88)	0.844 (204.51)	0.730 (284.19)
Log(TWL)	0.685 (321.63)	0.422 (57.45)	0.392 (58.80)	0.240 (40.98)	0.164 (43.73)
Log(TWMe)	0.032 (12.09)	-0.130 (-15.69)	0.115 (13.44)	0.054 (6.630)	-0.099 (-17.84)
Log(TWMi)	0.139 (66.31)	-0.050 (-8.08)	-0.024 (-3.88)	0.057 (9.157)	0.112 (26.25)
Log(TWH)	0.072 (55.91)	0.016 (3.99)	-0.094 (-24.32)	-0.044 (-12.40)	0.282 (98.36)
D_LA_WOK	-0.265 (-219.18)	-0.086 (-21.76)	-0.186 (-51.59)	-0.122 (-33.65)	-0.104 (-43.28)
Rho-squared (stage 3)	0.96	0.94	0.95	0.96	0.96

TABLE 3b: THREE-STAGE ESTIMATION RESULTS (Travel cost is entered as RENT+TC)

(t-scores in parentheses)

The next table, Table 4, presents the key elasticities evaluated at the estimates.

These are based on the estimates in Table 3a and 3b.

MODEL\VARIABLE	MEAN TRAVEL TIME	RENT
INCOME GROUP ALL	0.072	0.175
INCOME GROUP 1	0.042	-----
INCOME GROUP 2	0.037	0.354
INCOME GROUP 3	0.033	0.416
INCOME GROUP 4	0.038	0.009

TABLE 4a: Elasticity calculations from the estimated models of Table 3a

MODEL\VARIABLE	MEAN TRAVEL TIME	RENT
INCOME GROUP ALL	0.076	0.474
INCOME GROUP 1	0.047	0.624
INCOME GROUP 2	0.049	0.465
INCOME GROUP 3	0.042	0.388
INCOME GROUP 4	0.049	0.426

TABLE 4b: Elasticity calculations from the estimated models of Table 3b

It is important for the reader to understand the way these elasticities are calculated and their meanings. Suppose that, in the mode choice model, the mean travel time in a particular arrangement (i,j,m) increases by 1%, remaining constant in all other arrangements. Then the percent change in the probability for that choosing m given that (i,j) can be calculated from the mode choice probability:

$$P_{m|ij} = \frac{e^{\beta \ln(R_j + w_i(MTT_{ijm}) + (\text{other variables})ijm}}{\sum_n \Delta_{ijn} e^{\beta \ln(R_j + w_i(MTT_{ijn}) + (\text{other variables})ijn}}$$

This is done successively for each (i,j,m) and the resulting elasticities are averaged by weighting with the probabilities:

$$\eta_{MTTijm} = (1 - P_{m|ij}) \frac{\beta w_i(MTT_{ijm})}{R_j + w_i(MTT_{ijm})}, \ when\ \Delta_{ijm} = 1$$

$$\eta_{MTTij} = \sum_m \Delta_{ijm} P_{m|ij} \eta_{MTTijm}$$

$$\eta_{MTT} = \sum_{ij} \frac{N_{ij} \eta_{MTTij}}{\sum_{kl} N_{kl}}.$$

The rent elasticity is similarly calculated from the residence location choice model:

$$P_{j|i} = \frac{e^{U_j + \lambda \ln \sum_m \Delta_{ijm} e^{\beta \ln(R_j + w_i(MTT_{ijm}))}}}{\sum_k e^{U_k + \lambda \ln \sum_m \Delta_{ikm} e^{\beta \ln(R_k + w_i(MTT_{ikm}))}}}$$

$$\eta_{Rij} = \lambda\beta(1 - P_{j|i}) \sum_m \Delta_{ijm} P_{m|ij} \frac{R_j}{R_j + w_i(MTT_{ijm})}$$

$$\eta_R = \lambda\beta \sum_{ij} P_i P_{j|i} (1 - P_{j|i}) \sum_m \Delta_{ijm} P_{m|ij} \frac{R_j}{R_j + w_i(MTT_{ijm})}$$

5. REFERENCES

Anas, A., 1981. “The estimation of multinomial logit models of joint location and travel mode choice from aggregated data”, *Journal of Regional Science*, 21, 2, 223-242.

Anas, A. and Liu, Y., 2007. “A Regional Economy, Land Use and Transportation Model (RELU-TRAN): formulation, algorithm design and testing”, *Journal of Regional Science*, 47,32,415-455.

Anas, A. and Chu, C., 1984. “Discrete choice models and the housing price and travel to work elasticities of location demand,” *Journal of Urban Economics,* 15, 107-123.

Train, K., 2003. *Discrete Choice Methods with Simulation (first edition),* Cambridge University Press.

Train, K., 2009. *Discrete Choice Methods with Simulation (second edition),* Cambridge University Press.