INSTITUTO MUNICIPAL DE INVESTIGACION Y PLANEACION
Benjamin Franklin 4185 y Estocolmo. Circuito Pronaf. Cd. Juárez, Chih. Tels. 13-65-20 13-65-30 13-65-46 13-64-96

Tech Memo

DATE:	June 3, 1999.	**PROJECT:**	**Estudio Integral de Transporte (III)/ Multimodal Transportation Study:** Development of Travel Demand and Mobile Source Emissions Models for base year 1996, Juarez. (Contract No. 9880055000)
TO:	Carl Snow, Project Director / TNRCC Jim Yarbrough / EPA Jack Jones / TxDOT Zack Graham / TTI Luis Raul Cordova/SEMARNAP		
FROM:	Salvador Gonzalez-Ayala	**SERIAL:**	EITIII-02
SUBJ:	Progress under Task 3: Mode Split Model Application.		

Overview

Under the current project five specific tasks have been outlined for development:

1. Trip generation model application
2. Trip distribution model application
3. Mode split model application
4. Traffic assignment application
5. On-road mobile source emissions modeling

Having completed the trip generation part of the project (task 1), the conventional travel demand modeling methodology requires that trip distribution be the next step in the process. For the subsequent mode split step, this sequence has the advantage of facilitating the inclusion of the characteristics of the journey (such as the specific O-D pair); however, makes it more difficult to include characteristics of the trip maker as they may have already been aggregated in the O-D matrix. Under an alternative technique that considerably improves the distribution/mode-share characterization, the mode split model needs to be developed before the trip distribution step.

IMIP is exploring this new technique and thus, the present Tech Memo summarizes the work performed to accomplish task 3.

Background

Mode split analysis provides the market share of trips for every transportation mode under analysis. In the case of Juarez, and due to their apparent significance within the local transportation system, three generic modes have been defined:

1) Individual motorized transportation (AUTO).
   In addition to cars includes taxis, motorcycles, vans, trucks, and commercial vehicles.
2) Collective motorized transportation (BUS).
   Includes public transit, special bus service (private corporations), and school bus service.
3) Non-motorized transportation (WALK).
   Includes walking and bicycle use.

The shares for these three travel modes are expressed as a percentage of the total daily person trips between each traffic analysis zone (TAZ).

Discrete Choice Modeling

The modern approach to mode split analysis is to develop disaggregate demand algorithms known as discrete choice models. In contrast to aggregate demand (first-generation) models which are based on observed relations for groups of travelers, or average patterns at the TAZ level, discrete choice models are based on observed choices made by individual travelers, providing the setting for a more realistic characterization. In addition, discrete choice models tend to be more efficient in terms of information usage and thus usually require smaller sample sizes.

In general, discrete choice models yield the probability of individuals choosing a given transportation mode, based on a relative measure of its attractiveness to that of the other modes in the option set. This measure of attractiveness is provided through the development of a utility function for each of the modes. Because of their probabilistic nature and their dependence on utility functions, discrete choice models are also known as random utility models.

$$P_{qj} = f(U_j, U_k, U_l, \ldots) \qquad \textbf{(Eq. 1)}$$

Where:

P_{qj}: Probability of individual q choosing mode j.

U_j, U_k, U_l: Utility functions for modes $j, k, l, \ldots$

The utility function for each alternative mode usually depends on the characteristics of the individual and the attributes of the mode. This function is usually a linear-in-the-parameters combination of variables. Thus, an example of a utility function, say for a mode j, usually would have the following form:

$$U_j = \beta_0 + \beta_{x1}{*}X_1 + \beta_{x2}{*}X_2 + \ldots + \beta_{xn}{*}X_n + \beta_{s1}{*}S_1 + \beta_{s2}{*}S_2 + \ldots + \beta_{sm}{*}Sm \qquad \textbf{(Eq. 2)}$$

Where:

U_j: Utility function for mode j.

$X_1, X_2, \ldots, Xm$: Mode attributes.

$S_1, S_2, \ldots, Sm$: Characteristics of the individual.

β_0: Alternative-specific constant.

$\beta_{x1}, \beta_{x2}, \ldots, \beta_{xn}$: Parameters for mode attributes.

$\beta_{s1}, \beta_{s2}, \ldots, \beta_{sm}$: Parameters for characteristics of the individual.

In a process usually known as model calibration (or model estimation), the numerical values of the parameters (β's) are obtained by fitting the discrete choice model to the available data, through the use of the maximum-likelihood method.

This probabilistic approach provides a very convenient framework, since in practice the option (mode) with the highest utility will not necessarily be the one chosen. On the one hand the decision-maker does not posses perfect information about the transportation system (all attributes for all options); in fact, it is quite unlikely that two decision-makers have the same knowledge of the system's attributes. Moreover, the analyst searching for the utility function may not know all the relevant characteristics of the individual nor the true values of each mode attributes.

Therefore under these premises, the mode with the highest utility is not certain to be the one chosen, but instead has the highest probability of being chosen; in any case, the higher the utility of a mode, the higher the probability of being chosen.

The Multinomial Logit (MNL) model

The most popular discrete choice model in current practice is by far the multinomial logit (or MNL), which rewriting eq.1 has the following form:

$$P_{qj} = f(U_1, U_2, \ldots, U_N) = \frac{e^{U_j}}{\sum_{n=1}^{N} e^{U_n}} \qquad \textbf{(Eq. 3)}$$

Specifying the MNL model

The specification of the MNL model consists of building one or more forms of the utility function, based on the available data and a theoretical interpretation of the relations between the variables. This process involves identifying the variables and variable types for the utility functions, as well as any transformations or interactions of these variables that might seem appropriate.

Adaptation of the household database

As an initial step to develop the utility functions for the MNL model, trip information from the household-survey database was restructured into a compacted format through the use of several programming codes, to allow for a compatible and expedite exchange with the statistical analysis tools of TransCAD. Among other things, trips were rearranged as one record per trip and chained under the conventional purpose and dwell time criteria; external trips were identified and left out. In addition, the nine trip modes specified on the household survey were regrouped into the 3 generic ones previously mentioned (AUTO, BUS, WALK).

Anticipating the possible functional forms and variables to use from similar models, information from sources other than the household survey was included. In this regard network skims obtained for each of the three modes were added to each trip record regardless of the actual mode used, but consistent with the trip's O-D pair. Moreover, acknowledging the widely successful use of the Cost/Wage rate variable in discrete-mode choice models, the equivalent interaction term TravelCost/HHincome was added to the table for both AUTO and BUS modes and labeled as a travel cost index. Table 1, shows the final outline of fields on the simplified file. This data file is in dBASE format (under the name MNLall.dbf), and has a total of 27,688 records.

Table 1. Final set of fields for MNL analysis, from HH survey trip records and network skims.

FIELD NAME	DESCRIPTION
UNIQUE	Trip record unique number
ISN1	Survey conducted by Elementary or Junior High
ISN2	Household sample no. (unique by Elementary or Jr High)
IPN	Household member number
ITP2	Trip number
PURP	Trip purpose (HBW=1, HBNW=2, NHB=3)
IZN	Trip origin TAZ
IZN2	Trip destination TAZ
MNLmode	Mode used (AUTO, BUS, WALK)
IncVal	Household income per day (1996 pesos)
NT_aT	Network travel time by AUTO (in minutes)
NT_bT	Network travel time by BUS (in minutes)
NT_wT	Network travel time by WALK (in minutes)
AUTCST	Network out-of-pocket travel cost by AUTO (in 1996 pesos)
AUTCST_IDX	Travel cost index by AUTO = AUTCST/IncVal
NT_F2	Network out-of-pocket travel cost by BUS (in 1996 pesos)
BUSCST_IDX	Travel cost index by BUS = NT_F2/IncVal
NT_wL	Network travel length by WALK (in kilometers)

In evaluating the variables to use in a utility function, it is important to consider the future availability of this information or ease to forecast it under future scenarios. Thus, it is noticeable from Table 1 that time sensitive variables such as the distribution of age or gender of the decision-makers among others, although apparently relevant, were left out at this time in order to simplify the forecasting process. In case the utility functions failed to provide a good fit, an attempt to include them would then be warranted.

Functional form

Finding the functional form of the utility functions is an empirical process where the analyst fits several possible models to the data, and then tests and compares these to evaluate the best one. As an initial step in this direction, a selection of seemingly relevant fields from Table 1 that depicted characteristics of the individual and attributes of the modes were arranged to emulate the structure of eq. 2. The components of the resulting functions fall under the following categories:

a) Alternative specific variables.
From the standpoint of model specification, alternative specific variables should be those that would clearly vary across modes, that is, where recorded trip information on the variable is categorized separately under the different modes, and thus has in general different values under different modes. This is the case of all network skims on Juarez, which were obtained by specific modes: travel times, travel lengths, and travel costs, as well as travel cost indexes. If statistically significant, the inclusion of the interaction terms in the form of travel cost indexes (travel cost/HHincome), would have the econometric interpretation that lower household incomes gain disutility more quickly with higher travel costs.

The resulting parameters for an alternative specific variable should show different numerical values under each mode's utility function where the variable is specified.

b) Generic variables.
Due to the nature of mode split modeling, it appears instinctive to try to have all variables as alternative specific. In practice this is seldom possible since not all relevant information is always available for the different modes. Generic variables are those where recorded trip information about the variable is categorized under a single field for all modes, and thus, it shows the same value for all modes. Income as well as other characteristics of the individual (decision-maker) are typically generic variables, although sometimes mode attributes could also be of this type when the information available is presented aggregated from all modes. On the other hand, it is important to avoid specifying fields disaggregated by mode as generic variables, since this would diminish the predictive potential of the model. This would only be recommended if for future evaluation scenarios the information to feed to the model is not differentiated by mode.

An important restriction with generic variables is that such variables can not be included in all utility functions: at the most, on all but one of the modes could these variables be a part of. The utility in which the variable is not included is considered the base mode for this variable, and thus, only the relative impact of the variable is known. By itself and everything else being equal, this condition in turn yields the probability of the base mode being chosen as well as the probability of not being chosen, but nothing about the other modes can be explained. Obviously, this is a drawback only if more than two alternative modes are under evaluation.

For the Juarez case household income was chosen as one of the generic variables, with AUTO as the base mode. This would only allow a comparison between AUTO, and non-auto mode (BUS, WALK aggregated). To partially overcome this imperfect dichotomy, walking length was in addition selected as a complement generic variable, with WALK now as the base mode, accordingly forcing discrimination between a motorized mode (AUTO, BUS aggregated) and WALK mode.

The resulting parameters for a generic variable should show the same numerical values under the utility functions of all modes where the variable is specified.

c) Alternative specific constants.
In addition to the previous variable types, it is always recommended to include alternative specific constants (ASC) when building utility functions, since very conveniently represent the net influence of variables not explicitly included in the model. The only restriction is that the number of ASCs never exceeds the number of alternative modes considered in the model minus one; it is of no importance which mode's utility function is the one without an ASC, since the resulting predictions will be the same. For the Juarez case, ASCs were specified for the BUS and WALK modes.

If the model is properly specified[1], ASCs sould be of different numerical value for the different modes.

In summary 11 different parameters were incorporated for the initial model, four of these would be evaluated for the AUTO utility function, six of these for the BUS utility function, and three of these for the

[1] Technically, it is possible to force utility functions of different modes to yield exactly the same ASC value, but obviously this would weaken the main advantage of using ASCs.

WALK utility function. These parameters were based on 9 of the fields depicted in Table 1. Table 2 summarizes the main characteristics of the components of the utility functions.

Table 2. Description of components of the utility functions.

COMPONENTS OF UTILITY FUNCTIONS				REFERENCE
PARAMETER	VARIABLE	DESCRIPTION	TYPE	DATABASE FIELD
β_{0b}			ASC (BUS)	
β_{0w}			ASC (WALK)	
β_{1a}	TT_a	Travel time (AUTO)	Alternative specific variable	NT_aT
β_{1b}	TT_b	Travel time (BUS)	Alternative specific variable	NT_bT
β_{1w}	TT_w	Travel time (WALK)	Alternative specific variable	NT_wT
β_2	HHINC	Household income	Generic variable	IncVal
β_3	WLK_D	Walking distance	Generic variable	NT_wL
β_{4a}	TC_a	Travel cost (AUTO)	Alternative specific variable	AUTCST
β_{4b}	TC_b	Travel cost (BUS)	Alternative specific variable	NT_F2
β_{5a}	TC_a_IDX	Travel cost index (AUTO)	Alternative specific variable	AUTCST_IDX
β_{5b}	TC_b_IDX	Travel cost index (BUS)	Alternative specific variable	BUSCST_IDX

Thus, the utility functions initially suggested for the 3 generic modes have the following form:

AUTO mode $$U_a = \beta_{1a} * TT_a + \beta_3 * WLK\_D + \beta_{4a} * TC_a + \beta_{5a} * TC_a\_IDX \quad \textbf{(Eq. 4)}$$

BUS mode $$U_b = \beta_{0b} + \beta_{1b} * TT_b + \beta_2 * HHINC + \beta_3 * WLK\_D + \beta_{4b} * TC_b + \beta_{5b} * TC_b\_IDX \quad \textbf{(Eq. 5)}$$

WALK mode $$U_w = \beta_{0w} + \beta_{1w} * TT_w + \beta_2 * HHINC \quad \textbf{(Eq. 6)}$$

Once the initial utility functions have been assembled, the next step would then be to test and adjust the functional form through a process better known as model calibration.

Calibration of the MNL model

Also referred to as estimation of the MNL model, calibration is the process by which the MNL model is fitted to the data through the use of the maximum-likelihood principles and methodology, and thus, the parameters of the utility functions are computed. Moreover, under a trial-and-error procedure the initial functional form of the utility functions can be further adjusted and polished until an optimum fit is achieved.

Calibration of the MNL model can be accomplished through the statistical analysis tools of TransCAD. In this regard TransCAD provides an interface to input the utility functions, consisting of a fixed-format binary table (file extension *.bin*), also known as the TransCAD MNL table. Table 3 shows the format of the MNL table developed to input the utility functions previously indicated by equations 4, 5, and 6.

Table 3. The TransCAD MNL table used for estimation of the initial model.

Alternatives	ASCb	ASCw	TTa	TTb	TTw	HHINC	WLK_D	TCa	TCb	TCa_IDX	TCb_IDX
AUTO			NT_aT				NT_wL	AUTCST		AUTCST_IDX	
BUS	ONE			NT_bT		IncVal	NT_wL		NT_F2		BUSCST_IDX
WALK		ONE			NT_wT	IncVal					

The first column of the TransCAD MNL table identifies the different modes. The top row indicates the name of the variables, while the following rows are used to specify the fields of the source data (with the exception of the label "ONE", which indicates an ASC). Obviously in addition to the MNL table, the source data (e.i. MNLall.dbf) also needs to be open in TranCAD. A blank cell simply indicates that the variable is not included on the respective utility function. In the case of MNL models with more than two modes such as the one for Juarez, it is quite simple to identify the two types of variables: a) alternative specific variables are those that show only one field in the entire column, while b) generic variables are those that have more than one cell filled (with the same field name) in the entire column.

The output of the calibration procedure of TransCAD is very similar to that of other statistical packages for econometric analysis. The numerical value of the parameters are in fact provided on the same TransCAD MNL table, on the next row following the last utility function; but a more detailed output is provided on an external file, readable by any text editor (usually Notepad). Table 4a shows such output for the initial model.

Table 4a. MNL estimation output for initial model (all trip purposes combined).

Valid Cases:	27688		
Choice Distribution			
AUTO	13895	50.2%	
BUS	6673	24.1%	
WALK	7120	25.7%	
Maximum likelihood reached at iteration 16			

PARAMETER	**ESTIMATE**	**STD. ERR**	**t Test**
ASCb	-0.123754	0.044465	-2.783145
ASCw	1.805894	0.033248	54.316502
TTa	-0.020104	0.006789	-2.961252
TTb	0.012748	0.002160	5.900989
TTw	0.174938	0.017605	9.936688
HHINC	-0.003392	0.000084	-40.615230
WLK_D	2.962409	0.214364	13.819546
TCa	0.026385	0.016366	1.612248
TCb	-0.229897	0.022597	-10.173606
TCa_IDX	0.041937	0.068104	0.615777
TCb_IDX	-0.176026	0.123429	-1.426130

$l(0)$:	-30418.377049
$l(\beta)$:	-22402.292200
$-2(l(0)-l(\beta))$:	16032.169698
ρ^2 :	0.263528
ρ_{adj}^2 :	0.263166

The interest on this output is centered on three pieces of information:

a) t-test of parameter value

The t-test simply provides the significance level for rejecting the null hypothesis of having the true parameter value equal zero (rejecting that the parameter has no influence on the utility). A t-test value of ±1.96 means that the parameter is different than zero at a confidence level of about 95%. At a t-test value of ±1.50 the confidence level drops to 85%, which is the lower limit usually recommended to consider a parameter significant.

b) Sign of parameter value

The sign of the parameter allows judging if the variable conforms to a priori notions or theory about its overall behavior with respect to the utility. Under current practice it is recommended to keep relevant policy variables with a correct sign even if these fail the t-test. The reason being that the estimated parameters are the best approximation available for their real values, and the lack of significance may very well be the result of lack of enough data. On the other hand, variables with wrong sign should always be dropped from the model, even if the parameters pass the significance level test. Under this criteria TTb, TTw, TCa, and TCa_IDX (shaded rows) need to be dropped despite several significant t-test values, since with the exception of WLK_D, all of the model variables are expected to have negative signs.

c) Asymptotic, adjusted, and adjusted to market share ρ^2

The asymptotic[2] rho-squared (ρ^2) index is a similar concept to that of the coefficient of determination (R^2) obtained for linear regression models through least-squares, in that it provides

[2] The term asymptotic simply means that it is only valid for large samples.

a value between 0 and 1 as a measure of goodness-of-fit. Also similar to R^2, ρ^2 requires an adjustment to prevent increasing its value just by the inclusion of each new variable to the model (refer to appendix A for the different equation forms of ρ^2). In contrast to R_{adj}^2 though, ρ_{adj}^2 should mainly be used to compare the relative fit between models, since currently there are no intuitive interpretations nor general guidelines for when a ρ_{adj}^2 value is sufficiently high; values of 0.4 may in fact be excellent fits. In addition, the ρ_{adj}^2 value tends to vary with the proportion of individuals in the sample choosing each mode. To solve this difficulty, Ortuzar et al[3] suggest a market share adjustment (ρ_m^2) similar to the following:

$$\rho_m^2 = 1 - \frac{l(\beta) - k}{l(C)} \qquad \textbf{(Eq. 7)}$$

where: $l(\beta)$: Log-likelihood at end.
$l(C)$: Log-likelihood from market share.
k: Number of parameters in the model.

The component k is an addition to the original equation, but as in the case of ρ_{adj}^2 prevents the index to increase just by the sole inclusion of variables. The log-likelihood at end is the one obtained from the sample data at convergence and is one of the outputs of the MNL software. The log-likelihood from market share is obtained similar to the log-likelihood at zero $l(0)$, the only difference is that the usual null hypothesis of all β=0 includes only variables and not ASCs, thus a simple programming code was designed additionally to compute the value (Appendix B).

The three estimates of ρ^2 thus were obtained for the Juarez case so together could provide a feel for the goodness-of-fit of the models. As complement to Table 4a, the values of $l(C)$=-45435.95 and ρ_m^2=0.5067 were therefore computed for the initial model.

Before dropping any variables from the initial model, an attempt was made to measure the influence of the purpose of trips in mode preference. The best way thought to go about doing this was to disaggregate the data table (MNLall.dbf) into 3 separate tables categorized by the purpose of the recorded trips. The result of this process is shown in Tables 4b, 4c, and 4d. Again, variables with incorrect sign are shown shaded as candidates to be dropped from the models.

Table 4b. MNL estimation output for initial model, under purpose 1 only (HBW).

Valid Cases:	8178			
Choice Distribution				
AUTO	4718	57.7%		
BUS	2700	33.0%		
WALK	760	9.3%		
Maximum likelihood reached at iteration 15				
PARAMETER		**ESTIMATE**	**STD. ERR**	**t Test**
ASCb		0.024819	0.075448	0.328958
ASCw		0.721676	0.077296	9.336504
TTa		-0.008449	0.010045	-0.841143
TTb		0.018197	0.003444	5.283904
TTw		-0.009202	0.047546	-0.193539
HHINC		-0.003257	0.001540	-21.080326
WLK_D		0.438800	0.575012	0.763115
TCa		0.050941	0.021845	2.331982
TCb		-0.228099	0.034554	-6.601287
TCa_IDX		-0.221111	0.195632	-1.130237
TCb_IDX		-0.549472	0.422503	-1.300517
$l(0)$:		-8984.451297		
$l(\beta)$:		-6441.300475	$l(C)$:	-10889.604929
$-2(l(0)-l(\beta))$:		5086.301643		
ρ^2:		0.283061		
ρ_{adj}^2:		0.281837	ρ_m^2:	0.407481

[3] Juan de Dios Ortuzar, Luis G. Willumsen, "Modelling Transport", Second Edition (John Wiley & Sons/1994).

Table 4c. MNL estimation output for initial model, under purpose 2 only (HBNW).

Valid Cases: 17158

Choice Distribution		
AUTO	7622	44.4%
BUS	3578	20.9%
WALK	5958	34.7%

Maximum likelihood reached at iteration 16

PARAMETER	ESTIMATE	STD. ERR	t Test
ASCb	-0.106965	0.060672	-1.763003
ASCw	2.050570	0.040947	50.078362
TTa	-0.021376	0.011539	-1.852429
TTb	0.004717	0.003046	1.548767
TTw	0.179937	0.020791	8.654491
HHINC	-0.003337	0.000106	-31.567302
WLK_D	3.114231	0.253614	12.279429
TCa	-0.032865	0.032632	-1.007151
TCb	-0.223300	0.032475	-6.876151
TCa_IDX	0.125993	0.078165	1.611899
TCb_IDX	-0.003326	0.135718	-0.024509

$l(0)$:	-18849.989649		
$l(\beta)$:	-13966.744501	$l(C)$:	-27098.986568
$-2(l(0)-l(\beta))$:	9766.490296		
ρ^2 :	0.259058		
ρ_{adj}^2 :	0.258475	ρ_m^2 :	0.484197

Table 4d. MNL estimation output for initial model, under purpose 3 only (NHB).

Valid Cases: 2352

Choice Distribution		
AUTO	1555	66.1%
BUS	395	16.8%
WALK	402	17.1%

Maximum likelihood reached at iteration 16

PARAMETER	ESTIMATE	STD. ERR	t Test
ASCb	-0.377135	0.172439	-2.187062
ASCw	1.069239	0.119821	8.923651
TTa	-0.003994	0.025412	-0.157190
TTb	0.021783	0.008800	2.475413
TTw	0.392373	0.069876	5.615237
HHINC	-0.003970	0.000329	-12.063855
WLK_D	5.384212	0.849360	6.339139
TCa	0.058272	0.056280	1.035384
TCb	-0.337115	0.090200	-3.737402
TCa_IDX	-0.096410	0.203412	-0.473966
TCb_IDX	-0.636125	0.421704	-1.508466

$l(0)$:	-2583.936103		
$l(\beta)$:	-1661.953314	$l(C)$:	-3307.899755
$-2(l(0)-l(\beta))$:	1843.965578		
ρ^2 :	0.356813		
ρ_{adj}^2 :	0.352556	ρ_m^2 :	0.494255

In order to run these separate MNL estimations by trip purpose, the main data source MNLall.dbf was copied into 3 additional tables MNLprp1.dbf, MNLprp2.dbf, and MNLprp3.dbf. Each of these purpose-specific tables has the same field structure as that from MNLall.dbf (refer to Table 1). Originally this procedure was preferred to that of including an additional variable for purpose into the utility functions to avoid the restrictions imposed when specifying generic variables, since there is no intuitive arrangement to target a specific mode as base-alternative over the others.

Accordingly, the resulting ρ_m^2 values from the purpose-specific initial models seem slightly lower than the one obtained from the all-purpose one, suggesting that trip purpose might not have all that influence after all. Yet, in the all-purpose model more relevant policy variables needed to be dropped due to wrong signs, while on the purpose-specific models some of these could stay, at least for some of the purposes. Since the difference in the goodness-of-fit ratios seemed small, at this stage it was given a higher priority to keep important variables in the models.

An additional series of runs were generated gradually dropping wrong-sign variables from all four models, and carefully reviewing the impact of this on the re-accommodation of parameter signs and the re-computation of the ρ_m^2 indexes. As a final result, Tables 5a, 5b, 5c, and 5d show the modified models.

Table 5a. MNL estimation output for final all-purpose model (all trip purposes combined).

Valid Cases:	27688		
Choice Distribution			
AUTO	13895	50.2%	
BUS	6673	24.1%	
WALK	7120	25.7%	
Maximum likelihood reached at iteration 12			
PARAMETER	**ESTIMATE**	**STD. ERR**	**t Test**
ASCb	-0.132054	0.043665	-3.024265
ASCw	1.811495	0.032721	55.361450
TTa	-0.037878	0.002327	-16.280008
HHINC	-0.003434	0.000083	-41.226007
WLK_D	0.897564	0.014258	62.952745
TCb	-0.169312	0.019704	-8.592895
TCb_IDX	-0.238357	0.084089	-2.834576
$l(0)$:	-30418.377049		
$l(\beta)$:	-22463.284996	$l(C)$:	-45544.676941
$-2(l(0)-l(\beta))$:	15910.184106		
ρ^2 :	0.261523		
ρ_{adj}^2 :	0.261292	ρ_m^2 :	0.506632

Table 5b. MNL estimation output for final model, under purpose 1 only (HBW).

Valid Cases:	8178		
Choice Distribution			
AUTO	4718	57.7%	
BUS	2700	33.0%	
WALK	760	9.3%	
Maximum likelihood reached at iteration 12			
PARAMETER	**ESTIMATE**	**STD. ERR**	**t Test**
ASCw	0.705800	0.072850	9.688408
TTa	-0.030422	0.003718	-8.183104
TTw	-0.012322	0.046979	-0.262277
HHINC	-0.003252	0.000142	-22.976061
WLK_D	0.481111	0.568684	0.846007
TCb	-0.143898	0.023415	-6.145489
TCa_IDX	-0.002336	0.124406	-0.018777
$l(0)$:	-8984.451297		
$l(\beta)$:	-6456.796767	$l(C)$:	-10852.608319
$-2(l(0)-l(\beta))$:	5055.309060		
ρ^2 :	0.281337		
ρ_{adj}^2 :	0.280557	ρ_m^2 :	0.404402

Table 5c. MNL estimation output for final model, under purpose 2 only (HBNW).

Valid Cases:	17158	
Choice Distribution		
AUTO	7622	44.4%
BUS	3578	20.9%
WALK	5958	34.7%

Maximum likelihood reached at iteration 12

PARAMETER	ESTIMATE	STD. ERR	t Test
ASCb	-0.102658	0.060613	-1.693670
ASCw	2.057398	0.040778	50.454096
TTa	-0.028589	0.011210	-2.550273
HHINC	-0.003356	0.000105	-32.049031
WLK_D	0.948167	0.018654	50.829423
TCa	-0.040292	0.030934	-1.302530
TCb	-0.209506	0.028264	-7.412524
$l(0)$:	-18849.989649		
$l(\beta)$:	-14006.703442	$l(C)$:	-27143.955318
$-2(l(0)-l(\beta))$:	9686.572414		
ρ^2 :	0.256938		
ρ_{adj}^2 :	0.256567	ρ_m^2 :	0.483727

Table 5d. MNL estimation output for final model, under purpose 3 only (NHB).

Valid Cases:	2352	
Choice Distribution		
AUTO	1555	66.1%
BUS	395	16.8%
WALK	402	17.1%

Maximum likelihood reached at iteration 13

PARAMETER	ESTIMATE	STD. ERR	t Test
ASCb	-0.375754	0.169519	-2.216583
ASCw	1.090168	0.117649	9.266311
TTa	-0.030674	0.008640	-3.550425
HHINC	-0.004103	0.000330	-12.444948
WLK_D	0.733198	0.047729	15.361571
TCb	-0.241334	0.080463	-2.999328
TCa_IDX	-0.098873	0.206062	-0.479820
TCb_IDX	-0.666166	0.427296	-1.559028
$l(0)$:	-2583.936103		
$l(\beta)$:	-1679.729070	$l(C)$:	-3330.719318
$-2(l(0)-l(\beta))$:	1808.414065		
ρ^2 :	0.349934		
ρ_{adj}^2 :	0.346838	ρ_m^2 :	0.493284

Similar to the initial models, the final models yielded slightly lower ρ_m^2 values from the purpose-specific versions compared to the all-purpose one. Again, this suggests that trip purpose is not a significant factor influencing the mode choice in Juarez. An interesting result nevertheless from this exercise was the improvement of the t-test values for some variables, and better yet, the change to a correct sign of several relevant policy variable parameters.

Trip purpose might not be a relevant factor for mode choice, but in any case there is an additional need to discriminate by trip purpose since the other steps in the modeling process do. Thus, the purpose-specific models will be used for model evaluation and future scenario forecasting.

Limitations of household income variable

From the shown results, it can be seen that one of the more significant variables for mode choice in Juarez is in fact household income (HHINC). Unfortunately its use poses a problem when forecasting mode choice for NHB trips, since at that stage there is no way to know the household TAZ of the trip maker under the NHB trip purpose. This is not a problem for home-based trips, since the trip-production TAZ is always the household TAZ, and thus HHINC can be easily established. The options then would either be to drop HHINC and related interaction terms from the all-purpose model, or use purpose-specific models, and drop HHINC and related interaction terms only from the NHB one. The second seems to be a better trade-off, since NHB trips account for a minor fraction of urban trips, and thus the error introduced to the modeling would be limited. Table 6 shows the modified version of the NHB model without the income variable (HHINC) and related interaction terms (TCa_IDX and TCb_IDX).

Table 6. MNL estimation output for NHB modified model (no income variable nor interaction terms).

Valid Cases:	2352			
Choice Distribution				
AUTO	1555	66.1%		
BUS	395	16.8%		
WALK	402	17.1%		
Maximum likelihood reached at iteration 10				
PARAMETER		**ESTIMATE**	**STD. ERR**	**t Test**
ASCb		-1.208421	0.156168	-7.737939
ASCw		0.332743	0.099863	3.332005
TTa		-0.036266	0.008351	-4.342895
WLK_D		0.769321	0.047700	16.128265
TCb		-0.274708	0.078408	-3.503577
$l(0)$:		-2583.936103		
$l(\beta)$:		-1793.859802	$l(C)$:	-2673.985441
$-2(l(0)-l(\beta))$:		1580.152602		
ρ^2 :		0.305765		
ρ_{adj}^2 :		0.303830	ρ_m^2 :	0.327274

As expected, due to the significance of household income as an explanatory variable, the modified model strength (ρ_m^2) decreased notoriously.

In summary, the following set of equations will be used as the functional forms of the utility functions for the MNL models, in accordance with the base format presented under Eqs. 4, 5, and 6.

1) Calibrated utility functions for HBW trip-purpose:

AUTO $U_a = - 0.0304*TT_a + 0.4811*WLK\_D - 0.0023*TC_a\_IDX$ (**Eq. 8**)

BUS $U_b = - 0.0033*HHINC + 0.4811*WLK\_D - 0.1439*TC_b$ (**Eq. 9**)

WALK $U_w = 0.7058 - 0.0123*TT_w - 0.0033*HHINC$ (**Eq. 10**)

2) Calibrated utility functions for HBNW trip-purpose:

AUTO $U_a = - 0.0286*TT_a + 0.9482*WLK\_D - 0.0403*TC_a$ (**Eq. 11**)

BUS $U_b = - 0.1027 - 0.0034*HHINC + 0.9482*WLK\_D - 0.2095*TC_b$ (**Eq. 11**)

WALK $U_w = 2.0574 - 0.0034*HHINC$ (**Eq. 12**)

3) Calibrated utility functions for NHB trip-purpose:

AUTO $U_a = - 0.0363*TT_a + 0.7693*WLK\_D$ (**Eq. 13**)

BUS $U_b = - 1.2084 + 0.7693*WLK\_D - 0.2747*TC_b$ (**Eq. 14**)

WALK $U_w = 0.3327$ (**Eq. 15**)

Evaluation of the MNL model

Model evaluation is the process by which the calibrated MNL model is used to estimate the mode share of a trip or set of trips, given the numerical values of the explanatory variables in the model.

MNL evaluation using aggregate O-D based data

Already it has been underlined the advantages in terms or reliability and data requirements of working with disaggregate models as opposed to aggregate models. But while a disaggregate model allows the estimation of individual choice probabilities, normally the transportation modeling effort is more interested in the prediction of aggregate travel behavior.

Similarly, for the Juarez case the final objective of the modal split step was to develop mode shares between O-D pairs (in addition, now these need to be for each trip purpose). Since the data used to calibrate the MNL model was disaggregate at the individual level, the question was then how to use this model to obtain mode shares between zones. Mode attributes (TTa, TTb, TTw, WLK_D, TCa, and TCb) would depend only on the specific O-D pair and are readily available in matrix format as O-D pair skims, so in this regard there are no major complications. The issue is more on the side of the characteristics of the individual (i.e. HHINC and related interaction terms TCa_IDX, and TCb_IDX), and the proper way to aggregate their collective impact at the zone level.

One instinctive answer would be to obtain an average value of individual characteristics for the entire TAZ and with it run the MNL. This is known as the “naive aggregation” approach, since the MNL is not linear (instead it has an S-shape) and thus would yield a bias. The strictly correct procedure is to run the MNL for every individual in the TAZ, and then compute the average of the individual mode shares for the entire TAZ. Off course the down side of this theoretically correct approach is the large data set requirements and consequently the heavy computation volume.

A middle solution known as the “classification approach” is to aggregate individual units of the TAZ into a manageable number of groups according to ranges of individual characteristics, compute MNLs for each of the ranges, and then obtain a weighted average of the mode share for the TAZ. This weighted average is proportional to the number of individuals in each group.

$$W_j = \sum_{g=1}^{G} \frac{N_g}{N_T} P(j|X_g) \qquad \textbf{(Eq. 16)}$$

Where:
- W_j: Share of the population choosing mode j.
- G: Total number of ranges (groups).
- $P(j|X_g)$: Probability of an individual in range g choosing mode j.
- N_g: Number of individuals in range g.
- N_T : Total number of individuals in all ranges.

Under this approach the issue now became the proper selection of ranges of individual characteristics, to conform the groups. An obvious method was to analyze market-segmenting variables that present the greatest variance. In this regard, the high correlation between trip production rates and auto-bus mode split was used as the main criteria, and thus the six HHINC ranges suggested in the trip generation step were used as an initial approach:

Range Code	Household income range (1996 daily min wages)	Household income range (1996 pesos)	Range avg (1996 pesos)
R1	0 to 2	\$ 0.00 to \$ 52.96 /day	\$ 26.48 /day
R2	2 to 6	\$ 52.96 to \$158.88 /day	\$105.92 /day
R3	6 to 10	\$158.88 to \$264.80 /day	\$211.84 /day
R4	10 to 12	\$264.80 to \$317.76 /day	\$291.28 /day
R5	12 to 18	\$317.76 to \$476.64 /day	\$397.20 /day
R6	18 to 34	\$476.64 to \$900.32 /day	\$688.32 /day

This format facilitated the computational process, since TAZ information already was tailored to include household income distributions according to these ranges. Thus, an important simplification to the whole process was to assume that on any given TAZ the percentage distribution of trip-makers by HHINC was the same as the percentage distribution of households by HHINC, which as a preliminary approach seemed reasonable. As such for each TAZ, the percentage of households in each HHINC range would substitute the term N_g/N_T in Eq. 16. In order to provide this information to the MNL evaluation process, a new table was generated with fields TAZ, R1_PCT to R6_PCT, and R1_INC to R6_INC, indicating for each TAZ the percentage of households and average household income under each of the income ranges. The table was generated as a dBASE file with the name inc_dist.dbf.

Adaptation of skim matrices to run MNL evaluation

A final requirement to run the MNL evaluation for the Juarez case, was to provide information for the travel cost indexes TCa_IDX and TCb_IDX according to the six household income ranges. The best way to accomplish this was to add a set of fields to the skim matrices for the AUTO and BUS generic modes (Aut96.mtx, and Bus96.mtx). These fields were labeled R1_IDX to R6_IDX respectively for each of the six income ranges, and were computed simply by dividing the respective matrix cell for travel cost by each of the six income averages.

MNL evaluation with TransCAD

Evaluation of the MNL model can be accomplished through TransCAD, in which case the same TransCAD MNL tables (*.bin* files) developed for model calibration must be used; the only difference being that each table cell instead of indicating the data source reference to develop the MNL model, now will indicate the source of the data to forecast mode share. In addition, the TransCAD MNL table must include an additional row with the numerical values of each of the parameters. Tables 7, 8, and 9 show the TransCAD MNL tables used to evaluate the models for the three specific trip purposes. These tables are consistent with equations 8 to 15.

Table 7. The TransCAD MNL table used for evaluation of final HBW model.

Alternatives	ASCb	ASCw	TTa	TTb	TTw	HHINC	WLK_D	TCa	TCb	TCa_IDX	TCb_IDX
AUTO			Aut96|TIME				Wlk96|DIST			Aut96|Rn_IDX	
BUS						Rn_INC	Wlk96|DIST		Bus96|COST		
WALK		ONE			Wlk96|TIME	Rn_INC					
Estimates		0.7058	-0.0304		-0.0123	-0.0033	0.4811		-0.1439	-0.0023	

Note: One Transcad MNL table like this for each of the six income ranges.

Table 8. The TransCAD MNL table used for evaluation of final HBNW model.

Alternatives	ASCb	ASCw	TTa	TTb	TTw	HHINC	WLK_D	TCa	TCb	TCa_IDX	TCb_IDX
AUTO			Aut96|TIME				Wlk96|DIST	Aut96|COST			
BUS	ONE					Rn_INC	Wlk96|DIST		Bus96|COST		
WALK		ONE				Rn_INC					
Estimates	-0.1027	2.0574	-0.0286			-0.0034	0.9482	-0.0403	-0.2095		

Note: One Transcad MNL table like this for each of the six income ranges.

Table 9. The TransCAD MNL table used for evaluation of final NHB model.

Alternatives	ASCb	ASCw	TTa	TTb	TTw	HHINC	WLK_D	TCa	TCb	TCa_IDX	TCb_IDX
AUTO			Aut96|TIME				Wlk96|DIST				
BUS	ONE						Wlk96|DIST		Bus96|COST		
WALK		ONE									
Estimates	-1.2084	0.3327	-0.0363				0.7693		-0.2747		

Note: Only one Transcad MNL table like this is required, since income was dropped as a variable.

The labels Rn_INC and Rn_IDX shown on Tables 7 and 8 are not actually the exact names required on the TransCAD MNL tables; the letter "n" is used here simply to abbreviate the use of numbers 1 to 6 on the respective source fields. The actual field labels on the TransCAD MNL tables had the number for the household income range being evaluated.

To run the TransCAD MNL evaluation for each household income range and each trip purpose, the respective TransCAD MNL table (income and purpose specific) must be opened, as well as the table Inc_dist.dbf, and the network-skim matrices Aut96.mtx, Bus96.mtx, and Wlk96.mtx. An example of this procedure is shown in Figure 1 for purpose 1 (HBW) and for household income range 1; thus for this specific case the resulting matrix has been labeled P1_r1.mtx.

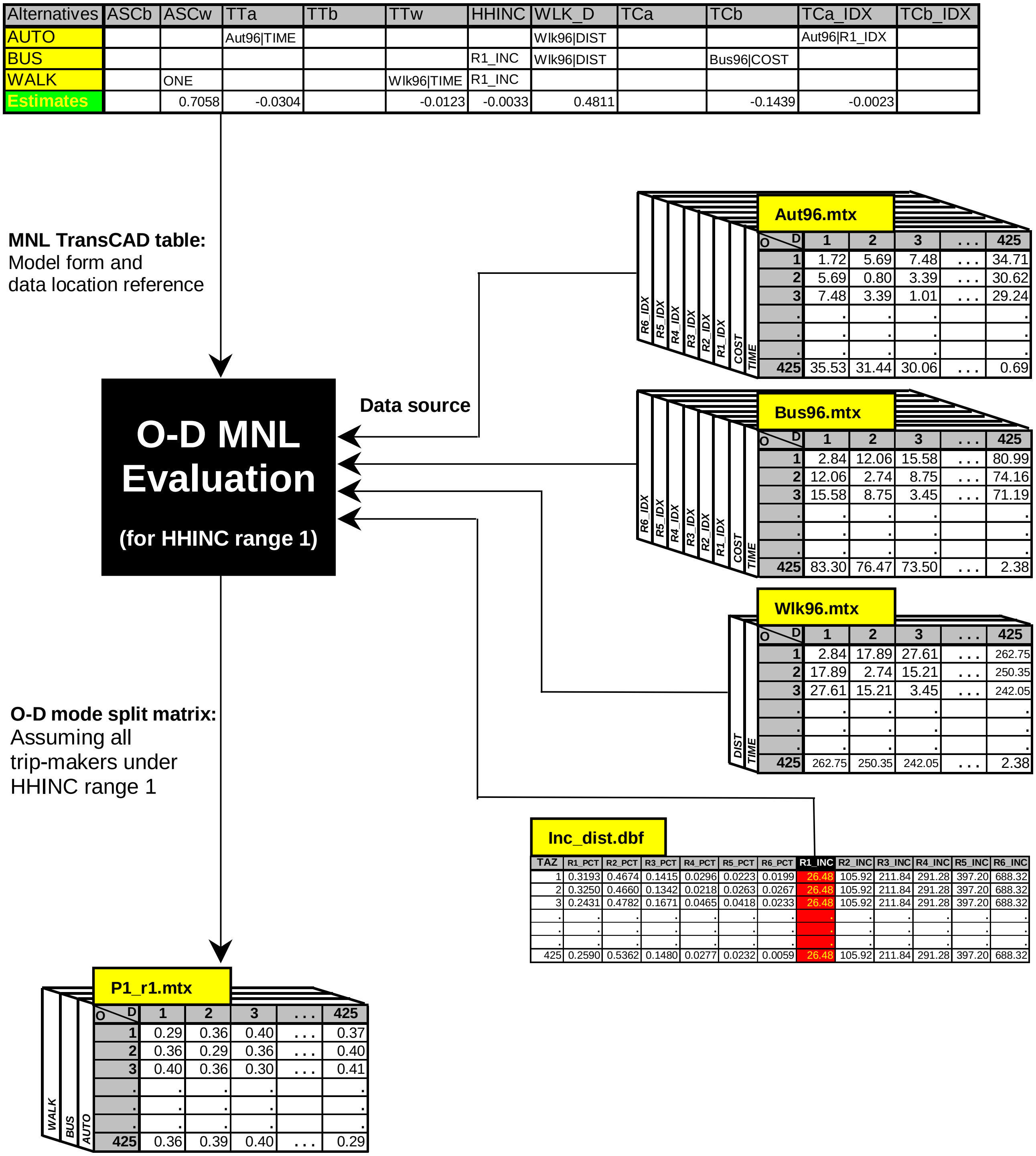


Figure 1. Flow chart for MNL evaluation. Example for purpose 1, income range 1.

A total of 13 different TransCAD MNL tables were specified: six income ranges times two trip purposes (HBW and HBNW), plus an additional one without income ranges (as shown in Table 9, NHB purpose did not include income for reasons previously discussed). In turn these TransCAD MNL tables yielded 13 O-D mode share matrices.

The data contained in these matrices in fact provided the numerical values of the component P(*j*|X*g*) of Equation 16. For the summation process of Equation 16 two programming codes were designed (Appendix C) and executed. One aggregates an O-D matrix with productions assumed at the origins (HHINC distribution from row TAZs). The other aggregates an O-D matrix with productions assumed at the destinations (HHINC distribution from column TAZs). The flow charts for the aggregation process and estimation of final O-D matrices are shown on appendices D1 to D3 for each of the three trip purposes.

As a result five final matrices were developed with the 1996 mode split between the O-D pairs: Two for HBW (refer to appendix D1), two for HBNW (refer to appendix D2), and one for NHB (refer to appendix D3). For the NHB trip purpose the final mode split matrix can be viewed either as an O-D or P-A matrix, since in this case origin is synonymous to production, and destination to attraction. Moreover, considering that household income was not used under this trip purpose, no aggregation process was required since the MNL evaluation directly provides the final O-D matrix.

The complete final matrices are not shown as part of this document due the volume of information, but are available for review, as well as the base data sources.

Appendix A

Forms of the likelihood ratio index (ρ^2).

Asymptotic rho-squared index (ρ^2)

$$\rho^2 = 1 - \frac{\boldsymbol{l}(\beta)}{\boldsymbol{l}(0)}$$

Adjusted rho-squared index (ρ_{adj}^2)

$$\rho_{adj}^2 = 1 - \frac{\boldsymbol{l}(\beta) - k}{\boldsymbol{l}(0)}$$

Market share adjusted rho-squred index (ρ_m^2)

$$\rho_m^2 = 1 - \frac{\boldsymbol{l}(\beta) - k}{\boldsymbol{l}(C)}$$

where: $\boldsymbol{l}(\beta)$: Log-likelihood at end.
$\boldsymbol{l}(0)$: Log-likelihood at zero (all β=0).
$\boldsymbol{l}(C)$: Log-likelihood from market share (all β=0 except ASCs).
k: Number of parameters in the model.

.

Appendix B

VisualBasic code (MS-Access) to compute $I(\beta)$, $I(0)$, and $I(C)$.

```vb
Option Compare Database
Option Explicit

Private Sub Command0_Click()
Dim TAB1 As Recordset
Dim MyBASE As Database
Dim j As Long
Dim P, P0, PC As Double
Dim TTa, TTb, TTw, HHINC, WLK_D, TCa, TCb, TCa_IDX, TCb_IDX As Double
Dim AscB, AscW, Btta, Bttb, Bttw, Binc, Bw_d, Btca, Btcb, Baid, Bbid As Double
Dim Ua, UaC, Ub, UbC, Uw, UwC, logP, logP0, logPC, lp, lp0, lpC As Double
Dim modo, base As String

Set MyBASE = CurrentDb()

'Set TAB1 = MyBASE.OpenRecordset("MNLall", DB_OPEN_TABLE)
'base = "all"
'Set TAB1 = MyBASE.OpenRecordset("MNLprp1", DB_OPEN_TABLE)
'base = "purp 1"
'Set TAB1 = MyBASE.OpenRecordset("MNLprp2", DB_OPEN_TABLE)
'base = "purp 2"
Set TAB1 = MyBASE.OpenRecordset("MNLprp3", DB_OPEN_TABLE)
base = "purp 3"

DoCmd.Hourglass False
DoCmd.Hourglass True

j = 0
lp = 0
lp0 = 0
lpC = 0

If base = "all" Then
    AscB = -0.132054
    AscW = 1.811495
    Btta = -0.037878
    Bttb = 0
    Bttw = 0
    Binc = -0.003434
    Bw_d = 0.897564
    Btca = 0
    Btcb = -0.169312
    Baid = 0
    Bbid = -0.238357
End If
If base = "purp 1" Then
    AscB = 0
    AscW = 0.7058
    Btta = -0.030422
    Bttb = 0
    Bttw = -0.012322
    Binc = -0.003252
    Bw_d = 0.481111
    Btca = 0
    Btcb = -0.143898
    Baid = -0.002336
    Bbid = 0
End If
If base = "purp 2" Then
    AscB = -0.102658
    AscW = 2.057398
    Btta = -0.028589
    Bttb = 0
    Bttw = 0
    Binc = -0.003356
    Bw_d = 0.948167
    Btca = -0.040292
    Btcb = -0.209506
    Baid = 0
    Bbid = 0
End If
```

```
If base = "purp 3" Then
    AscB = -1.208421
    AscW = 0.332743
    Btta = -0.036266
    Bttb = 0
    Bttw = 0
    Binc = 0
    Bw_d = 0.769321
    Btca = 0
    Btcb = -0.274708
    Baid = 0
    Bbid = 0
End If

TAB1.MoveFirst
Do While Not TAB1.EOF
    modo = TAB1!MNLmode
    TTa = TAB1!NT_aT
    TTb = TAB1!NT_bT
    TTw = TAB1!NT_wT
    HHINC = TAB1!IncVal
    WLK_D = TAB1!NT_wL
    TCa = TAB1!AUTCST
    TCb = TAB1!nt_f2
    TCa_IDX = TAB1!AUTCST_IDX
    TCb_IDX = TAB1!BUSCST_IDX
    Ua = Btta * TTa + Bw_d * WLK_D + Btca * TCa + Baid * TCa_IDX
    Ub = AscB + Bttb * TTb + Binc * HHINC + Bw_d * WLK_D + Btcb * TCb + Bbid * TCb_IDX
    Uw = AscW + Bttw * TTw + Binc * HHINC
    UaC = 0
    UbC = AscB
    UwC = AscW
    If modo = "AUTO" Then
        P = Exp(Ua) / (Exp(Ua) + Exp(Ub) + Exp(Uw))
        P0 = Exp(0) / (Exp(0) + Exp(0) + Exp(0))
        PC = Exp(UaC) / (Exp(UaC) + Exp(UbC) + Exp(UwC))
    End If
    If modo = "BUS" Then
        P = Exp(Ub) / (Exp(Ua) + Exp(Ub) + Exp(Uw))
        P0 = Exp(0) / (Exp(0) + Exp(0) + Exp(0))
        PC = Exp(UbC) / (Exp(UaC) + Exp(UbC) + Exp(UwC))
    End If
    If modo = "WALK" Then
        P = Exp(Uw) / (Exp(Ua) + Exp(Ub) + Exp(Uw))
        P0 = Exp(0) / (Exp(0) + Exp(0) + Exp(0))
        PC = Exp(UwC) / (Exp(UaC) + Exp(UbC) + Exp(UwC))
    End If

    logP = Log(P)
    logP0 = Log(P0)
    logPC = Log(PC)
'    MsgBox "P" & j + 1 & "= " & P
'    MsgBox "LogP" & j + 1 & "= " & logP
    lp = lp + logP
    lp0 = lp0 + logP0
    lpC = lpC + logPC
    j = j + 1
    TAB1.MoveNext
    If j = 1000 Then MsgBox "van 1,000"
    If j = 10000 Then MsgBox "van 10,000"
    If j = 150000 Then MsgBox "van 150,000"
Loop

TAB1.Close
DoCmd.Hourglass False
MsgBox "Terminó!...Viajes Totales:" & j
MsgBox "l(B)= " & lp
MsgBox "l(0)= " & lp0
MsgBox "l(C)= " & lpC

End Sub
```

Appendix C

VisualBasic code (MS-Access) to compute final O-D mode split matrices

```
Option Compare Database
Option Explicit

Private Sub Command0_Click()
Dim TAB0, TAB1, TAB2, TAB3, TAB4, TAB5, TAB6, TABF As Recordset
Dim MyBASE As Database
Dim n, rcindex, rcindex1 As Integer
Dim j As Long
Dim AUTO, BUS, WALK As Double
Dim base As String

Set MyBASE = CurrentDb()

'base = "purp 1"
base = "purp 2"

If base = "purp 1" Then
    Set TAB0 = MyBASE.OpenRecordset("Inc_dist", DB_OPEN_TABLE)
    Set TAB1 = MyBASE.OpenRecordset("P1_R1", DB_OPEN_TABLE)
    Set TAB2 = MyBASE.OpenRecordset("P1_R2", DB_OPEN_TABLE)
    Set TAB3 = MyBASE.OpenRecordset("P1_R3", DB_OPEN_TABLE)
    Set TAB4 = MyBASE.OpenRecordset("P1_R4", DB_OPEN_TABLE)
    Set TAB5 = MyBASE.OpenRecordset("P1_R5", DB_OPEN_TABLE)
    Set TAB6 = MyBASE.OpenRecordset("P1_R6", DB_OPEN_TABLE)
    Set TABF = MyBASE.OpenRecordset("MtrxP1_O", DB_OPEN_TABLE)
End If
If base = "purp 2" Then
    Set TAB0 = MyBASE.OpenRecordset("Inc_dist", DB_OPEN_TABLE)
    Set TAB1 = MyBASE.OpenRecordset("P2_R1", DB_OPEN_TABLE)
    Set TAB2 = MyBASE.OpenRecordset("P2_R2", DB_OPEN_TABLE)
    Set TAB3 = MyBASE.OpenRecordset("P2_R3", DB_OPEN_TABLE)
    Set TAB4 = MyBASE.OpenRecordset("P2_R4", DB_OPEN_TABLE)
    Set TAB5 = MyBASE.OpenRecordset("P2_R5", DB_OPEN_TABLE)
    Set TAB6 = MyBASE.OpenRecordset("P2_R6", DB_OPEN_TABLE)
    Set TABF = MyBASE.OpenRecordset("MtrxP2_O", DB_OPEN_TABLE)
End If

DoCmd.Hourglass False
DoCmd.Hourglass True
DoCmd.Hourglass False
DoCmd.Hourglass True
TAB0.Index = "inc_idx"
TAB1.Index = "r1_idx"
TAB2.Index = "r2_idx"
TAB3.Index = "r3_idx"
TAB4.Index = "r4_idx"
TAB5.Index = "r5_idx"
TAB6.Index = "r6_idx"

j = 0
n = 0

TAB0.MoveFirst
TAB1.MoveFirst
TAB2.MoveFirst
TAB3.MoveFirst
TAB4.MoveFirst
TAB5.MoveFirst
TAB6.MoveFirst
Do While Not TAB1.EOF
    rcindex = TAB1!rcindex
    rcindex1 = TAB1!rcindex1
    AUTO = TAB0!R1_PCT * TAB1!AUTO + TAB0!R2_PCT * TAB2!AUTO + TAB0!R3_PCT * TAB3!AUTO _
           + TAB0!R4_PCT * TAB4!AUTO + TAB0!R5_PCT * TAB5!AUTO + TAB0!R6_PCT * TAB6!AUTO
    BUS = TAB0!R1_PCT * TAB1!BUS + TAB0!R2_PCT * TAB2!BUS + TAB0!R3_PCT * TAB3!BUS _
           + TAB0!R4_PCT * TAB4!BUS + TAB0!R5_PCT * TAB5!BUS + TAB0!R6_PCT * TAB6!BUS
    WALK = TAB0!R1_PCT * TAB1!WALK + TAB0!R2_PCT * TAB2!WALK + TAB0!R3_PCT * TAB3!WALK _
           + TAB0!R4_PCT * TAB4!WALK + TAB0!R5_PCT * TAB5!WALK + TAB0!R6_PCT * TAB6!WALK
```

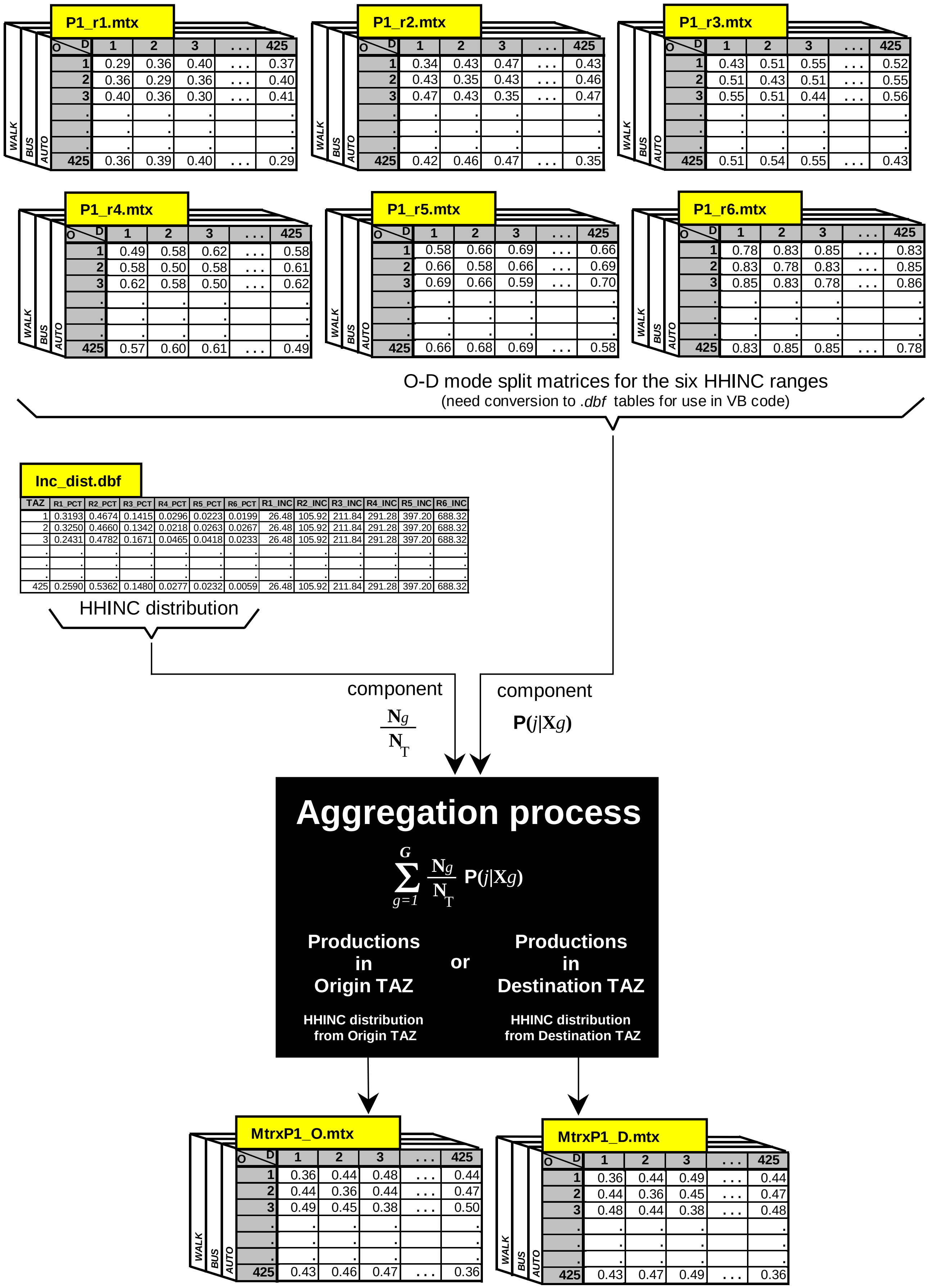
```
    With TABF
        .AddNew
        !rcindex = rcindex
        !rcindex1 = rcindex1
        !AUTO = AUTO
        !BUS = BUS
        !WALK = WALK
        .Update
    End With

    j = j + 1
    n = n + 1

If n = 425 Then
        TAB0.MoveNext
        n = 0
End If
    TAB1.MoveNext
    TAB2.MoveNext
    TAB3.MoveNext
    TAB4.MoveNext
    TAB5.MoveNext
    TAB6.MoveNext
    If j = 1000 Then MsgBox "van 1,000"
    If j = 10000 Then MsgBox "van 10,000"
    If j = 150000 Then MsgBox "van 150,000"
Loop

TAB0.Close
TAB1.Close
TAB2.Close
TAB3.Close
TAB4.Close
TAB5.Close
TAB6.Close
TABF.Close
DoCmd.Hourglass False
MsgBox "Terminó!...Viajes Totales:" & j

End Sub
```

```
Option Compare Database
Option Explicit

Private Sub Command0_Click()
Dim TAB0, TAB1, TAB2, TAB3, TAB4, TAB5, TAB6, TABF As Recordset
Dim MyBASE As Database
Dim n, rcindex, rcindex1 As Integer
Dim j As Long
Dim AUTO, BUS, WALK As Double
Dim base As String

Set MyBASE = CurrentDb()

'base = "purp 1"
base = "purp 2"

If base = "purp 1" Then
    Set TAB0 = MyBASE.OpenRecordset("Inc_dist", DB_OPEN_TABLE)
    Set TAB1 = MyBASE.OpenRecordset("P1_R1", DB_OPEN_TABLE)
    Set TAB2 = MyBASE.OpenRecordset("P1_R2", DB_OPEN_TABLE)
    Set TAB3 = MyBASE.OpenRecordset("P1_R3", DB_OPEN_TABLE)
    Set TAB4 = MyBASE.OpenRecordset("P1_R4", DB_OPEN_TABLE)
    Set TAB5 = MyBASE.OpenRecordset("P1_R5", DB_OPEN_TABLE)
    Set TAB6 = MyBASE.OpenRecordset("P1_R6", DB_OPEN_TABLE)
    Set TABF = MyBASE.OpenRecordset("MtrxP1_D", DB_OPEN_TABLE)
End If
If base = "purp 2" Then
    Set TAB0 = MyBASE.OpenRecordset("Inc_dist", DB_OPEN_TABLE)
    Set TAB1 = MyBASE.OpenRecordset("P2_R1", DB_OPEN_TABLE)
    Set TAB2 = MyBASE.OpenRecordset("P2_R2", DB_OPEN_TABLE)
    Set TAB3 = MyBASE.OpenRecordset("P2_R3", DB_OPEN_TABLE)
    Set TAB4 = MyBASE.OpenRecordset("P2_R4", DB_OPEN_TABLE)
    Set TAB5 = MyBASE.OpenRecordset("P2_R5", DB_OPEN_TABLE)
    Set TAB6 = MyBASE.OpenRecordset("P2_R6", DB_OPEN_TABLE)
    Set TABF = MyBASE.OpenRecordset("MtrxP2_D", DB_OPEN_TABLE)
End If

DoCmd.Hourglass False
DoCmd.Hourglass True
DoCmd.Hourglass False
DoCmd.Hourglass True
TAB0.Index = "inc_idx"
TAB1.Index = "r1_idx"
TAB2.Index = "r2_idx"
TAB3.Index = "r3_idx"
TAB4.Index = "r4_idx"
TAB5.Index = "r5_idx"
TAB6.Index = "r6_idx"

j = 0
n = 0

TAB0.MoveFirst
TAB1.MoveFirst
TAB2.MoveFirst
TAB3.MoveFirst
TAB4.MoveFirst
TAB5.MoveFirst
TAB6.MoveFirst
Do While Not TAB1.EOF
    rcindex = TAB1!rcindex
    rcindex1 = TAB1!rcindex1
    AUTO = TAB0!R1_PCT * TAB1!AUTO + TAB0!R2_PCT * TAB2!AUTO + TAB0!R3_PCT * TAB3!AUTO _
           + TAB0!R4_PCT * TAB4!AUTO + TAB0!R5_PCT * TAB5!AUTO + TAB0!R6_PCT * TAB6!AUTO
    BUS = TAB0!R1_PCT * TAB1!BUS + TAB0!R2_PCT * TAB2!BUS + TAB0!R3_PCT * TAB3!BUS _
           + TAB0!R4_PCT * TAB4!BUS + TAB0!R5_PCT * TAB5!BUS + TAB0!R6_PCT * TAB6!BUS
    WALK = TAB0!R1_PCT * TAB1!WALK + TAB0!R2_PCT * TAB2!WALK + TAB0!R3_PCT * TAB3!WALK _
           + TAB0!R4_PCT * TAB4!WALK + TAB0!R5_PCT * TAB5!WALK + TAB0!R6_PCT * TAB6!WALK
    With TABF
        .AddNew
```

```
        !rcindex = rcindex
        !rcindex1 = rcindex1
        !AUTO = AUTO
        !BUS = BUS
        !WALK = WALK
        .Update
    End With

    j = j + 1
    n = n + 1

    If n = 425 Then
        TAB0.MoveFirst
        n = 0
    Else
        TAB0.MoveNext
    End If
    TAB1.MoveNext
    TAB2.MoveNext
    TAB3.MoveNext
    TAB4.MoveNext
    TAB5.MoveNext
    TAB6.MoveNext
    If j = 1000 Then MsgBox "van 1,000"
    If j = 10000 Then MsgBox "van 10,000"
    If j = 150000 Then MsgBox "van 150,000"
Loop

TAB0.Close
TAB1.Close
TAB2.Close
TAB3.Close
TAB4.Close
TAB5.Close
TAB6.Close
TABF.Close
DoCmd.Hourglass False
MsgBox "Terminó!...Viajes Totales:" & j

End Sub
```

Appendix D1

Flow chart for aggregation process/estimation of final O-D mode split matrices, under trip purpose 1 (HBW).


P1_r1.mtx
WALK BUS AUTO
O D 1 2 3 . . . 425
1 0.29 0.36 0.40 . . . 0.37
2 0.36 0.29 0.36 . . . 0.40
3 0.40 0.36 0.30 . . . 0.41
425 0.36 0.39 0.40 . . . 0.29
P1_r2.mtx
1 0.34 0.43 0.47 . . . 0.43
2 0.43 0.35 0.43 . . . 0.46
3 0.47 0.43 0.35 . . . 0.47
425 0.42 0.46 0.47 . . . 0.35
P1_r3.mtx
1 0.43 0.51 0.55 . . . 0.52
2 0.51 0.43 0.51 . . . 0.55
3 0.55 0.51 0.44 . . . 0.56
425 0.51 0.54 0.55 . . . 0.43
P1_r4.mtx
1 0.49 0.58 0.62 . . . 0.58
2 0.58 0.50 0.58 . . . 0.61
3 0.62 0.58 0.50 . . . 0.62
425 0.57 0.60 0.61 . . . 0.49
P1_r5.mtx
1 0.58 0.66 0.69 . . . 0.66
2 0.66 0.58 0.66 . . . 0.69
3 0.69 0.66 0.59 . . . 0.70
425 0.66 0.68 0.69 . . . 0.58
P1_r6.mtx
1 0.78 0.83 0.85 . . . 0.83
2 0.83 0.78 0.83 . . . 0.85
3 0.85 0.83 0.78 . . . 0.86
425 0.83 0.85 0.85 . . . 0.78
O-D mode split matrices for the six HHINC ranges
(need conversion to .dbf tables for use in VB code)
Inc_dist.dbf
TAZ R1_PCT R2_PCT R3_PCT R4_PCT R5_PCT R6_PCT R1_INC R2_INC R3_INC R4_INC R5_INC R6_INC
1 0.3193 0.4674 0.1415 0.0296 0.0223 0.0199 26.48 105.92 211.84 291.28 397.20 688.32
2 0.3250 0.4660 0.1342 0.0218 0.0263 0.0267 26.48 105.92 211.84 291.28 397.20 688.32
3 0.2431 0.4782 0.1671 0.0465 0.0418 0.0233 26.48 105.92 211.84 291.28 397.20 688.32
425 0.2590 0.5362 0.1480 0.0277 0.0232 0.0059 26.48 105.92 211.84 291.28 397.20 688.32
HHINC distribution
component
Ng / NT
component
P(j|Xg)
Aggregation process
Σ g=1..G Ng/NT P(j|Xg)
Productions in Origin TAZ
or
Productions in Destination TAZ
HHINC distribution from Origin TAZ
HHINC distribution from Destination TAZ
MtrxP1_O.mtx
1 0.36 0.44 0.48 . . . 0.44
2 0.44 0.36 0.44 . . . 0.47
3 0.49 0.45 0.38 . . . 0.50
425 0.43 0.46 0.47 . . . 0.36
MtrxP1_D.mtx
1 0.36 0.44 0.49 . . . 0.44
2 0.44 0.36 0.45 . . . 0.47
3 0.48 0.44 0.38 . . . 0.48
425 0.43 0.47 0.49 . . . 0.36

Appendix D2

Flow chart for aggregation process/estimation of final O-D mode split matrices, under trip purpose 2 (HBNW).

P2_r1.mtx (WALK, BUS, AUTO)

O \ D	1	2	3	. . .	425
1	0.13	0.26	0.35	. . .	0.33
2	0.26	0.13	0.24	. . .	0.37
3	0.35	0.24	0.14	. . .	0.38
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	0.33	0.36	0.38	. . .	0.13

P2_r2.mtx (WALK, BUS, AUTO)

O \ D	1	2	3	. . .	425
1	0.16	0.31	0.41	. . .	0.40
2	0.31	0.17	0.29	. . .	0.43
3	0.41	0.29	0.17	. . .	0.45
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	0.39	0.43	0.44	. . .	0.16

P2_r3.mtx (WALK, BUS, AUTO)

O \ D	1	2	3	. . .	425
1	0.22	0.39	0.50	. . .	0.48
2	0.39	0.22	0.37	. . .	0.52
3	0.50	0.37	0.23	. . .	0.54
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	0.48	0.52	0.53	. . .	0.22

P2_r4.mtx (WALK, BUS, AUTO)

O \ D	1	2	3	. . .	425
1	0.27	0.46	0.56	. . .	0.55
2	0.46	0.27	0.44	. . .	0.59
3	0.56	0.44	0.28	. . .	0.60
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	0.54	0.58	0.59	. . .	0.27

P2_r5.mtx (WALK, BUS, AUTO)

O \ D	1	2	3	. . .	425
1	0.34	0.54	0.65	. . .	0.64
2	0.54	0.35	0.53	. . .	0.67
3	0.65	0.53	0.36	. . .	0.68
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	0.63	0.66	0.68	. . .	0.34

P2_r6.mtx (WALK, BUS, AUTO)

O \ D	1	2	3	. . .	425
1	0.58	0.76	0.83	. . .	0.82
2	0.76	0.58	0.75	. . .	0.84
3	0.83	0.75	0.60	. . .	0.85
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	0.82	0.84	0.85	. . .	0.58

O-D mode split matrices for the six HHINC ranges
(need conversion to *.dbf* tables for use in VB code)

Inc_dist.dbf

TAZ	R1_PCT	R2_PCT	R3_PCT	R4_PCT	R5_PCT	R6_PCT	R1_INC	R2_INC	R3_INC	R4_INC	R5_INC	R6_INC
1	0.3193	0.4674	0.1415	0.0296	0.0223	0.0199	26.48	105.92	211.84	291.28	397.20	688.32
2	0.3250	0.4660	0.1342	0.0218	0.0263	0.0267	26.48	105.92	211.84	291.28	397.20	688.32
3	0.2431	0.4782	0.1671	0.0465	0.0418	0.0233	26.48	105.92	211.84	291.28	397.20	688.32
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
425	0.2590	0.5362	0.1480	0.0277	0.0232	0.0059	26.48	105.92	211.84	291.28	397.20	688.32

HHINC distribution

component $\frac{N_g}{N_T}$

component $P(j|X_g)$

Aggregation process

$$\sum_{g=1}^{G} \frac{N_g}{N_T} P(j|X_g)$$

Productions in Origin TAZ or **Productions in Destination TAZ**

HHINC distribution from Origin TAZ / **HHINC distribution from Destination TAZ**

MtrxP2_O.mtx (WALK, BUS, AUTO)

O \ D	1	2	3	. . .	425
1	0.17	0.32	0.42	. . .	0.41
2	0.33	0.18	0.31	. . .	0.44
3	0.44	0.32	0.20	. . .	0.47
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	0.40	0.44	0.45	. . .	0.17

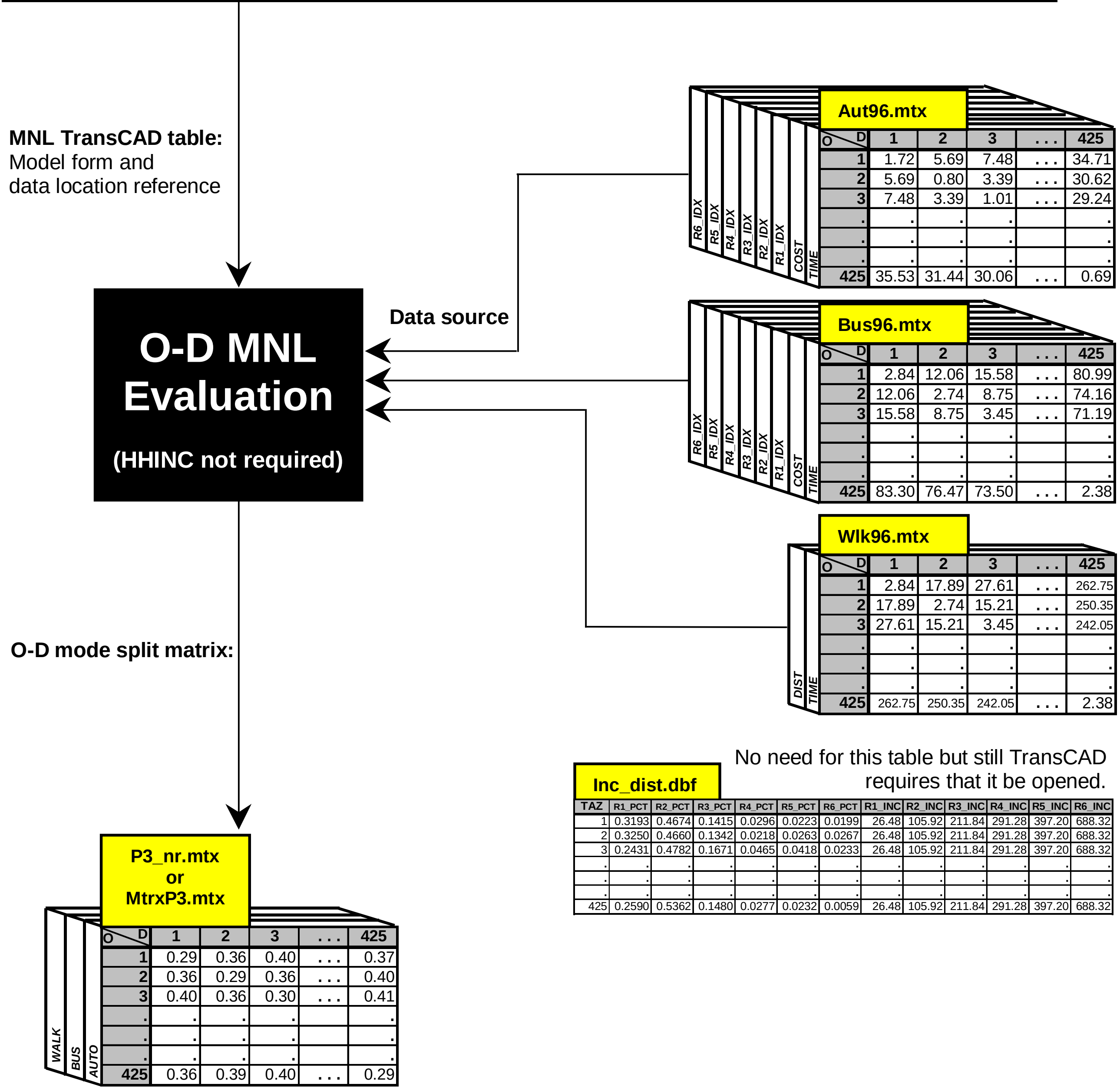
MtrxP2_D.mtx (WALK, BUS, AUTO)

O \ D	1	2	3	. . .	425
1	0.17	0.33	0.44	. . .	0.41
2	0.32	0.18	0.32	. . .	0.44
3	0.42	0.31	0.20	. . .	0.46
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	0.40	0.44	0.47	. . .	0.17

Appendix D3

Flow chart for estimation of final O-D mode split matrix, under trip purpose 3 (NHB).

Alternatives	ASCb	ASCw	TTa	TTb	TTw	HHINC	WLK_D	TCa	TCb	TCa_IDX	TCb_IDX
AUTO			Aut96|TIME				Wlk96|DIST				
BUS	ONE						Wlk96|DIST		Bus96|COST		
WALK		ONE									
Estimates	-1.2084	0.3327	-0.0363				0.7693		-0.2747		



Aut96.mtx

O \ D	1	2	3	...	425
1	1.72	5.69	7.48	...	34.71
2	5.69	0.80	3.39	...	30.62
3	7.48	3.39	1.01	...	29.24
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	35.53	31.44	30.06	...	0.69

Bus96.mtx

O \ D	1	2	3	...	425
1	2.84	12.06	15.58	...	80.99
2	12.06	2.74	8.75	...	74.16
3	15.58	8.75	3.45	...	71.19
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	83.30	76.47	73.50	...	2.38

Wlk96.mtx

O \ D	1	2	3	...	425
1	2.84	17.89	27.61	...	262.75
2	17.89	2.74	15.21	...	250.35
3	27.61	15.21	3.45	...	242.05
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	262.75	250.35	242.05	...	2.38

Inc_dist.dbf

TAZ	R1_PCT	R2_PCT	R3_PCT	R4_PCT	R5_PCT	R6_PCT	R1_INC	R2_INC	R3_INC	R4_INC	R5_INC	R6_INC
1	0.3193	0.4674	0.1415	0.0296	0.0223	0.0199	26.48	105.92	211.84	291.28	397.20	688.32
2	0.3250	0.4660	0.1342	0.0218	0.0263	0.0267	26.48	105.92	211.84	291.28	397.20	688.32
3	0.2431	0.4782	0.1671	0.0465	0.0418	0.0233	26.48	105.92	211.84	291.28	397.20	688.32
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
425	0.2590	0.5362	0.1480	0.0277	0.0232	0.0059	26.48	105.92	211.84	291.28	397.20	688.32

P3_nr.mtx or MtrxP3.mtx

O \ D	1	2	3	...	425
1	0.29	0.36	0.40	...	0.37
2	0.36	0.29	0.36	...	0.40
3	0.40	0.36	0.30	...	0.41
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
425	0.36	0.39	0.40	...	0.29

No need for aggregation process under this trip purpose,
since HHINC was not used.
This is the final O-D mode split matrix for purpose 3 (NHB).