USING SENSITIVITY ANALYSIS TO GUIDE TRAVEL DATA COLLECTION IN SMALL- AND MEDIUM-SIZED COMMUNITIES

Kevin M. Ford
Purdue University
School of Civil Engineering
550 Stadium Mall Drive
West Lafayette, Indiana 47907-2051

Telephone: (765) 494-2204
FAX: (765) 496-7996
Email: KMFord@purdue.edu

Jon D. Fricker
Purdue University
School of Civil Engineering
550 Stadium Mall Drive
West Lafayette, Indiana 47907-2051
Email: fricker@purdue.edu

Submittal Date: 22 July 2009

4,256 words + 12 tables = 7,256 words

ABSTRACT

Small and medium-sized communities often lack data that are complete, current, representative of the community, and appropriate for the travel demand models/software to be applied. Borrowing data from other communities is risky. Published data reveal a wide range of values for trip rates and trip lengths, for example. The "transferability" of any single value taken from another study area is questionable. A statistical update of borrowed data using a local sample may be advisable. However, our experience with a statewide model taught us that the model structure makes the model results more sensitive to changes in some input values than others. The limited resources available for data collection can be guided by knowledge of which input data need to be more carefully verified. A case study of one small MPO will be presented.

USING SENSITIVITY ANALYSIS TO GUIDE TRAVEL DATA COLLECTION IN SMALL- AND MEDIUM-SIZED COMMUNITIES

1. INTRODUCTION

Efficient data collection is critical for small- and medium-sized planning communities that lack significant resources. To expend these resources wisely, it would be helpful to know which input data (e.g., regression variables) or model parameters (e.g., coefficients in regression equations) have the greatest impact on the model output. With this knowledge, planners can better focus their efforts to refine data or parameter values. For instance, should a planning agency spend money to obtain trip length data for trip distribution or instead invest resources in obtaining volume-delay parameters to replace standard values? To help answer such questions, sensitivity analysis is used in this research.

In the literature, sensitivity analyses have often focused on how the output changes due to variations of input data within each step of the planning process. For instance, in a study by Barton-Aschman, Inc. and Cambridge Systematics, Inc. (*1*) the effects of varying utility function inputs on mode choice outputs were observed. In the same study, the sensitivity of trip generation outputs caused by the variation of socio-economic inputs was also examined. Similarly, Fehr and Peers (*2*) detailed the effect of varying the capacity input within the link performance function on the trip assignment output.

Zhao and Kockelman (*3*) investigated the propagation of uncertainty through the four-step travel demand model by using Monte Carlo simulation and sensitivity analysis with coefficients in each step having a lognormal distribution with standard deviation fixed to 0.30 of the mean of the coefficients. In our research, we assumed the role of a travel demand modeler faced with a need to borrow data and/or parameter values, or to collect data locally. The range of values we used was defined by the information we found that could be used in our travel demand model.

For the sensitivity analysis in this study, a two-tiered approach is recommended. First, the sensitivity of the output to borrowing various model parameters is discussed. By knowing which parameters are the most critical to the output, a decision can be made to decide which inputs to study further. Secondly, the output sensitivity due to varied input values is analyzed. The effects of these varied inputs can be tracked throughout the entire model and not just within the corresponding planning step. As a result of this two-tiered approach, the sensitivity of the output to model selection and input variation can be quantified.

To accomplish this, it is recommended to collect as many different parameters as possible within each of the four standard planning steps and then to apply local data. For instance, a database of regression equations or link performance functions can be built. When such equations and functions are applied to local input data, various outcomes are produced. In doing so, a range of results are created, from which a planner can make a more informed model selection, as well as learn what the consequences are from borrowing various models. The volatility of this range of outputs can help guide data collection. For instance, if the results do not vary significantly when applying various parameters, then data collection should be focused elsewhere. To further guide data collection, we can also assess the volatility of the selected model(s) by observing the change in output caused by multiple inputs.

To help demonstrate this approach, a sample sensitivity analysis will be presented using data from the Columbus Area Metropolitan Planning Organization (CAMPO), a small Indiana MPO. This study area, with a population of 73,900, as of the 2000 Census, is located approximately 46 miles south of Indianapolis.

To assess sensitivity for the CAMPO study area or any other planning jurisdiction, the following procedure is recommended at each of the standard four planning steps:

1. Build a database of equations/models with varying parameters.
2. Apply local data to develop a range of outputs.
3. Identify the sets of parameters yielding the minimum, average, and maximum outputs based on a predetermined evaluation criterion.
4. Select a model that is deemed most appropriate for the study area. For this research, two selections will be made:
    a. The ‘original selection’ represents the models chosen independently of this sensitivity analysis, based on the characteristics of the borrowed source;

   b. The ‘revised selection’ represents the models chosen based on the results of this sensitivity analysis using local data.
5. Where appropriate, apply a distribution to the input data and evaluate how the output varies based on different input confidence levels.
6. Carry on the selected, minimum, average, and maximum cases, as well as any input variations, to the next planning step.

With this procedure, it is possible to assess the impact of applying each set of parameters and input data values on the output for five different cases.

Before applying the six steps, a planner first needs to define what equations/models to assess and which evaluation criteria should be used to determine the minimum, average, and maximum cases. Such decisions are left to the planners, but recommendations are given in TABLE 1.

TABLE 1 Recommended Models and Evaluation Criteria for Each Planning Step within a Sensitivity Analysis.

Planning Step	Equation/Model	Evaluation Criteria
Trip Generation Cross-Classification Regression	Trip Rate Matrices P and A Equations	Total Trips by trip purpose
Trip Distribution	Friction Factor Functions, Vehicle Occupancy, and Time of Day Factors	Average Trip Length
Mode Choice	Utility Functions	Percent Auto Trips
Trip Assignment	Link Performance Functions	Peak Hour VMT, PRMSE

All of these steps, except mode choice, are evaluated in this case study.

2. TRIP GENERATION

Planners primarily rely on two basic approaches when determining the number of trips generated in a community: (1) Cross-Classification and (2) Regression. While regression is more commonly used, both methods will be analyzed for sensitivity in this study.

2.1 Trip Generation by Cross-Classification

Assuming that a planning organization knows how its households are distributed by two stratifying variables, a trip rate matrix is then needed to predict trips. To assess the sensitivity involved in borrowing trip rates, a database of trip matrices (*4*) was created, with the vast majority coming from Appendix 2 of NCHRP 365 (*5*). Next, each set of trip rates was applied to the CAMPO region household ‘membership matrix’, defined as the distribution of houses as stratified by household size and vehicle ownership. The resulting range of predicted trips is shown in

TABLE **2**.

TABLE 2 Cross-classification -- Range of Trips Predicted with Varied Parameters

Trip Rate Source	Total Trips Predicted	Average Daily Person Trips per Household	
Seattle	141,971	4.7	Minimum
Vancouver	167,824	5.5	
Houston	199,604	6.6	
New Jersey	224,525	7.4	
San Francisco	236,471	7.8	
Austin	240,091	7.9	
Nashua	241,362	7.9	
Phoenix	254,621	8.4	Average
Reno	257,395	8.5	
Albany	257,272	8.5	
Atlanta	268,325	8.8	
St. Louis	270,319	8.9	
Charlotte	282,924	9.3	
Pittsburgh	283,682	9.3	
NHTS Transferability Program – CAMPO	300,954	9.9	Original and Revised Selections
San Diego	412,983	13.6	Maximum

Depending on the rates that were "borrowed", the number of trips predicted for the CAMPO region varies from 141,971 with Seattle trip rates to 412,983 with San Diego trip rates. The average daily person-trips per household vary between 4.7 and 13.6. Such large variations indicate that caution must be applied when deciding which of the published trip rates are most appropriate to be transferred to a study area. This may in part be due to the size of the geographic area from which the rates are transferred, but also due to sampling errors within the surveys collected in each study area. For instance, the Seattle rates from Appendix A of the NCHRP 365 (*5*) seem quite low, ranging from 3.86 to 5.60, despite the variations in household size and vehicles available. If local travel survey data on average trip rate is not available, the volatility in the outputs from borrowed models justifies an effort to collect local data.

Having assessed the sensitivity involved in choosing a model, the sensitivity of our selected model to variations in input data will be focused on. In this case, our 'original' and 'revised' selections happen to be the same. This is due to the NHTS Transferability program's use of local data combined with census tracts of similar land use types and income levels (*6*). By quantifying how sensitive the model output is to varied input data, a planner can better guide potential data collection. To do this, it would be helpful to have statistical distributions for each cell in the household 'membership matrix' (defined as how households are distributed by two stratifying variables). These distributions could then quantify the sampling uncertainty or volatility of each cell.

To establish a distribution for household membership, the American Community Survey (ACS) is recommended. The ACS provides 'margin of error' data which represents the sampling error at the 90% confidence level based on the assumption that the data are normally distributed (*7*). With this information, the standard deviation of each cell can be calculated ('margin of error'/1.645) and other confidence levels explored. As applied to CAMPO, the mean (based on local travel demand surveys) and standard deviation for each household 'membership' cell can be seen in Table 3.

TABLE 3 Cross-classification – Household 'Membership' Means and Standard Deviations

Input	Mean	Standard Deviation
1-Person Household:		
No Vehicle Available	1,281	672
1 Vehicle Available	4,326	1,048
2 Vehicles Available	1,464	705
3 Vehicles Available	47	97
4+ Vehicles Available	52	58
2-Person Household:		
No Vehicle Available	295	453
1 Vehicle Available	2,058	1,077
2 Vehicles Available	6,871	1,220
3 Vehicles Available	1,274	552
4+ Vehicles Available	386	264
3-Person Household:		
No Vehicle Available	55	229
1 Vehicle Available	1,193	773
2 Vehicles Available	1,761	604
3 Vehicles Available	1,574	731
4+ Vehicles Available	418	371
4+-Person Household:		
No Vehicle Available	6	180
1 Vehicle Available	935	744
2 Vehicles Available	3,044	1,088
3 Vehicles Available	2,093	887
4+ Vehicles Available	1,248	519

When the standard deviation is larger than the mean, a lower bound of zero should be used to avoid applying negative inputs. With a distribution for the household 'membership matrix', the planner can start exploring how total trip predictions vary with different inputs. For instance, if the lower and upper bounds of the 90% confidence interval are applied to the household 'membership matrix', how would the output change? With CAMPO data, the difference between the two bounds was found to range from 199,415 total trips to 410,482 total trips, as compared to the 300,954 total trip value found using our current data or mean (

TABLE **4**).

TABLE 4 Cross-classification -- Range of Trips Predicted with Varied Inputs

Input Variable	90% Lower Bound	Mean	90% Upper Bound	Range
Total:	199,415	300,964	410,482	211,067
No Vehicle Available	2,804	7,867	20,674	17,870
1 Vehicle Available	35,810	63,767	91,724	55,914
2 Vehicles Available	102,789	134,723	166,657	63,868
3 Vehicles Available	41,265	64,723	88,404	47,140
4+ Vehicles Available	16,747	29,885	43,023	26,276
1-Person Household:	22,571	32,174	42,001	19,430
No Vehicle Available	2,804	4,932	7,060	4,257
1 Vehicle Available	15,000	18,734	22,467	7,467
2 Vehicles Available	4,739	7,848	10,957	6,218
3 Vehicles Available	0	295	813	813
4+ Vehicles Available	28	365	703	675
2-Person Household:	64,819	87,641	111,045	46,226
No Vehicle Available	0	2,211	5,003	5,003
1 Vehicle Available	9,124	16,023	22,922	13,798
2 Vehicles Available	47,503	55,628	63,754	16,252
3 Vehicles Available	6,760	10,503	14,246	7,485
4+ Vehicles Available	1,431	3,275	5,119	3,688
3-Person Household:	34,054	58,628	85,825	51,771
No Vehicle Available	0	626	3,876	3,876
1 Vehicle Available	6,402	13,700	20,999	14,597
2 Vehicles Available	14,744	20,542	26,340	11,596
3 Vehicles Available	11,537	18,671	25,805	14,268
4+ Vehicles Available	1,371	5,088	8,805	7,434
4-Person Household:	77,971	122,521	171,611	93,640
No Vehicle Available	0	97	4,735	4,735
1 Vehicle Available	5,285	15,310	25,336	20,051
2 Vehicles Available	35,803	50,704	65,605	29,802
3 Vehicles Available	22,967	35,253	47,540	24,573
4+ Vehicles Available	13,916	21,156	28,396	14,479

This wide range of outputs indicates the caution that must be taken when deciding on inputs.

To guide the data collection for these inputs, it is recommended to identify those cells that exhibit the most volatility in terms of trip productions. For CAMPO data, the most volatile cell was determined to be the 4+-person, 2-vehicle cell, which varied by nearly 30,000 trips, when applying the lower and upper bounds of the 90% confidence interval. Therefore, data collection should be focused on households with these characteristics.

2.2 Trip Generation by Regression

For those communities without sufficient data for cross-classification, regression equations are commonly used. These equations are typically separated by trip purpose and by productions and attractions. For this research, the home-based work (HBW), home-based other (HBO), and non-home-based (NHB) trip productions and attractions will be evaluated for sensitivity.

To test sensitivity, the database of regression equations was first created. (*4*) Of these regression models, all those whose input variables matched the socio-economic data available at the traffic analysis zone level in the CAMPO study area were included in the sensitivity analysis. After each regression equation was run through a travel demand model, the productions and attractions were balanced. The results with the minimum, average, and maximum total trips, along with the 'original' and 'revised selections' are displayed in TABLE 5 for productions and TABLE 6 for attractions.

TABLE 5 Regression -- Range of Trips Produced by Purpose

Model	HBW	HBO	NHB
Minimum	31,170	73,825	42,985
Average	57,774	113,883	57,667
Original Selection	32,719	79,515	96,838
Revised Selection	37,205	159,914	96,838
Maximum	76,153	218,493	96,838
% Difference (Range/Maximum)	59%	66%	56%

TABLE 6 Regression -- Range of Trips Attracted by Purpose

Model	HBW	HBO	NHB
Minimum	31,194	41,481	44,412
Average	58,882	73,227	51,507
Original Selection	56,790	79,515	96,838
Revised Selection	36,089	161,195	96,838
Maximum	82,493	161,195	121,504
% Difference (Range/Maximum)	62%	74%	68%

The 'original selection' models are those from Madisonville, Kentucky as reported in NCHRP 167 (*8*). The characteristics of this city are believed to best match those of the CAMPO region. The 'revised selection' models were selected based on the trip rates calculated for each purpose from the NHTS Transferability program (*6*).

Once each case was balanced, the range of total trips predicted for CAMPO was found to vary from 128,554 to 378,338; and the range of average daily person-trips per household was found to vary from 4.2 to 12.5. As with cross-classification, this range is indicative of the care that must be taken when borrowing trip generation models. It should also be noted that the 'original selection' predicted approximately 70,000 fewer trips than the 'revised selection'. Therefore, borrowing a model based solely on the source may not lead to the most representative results.

While the output is deemed sensitive to the regression parameters, we will now assess the sensitivity of the output to the input data. For this analysis, the 'revised selection' model will be studied. As with the cross-classification method, we can create a normal distribution for each input variable with the 'margin of error' values provided by the ACS. The mean (based on local survey data) and standard deviation, calculated from the ACS, can be seen in TABLE 7 for each input variable.

TABLE 7 Regression -- Variable Input Means and Standard Deviations

Input	Mean	Standard Deviation
Total Employment	44,814	2,099
Retail Employment	11,413	4,767
# of Vehicles	58,792	3,134
# of 0 Vehicle Households	1,736	836
# of 1 Vehicle Households	8,575	1,610
# of 2 Vehicle Households	12,989	1,447
# of 3+ Vehicle Households	7,081	1,771
# of Households	30,380	1,546
Median Household Income	$47,371	$6,042

As before, it is possible analyze how the output changes by applying data at the lower and upper bounds of their respective 90% confidence intervals. In doing so, the total trips, in

TABLE **8**

TABLE 8, were found to vary between 264,784 and 322,690 total trips as compared to the 293,733 value obtained using the current data.

TABLE 8 Regression -- Range of Balanced Trips Predicted with Varied Inputs

Model	HBW	HBO	NHB	Total	Average Daily Person Trip Rates per Household
90% Lower Bound	33,271	137,375	94,138	264,784	9.1
Mean	36,341	160,554	96,838	293,733	9.7
90% Upper Bound	39,412	183,733	99,546	322,690	10.2
Range	6,141	46,357	5,408	57,906	1.1
% Difference (Range/Lower Bound)	19%	34%	6%	22%	12%

This range is more acceptable than that found during cross-classification, yet a narrower variation would be preferred.

With more focused data collection, a planner can be even more confident in the number of total trips generated. To guide the collection, it is recommended to focus on the most volatile model in terms of the range of trips predicted between the two bounds. For CAMPO it can be determined that HBO is the most volatile model. This is due to the combination of HBO being the most common trip purpose and the use of retail employment as an input variable in the attraction model. Therefore, data collection should be focused on obtaining accurate retail employment values, particularly in those zones with the highest concentrations of retail.

To compare the difference in outputs between cross-classification and regression, the total balanced trip results of the two approaches are shown in TABLE 9.

TABLE 9 Comparing Cross-Classification and Regression Techniques for Trip Generation

Model	Cross-Classification		Regression	
	Total Trips	Average Daily Person Trips per Household	Total Trips	Average Daily Person Trips per Household
Minimum	141,971	4.7	128,554	4.2
Average	254,621	8.4	206,470	6.8
Original Selection	300,957	9.9	221,107	7.3
Revised Selection	300,957	9.9	293,733	9.7
Maximum	412,983	13.6	378,338	12.5
Range	271,012	8.9	249,784	8.2

Both methods yield similar total trip results for the 'revised selection' case. The cross-classification method was found to be much more sensitive to input data variations and slightly more sensitive to parameter variations. With cross-classification, the sampling error involved with travel demand surveys was found to be more significant than the difference among the borrowed trip rates. For regression, the influence of the parameter values on the output was stronger than the inputs. To reduce the volatility of regression results, more data is needed to calibrate the models.

Because regression is considered to be the more common trip generation method used by small- and medium-sized planning organizations, the regression results are carried forward in the remaining planning steps for this study.

3. TRIP DISTRIBUTION

With the five cases from trip generation by regression, as well as the two input variations from the 'revised selection' case, trip distribution was then applied. To assess sensitivity, variations within the friction factor function were analyzed. In particular, the a, b, and c parameters in the Tanner form of this function, as seen in Equation 1, were varied. The impedance for this study is assumed to be the free-flow travel time between zones, t_{ij}.

$$Friction\ \ Factor = a * t_{ij}^{b} * e^{(c * t_{ij})} \tag{1}$$

Note that the 'a' parameter has no effect when used in the gravity model and is only used to scale up the friction factors, so they are more manageable to work with. Also, the more recognizable exponential form of this function is obtained by setting 'c' to zero and 'b' to -2.

After compiling a database of published Tanner functions (*4*), each of the five cases was run for every collected Tanner function. The average trip length results for each equation, trip type, and case are shown in TABLE 10, with the highlighted values to be carried on to the trip assignment step.

TABLE 10 Average Trip Lengths in Minutes by Impedance Function

	Minimum			Average			Original Selection		
Equation	HBW	HBO	NHB	HBW	HBO	NHB	HBW	HBO	NHB
1	13.7	16.3	14.5	11.0	15.9	12.2	13.7	16.2	13.2
2	13.7	16.3	14.5	11.0	16.0	12.2	13.7	16.3	13.2
3	13.5	16.3	14.1	10.9	15.9	11.9	13.5	16.2	12.8
4	13.0	15.6	11.0	10.5	15.2	9.2	13.0	15.5	9.9
5	14.5	12.2	14.4	11.7	12.1	12.3	14.5	12.1	13.2
6	13.6	16.3		11.0	15.9		13.5	16.2	
7	13.2			10.6			13.2		
8	13.3			10.9			13.3		
9	13.0			10.6			13.0		
10	12.8			10.3			12.8		
Range	1.7	4.1	3.5	1.4	3.9	3.1	1.7	4.1	3.3

	Revised Selection			Maximum		
Equation	HBW	HBO	NHB	HBW	HBO	NHB
1	11.4	14.7	13.2	11.3	14.7	12.8
2	11.4	14.7	13.2	11.3	14.7	12.8
3	11.3	14.7	12.8	11.2	14.7	12.4
4	10.8	14.1	9.9	10.8	14.1	9.6
5	12.1	11.5	13.2	12.1	11.5	12.8
6	11.3	14.7		11.3	14.7	
7	11.0			10.9		
8	11.2			11.2		
9	10.9			10.9		
10	10.6			10.6		
Range	1.5	3.2	3.3	1.5	3.2	3.3

After applying the various Tanner functions to the CAMPO data, it can be seen that the HBW models vary the least, being within 1.7 minutes of each other for every case. The most volatile collection of Tanner functions is for the HBO trip purpose, which varies by up to 4.1 minutes for two of the cases; however, the NHB trip purpose tanner functions were found to be the most volatile for the 'revised selection' and 'maximum' cases. Therefore, if data is to be collected for this step, then trip length information for HBO and NHB should be obtained so as to calibrate the Tanner function for this purpose. By varying the 'a', 'b', and 'c' parameters, planners may be able to better represent the locally collected travel times.

The 'original selections' for this step are the models taken from an FHWA study (*9*), in part because it is based on experience gained from analyzing various MPO models across the country. The 'revised selections' were chosen based on the average HBW trip length reported for the area by the

ACS, and the NCHRP 365 recommendation that the HBO and NHB trip lengths be approximately 75 and 85% of the HBW trip length (*5*). The total average trip lengths between the selections differ by less than one minute.

Because free flow travel time is considered to be a reliable input, no distributions will be applied in this step. Instead, we will track the effect on the average trip length caused by the two input variations examined during trip generation by regression. By applying the results from the 90% confidence bounds to the 'revised' Tanner function selection, it was found that the average trip length for CAMPO varies by approximately 4 seconds for HBW and HBO and 3 seconds for NHB. Therefore, it can be determined that the variation of trip generation inputs has a negligible effect on average trip length.

The decision to spend resources at all on calibrating the Tanner function should be based on the planner's personal preference. If a planner is willing to accept an error of x minutes in average trip length and the variation for each trip purpose is below this value, then no resources need be expended. Average trip length is a value that is not readily available to many small MPOs. If the trip assignment step results indicate a systematic overloading or under-loading of links, and trip generation is not the reason, then the value of average trip length must be further researched.

3. TRIP ASSIGNMENT

Transit ridership is very low in CAMPO, so the mode choice step can be skipped. The user equilibrium method will be applied for trip assignment. To assess sensitivity, the α and β parameters within the link performance function, shown in the standard form in Equation 2, are to be varied.

$$t = t_0\left(1+\alpha\left(\frac{v}{c}\right)^{\beta}\right) \tag{2}$$

where: $t \equiv$ Congested travel time,
$t_0 \equiv$ Free-flow travel time, and
$v/c \equiv$ volume to capacity ratio

A database of link performance functions (LPFs) applying to all link types was first accumulated (*4*).

With a collection of equations, each model can be applied to the trip distribution results for the five cases and two carry-over input variations from trip generation by regression. The results of this application in terms of total VMT during the peak hour can be seen in TABLE 11.

TABLE 11 Total Peak-Hour VMT by Link Performance Function

Equation	Minimum	Average	Original Selection	Revised Selection	Maximum
1	64,724	104,507	100,252	110,847	192,766
2	64,932	105,371	100,984	111,838	200,187
3	64,927	105,292	100,956	112,096	216,596
4	64,720	104,705	100,506	111,357	195,993
5	64,704	105,507	100,536	111,329	195,608
6	64,841	105,601	100,935	111,751	203,161
% Difference (Range/Minimum)	0.4	1.0	0.7	1.1	12.4

The results for total VMT vary from 64,704 to 216,596 between the cases. These values differ by approximately a factor of four, showing the consequences from choosing different models.

As for the selected cases, the 'original selection' uses the standard FHWA parameter values of α=0.15 and β=4.0, and the 'revised selection' was made based on VMT data obtained from the NHTS transferability program. (*6*) The results of these selections differ by 11,843 total peak-hour VMT, which is a percent difference of 12%. This finding further highlights the consequences of making model selections based solely on the parameter source.

With regards to the variation of the LPF within each case, there appears to be only a slight effect on the total VMT. For every case except the maximum, the LPFs result in outputs that differ by less than 1.1%. This suggests that collecting VMT data for the purpose of refining the LPF parameters is not critical in most cases for CAMPO. To confirm this supposition, there is need to examine the loadings at the link level. Before the standard calculation of deviation in terms of the percent root mean squared error (PRMSE) between modeled and observed link volumes is done, a check for unusually large link flow rates ought to be made.

The sensitivity of PRMSE to a variation in LPF parameters is studied in TABLE 12.

TABLE 12 Peak Hour PRMSE by Link Performance Function

Equation	Minimum	Average	Original Selection	Revised Selection	Maximum
1	33.55	52.28	75.01	46.31	20.79
2	68.51	26.36	35.18	27.46	12.58
3	32.60	13.46	16.33	12.37	6.73
4	19.52	30.86	47.33	40.28	15.07
5	10.04	32.31	40.75	41.34	12.78
6	24.33	27.77	33.71	28.50	13.06
Range	58.5	38.8	58.7	33.9	14.1

While transferring LPF parameters was found to have a negligible effect on peak hour VMT, the variation of parameters significantly affects the model fit to collected traffic counts. PRMSE varies by up to 60% across parameter combinations. The 'original' and 'revised' LPF selection differ by over 60%, again showing the caution planners must take when selecting model parameters. The model with the best fit, according to the lowest PRMSE, is the 'maximum' case, however, the trips per household for this model were considered to be unreasonable for the CAMPO area.

As with trip distribution, the inputs for trip assignment, such as capacity values, are deemed reliable and therefore it would not be beneficial to apply statistical distributions. Instead, the sensitivity of the output due to applying the two 90% input bounds from trip generation was examined. In terms of total VMT, the two variations resulted in values of 101,121 and 123,278 as compared to the 112,096 value obtained using the current data. Therefore, the trip generation inputs are significant when it comes to total VMT resulting in a percent difference around 22%, almost exactly the same value found for the percent difference of total trips generated for the two bounds. Congestion effects on route choice are likely the reason the two values differ slightly. The difference, due to inputs, provides further proof that data collection is indeed critical at the trip generation step.

4. CONCLUSIONS AND RECOMMENDATIONS

From this study, three basic lessons are learned: (1) sensitivity analyses can guide data collection, (2) it is possible to quantify output volatility by fitting distributions to input data, and (3) there can be significant consequences from selecting different models. For the CAMPO region in particular, it was learned that:

- Borrowing parameter data from communities with similar socio-economic and geographic characteristics does not guarantee that the data are the most appropriate for the area being studied;
- Cross-classification outputs are more sensitive to varied input data than regression outputs, while only slightly more sensitive to varied parameters;
- To increase the confidence in trip generation outputs by regression, careful attention to the accuracy of retail employment data should be given, particularly in the central business district (CBD) or other employment centers;
- The variation of trip generation inputs have a negligible effect on average trip length;
- To increase confidence in trip distribution outputs, emphasis should be placed on the accuracy of HBO and NHB trip lengths, so as to better calibrate the Tanner Function.
- The link performance function (LPF) parameters are not that critical to the total peak hour VMT outputs, however they are extremely critical to the model fit, and therefore LPF calibration data should be collected.

Such determinations allow for the more efficient use of resources. For this reason, among others, sensitivity analyses are strongly recommended (*10*).

With the input distributions and lessons learned in this research, future studies will focus on how to manage the risk of programming less critical capacity-building projects due to uncertain travel demand model inputs and parameters.

REFERENCES

1 Barton-Aschman Associates, Inc. and Cambridge Systematics, Inc.. Model Validation and Reasonableness Checking Manual. Prepared for the Travel Model Improvement Program of the Federal Highway Administration, Publication No. FHWA-EP-01-023. Arlington, TX, February 1997. Retrieved May 2, 2008 from http://tmip.fhwa.dot.gov/resources/clearinghouse/docs/mvrcm/mvrcm.pdf.

2 Fehr and Peers, "Las Vegas Travel Demand Model Guidelines for Estimation, Calibration, and Validation", Regional Transportation Commission of Southern Nevada, 2005.

3 Zhao, Yong, and Kara M. Kockelman. "The propagation of the uncertainty through travel demand forecasts: An exploratory analysis." In *Annals of Regional Science* (2002) 36:145-163

4 Ford, Kevin M., "Uncertainty and Risk Management in Travel Demand Modeling", M.S. Thesis, School of Civil Engineering, Purdue University, West Lafayette IN, May 2009.

5 Martin, William A. and Nancy A. McGuckin, "Travel Estimation Techniques for Urban Planning", National Cooperative Highway Research Program Report 365. Prepared for the Transportation Research Board, National Research Council. National Academy Press: Washington D.C. 1998.

6 Hu, Patricia S., Timothy Reuscher, Richard L. Schmoyer, and Shih-Miao Chin, "Transferring 2001 National Household Travel Survey". Prepared for the Federal Highway Administration, May 2, 2007.

7 United States Census Bureau: American Community Survey (2007). *Accuracy of the Data (2007)*. http://www.census. gov/acs/www/Downloads/ACS/ accuracy2007.pdf.

8 Grecco, W.L., J. Spencer, F.J. Wegmann, and A. Chatterjee, T, "Transportation Planning for Small Urban Areas", National Cooperative Highway Research Program Report 167. Prepared for the Transportation Research Board, National Research Council. National Academy Press: Washington D.C. 1976.

9 Ismart, Dane, "Calibration and Adjustment of System Planning Models", Federal Highway Administration Publication No. FHWA-ED-90-015, December 1990. Retrieved May 21, 2008 from http://ntl.bts.gov/DOCS/377CAS.html.

10 Committee for Determination of the State of the Practice in Metropolitan Area Travel Forecasting, "Special Report 288: Metropolitan Travel Forecasting – Current Practice and Future Direction", Transportation Research Board, National Research Council. National Academy Press: Washington D.C. 2007.