Sample Weighting and Expansion
Part II: Average Weekday Weights

California Household Travel Survey 2012/13
for the
San Francisco Bay Area

Planning Section
Metropolitan Transportation Commission
101 Eighth Street
Oakland, California 94607

October 2013

Table of Contents

List of Tables and Figures

Tables in Body of Report

Appendices

I. INTRODUCTION

This working paper is the second in a series for documenting procedures and results of the year 2012/2013 California Household Travel Survey conducted in the San Francisco Bay Area (CHTS12/13). The purpose of this working paper, *Sample Weighting and Expansion: Part II: Average Weekday Weights,* is to describe procedures for weighting and expanding CHTS12/13 household and person files for the "weekday sample" (that is, persons providing travel data on an assigned weekday).

Four sets of weights are envisioned for this study:

1) Average Daily weights (for the combined samples);
2) Average Weekday weights (for the Monday through Friday samples);
3) Average Saturday weights (for the Saturday sample); and
4) Average Sunday weights (for the Sunday sample).

Working papers such as this report tend to be a "work in progress" and may be updated to incorporate other improvements, clarifications and analyses. Please check with MTC to obtain the most current version of this and other working papers.

Background on "what is weighting" and "what is raking" and previous household travel surveys is covered in Part I of this three part report on sample weighting and expansion. It is suggested that the Part I report is reviewed before Part II (weekday weights), and Part III (Saturday and Sunday weights).

The Bay Area portion of the California Household Travel Survey (2012/13) was a one-day travel/activity data from 9,719 households. Data was collected between February 1, 2012 and January 31, 2013. Of the 9,719 sample households, 8,086 provide weekday travel data; 717 provide Saturday data; and 916 households provide Sunday data.

Table 1
CHTS 2012/13 Bay Area Sample Households by Day of Week

Day of Week	Sample Households	% of Total
Monday	775	8.0%
Tuesday	2,149	22.1%
Wednesday	2,124	21.9%
Thursday	2,160	22.2%
Friday	878	9.0%
Saturday	717	7.4%
Sunday	916	9.4%
TOTAL	9,719	100.0%
Weekday Total	8,086	83.2%
Weekend Total	1,633	16.8%
Tuesday-Thursday Total	6,433	66.2%

The proposed raking scheme for the Bay Area CHTS2012/13 WEEKDAY sample has seven raking levels:

1) County (9) by Tenure (2) by Race/Ethnicity of Householder (5);
2) PUMA (55) by Tenure (2) by Minority Status of Householder (2);
3) County (9) by Tenure (2) by Workers in Household (4);
4) County (9) by Tenure (2) by Vehicles in Household (4);
5) PUMA (55) by Tenure (2) by Age of Householder (5);
6) County (9) by Number of Persons Age 20-29 in Household (3); and
7) PUMA (55) by Tenure (2) by Household Size (5).

This is the same raking strategy as used to prepare the "average daily" or "combined sample" weights, documented in Part I.

II. EXPLORING CENSUS AND SURVEY CHARACTERISTICS

The purpose of this section is to compare census "marginal control total" data to the corresponding patterns from the CHTS 2012/13 weekday sample in the Bay Area. This will highlight the critical biases in the survey that are correctable using appropriate weighting schemes.

Detailed data tables included in Appendix A are reported in this section. As appropriate, census data sources (Census 2010 "short form" data versus American Community Survey 2007/11) are cited.

More detailed analyses of the "combined sample" relative to Census and ACS data is included in the Part I report. This Part II report focuses on the raking levels that compare the WEEKDAY sample to the same census/ACS data.

Table 2 summarizes the raking level, and corresponding appendix table, used in developing the various raking/weight models.

Table 2
Raking Levels and Corresponding Appendix Tables

Table	Raking Level Description
A.3	County (9) by Tenure (2) by Race/Ethnicity (5)
A.8	PUMA (55) by Tenure (2) by Minority Status (2)
A.4	County (9) by Tenure (2) by Workers in Household (4)
A.5	County (9) by Tenure (2) by Vehicles in Household (4)
A.7	PUMA (55) by Tenure (2) by Age of Householder (5)
A.9	County (9) by Number of Persons Age 20 to 29 (3)
A.6	PUMA (55) by Tenure (2) by Household Size (5)

The ordering of the raking levels is not important, except for the final, last raking level. The "last rake" will show the best fit, comparing the modeled, expanded households to the "marginal control totals." For previous and current MTC weighting approaches, the focus is on obtaining accurate estimates of households by geography by household size, in order to ensure the best approximation of total household population.

III. VALIDATION AND EVALUATION OF WEIGHTING METHODS

Six weighting methods are examined in this section. Detailed data tables are included as Appendices B and C.

The study consultant developed a set of weights, which are denoted as "Model #0" weights in this overall study. MTC staff developed multiple sets of raking models/weights, denoted as Model #1, Model #1c, Model #2, Model #2c1, Model #2c2, Model #3, Model #3c1 and Model #3c2. The "c" stands for "constrained".

The Model #1 and #2 series are for the "combined sample" also known as the "average daily sample." The Model #3 series of weights are for the "weekday sample". (Model #4 is for the Saturday weights; Model #5, for the Sunday weights.)

Model #3 weights were constrained using floors and ceilings. Two options were produced:

1) Model #3c1, constraining the weights to 85.0 floor and 2500.0 ceiling; and
2) Model #3c2, constraining the weights to 61.0 floor and 3000.0 ceiling.

Table 3
Range of Average Weekday Sample Weights by Weighting Model

	Median	Mean	Minimum	Maximum
Model #3	164.5	322.54	2.91E-08	4,772.94
Model #3c1	164.5	322.65	85.00	2,500.00
Model #3c2	164.5	322.61	61.00	3,000.00

Appendix B summarizes the household level validation of Models #3, #3c1 and #3c2, at the county and regional level. Regional level results are provided in Table B.1.1 through B.1.6. Model #3 perfectly matches the Census 2010 count of households by household size, given that the last raking level is household size by tenure (Table B.1.1.) Model #3 tenure is slightly different than Census 2010 control totals due to the need for collapsing categories at several PUMA/tenure/household size combinations. Comparison for the other household categories (householder age, householder race/ethnicity,

workers/household, vehicles/household) show reasonable matches, Model #3 relative to Census 2010 or ACS data.

Appendix C summarizes the person level validation of Models #3, #3c1 and #3c2, at the county and regional level. Household population by race/ethnicity is not easily derivable from Census 2010 standard tabulations (Table C.1). (Census 2010 data is published for household population white, not-Hispanic; and Hispanic/Latino, but not for the other three minority groups.)

Household population by the eight "person types" used in the current travel model system are shown in Tables C.5 and C.6.1-C.6.9. Models #3, #3c1 and #3c2 show very reasonable estimates of workers, non-workers, and students, relative to Census/ACS observed totals.

The last set of tables in Appendix C review the distribution of student population by age and by school level. Model #3c2 does a good job in estimating high school (-0.3 percent) enrollment; but over-estimates undergraduate college (+11.1%), grade school (K-8) (+10.3 percent); and under-estimates graduate school enrollment (-27.4 percent) (Table C.7.3).

The last set of tables in Appendix C, Tables C.12 through C.16, examines average (mean) characteristics by market segment: average household size, average workers per household, average students per household, and the average age of persons in household. The rates are shown for the consultant-provided "Model #0", Saturday (Model #4c1), Sunday (Model #5c1) and the three weekday models (#3, #3c1, #3c2).

Data is compared to American Community Survey 2007/11, from the ACS PUMS records. Data is summarized for the ACS and the six raking models. These tables are interesting to show how these regional mean characteristics vary by the different household classifications as used in this study.

Appendix D is a summary of the "person correction factors" used to adjust the household weights for the very large (five-or-more person) households. This technique was also used in adjusting the 1990 and 2000 Bay Area Travel Surveys. This is a fix since the average size of 5+ persons in the Census (5.979 persons/5+ person household) is slightly higher than the average size in the weighted weekday survey (5.504 persons/5+ person household). These correction factors are calculated at the PUMA-of-residence, and may range from 0.91 to 1.31. These person correction factors are applied equally to all members within the 5+ person household, ensuring that the final person factors do not vary within each sample household.

IV. NEXT STEPS

The recommendation is that Model #3c2 weights are the final MTC weights on the Bay Area WEEKDAY sample households in the CHTS 2012/13 database. This is for the "weekday" sample (N=8,086 sample households) for households providing travel data for Monday through Friday through the 2012/13 survey period.

The next step in this analysis is to prepare separate weights for the "Saturday sample households" (N=717); and the "Sunday sample households" (N=916). It is envisioned that the "combined sample weights" will be used when estimating auto ownership level models, and other analyses focusing on the demographics of the household. For purposes of estimating aggregate "average weekday" travel characteristics, only data from the "weekday sample households" would be used.

Future steps in the analysis of the CHTS 2012/13 travel survey for the Bay Area include detailed processing of the unlinked trip records to produce a linked trip, tour and sub-tour files. The product of this trip linking/chaining process will be both traditional linked trip files, as used in trip-based travel demand models; and tour-based travel files, for supporting the current and future generation of MTC travel behavior models.

Procedures to impute missing values will be documented in separate technical reports. Other "data cleaning" notes will be included in MTC staff notes and technical documentation.

New weighting/raking methods have also been developed for the entire statewide CHTS 2012/13 databases. Technical reports documenting these methods will be produced. In addition, the recommended household and person-level weights will be extracted and provided to CHTS 2012/13 data users. Appropriate metadata will be developed to assist the data user.

Further research on raking methods will be undertaken as time permits. Options may include simplifying some of the three-dimensional raking schemes (e.g., omitting in tenure in the PUMA by tenure by household size) to analyze the impacts on extreme weights, and raking model closure.

Procedures to impute missing trips and tours may be required, and will probably be included in future technical reports.