FINAL REPORT

to

THE FLORIDA DEPARTMENT OF TRANSPORTATION
SYSTEM PLANNING OFFICE

on Project

Procedure for Forecasting Household Characteristics for Input to Travel-demand Models

FDOT Contract No. BD545, RPWO #79
(UF Project 00064011)

December 31, 2008

Transportation Research Center
The University of Florida

DISCLAIMER

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation.

METRIC CONVERSION CHART

U.S. UNITS TO METRIC (SI) UNITS

LENGTH

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
in	Inches	25.4	millimeters	mm
ft	Feet	0.305	meters	m
yd	Yards	0.914	meters	m
mi	Miles	1.61	kilometers	km

METRIC (SI) UNITS
TO U.S. UNITS

LENGTH

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
mm	millimeters	0.039	inches	in
m	Meters	3.28	feet	ft
m	Meters	1.09	yards	yd
km	kilometers	0.621	miles	mi

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Procedure for Forecasting Household Characteristics for Input to Travel-demand Models	5. Report Date December 31, 2008
	6. Performing Organization Code UF-TRC
7. Author(s) Sivaramakrishnan Srinivasan, Lu Ma, and Karun Yathindra	8. Performing Organization Report No. TRC-FDOT-64011-2008
9. Performing Organization Name and Address Transportation Research Center University of Florida 512 Weil Hall, PO Box 116580 Gainesville, FL 32611-6580	10. Work Unit No. (TRAIS)
	11. Contract or Grant No. FDOT Contract BD545, RPWO #79
12. Sponsoring Organization Name and Address Florida Department of Transportation 605 Suwannee Street, MS 30 Tallahassee, FL 32399	13. Type of Report and Period Covered Final Report
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Over the past several years, there has been a growing interest in the development of disaggregate (individual- or household-level) travel-demand models. In the case of Florida, this is evident from their efforts to incorporate socio-demographic variables (*i.e.,* household characteristics) within the FSUTMS structure via "lifestyle" trip production models. However, the lack of a systematic procedure to forecast the household characteristics (*i.e.,* the lifestyle variables) required by such disaggregate travel-demand models has been recognized as an important impediment to furthering these efforts for state-wide adoption. In this context, the broad focus of this research is to contribute towards the development of methodology for comprehensively forecasting all traveler characteristics required as inputs to travel-demand forecasting models. This procedure is also referred to as synthetic population generation (SPG) in the literature. The following are the objectives of this study: (1) Develop a population synthesis procedure that accommodates both household- and person-level controls. (2) Assess the importance of controlling for person-level attributes by comparing the populations synthesized with only household level controls with the one synthesized using both household and person-level controls. (3) Validate the population synthesis by comparing the synthesized and true populations for a target year (by "back-casting"). (4) Compare the predicted trip rates from trip generation models applied to both true and synthesized populations.

17. Key Words Population synthesis, disaggregate travel-demand models, census data	18. Distribution Statement No restrictions.

19. Security Classif. (of this report) Unclassified.	20. Security Classif. (of this page) Unclassified.	21. No. of Pages 77	22. Price NA

ACKNOWLEDGMENTS

We acknowledge the support extended by the project manager, Vidya Mysore of the Systems Planning Office of FDOT during the project. Yongqiang Wu, formerly of the FDOT Systems Planning Office, also provided useful inputs. The authors would also like to acknowledge the contributions of Abishek Komma, a graduate student in the Department of Civil and Coastal Engineering at the University of Florida.

EXECUTIVE SUMMARY

Over the past several years, there has been a growing interest in the development of disaggregate (individual- or household-level) travel-demand models. This interest is motivated by several factors such as (1) reduction of aggregation errors, (2) ensure sensitivity to demographic shifts like the ageing of the population, (3) capture differential sensitivity and response of travelers to policy actions, and (4) address special travel-needs of certain population groups.

The recognition of the above-described issues by Florida Department of Transportation is evident from their efforts to incorporate socio-demographic variables (*i.e.,* household characteristics) within the Florida Standard Urban Transportation Model Structure (FSUTMS). Specifically, the Tampa Bay and the South-East Florida regions have developed "lifestyle" trip production models. However, the lack of a systematic procedure to forecast the household characteristics (*i.e.,* the lifestyle variables) required by such disaggregate travel-demand models has been recognized as an important impediment to furthering these efforts for state-wide adoption.

In this context, the broad focus of this research is to contribute towards the development of methodology for comprehensively forecasting all traveler characteristics required as inputs to travel-demand forecasting models. This procedure is also referred to as synthetic population generation (SPG) in the literature.

The state-of-the-practice approach to population synthesis involves the use of the Iterative Proportional Fitting (IPF) method. While there have been several applications of this approach, the following issues still remain. First, the number of controls used in the synthesis of the population has been limited. In particular, most practical applications do not control for person-level attributes such as age and gender. Second, documentation of the validation of the procedure, especially in the context of a target year population is limited. Third, there does not seem to be any comparison of the travel patterns predicted using true populations with those predicted using synthetic populations.

This research addresses the issues identified above. A new greedy-heuristic data-fitting algorithm is developed that can be used to synthesize population with a large number of control tables both at household- and person-levels. The procedure is implemented in GAUSS, a matrix programming language. The code was used to synthesize the year 2000 population for 13 census tracts of varying populations and areas in Florida. Two sets of populations were estimated – the first with only household-level controls and the second with both household- and person-level controls. Validation analysis indicates that the second synthesized population matches the true distributions better. In fact, the extent of mismatch with the (uncontrolled) person-level tables is significant with the first population (i.e., synthesized with only household controls).

As a second step, the populations of 1990 were synthesized for the same 13 census tracts. Once again, two sets of populations were synthesized. One used the year-2000 population synthesized with only household-level controls as the seed data whereas the second used the year-2000 population synthesized with both household- and person-level controls as the seed data. The aggregate characteristics of the synthesized populations were compared with several control tables from the 1990 US Census. Once again, the results indicate that the use of both person- and household- controls in the base year synthesis leads to more accurate population estimates for the target year. Overall, the analysis highlights the value of a methodology that incorporates both controls in population synthesis.

Finally, travel estimates obtained by applying trip-generation models to the true population were compared with those obtained by applying the same models to a synthetic population. Trip generation models (household-level and person-level) were estimates using the weekday, national sample from the National Household Travel Survey of 2000. Subsequently, the estimated models were applied to the Florida sample of the survey data (i.e., the true population) to predict the travel estimates. The population characteristics of the Florida sample were also synthesized and the models were applied to these synthesized populations. The analyses provide some evidence in favor of disaggregate models. Specifically, for two trip purposes (home-based other and non-home-based), we find that the disaggregate models can perform just as good (if not better) as the aggregate models. For the same trip purposes, we also find that the travel estimates obtained by applying the models to the synthetic population are as accurate as the ones obtained by applying the same model to the true population. Thus, the need to synthesize the population characteristics does not necessarily deteriorate the trip-generation predictions (from linear-regression models) substantially. The results for the home-based work trip purpose highlights the need for choosing the appropriate econometric structure when developing disaggregate models and the right control variables for the population synthesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

Traditionally, travel-demand models have required few, aggregate data inputs (such as zonal population, number of households, and employment levels) for forecasting. However, over the past several years, there has been a growing interest in the development of disaggregate (individual- or household-level) travel-demand models. This interest is motivated by several factors including (1) reduction of aggregation errors, (2) ensuring sensitivity to demographic shifts like the ageing of the population, (3) capturing differential sensitivity and response of travelers to policy actions, and (4) addressing special travel-needs of certain population groups.

The recognition of the above-described issues by Florida Department of Transportation is evident from their efforts to incorporate socio-demographic variables (*i.e.,* household characteristics) within the Florida Standard Urban Transportation Model Structure (FSUTMS). Specifically, the Tampa Bay and the South-East Florida regions have developed "lifestyle" trip-production models. However, the lack of a systematic procedure to forecast the household characteristics (such as the lifestyle variables) required by such disaggregate travel-demand models has been recognized as an important impediment to furthering these efforts for state-wide adoption. In this context, the broad focus of this research is to contribute towards the development of methodology for comprehensively forecasting all traveler characteristics required as inputs to travel-demand forecasting models. This procedure is also referred to as synthetic population generation (SPG) in the literature.

The state-of-the-practice approach to population synthesis involves the use of the Iterative Proportional Fitting (IPF) method (Beckman *et al.*, 1996). While there have been several applications of this approach, the following issues still remain. First, there number of controls used in the synthesis of the population has been limited. In particular, most practical applications do not control for person-level attributes such as age and gender. Second, documentation of the validation of the procedure, especially in the context of a target year population is limited. Third, there does not seem to be any comparison of the travel patterns predicted using true populations with those predicted using synthetic populations.

In the light of the above discussions, the following are the objectives of this study:

1. Develop a population synthesis procedure that accommodates both household- and person-level controls.
2. Assess the importance of controlling for person-level attributes by comparing the populations synthesized with only household-level controls with the one synthesized using both household- and person-level controls.
3. Validate the population synthesis by comparing the synthesized- and true-populations for a target year (by "back-casting")
4. Compare the predicted trip rates from trip generation models applied to both true- and synthesized- populations.

Overall, this study contributes towards enhancing the population-synthesis procedure. A heuristic data-fitting approach is developed that is able to handle both household- and person-level control tables. The algorithm is implemented using the GAUSS programming language. Several validation results are presented to establish the superiority of the proposed method over the state-of-practice approaches.

The rest of this document is organized as follows. Chapter 2 presents an overview of the methods available for population synthesis. Chapter 3 discusses our method for synthesizing the base-year population. This chapter also includes the results from the application of the procedure to generate the population characteristics for 13 census tracts in Florida for the year 2000. Chapter 4 presents our approach to synthesize the target year population. The procedure is applied to generate the population characteristics of the same 13 census tracts for the year 1990 and these are compared to the true values from the 1990 US Census. Chapter 5 presents trip-generation models estimates using the 2001 NHTS datasets. In addition, the population characteristics of the Florida data are also synthesized using the same survey data. Subsequently, these models are used to predict the volume of trips for Florida using both the "true" population characteristics available directly from the survey and the synthesized population characteristics. Chapter 6 summarizes this report and identifies the major conclusions.

CHAPTER 2 METHODS FOR POPULATION SYNTHESIS

This chapter presents an overview of the methods currently available for population synthesis. Section 2.1 presents a conceptual overview of the overall synthesis procedure. Although the intent is generally to generate a population for a future year (called the target-year in the rest of this document), the synthesis procedure begins with generating a population for a current year (called the base-year in the rest of this document). Section 2.2 discusses methods for synthesizing base-year population whereas Section 2.3 describes the methods for target-year population synthesis. Finally, Section 2.4 presents a summary and identifies the primary contributions of this research.

2.1 Conceptual Overview of the Population-Synthesis Procedure

A conceptual overview of the population-synthesis procedure is presented in Figure 1. Broadly, the first step in this procedure is to generate the population for the *base* year. For the purposes of this document, *base year* is defined as the most recent census year in the past (currently, this would be year 2000). This base-year population then forms an input in the synthesis of the population for any *target* year. A *target* year is defined as any year beyond the base year and may or may not be a year for which the decennial census has been planned. That is, if the base year is 2000, years 2003, 2010, and 2025 would all qualify as target years.

The synthesis of the base-year population is performed using data-fusion techniques. Broadly, aggregate control-tables (often at the census-tract level) are fused with disaggregate data on population characteristics (seed data) available for a sample of households in the PUMA to which the census tract belongs. The result is a synthetic population for the base year comprising households drawn from the corresponding PUMA such that the aggregate characteristics are consistent with the control tables for the census tract. Details of this data-fusion procedure are described in Section 2.2

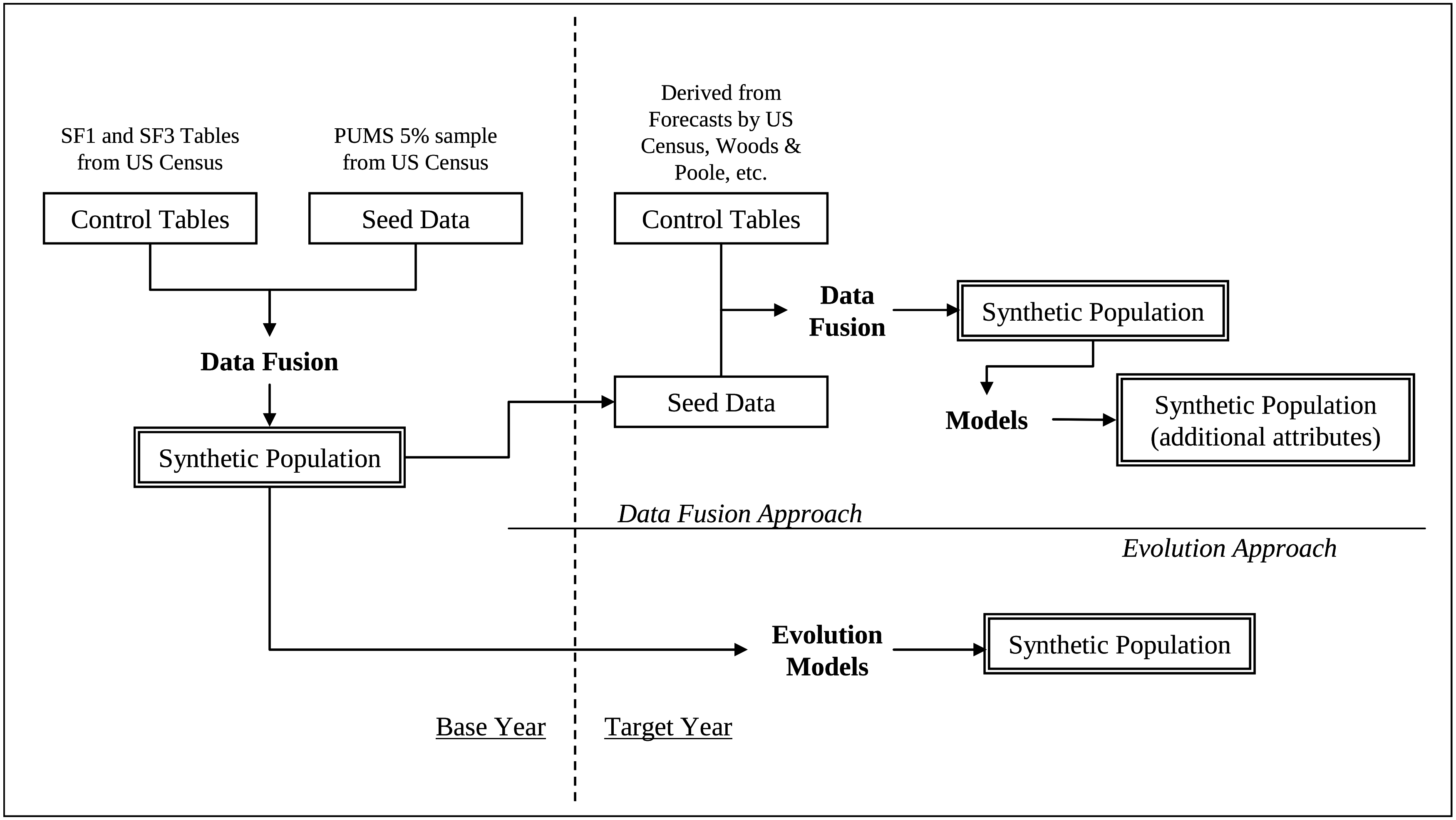


Figure 2.1 Conceptual Framework of the Population Synthesis Procedur

Given the base-year synthetic population, there are two broad approaches to generating the target-year population. The first methodology (*Data Fusion Approach* - discussed in Section 2.3.1) involves the application of the data-fusion techniques similar to the one for the base-year synthesis. The base- year population serves as the seed data in this process. This methodology may also involve the use of statistical models to generate attributes that are not directly synthesized by the fusion approach. The second approach (*Evolution Approach* - discussed further in Section 2.3.2) involves "growing" each base-year household over time to determine its characteristics at the target year. This involves modeling complex phenomenon such as household formation, dissolution, and migration.

2.2 Synthesizing the Base-Year Population

The state-of-the-practice approach to *base-year* population synthesis involves fusing aggregate control tables with disaggregate seed data. Control tables are one-way or multi-way marginal distributions. Each of these tables corresponds to the joint distribution of a *subset* of the required population attributes. Typically, these distribution tables are available from the census SF1 and SF3 files and at the spatial resolution of census block groups or census tracts. The population is synthesized at the spatial resolution of the control tables (this is referred to as the "synthesis area" in the rest of this document). The seed dataset comprises a sample of population records with each household/person characterized by *all* the attributes of interest. The location of these households is typically known only at a more aggregate spatial scale (in contrast to the finer spatial resolution of the control tables). Typically, such household-level information is obtained from the US census Public Use Microdata Samples (PUMS) and the location is defined in terms of the Public Use Microdata Areas (PUMAs).

The state-of-the-practice data fusion procedure involves two major steps. First, a joint multi-way distribution of all attributes of interest is generated using the Iterative Proportional Fitting (IPF) procedure (conceptually, the procedure is analogous to the Fratar balancing technique; detailed algorithm of the IPF procedure is available from Beckman *et al.*, 1996). The IPF procedure ensures that, when the multi-way distribution is appropriately aggregated, the results match the marginal distributions provided by the control tables (the extent of "matching" depends on the tolerance used). The result of this iterative procedure is a multi-way distribution table that provides the number of households of each type in the synthesis area. In the second step, individual household records are drawn from the seed dataset using monte-carlo simulation so as to satisfy the joint multi-way distributions.

This methodology has been applied to support travel-demand modeling in several areas such as Portland Metro, San Francisco, New York, Columbus, Atlanta, Sacramento, Bay Area, and Denver. Bradley and Bowman (2006) and Bowman (2004) provide a general overview of these applications. The Sacramento application is available in Bowman and Bradley (2006) and the Atlanta application and validation results are presented in Bowman and Rousseau (2006).

All the applications discussed thus far control for only household-level attributes. Guo and Bhat (2007) provide an extension to incorporate both household- and person-level controls in the IPF-based population-synthesis procedure. Broadly, this procedure begins with generating the household-level and person-level multidimensional tables independently. Next, households are drawn from the seed data based on the household-level multi-way distributions. The households are retained as long as they do not violate any person-level distributions (subject to tolerance criteria). The authors applied their procedure to the Dallas-Fort Worth area and demonstrated that the synthesized population matches more closely with the true population if both household and person-level controls are incorporated.

2.3 Synthesizing the Target-Year Population

The procedure for generating the population characteristics for a base year was described previously. In this section, we discuss methods for generating these characteristics for the *target year*. As indicated in Figure 1, there are two major classes of methods: (1) The Data Fusion Approach, and (2) The Evolution Approach. These are discussed in Sections 2.3.1 and 2.3.2 respectively.

2.3.1 The Data-Fusion Approach

The data-fusion approach for the synthesis of the target-year population is conceptually similar to the one used for the generating the base-year population. Once again, aggregate control tables and disaggregate seed data are the inputs.

The control-tables represent the aggregate socio-economic-mobility characteristics of the synthesis area in the target year. There are two key differences between the control tables used in the base-year synthesis and those used in the target-year synthesis. First, for the target year, the number of controls available is limited (and often multi-dimensional controls may not be available). In contrast, the base year would have several (and multi-dimensional) controls from the Census data. Second, the control tables for the target year may not even be available at the synthesis-area level and may have to be derived from more aggregate spatial units (such as the county).

The structure of the seed data for the target-year population synthesis is the same as the one for the base year. This is because the synthesized base-year population is taken as the seed data. The reader will note that the seed data for the base year are at the PUMA level, but from the same year which is in contrast to the seed data for the target year which are from the same census tract but are from the base year.

The methodology used for the population synthesis is predominantly the same as the one used in the base year. However, some of the attributes of interest may not be directly synthesized due to lack of control data. For these cross-section models can be used. A classic example of an attribute which is forecasted in such a manner is automobile ownership [see for example, the Oregon2 Model (Hunt *et al.*, 2004) or the SACOG model (Bowman and Bradley, 2007)]. Typically, US census does not provide projections of aggregate auto-ownership levels for any future year for use in a data-fusion approach. However, it is possible to develop cross sectional models of auto ownership (as a function of household characteristics, land use patterns, transportation system characteristics, etc.) using data from local household travel surveys or the PUMS. Thus, once the appropriate socio-economic characteristics for a forecast year have been determined using data-fusion techniques, the cross-sectional model can be applied to each household to generate the auto-ownership levels.

2.3.2 Evolution Approach

In this method, each household in the base-year synthetic population database is evolved or aged though time to determine its characteristics for any future year. This involves the development of a system of models that describe the common demographic/economic transitions that take place over the life-cycle of a household. These transitions include processes such as ageing, births, deaths, formation (marriage) and dissolution (divorce) of households, employment and education choices, children moving out of the household, automobile ownership decisions, and emigration from or immigration to the study region. Some of the currently available model systems that adopt such an approach include MIDAS (Goulias and Kitamura, 1996), MASTER (Mackett, 1990), CEMSELTS (Eluru *et al.*, 2008), DEMOS (Sundararajan and Goulias, 2003), and the HA module of the Oregon2 model system (Hunt *et al.*, 2003). Such methods are appealing as they try to mimic the real processes households go though and model behavioral decisions made at different stages of the life cycle. However, as identified by Eluru *et al.*, (2008), limited theoretical knowledge on the complex socio-economic evolution processes and the minimal availability of relevant data at the household level limit our ability to specify and estimate good models of household evolution.

2.4 Summary

A review of the recent literature indicates several studies aimed at generating the disaggregate socio-economic-mobility characteristics of the population. For the base year synthesis, the IPF-based methodology is most widely adopted. However, the number of controls used appears to be rather limited. In particular, most studies do not control for person-level attributes such as age and gender. Consequently, the synthesized base-year population may not accurately reflect these distributions. In turn, the accuracy of the target year population could also be affected as the synthesized base-year data forms a key input to the target-year synthesis (irrespective of whether the methodology is data-fusion or evolution). Further, if the data-fusion methodology is used for target-year synthesis (again, this appears to be the popular state-of-practice approach), some of the controls may be available at the person-level instead of all the controls being available at the household-level. This necessitates a population-synthesis procedure that is able to handle both household- and person-level controls.

In the light of the above discussion, this research develops a methodology for synthesizing the population characteristics by controlling for both household and person-level attributes. The research also validates the approach for both the base-year and target-years.

CHAPTER 3 SYNTHESIZING THE BASE-YEAR POPULATION

This chapter describes the procedure for synthesizing the base-year population. Section 3.1 identifies the input-data requirements, Section 3.2 describes the algorithm, and Section 3.3 presents the results of the application of the procedure for synthesizing the population for 13 census tracts in Florida.

3.1 Data

There are two major types of data required as inputs for synthesizing the base-year population: The “Control Tables” and the “Seed Data”. The former are discussed in Section 3.1.1 and the latter in Section 3.1.2

3.1.1 Control Tables

Control tables are one-way or multi-way marginal-distribution tables. Each of these tables corresponds to the joint distribution of a *subset* of the required population attributes. In this research, we use the distribution tables available from the US census SF1 and SF3 files. As already defined in Chapter 2, the spatial resolutions at which these data are available are called as the synthesis areas (i.e., the population is synthesized at this spatial unit). In this study, the census tracts are the synthesis areas.

Table 3.1 identifies twelve control tables (nine two-dimensional tables and three one-dimensional tables) used in this study. These controls cover most of the important socio-economic-mobility attributes commonly used in travel modeling (in the context of Florida, it is also useful to distinguish between the travel patterns of seasonal- and permanent- residents. However, the data from the US census does not provide such a distinction and hence we are unable to address this issue in this research). The categorical values that these attributes take are also listed in the table. For example, the dwelling unit can be either single-family or multi-family. Further, there is variability in the *universe* for which these control tables are defined. For instance, some attributes (household size, tenure, dwelling unit type, household structure, number of automobiles, and income) are defined for all households while others (age distribution of children, number of workers) are defined only for family households. Attributes such as age, gender, ethnicity, and citizenship are defined for all persons *including* those in group quarters

(This makes it necessary to simultaneously synthesize the populations in households and group quarters). The number of working hours per week is provided only for persons 16 years and older as it is necessarily zero for persons of age 15 or lesser. The last control table (multi-way distribution of age and gender) is defined for the population in group-quarters. One of the major strengths of the proposed population-synthesis algorithm is its ability to deal with such control tables from different "universes".

		Dimension 1		Dimension 2		
S. No	**Universe**	**Attribute**	**Categories**	**Attribute**	**Categories**	**SF Table Used**
1	Households	TENURE	Own, Rent	HHSIZE	1,2,3,4,5,6,7+	H15(SF1)
2	Households	TENURE	Own, Rent	DUTYPE	Single Family, Multi-Family	H32(SF3) adjusted by H15(SF1)
3	Households	TENURE	Own, Rent	NUMAUTO	0,1,2,3,4,5+	H44(SF3) adjusted by H15(SF1)
4	Households	HHSTRUCT	Family, Non-Family	HHSIZE	1,2,3,4,5,6,7+	P26(SF1)
5	Families	HHSTRUCT	Married couple, Other family	CHAGE[1]	None, Only <6 years, Only >=6 years, Both <6 years and >= 6 years	P34(SF1)
6	Families	HHSTRUCT	Married couple, Other family	NUMWORK[2]	0,1,2, 3+	P48(SF3) adjusted by P34(SF1)
7	Households	INCOME	< 30K, 30-50K, 50-75K, 75-125K, more than 125K		*NA*	P52(SF3) adjusted by P7(SF1)
8	Total Population	ETHNICITY	White, Black, Other, and Multiple Race		*NA*	P7(SF1)
9	Total Population	GENDER	Male, Female	AGE	0-5, 6-15, 16-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, over 75	P12+P14(SF1)
10	Total Population	CITIZEN	Native, Naturalized, Non Citizen		*NA*	P21(SF3)
11	Population >=16 years	GENDER	Male, Female	WRKHOURS[3]	0,1-14, 15-35, more than 35	P47(SF3) adjusted by P12+P14(SF1)
12	Population in Group Quarters	GENDER	Male, Female	AGE	0-17, 18-64, over 65	P38(SF1)

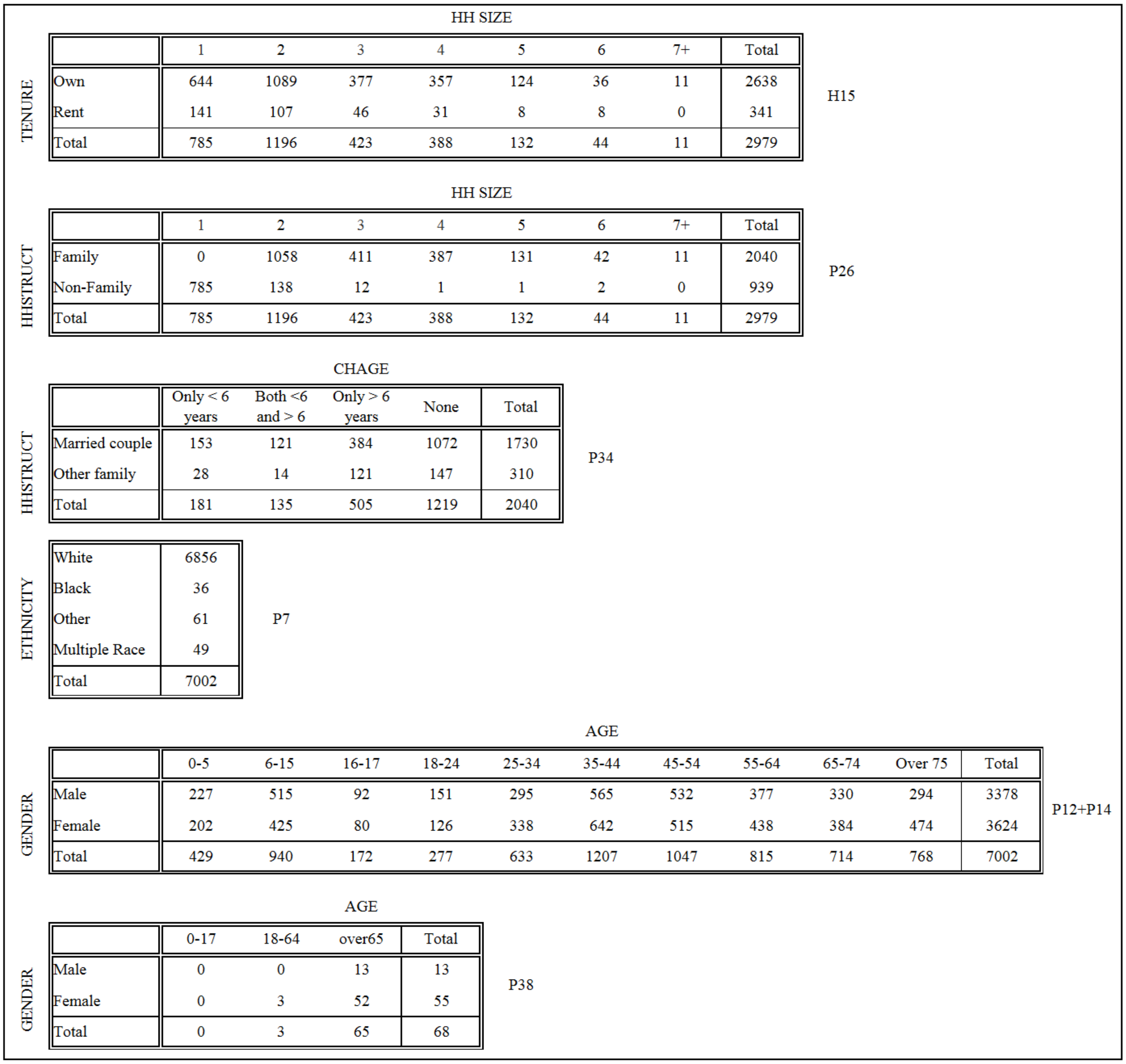
[1] Age distribution of "own children" in the household

[2] Number of workers (more than 0 hours per week in 1999)

[3] Hours of work per week in 1999

Table 3.1 Control Tables from SF1 and SF3

Table 3.1 also identifies the specific US Census SF table from which each of the control tables is drawn. Six (1, 4, 5, 8, 9, and 12) of the control tables are from SF1 tables and the remaining (2, 3, 6, 7, 10, and 11) are from SF3. All the twelve tables are presented for an arbitrary census tract in Figures 3.1 (the six SF1 Tables) and 3.2 (the six SF3 tables). The data in SF1 tables are based on complete count whereas those in SF3 are expanded from a sample. While it may be desirable to rely on the SF1 tables from the standpoint of accuracy, the SF3 tables provide several important attributes like dwelling-unit type, auto ownership, presence of children, and employment status which are not available in SF1 tables. Thus, it becomes necessary to draw from both SF1 and SF3 files. However, this leads to a situation in which there may be inconsistencies in the value of the attributes that the present in both tables. For example, based on the SF1 table “H15” that cross tabulates tenure against household size (see the column labeled “Total” in first control table in Figure 3.1), we find that 2638 households own their home and 341 rent. However, from the SF3 table “H32” that cross tabulates tenure against dwelling-unit type, we find that 2640 households own their home and 339 rent (see the column labeled “Total” in first control table in Figure 3.2). In order to reconcile these differences, an adjustment procedure is implemented that scales the SF3 values to match the corresponding aggregate controls form SF1. Note that the SF3 values are estimates and hence are “corrected” to match the SF1 totals which is expected to be more accurate. The SF1 table(s) used for the adjustment of each of the SF3 tables is also identified in Table 3.1. The adjusted SF3 control tables corresponding to the raw SF3 tables in Figure 3.2 are presented in Figure 3.3. The adjustment procedure is illustrated with a numerical example in Appendix A.



HH SIZE

TENURE

	1	2	3	4	5	6	7+	Total
Own	644	1089	377	357	124	36	11	2638
Rent	141	107	46	31	8	8	0	341
Total	785	1196	423	388	132	44	11	2979

H15

HH SIZE

HHSTRUCT

	1	2	3	4	5	6	7+	Total
Family	0	1058	411	387	131	42	11	2040
Non-Family	785	138	12	1	1	2	0	939
Total	785	1196	423	388	132	44	11	2979

P26

CHAGE

HHSTRUCT

	Only < 6 years	Both <6 and > 6	Only > 6 years	None	Total
Married couple	153	121	384	1072	1730
Other family	28	14	121	147	310
Total	181	135	505	1219	2040

P34

ETHNICITY

White	6856
Black	36
Other	61
Multiple Race	49
Total	7002

P7

AGE

GENDER

	0-5	6-15	16-17	18-24	25-34	35-44	45-54	55-64	65-74	Over 75	Total
Male	227	515	92	151	295	565	532	377	330	294	3378
Female	202	425	80	126	338	642	515	438	384	474	3624
Total	429	940	172	277	633	1207	1047	815	714	768	7002

P12+P14

AGE

GENDER

	0-17	18-64	over65	Total
Male	0	0	13	13
Female	0	3	52	55
Total	0	3	65	68

P38

Figure 3.1 Sample SF1 Control Tables

DUTYPE

TENURE		Single-Family	Multi-Family	Total
	Own	2279	361	2640
	Rent	164	175	339
	Total	2443	536	2979

H32

NUMAUTOS

TENURE		0	1	2	3	4	5+	Total
	Own	88	967	1305	250	30	0	2640
	Rent	24	183	120	12	0	0	339
	Total	112	1150	1425	262	30	0	2979

H44

NUMWORK

HHSTRUCT		0	1	2	3+	Total
	Married couple	324	376	950	128	1778
	Other family	57	113	83	31	284
	Total	381	489	1033	159	2062

P48

INCOME	Less than 30,000	786
	30,000 - 49,999	555
	50,000 - 74,999	672
	75,000 - 124,999	651
	125,000 or more	310
	Total	2974

P52

CITIZEN	Native	6462
	Naturalized	264
	Non-Citizen	276
	Total	7002

P21

WRKHOURS

GENDER		0 hours	1-14 hours	15-34 hours	>= 35 hours	Total
	Male	623	51	325	1654	2653
	Female	1290	119	563	1017	2989
	Total	1913	170	888	2671	5642

P47

Figure 3.2 Sample SF3 Control Tables

DUTYPE

		Single-Family	Multi-Family	Total
TENURE	Own	2277	361	2638
	Rent	165	176	341
	Total	2442	537	2979

H32

NUMAUTOS

		0	1	2	3	4	5+	Total
TENURE	Own	88	966	1304	250	30	0	2638
	Rent	24	184	121	12	0	0	341
	Total	112	1150	1425	262	30	0	2979

H44

NUMWORK

		0	1	2	3+	Total
HHSTRUCT	Married couple	315	366	924	125	1730
	Other family	62	123	91	34	310
	Total	377	489	1015	158	2040

P48

INCOME	Less than 30,000	787
	30,000 - 49,999	556
	50,000 - 74,999	673
	75,000 - 124,999	652
	125,000 or more	311
	Total	2979

P52

CITIZEN	Native	6462
	Naturalized	264
	Non-Citizen	276
	Total	7002

P21

WRKHOURS

		0 hours	1-14 hours	15-34 hours	>= 35 hours	Total
GENDER	Male	619	51	323	1643	2636
	Female	1293	119	565	1020	2997
	Total	1912	170	887	2663	5633

P47

Figure 3.3 Sample Adjusted SF3 Control Tables

3.1.2 Seed Data

The seed data comprises a sample of households characterized by *all* the attributes of interest. These attributes may be at the household- (such as household size and number of vehicles) or person- levels (such as age and gender). The location of these households is typically known only at an aggregate spatial scale (called as the seed area). Typically, the seed data are obtained from the US Census Public Use Microdata Samples (PUMS) and the location is defined in terms of the Public Use Microdata Areas (PUMAs). As discussed in Chapter 2, the data-fusion approach to population synthesis involves drawing households from a PUMA so as to generate a population for a census tract (located in the corresponding PUMA) that is consistent with the tract-level controls identified in the previous section.

The PUMS data are summarized in Tables 3.2 (Household-level attributes) and 3.3 (Person-level attributes). The tables identify the variable names as provided in the PUMS database and provide brief descriptions of the attributes. In addition, the tables also identify the categorical values that each attribute takes in the raw data and the aggregation scheme used to generate the “required categories” (i.e., consistent with those in the control tables). For instance, the BLDGSZ variable (Table 3.2), which captures the type of dwelling unit, can be one of ten different categories in the raw data. However, in the control table only two types of dwelling units are identified (“single family” and “multi-family”). These tables list and describe only those attributes of the seed data that are explicitly controlled at the census-tract level (See Table 3.1). The PUMS also provides several other characteristics of the households and persons.

Table 3.2 Household-level Attributes of Interest from the PUMS Data

Attribute Name	Description	PUMS Variable Name (Field)	PUMS Categories	Required Categories
HHID	Common identifier for each unit and all individuals in the unit	SERIALNO (H2-8)	0000001-9999999	Continuous integer numbering
DUTYPE	Size of the residential unit (This variable is left blank for Group Quarters)	BLDGSZ (H115-116)	01 A mobile home 02 A one-family house detached from any other house 03 A one-family house attached to one or more houses 10 Boat, RV, van, etc.	Single-family dwelling unit
			04 A building with 2 apartments 05 A building with 3 or 4 apartments 06 A building with 5 to 9 apartments 07 A building with 10 to 19 apartments 08 A building with 20 to 49 apartments 09 A building with 50 or more apartments	Multi-family dwelling unit
TENURE	Home ownership (Has a value of 0 for Group Quarters)	TENURE (H113)	1 Owned by you or someone in this household with a mortgage or loan 2 Owned by you or someone in this household free and clear (without a mortgage or loan)	Own
			3 Rented for cash rent 4 Occupied without payment of cash rent	Rent
HHSIZE	Number of persons in household (Has a value of 1 for Group Quarters)	PERSONS (H106-107)	1 to 97	Same as PUMS
HHSTRUCT	Household Structure(Has a value of 0 for Group Quarters)	HHT (H213)	1 Family household: Married-couple	Married couple household
			2 Family household: Male householder, no wife present 3 Family household: Female householder, no husband present	Other family household
			4 Nonfamily household: Male householder, living alone 5 Nonfamily household: Male householder, not living alone 6 Nonfamily household: Female householder, living alone 7 Nonfamily household: Female householder, not living alone	Non-family household
INCOME	This includes the income of the householder and all other individuals 15 years old and over in the household, whether they are related to the householder or not. (Has a value of 0 for Group Quarters)	HINC (H251-258)	-0059999-,…, 99999999+	Non-negative continuous values
NUMAUTO	The number of passenger cars, vans, and pickup or panel trucks of 1-ton capacity or less kept at home and available for the use of household members. (This variable is left blank for Group Quarters)	VEHICL (H134)	0-5 and 6+	Same as PUMS

Table 3.3 Person-level Attributes of Interest from the PUMS Data

Attribute Name	Description	PUMS Variable Name (Field)	PUMS Categories	Required Categories
HHID	Common identifier for each unit and all individuals in the unit	SERIALNO (P2-8)	0000001-9999999	Continuous integer numbering
PERSID	Common identifier for all individuals in the unit	PNUM (P9-10)	1 through 97	Continuous integer numbering
AGE	Age	AGE (P25-26)	<1,1-89,90,90+	Continuous integers
OWNCHILD	Is person an own child in the household	OC (P20)	0 Not an own child under 18 years 1 Yes, own child under 18 years	Same as PUMS
GENDER	Gender	SEX (P23)	Male/Female	Male/Female
ETHNICITY	Ethnicity of Person	RACE1 (P38)	1 White alone	White alone
			2 Black or African American alone	Black alone
			3 American Indian alone 4 Alaska Native alone 5 American Indian and Alaska Native tribes specified, and American Indian or Alaska Native, not specified, and no other races 6 Asian alone 7 Native Hawaiian and Other Pacific Islander 8 Some other race alone	Other Ethnicity alone
			9 Two or more major race groups	Multiple Race
WRKHOURS	Average hours of work per week in 1999 (Is 0 if person is aged 15 or below or person was not working in 1999)	HOURS (P241-242)	0-99	Continuous
CITIZEN	Citizenship status	CITIZEN (P76)	1 Yes, born in the United States 2 Yes, born in Puerto Rico, Guam, U.S. Virgin Islands, American Samoa, or Northern Marianas 3 Yes, born abroad of American parent or parents	Native
			4 Yes, U.S. citizen by naturalization	Naturalized
			5 No, not a citizen of the United States	Non citizen

3.2 Methodology

This section outlines the procedure for synthesizing the base-year population iteratively (for any census tract) based on the inputs described in the previous sections. Broadly, this procedure involves selecting a set of households from the PUMS data in such a way that the tract-level controls are satisfied. One household is selected in each iteration of the procedure.

The first step of the population-synthesis procedure involves “pre-treatment” of the PUMS data. As the PUMS data represent only a 5% sample of the overall population, it is possible that there are certain types of households (especially “rare” households) which are represented in the tract-level control tables but are not present in the PUMS data from the corresponding PUMA. For example, a control table may indicate few “4-persons and 1-car” households to be present in a census tract. However, the 5% PUMS data from the PUMA to which this tract belongs may not have any such households. The pre-treatment procedure simply augments the PUMS data for each PUMA by adding such missing household types from other PUMAs. Our current procedure ensures that each PUMA has at least one household that satisfies each cell (independently) in each of the twelve control tables identified in Table 3.1. One household of the missing type is borrowed (arbitrarily, in the current implementation) from some other PUMA to satisfy this requirement. Overall, the pre-treatment procedure broadly ensures consistency between the seed data and the control tables and, therefore, it would always be possible to find a household in the seed data towards satisfy each cell of the control tables.

The next step involves the initialization of the *count tables* which are used to track the number of households of each type (defined in terms of the control attributes identified in the previous sections) that have already been "selected" until that point. The initialization involves setting the cell values of all the count tables to zero as no household has been selected for the target area. There are as many count tables as there are control tables. Together, the count- and control- tables enable us to assess whether the control targets have been achieved (i.e., if the value in a cell of the count table is less than the corresponding value in the control table, then the target has not been achieved for that cell).

Once the PUMS data have been pre-treated and the count tables initialized, the population of the census tract is synthesized in an iterative fashion. Specifically, one household is added to the population in each iteration and the count tables are appropriately updated. The selection of the household to be added is based on the relative fitness values of all the households in the PUMS data. The fitness of a household *i* in iteration *n* $\left(F^{in}\right)$ is calculated using the following formula:

$$F^{in} = \sum_{j=1}^{J}\left[\frac{1}{e_j^i}\sum_{k=1}^{K_j}\left[\frac{\left(R_{jk}^{n-1}\right)^2}{T_{jk}} - \frac{\left(R_{jk}^{n-1} - HT_{jk}^i\right)^2}{T_{jk}}\right]\right]$$

$$\text{Where}: \; R_{jk}^{n-1} = T_{jk} - CT_{jk}^{n-1}$$

In the above formula, *j* is an index representing the control (and the corresponding count) tables and J is the total number of control (or count) tables. For example, *j* =1 could represent the joint distribution of household size against tenure (see "H15" in Table 3.2); *j* = 2 could represent the joint distribution of household size against household structure (see “P26” in Table 3.2); and so on.

For each control (count) table *j, k* is an index representing the different cells in that table. For example, in table *j* = 1 (“H15” in Table 3.2), *k* has values from 1 though 14 (Note: K_1 = 14 as K_j represents the number of cells in Table *j*) representing the 14 different cells (7 categories for household size multiplied by the 2 categories for tenure). Therefore, for this table, *k* = 1 represents the first cell (1 person / own household), *k* =2 represents the second cell (2 person / own household), and so on.

T_{jk} represents the value of cell *k* in control table *j*. For the census tract presented in Table 3.2, T_{11} =644, T_{12} =1089, and so on. This represents the target number of households of a particular type to be synthesized. In the case of the above example, the numbers indicate that 644 one-person, own-home households and 1089 two-persons, own-home households have to be synthesized.

CT_{jk}^{n-1} represents the value of cell *k* in count table *j* after iteration *(n-1)*. At initialization (*n=1*), all values of the count tables are set to zero. After each draw, the values of the cells in the count tables are updated based on the type of the household drawn. For example, if a one-person, own-home household is drawn in the first iteration, then CT_{11}^{2} will be 1.

Based on the above definitions, $R_{jk}^{n-1} = \left(T_{jk} - CT_{jk}^{n-1}\right)$ is the number of households/persons required to satisfy the target for cell *k* in control table *j* after iteration *(n-1)*. This is calculated as the difference between the corresponding cell values of the control and the count tables. At initialization (*n=1*), $R_{jk}^{n-1} = T_{jk}$ as the values in the count table are all zero.

HT_{jk}^{i} is the contribution of the *i*th household in the PUMS dataset (seed data) to the *k*th cell in control table *j*. For example, if the 1st record in the PUMS dataset is a two-person, own-home household then $HT_{11}^{1} = 0$, $HT_{12}^{1} = 1$, and so on.

e_j^i takes the value of 1 if control table *j* represents a household-level table. For person-level tables, e_j^i is the size of household *i*. This need to differentially scale the household- and person-tables using the above tern is because of the following reason. Addition of a household to the population will always change only one cell of any household-level count table. Further, the change will necessarily be one unit. However, as a household can have multiple persons, the addition of a household to the population can result in changes to multiple cells of the person-level count tables (for example, a household can add both a male and a female to the population thereby resulting in changes to two cells in a person-level count table that provides the distribution of gender). Further, it is also possible that the changes are more than one unit (for example, the household may have two males changing the cell corresponding to male in the count table for gender by two). Thus, the fitness contributions based on the person-level tables are scaled by the household size to capture the per-person contribution. If this were not done, it is possible that the larger households (i.e., with more people) are systematically selected in the early stages of the iteration.

Note that R_{jk}^{n-1} is the required number households to achieve the target in cell *k* of control table *j* in iteration (*n*-*1*) and $\left(R_{jk}^{n-1} - HT_{jk}^{i}\right)$ is the required number households to achieve the target in cell *k* of control table *j* if household *i* is added to the population in iteration *n*. Thus, the fitness of a household is related to the decrease in the required number of households of different types by adding that household into the population of the census tract. With the two fundamental terms, R_{jk}^{n-1} and $\left(R_{jk}^{n-1} - HT_{jk}^{i}\right)$, several functions can be constructed to calculate the overall fitness of the household. The one adopted in this study is presented in the formula above. A comparison of the performance of the algorithm for under different functional forms for the fitness calculations is an area of future study. In addition, it is also useful to note here that the present algorithm assumes that all control tables are equally important. If this is not the case (for example, if matching the household-size distribution is more important than matching the ethnicity distribution), weights can be added to reflect the relative importances of the different tables.

The higher the value of the fitness for a household, the greater is the contribution of this household towards satisfying the control targets. Therefore, it is desirable to add a household with higher fitness into the synthetic population of the census tract. In this research, a greedy-heuristic is employed. That is, the household which as the highest (positive) fitness value is added to the population. Thus, the household that contributes the most is chosen in each iteration and the count tables are suitably updated. The fitness values are re-calculated and the iterations continue.

It is useful to note here that if there are multiple households that have the same fitness value, the first one in the database containing the seed households is chosen. If a large number of control tables are used, it is unlikely that two dissimilar households will end up with exactly the same maximum-fitness value and hence the issue is not of practical importance. However, if the number of control tables is few, it is possible that two very different households have the same (maximum) fitness in any iteration. Further examination of this issue is an area for future studies.

It is also possible that the fitness of a certain household is determined to be negative. This will happen if adding this household into the population of the census tract will result in one or more values of the count tables exceeding the corresponding values of the control tables. Thus, a natural termination criterion for the algorithm is when all households in the seed data have negative fitness values.

If the fitness of a household is determined to be negative in any iteration, then it will remain negative in all subsequent iterations. Thus, once a household is found to have a negative fitness, it can simply be removed from consideration from all future iterations. This approach would provide benefits in terms of reduced run-times of the algorithm.

The above algorithm has been implemented using the matrix programming language, GAUSS. A numerical illustration of the application of the algorithm is presented in Appendix B.

3.3 Application

This section describes the application of the procedure for synthesizing the base year populations for 13 census tracts in Florida. These census tracts and the PUMAs to which they belong are identified in Table 3.4. The reader will note that the there are wide variations in the populations and the areas of the chosen census tracts and PUMAs. Further, these census tracts were chosen to represent some of the major urban regions in Florida were advanced travel-demand models are likely to be developed. Finally, for all these census tracts, the boundaries did not change between 1990 and 2000. This enables us to do the back-casting validations (discussed in the next Chapter).

Table 3.4 Characteristics of the Census Tracts Chosen for Analysis

Case ID	Census Tract ID	PUMA ID	County Name	Census Tract		PUMA	
				Population	Area (sq. miles)	Population*	Area (sq. mile)
1	0026	3502	Palm Beach	414	0.29	7150	50.13
2	0012	701	Leon	1030	0.61	7907	111.49
3	0273.09	2601	Pinellas	1606	8.84	5794	82.47
4	0215.03	2003	Seminole	1630	1.33	5004	60.42
5	0202	300	Okaloosa	1799	99.52	8851	1151.98
6	0101.24	4016	Miami-Dade	2257	3.07	4704	23.38
7	0142.02	1104	Duval	3770	0.74	6177	68.20
8	0016	3502	Palm Beach	3875	0.59	7150	50.13
9	0219.02	2001	Seminole	4513	2.10	7457	99.73
10	0019.06	3502	Palm Beach	7728	1.94	7150	50.13
11	0168.02	1106	Duval	8145	1.99	5124	83.23
12	9801	600	Jefferson	8894	315.31	7649	6412.43
13	0054.02	4011	Miami-Dade	9426	0.55	7689	17.45

* The PUMA population represents a 5% sample

For each census tract, two populations were synthesized. The first population was generated by using only the household-level control tables, is consistent with the state-of-the-practice (however, the number of control tables used is significantly more). The second population includes both person- and household- level control tables. From Table 3.5, the reader will note that the number of households and persons synthesized is very close to the actual number of households and persons in the census tracts. This is true for both the populations synthesized.

Table 3.5Aggregate Comparisons of the True and Synthesized Populations

Case ID	Households			Total Population			Group Quarters Population		
	True Values	Synthesized		True Values	Synthesized		True Values	Synthesized	
		Only HH-level controls	HH- and Person-level controls		Only HH-level controls	HH- and Person-level controls		Only HH-level controls	HH- and Person-level controls
1	129	128	132	414	405	408	160	160	158
2	474	474	475	1030	1030	1030	0	0	0
3	643	643	639	1606	1606	1601	55	55	51
4	593	593	597	1630	1626	1620	130	130	110
5	711	711	710	1799	1793	1795	0	0	0
6	581	580	583	2257	2245	2227	87	87	84
7	1992	1992	1981	3770	3774	3771	30	30	29
8	1606	1606	1600	3875	3894	3872	0	0	0
9	1862	1861	1862	4513	4489	4498	14	14	14
10	4170	4170	4157	7728	7728	7724	342	342	332
11	3529	3529	3514	8145	8141	8143	0	0	0
12	3128	3127	3156	8894	8846	8851	1034	1034	935
13	3720	3720	3714	9426	9405	9427	12	12	12

To compare the synthesized and the true populations, error-percentage measures were computed. For any control table, these values were calculated as follows: For each cell in control table, the absolute value of the difference between the target value (i.e., value of the cell in the control table) and the value of the corresponding cell in the count table at the convergence of the algorithm is calculated. These are then added across all the cells in the control table and expressed as a percentage of either the number of households in the census tract or the population depending on whether the control table is at the household- or the person-level. The percentage error represents the extent of misclassification in each table. The calculation is numerically illustrated in Figure 3.4.

Control table

T_{1k}	HHSize = 1	HHSize = 2
Own	1	5
Rent	2	2

Count table at convergence

CT_{1k}	HHSize = 1	HHSize = 2
Own	1	6
Rent	1	2

Absolute value of the difference

Abs. Diff	HHSize = 1	HHSize = 2
Own	0	1
Rent	1	0

2 of the 10 households are misclassified

% Error = (2/10 * 100) = 20%

Figure 3.4 Numerical Illustration of the Calculation of the Percentage-Error Measure

Tables 3.6 and 3.7 present the error percentages for each of the twelve control tables defined in Table 3.1. Table 3.6 presents the results for the population synthesized using only household controls whereas Table 3.7 presents the results for the population synthesized using both household- and person-controls. In Table 3.6, the reader will note that the values for the person-level tables are large. In Table 3.7, these values are significantly reduced as the person-level characteristics are explicitly controlled for. Between Tables 3.6 and 3.7, the former has slightly lower values for the error-percentage for the household-level tables. This is expected as the total number of controls is greater in the second population represented in Table 3.7 and as the number of controls increase, the ability to replicate each table decreases. However, the increase in the error percentages for the household tables appears smaller compared to the decrease in error percentages for the person tables.

Table 3.6 Error Percentages at Convergence: Population Synthesized with only Household-level Controls

Case ID	HH-level Control Tables								Person-level Control Tables			
	H15	P26	P34	H32	H44	P48	P52	P38*	P7	P12+P14	P21	P47
	HHSIZE	HHSIZE	CHAGE	DUTYPE	NUMAUTOS	NUMWORK	INCOME	AGE	ETHNICITY	AGE	CITIZEN	WRKHOURS
	TENURE	HHSTRUCT	HHSTRUCT	TENURE	TENURE	HHSTRUCT		GENDER		GENDER		GENDER
1	0.78%	0.78%	0.00%	0.83%	0.78%	8.32%	0.89%	0.00%	27.29%	81.88%	8.11%	60.10%
2	0.00%	0.00%	0.00%	0.17%	0.40%	1.00%	0.22%	NA	69.03%	33.11%	10.08%	43.84%
3	0.00%	0.00%	0.37%	0.18%	0.14%	0.32%	0.29%	0.00%	0.62%	20.80%	9.21%	11.51%
4	0.00%	0.00%	0.00%	0.14%	0.45%	0.34%	0.19%	0.00%	5.03%	14.72%	7.31%	11.86%
5	0.00%	0.00%	0.00%	0.01%	0.32%	0.17%	0.13%	NA	10.89%	15.01%	8.56%	12.83%
6	0.17%	0.52%	0.20%	0.17%	0.59%	0.55%	0.42%	0.00%	15.77%	18.17%	7.98%	13.19%
7	0.00%	0.00%	0.00%	0.05%	0.11%	0.14%	0.07%	0.00%	9.55%	19.73%	10.24%	17.16%
8	0.00%	0.00%	0.00%	0.11%	0.11%	0.24%	0.07%	NA	46.48%	16.23%	8.39%	13.97%
9	0.16%	0.27%	0.09%	0.09%	0.16%	0.29%	0.09%	0.00%	30.36%	14.00%	17.59%	16.61%
10	0.00%	0.00%	0.00%	0.02%	0.04%	0.18%	0.02%	0.00%	24.79%	32.48%	7.25%	20.77%
11	0.00%	0.00%	0.00%	0.03%	0.09%	0.12%	0.06%	NA	16.16%	19.64%	6.37%	12.29%
12	0.03%	0.10%	0.05%	0.05%	0.08%	0.15%	0.06%	0.00%	40.48%	21.36%	1.33%	21.23%
13	0.00%	0.00%	0.00%	0.02%	0.06%	0.09%	0.04%	0.00%	11.36%	21.55%	19.40%	7.63%

* Group Quarters (each person in a group-quarters residential unit is treated as a single-person "household")

NA = No Group Quarters Population in this Census Tract

Table 3.7 Error Percentages at Convergence: Population Synthesized with Household- and Person-level Controls

Case ID	HH-level Control Tables								Person-level Control Tables			
	H15	P26	P34	H32	H44	P48	P52	P38*	P7	P12+P14	P21	P47
	HHSIZE	HHSIZE	CHAGE	DUTYPE	NUMAUTOS	NUMWORK	INCOME	AGE	ETHNICITY	AGE	CITIZEN	WRKHOURS
	TENURE	HHSTRUCT	HHSTRUCT	TENURE	TENURE	HHSTRUCT		GENDER		GENDER		GENDER
1	3.88%	3.88%	5.13%	2.33%	2.80%	8.32%	2.33%	11.25%	2.42%	9.66%	2.63%	1.85%
2	0.21%	0.63%	0.99%	0.21%	0.76%	1.80%	0.43%	NA	0.78%	4.47%	3.83%	2.72%
3	1.56%	1.56%	0.74%	0.80%	0.65%	0.79%	0.78%	7.27%	0.31%	4.30%	1.65%	2.10%
4	1.01%	1.01%	0.73%	0.76%	1.09%	1.84%	0.72%	15.38%	0.61%	4.66%	0.43%	3.17%
5	0.70%	0.70%	0.75%	0.15%	0.37%	0.75%	0.14%	NA	0.44%	2.78%	0.22%	1.60%
6	0.34%	1.03%	1.57%	0.36%	1.14%	1.05%	0.84%	3.45%	1.51%	4.34%	0.67%	1.65%
7	1.15%	1.15%	1.44%	0.55%	0.55%	1.56%	0.55%	3.33%	0.08%	3.74%	0.03%	2.07%
8	1.25%	1.25%	1.27%	0.37%	0.40%	1.24%	0.37%	NA	0.08%	3.38%	0.08%	2.15%
9	0.32%	0.64%	0.35%	0.08%	0.16%	0.73%	0.12%	0.00%	0.60%	1.66%	0.33%	0.82%
10	0.94%	0.94%	1.81%	0.32%	0.31%	1.21%	0.31%	2.92%	0.13%	2.72%	0.08%	1.49%
11	0.88%	0.88%	0.64%	0.43%	0.44%	0.61%	0.43%	NA	0.05%	2.55%	0.78%	1.55%
12	1.02%	1.41%	1.51%	0.90%	0.90%	1.47%	1.08%	10.93%	0.48%	5.77%	0.48%	3.94%
13	0.54%	0.54%	0.76%	0.16%	0.18%	0.48%	0.16%	0.00%	0.01%	2.09%	0.22%	1.17%

* Group Quarters (each person in a group-quarters residential unit is treated as a single-person "household")

NA = No Group Quarters Population in this Census Tract

The previous tables presented the error-percentage values separately for each of the control tables. Table 3.8 presents, for each census tract, a single error-percentage value across all the household-level tables and another single value across all person-level tables. These single-values were calculated as weighted averages of the error percentages of the household-level and person-level tables respectively. The inverse of the number of cells in the table was used to weight the error-percentages for each of the tables. This is because the error is likely to be smaller if the number of cells in the table is fewer (or, in other words it is easier to replicate/match a table with just four cells more accurately as opposed to matching a table with twenty cells). These aggregate, household- and person-level error-percentage values are presented for each of the two populations. This table clearly indicates the significant decrease in the overall error of replicating person-level characteristics and the marginal increase in the overall error of replicating household-level characteristics when both household- and person-level characteristics are controlled for. Thus, the results highlight the need to control for both household- and person-characteristics in population synthesis.

Table 3.8 Aggregate, Household- and Person-level Error Percentages

Case ID	Synthesized with only HH-level Control Tables		Synthesized with HH- and Person-level Control Tables	
	All HH-level Control Tables	All Person-level Control Tables	All HH-level Control Tables	All Person-level Control Tables
1	1.73%	27.87%	3.79%	2.89%
2	0.26%	36.60%	0.66%	2.68%
3	0.22%	7.52%	0.89%	1.46%
4	0.17%	7.80%	0.96%	1.22%
5	0.08%	10.46%	0.42%	0.69%
6	0.34%	12.08%	0.84%	1.35%
7	0.06%	11.78%	0.90%	0.63%
8	0.09%	22.38%	0.75%	0.64%
9	0.14%	21.40%	0.28%	0.59%
10	0.04%	16.92%	0.73%	0.50%
11	0.05%	11.45%	0.55%	0.78%
12	0.07%	18.84%	1.14%	1.40%
13	0.03%	14.95%	0.34%	0.43%

Finally, we compared the predicted distributions of car ownership (number of cars per person) across the two synthetic populations. Note that there is no control table that explicitly provides the joint distribution of car ownership against household size. Thus, this analysis seeks to compare the populations on a joint distribution that is not explicitly controlled for. Table 3.9 presents these distributions for each census tract and for each of the two synthetic populations. The car ownership is categorized into five levels: no cars, less than 0.5 car per person, between 0.5 and 1 car per person, between 1 and 2 cars per person, and more than 2 cars per person. Table 3.9 presents the percentage of the synthesized households in each census tract falling under each category. For example, in the census tract 1, 5.47% of the households synthesized with only household-level controls were found to have 0 cars. In the case of households synthesized with both household- and person-level controls, 6.07% were found to have zero cars. Note that the percentages sum to 100 across the five car-ownership categories for each census tract. On comparing the distributions obtained form the two populations, we find that the population synthesized with only household-level controls, in general, has a lower percentage of households in the 0.51-1 cars-per-person category relative to the population synthesized with household- and person-level controls. For all the other car-ownership categories, the percentages form the former population are greater compared to the ones from the latter population. As the true relationship between the car-ownership and household size is not known for the census tracts, we compared the above distributions to the ones at the PUMA level. Table 3.10 presents the distribution of car-ownership per person for all the PUMAs in which the census tracts chosen for analysis lie. On examining the synthesized tract-level distributions from Table 3.9 with the observed PUMA-level distributions from table 3.10, we find that the population synthesized with both household- and person-level distributions match the PUMA-level distributions better. For example, one can see that the population synthesized with only household-level controls predicts a higher percentage of households with more than 1 car per person compared to the distributions observed at the PUMA-level. Thus, these results also reinforce the value of incorporating both household- and person-level control tables in the population synthesis.

Table 3.9 Predicted Distribution of Car Ownership in the Census Tracts

Case ID	Synthesized with Only HH-level Controls					Synthesized with HH- and Person-level Controls				
	0	**0.1 to 0.5**	**0.51 to 1**	**1.01 to 2**	**More than 2**	**0**	**0.1 to 0.5**	**0.51 to 1**	**1.01 to 2**	**More than 2**
1	5.47%	39.84%	47.66%	7.03%	0.00%	6.06%	34.85%	56.06%	3.03%	0.00%
2	24.05%	36.71%	28.06%	10.13%	1.05%	24.21%	33.26%	37.47%	3.79%	1.26%
3	0.00%	38.41%	45.72%	14.15%	1.71%	0.00%	35.05%	52.27%	12.52%	0.16%
4	9.27%	30.86%	42.83%	12.98%	4.05%	9.21%	23.79%	55.44%	10.55%	1.01%
5	8.02%	38.12%	33.47%	16.03%	4.36%	8.03%	32.54%	45.49%	11.55%	2.39%
6	7.07%	40.69%	40.86%	9.66%	1.72%	6.86%	31.73%	57.98%	2.40%	1.03%
7	16.52%	29.82%	40.06%	10.39%	3.21%	16.56%	26.96%	48.91%	5.40%	2.17%
8	12.89%	42.22%	32.00%	9.90%	2.99%	12.81%	38.31%	40.69%	6.56%	1.63%
9	10.64%	40.35%	35.20%	10.69%	3.12%	10.63%	35.71%	44.95%	7.95%	0.75%
10	6.00%	29.23%	45.32%	18.47%	0.98%	5.99%	22.37%	59.01%	12.48%	0.14%
11	2.24%	39.78%	39.44%	16.04%	2.49%	2.22%	33.92%	52.08%	10.50%	1.28%
12	8.83%	37.61%	34.89%	15.06%	3.61%	8.90%	33.05%	45.47%	10.99%	1.58%
13	34.01%	36.61%	21.56%	6.61%	1.21%	34.01%	38.26%	24.93%	2.15%	0.65%

Table 3.10 Observed Distribution of Car Ownership in the PUMAs

PUMA ID	**0**	**0.1 to 0.5**	**0.51 to 1**	**1.01 to 2**	**More than 2**
3502	14.26%	30.72%	50.05%	4.35%	0.62%
701	7.92%	24.07%	61.04%	6.29%	0.68%
2601	5.65%	32.09%	57.26%	4.76%	0.24%
2003	4.47%	30.60%	59.13%	5.34%	0.46%
300	4.00%	27.55%	59.33%	8.35%	0.77%
4016	5.36%	42.24%	50.03%	2.17%	0.20%
1104	3.56%	33.18%	57.82%	4.77%	0.67%
2001	3.44%	27.32%	61.68%	6.74%	0.82%
1106	6.47%	28.25%	59.01%	6.04%	0.24%
600	10.12%	30.11%	49.80%	9.13%	0.85%
4011	19.18%	37.70%	39.74%	2.97%	0.41%

CHAPTER 4 SYNTHESIZING TARGET-YEAR POPULATION

This chapter describes the synthesis of the 1990 populations for the same 13 census tracts for which the year 2000 populations were previously synthesized (described in Chapter 3). This "back-casting" exercise is intended to correspond to the forecasting that will be required for transportation planning. Thus, the populations synthesized for the year 2000 for the thirteen census tracts were used as the seed data. Further, the number of control tables used are limited (Note that these are consistent with the description of the target-year population synthesis procedure in Chapter 2). Once the populations were synthesized, they were compared with several other joint-distribution tables (those that were not used as control tables in the synthesis procedure) obtained from the 1990 SF1 and SF3 files to assess the accuracy of the synthesized 1990 populations.

The rest of this chapter is organized as follows. Section 4.1 presents the data and methodology for synthesizing the 1990 populations for the thirteen census tracts. Section 4.2 presents an analysis of the accuracy of the back-casts.

4.1 Data and Methodology

Table 4.1 presents a comparative assessment of the number of households and people in each of the census tracts in the years 1990 and 2000. The table also presents the change in the number of households/persons between the two years as a percentage of the corresponding year 2000 values. Overall, one can see that there is considerable variation in the percentage change in the number of households and population across the thirteen census tracts chosen for analysis. Thus, the back-casting validation presented in this chapter reflects a wide range of growth (or decline in this case as the target year is before the base year) scenarios from the base-year to the target-year. The areas of these census tracts remained the same across the two years.

Table 4.1 Characteristics of the Census Tracts in 1990 and 2000

Case ID	Census Tract ID	PUMA ID	County Name	Households			Population			Group Quarters Population		
				2000	1990	% Change	2000	1990	% Change	2000	1990	% Change
1	0026	3502	Palm Beach	129	208	61.24	414	550	32.85	160	138	-13.75
2	0012	701	Leon	474	491	3.59	1030	1094	6.21	0	0	NA
3	0273.09	2601	Pinellas	643	240	-62.67	1606	617	-61.58	55	11	-80.00
4	0215.03	2003	Seminole	593	556	-6.24	1630	1561	-4.23	130	112	-13.85
5	0202	300	Okaloosa	711	612	-13.92	1799	1592	-11.51	0	0	NA
6	0101.24	4016	Miami-Dade	581	429	-26.16	2257	1290	-42.84	87	0	-100.00
7	0142.02	1104	Duval	1992	1797	-9.79	3770	3683	-2.31	30	0	-100.00
8	0016	3502	Palm Beach	1606	1515	-5.67	3875	3423	-11.66	0	34	NA
9	0219.02	2001	Seminole	1862	1857	-0.27	4513	4469	-0.97	14	25	78.57
10	0019.06	3502	Palm Beach	4170	2274	-45.47	7728	4260	-44.88	342	0	-100.00
11	0168.02	1106	Duval	3529	2203	-37.57	8145	5409	-33.59	0	0	NA
12	9801	600	Jefferson	3128	2747	-12.18	8894	7634	-14.17	1034	205	-80.17
13	0054.02	4011	Miami-Dade	3720	3572	-3.98	9426	8855	-6.06	12	0	-100.00

The synthesis of the population for a *target* year uses the synthesized population from the *base* year as the seed data (See the detailed discussion in Chapter 2). For the back-casting exercise presented here, the base year is 2000 and the target year is 1990. As described in Section 2.3, two populations were synthesized for each census tract for the base year 2000 – the first was generated with only household-level control tables whereas the second used both household- and person-level control tables. Correspondingly, two populations (Pop-1 and Pop-2) were also synthesized for the target year (1990). The base-year population developed using only household-level controls were used as the seed data in generating Pop-1. The base-year population developed using both household- and person-level controls was used as the seed data in generating Pop-2. The synthesized base-year populations have income values measured in year-2000 dollars. Prior to use as seed data for the 1990 synthesis, the income values were converted to 1990 dollars using the Consumer Price Index (CPI) value of 0.759 (i.e., 100 dollars in 2000 = 75.9 dollars in 1990).

Five one-dimensional control tables (at the census-tract level) were used in the synthesis of the 1990 population. Two of these are at the household level (household size and dwelling-unit type) and two are at the person level (age and gender). The fifth control is a single value for the total population in group quarters. The reader will note that these socio-demographic-land-use attributes are the ones that are most likely to be available for a target year from sources such as the Woods and Poole projections. It is useful to acknowledge that the projections are likely to be available only at the county level. However, the population synthesis procedure discussed here requires the controls at the census-tract level (i.e., the spatial unit at which the population is being synthesized). The tract-level projects can be obtained by disaggregating county-level projections. The errors because of such a procedure are, however, not considered in this back-casting validation analysis. Rather, the intent is to assess the impact of the seed-data (synthesized with and without person-level controls in the base year) on the accuracy of the synthesized target year populations and when few target-year controls are used.

Several other joint distribution tables were also obtained from the 1990 SF1 and SF3 files (Table 4.2). While these are not used as control tables in the synthesis of the 1990 population, the accuracy of the synthesized population can be assessed against these true distributions. The SF3 control tables were adjusted using the same procedure as described in Chapter 3 to ensure consistency with the SF1 control tables.

Table 4.2 SF1 and SF3 Tables from 1990 Census

		Dimension 1		Dimension 2		SF Table Used
S. No	Universe	Attribute	Categories	Attribute	Categories	
1	Households	TENURE	Own, Rent	HHSIZE	1,2,3,4,5,6,7+	H18(SF1)
2	Households	HHSTRUCT	Family, Non-Family	HHSIZE	1,2,3,4,5,6,7+	P27(SF1)
3	Households	TENURE	Own, Rent	DUTYPE	Single Family, Multi-Family	H43(SF1)
4	Households	TENURE	Own, Rent	NUMAUTO	0,1,2,3,4,5+	H37(SF3) adjusted by H18(SF1)
5	Households	INCOME	< 30K, 30-50K, 50-75K, 75-125K, more than 125K	*NA*		P80(SF3) adjusted by H18(SF1)
6	Persons	GENDER	Male, Female	AGE	0-5, 6-15, 16-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, over 75	P12(SF1)
7	Persons	CITIZEN	Native, Naturalized, Non Citizen	*NA*		P37(SF3) adjusted by P12(SF1)
8	Persons 16 years and over	GENDER	Male, Female	WRKHOURS	0,1-14, 15-35, more than 35	P76(SF3) adjusted by P12(SF1)

4.2 Accuracy of the Back-Cast

As described in the previous section, two populations were synthesized for 1990. The base-year population developed using only household-level controls were used as the seed data in generating Pop-1. The base-year population developed using both household- and person-level controls was used as the seed data in generating Pop-2. In each case, five controls were used. Table 4.3 presents the error percentages at convergence (defined in Chapter 3) for each of these five controlled attributes. All the values are close to zero indicating that the synthesized populations (both Pop-1 and Pop-2) replicate the distribution of the controlled-attributes almost perfectly.

Table 4.3 Error Percentages at Convergence for the Controlled Tables

Case ID		Pop-1: Synthesized With Seed Data 1						Pop-2: Synthesized With Seed Data 2				
		GENDER	AGE	HHSIZE	DUTYPE	GQ-POP		GENDER	AGE	HHSIZE	DUTYPE	GQ-POP
1		3.45%	5.64%	2.40%	2.40%	7.97%		0.18%	0.91%	0.00%	0.00%	0.00%
2		0.91%	0.91%	0.81%	0.81%	NA		0.91%	0.91%	0.81%	0.81%	NA
3		0.32%	0.32%	0.00%	0.00%	0.00%		0.00%	0.32%	0.00%	0.00%	0.00%
4		0.13%	0.13%	0.00%	0.00%	0.00%		0.13%	0.13%	0.00%	0.00%	0.00%
5		0.57%	0.57%	0.49%	0.49%	NA		0.57%	0.57%	0.49%	0.49%	NA
6		0.62%	0.62%	0.47%	0.47%	NA		0.62%	0.62%	0.47%	0.47%	NA
7		0.08%	0.08%	0.11%	0.11%	NA		0.08%	0.08%	0.11%	0.11%	NA
8		0.03%	0.03%	0.13%	0.13%	0.00%		0.12%	0.12%	0.13%	0.13%	0.00%
9		0.22%	0.22%	0.27%	0.27%	0.00%		0.20%	0.20%	0.16%	0.16%	0.00%
10		0.00%	0.00%	0.00%	0.00%	NA		0.00%	0.00%	0.00%	0.00%	NA
11		0.04%	0.04%	0.00%	0.00%	NA		0.07%	0.07%	0.05%	0.05%	NA
12		0.30%	0.30%	0.29%	0.29%	0.49%		0.29%	0.29%	0.33%	0.33%	0.49%
13		0.34%	0.34%	0.34%	0.34%	NA		0.32%	0.32%	0.34%	0.34%	NA

Seed Data 1 = Base year population synthesized using only HH-level controls

Seed Data 2 = Base year population synthesized using HH- and Person-level controls

Table 4.4 presents the summary statistics on the error percentages at convergence for the eight uncontrolled distributions (identified in Table 4.2). Each column in this table corresponds to one of the joint distribution tables. For instance, the first column compares how the distribution of household size against tenure obtained from the synthesized population compares with the true distribution obtained from the SF1 tables. Table 4.4 also has two major blocks of rows, one for each of Pop-1 and Pop-2. Within each block, the mean, median, standard deviation, minimum, and maximum of the error percentages (across the thirteen census tracts) are presented for each control table (The error percentages for each of the census tracts and for each of the eight joint-distribution tables are presented in Tables 4.5 and 4.6). Overall, one finds that the average of the percentage errors ranges from about 9% to 42% (across all eight control tables and the two synthetic populations). The average of the errors is particularly high for the joint distribution of auto ownership against tenure. At the same time, the average errors for Pop-2 are generally lower (albeit marginally) compared to the average errors for Pop-1 for most of the control tables.

Table 4.4 Summary of Error Percentages at Convergence

		HH-level Tables					Person-level Tables		
		H15	**P26**	**H32**	**H44**	**P52**	**P12+P14**	**P21**	**P47**
		HHSIZE	HHSIZE	DUTYPE	NUMAUTOS	INCOME	AGE	CITIZEN	WRKHOURS
		TENURE	HHSTRUCT	TENURE	TENURE		GENDER		GENDER
Pop-1: Synthesized with Seed Data 1	Mean	20.74%	10.86%	14.69%	39.18%	26.77%	22.28%	19.80%	19.35%
	Median	19.54%	8.63%	13.62%	34.41%	22.00%	19.54%	18.40%	15.47%
	Std. Dev.	7.75%	9.24%	9.31%	19.26%	11.75%	11.98%	14.02%	13.81%
	Min.	6.67%	1.47%	3.24%	17.93%	14.57%	13.64%	0.38%	8.58%
	Max	40.87%	35.03%	40.87%	90.25%	50.45%	59.09%	54.21%	61.14%
Pop-2: Synthesized with Seed Data 2	Mean	19.79%	8.79%	11.46%	41.99%	19.38%	18.55%	15.23%	21.05%
	Median	20.10%	5.83%	12.37%	35.85%	16.45%	17.27%	10.10%	18.09%
	Std. Dev	8.62%	5.65%	9.01%	18.03%	12.91%	6.70%	13.55%	14.29%
	Min.	4.17%	2.96%	1.14%	22.81%	2.56%	7.23%	3.36%	8.05%
	Max	33.65%	21.99%	29.04%	96.50%	44.66%	32.74%	48.18%	63.30%

Seed Data 1 = Base year population synthesized using only HH-level controls

Seed Data 2 = Base year population synthesized using HH- and Person-level controls

Table 4.5 Error Percentages at Convergence by Census Tract for Pop-1

Case ID	HH-level Tables					Person-level Tables		
	H15	**P26**	**H32**	**H44**	**P52**	**P12+P14**	**P21**	**P47**
	HHSIZE	HHSIZE	DUTYPE	NUMAUTOS	INCOME	AGE	CITIZEN	WRKHOURS
	TENURE	HHSTRUCT	TENURE	TENURE		GENDER		GENDER
1	25.48%	16.83%	9.13%	90.25%	50.45%	59.09%	26.00%	61.14%
2	15.48%	35.03%	11.41%	25.04%	22.00%	20.66%	18.65%	10.15%
3	6.67%	2.50%	6.67%	51.72%	38.70%	28.85%	7.33%	15.47%
4	20.50%	8.63%	3.24%	34.41%	17.15%	16.78%	14.71%	14.98%
5	25.33%	1.47%	17.16%	28.74%	21.49%	25.06%	8.38%	27.36%
6	19.11%	13.52%	19.11%	48.93%	48.79%	22.33%	27.47%	24.53%
7	22.70%	13.13%	15.14%	17.93%	20.07%	16.75%	11.18%	18.00%
8	19.54%	10.17%	17.82%	41.39%	14.57%	13.64%	18.40%	17.98%
9	15.67%	2.21%	9.75%	27.15%	25.58%	13.87%	19.45%	8.58%
10	17.68%	19.53%	18.73%	19.69%	22.56%	22.54%	36.64%	19.77%
11	19.06%	3.63%	13.62%	44.97%	28.09%	15.05%	14.58%	9.88%
12	21.55%	6.63%	8.37%	28.47%	20.03%	15.44%	0.38%	14.31%
13	40.87%	7.89%	40.87%	50.62%	18.57%	19.54%	54.21%	9.45%

Table 4.6 Error Percentages at Convergence by Census Tract for Pop-2

Case ID	HH-level Tables					Person-level Tables		
	H15	**P26**	**H32**	**H44**	**P52**	**P12+P14**	**P21**	**P47**
	HHSIZE	HHSIZE	DUTYPE	NUMAUTOS	INCOME	AGE	CITIZEN	WRKHOURS
	TENURE	HHSTRUCT	TENURE	TENURE		GENDER		GENDER
1	33.65%	4.81%	18.27%	96.50%	16.45%	17.27%	48.18%	63.30%
2	24.44%	16.70%	3.26%	43.56%	2.56%	16.64%	7.50%	20.37%
3	4.17%	5.83%	4.17%	50.61%	42.79%	32.74%	4.26%	32.43%
4	28.42%	5.40%	3.96%	35.85%	10.37%	18.19%	10.10%	16.25%
5	20.10%	5.07%	1.14%	43.91%	21.16%	19.79%	3.36%	22.94%
6	16.32%	4.66%	16.32%	34.65%	29.51%	18.29%	21.74%	19.70%
7	24.04%	13.13%	16.58%	34.34%	12.42%	12.90%	3.86%	18.09%
8	31.68%	9.77%	29.04%	46.15%	6.97%	14.55%	23.74%	23.28%
9	10.29%	2.96%	1.35%	22.81%	18.00%	7.23%	6.67%	8.05%
10	14.51%	21.99%	16.97%	34.20%	22.89%	30.42%	16.55%	13.37%
11	21.56%	11.85%	21.38%	41.87%	44.66%	16.31%	13.80%	15.09%
12	14.96%	5.42%	4.19%	30.76%	9.77%	20.36%	4.61%	11.17%
13	13.16%	6.66%	12.37%	30.68%	14.43%	16.51%	33.59%	9.60%

Tables 4.4, 4.5, and 4.6 examined the error percentages for each of the joint-distribution tables. In contrast, Table 4.7 presents two aggregate error measures (one for household-level distributions and one for person-level distributions) for each census tract and for each population. As discussed in Chapter 3, these aggregate errors represent weighted averages across all the household- and person-level distributions. An examination of the differences in the error percentages across the two populations indicate that, in six out of the thirteen census tracts, Pop-2 had lower aggregate errors for both the household- and person-level joint distributions. In four other census tracts, the errors for Pop-2 were lower for at least one of the household- and person-level distributions. There were only three census tracts in which Pop-1 had lower errors for both the household- and person-level distributions. Further, the magnitudes of the positive error-percentage difference (i.e., the error percentage of Pop 2 subtracted from the error percentage of Pop 1) are generally higher than the magnitudes of the negative error percentages. Thus, these analysis indicate that the base-year population synthesized with both household- and person-level controls can provide a more accurate estimate of the target-year population compared to the base-year population synthesized with only household-level controls.

Table 4.7 Household- and Person-level Error Percentages by Census Tract

Case ID	All HH-level Tables			All Person-level Tables		
	Pop 1: Synthesized with Seed Data 1	Pop-2: Synthesized with Seed Data 2	Difference in the Percentages	Pop 1: Synthesized with Seed Data 1	Pop-2: Synthesized with Seed Data 2	Difference in the Percentages
1	33.89%	27.57%	**6.32%**	37.90%	48.86%	-10.96%
2	19.14%	11.68%	**7.47%**	16.76%	11.56%	**5.20%**
3	21.25%	21.49%	-0.24%	11.45%	13.99%	-2.54%
4	13.59%	12.52%	**1.07%**	14.98%	12.41%	**2.57%**
5	19.07%	14.75%	**4.32%**	14.69%	9.79%	**4.90%**
6	30.97%	21.25%	**9.73%**	26.24%	20.90%	**5.35%**
7	17.53%	17.96%	-0.43%	13.40%	8.25%	**5.15%**
8	19.14%	22.87%	-3.73%	17.83%	22.72%	-4.90%
9	16.41%	10.03%	**6.37%**	16.23%	7.06%	**9.16%**
10	19.96%	21.12%	-1.16%	31.11%	17.13%	**13.97%**
11	21.28%	29.80%	-8.52%	13.47%	14.36%	-0.90%
12	15.50%	10.38%	**5.12%**	5.29%	7.77%	-2.48%
13	31.99%	14.72%	**17.28%**	39.79%	26.01%	**13.78%**

Seed Data 1 = Base year population synthesized using only HH-level controls
Seed Data 2 = Base year population synthesized using HH- and Person-level controls

CHAPTER 5 ASSESSING THE ACCURACY OF DISAGGREGATE TRIP-GENERATION MODELS

Disaggregate models capture travel behavior of the fundamental decision-making units and include several explanatory variables (including socio-economic and mobility characteristics). Consequently, one may expect such models to provide more accurate predictions of the travel characteristics than aggregate models which include fewer explanatory variables. At the same time, for use in forecasting, disaggregate models require more inputs compared to aggregate models (as the number for explanatory variables are more in the former models). To generate such detailed inputs, population-synthesis procedures such as the one presented in this report are used. Thus, the accuracy of the socio-economic-mobility characteristics of the synthesized population (i.e., the inputs to the disaggregate models) is of particular interest. Specifically, if the synthesized population is an inaccurate representation of the true population, gains because of a disaggregate model could be offset by the errors in the synthesized population. In the light of the above discussion, the intent of this chapter is to compare the predictions from aggregate trip-generation models with those from disaggregate trip-generation models. Further, for each of the two types of models, the predictions when the model is applied to the true population are compared to the predictions when the model is applied to a synthesized population. The scope of this comparative analysis is limited to linear-regression-based trip-generation models. The extension of such an analysis for the evaluation of non-linear models such as the multinomial-logit for mode choice is identified as an important future avenue for study.

The rest of this chapter is organized as follows. Section 5.1 describes the aggregate- and disaggregate- trip-generation models estimated for this analysis. Section 5.2 presents and compares the synthesized populations with the true population. Section 5.3 examines the accuracy of model predictions when the aggregate and disaggregate models were applied to the true and synthesized populations.

5.1 Trip-Generation Models

Data from the 2001 National Household Travel Survey (NHTS) were used to estimate the aggregate and disaggregate trip-generation models used in this analysis. The week-day travel records from the entire national sample were used in the analysis. Further, pre-processing of the data was performed to eliminate those households for which we did not have the travel records for all persons in the household. Additional cleaning was also performed to remove cases with missing values for the explanatory variables of interest.

The final estimation sample comprised 34,257 persons from 14,170 households. Using the above dataset, linear-regression models were estimated for three trip purposes: Home-based Work (HBW), Home-Based Non-work (HBNW), and Non-Home-Based (NHB). Frequency distributions of the household- and person-level trip rates for each of the three trip purposes are presented in Tables 5.1 and 5.2 respectively. These tables are interpreted as follows: 6,030 out of 14,170 or 42.55% of all households in the estimation sample did not have any home-based work (HBW) trips (see first row in Table 5.1). Similarly, from the first row of Table 5.2, we see that 23,035 out of 34,257 or 67.24% of all persons in the same did not make any home-based work trips.

Table 5.1 Frequency Distribution of Household-level Trip Rates

# Trips / HH	**HBW**		**HBNW**		**NHB**	
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**
0	6030	42.55	1860	13.13	4105	28.97
1	1899	13.40	853	6.02	1835	12.95
2	3475	24.52	2281	16.10	1979	13.97
3	1057	7.46	763	5.38	1367	9.65
4	1175	8.29	2141	15.11	1305	9.21
5	217	1.53	542	3.82	799	5.64
6	201	1.42	1360	9.60	721	5.09
7	50	0.35	459	3.24	455	3.21
8	45	0.32	991	6.99	385	2.72
9	13	0.09	306	2.16	283	2.00
10	4	0.03	631	4.45	239	1.69
11+	4	0.03	1983	13.99	697	4.92
Total	**14170**	**100**	**14170**	**100**	**14170**	**100**

Table 5.2 Frequency Distribution of Person-level Trip Rates

# Trips / Person	HBW		HBNW		NHB	
	Freq.	**%**	**Freq.**	**%**	**Freq.**	**%**
0	23035	67.24	8057	23.52	16407	47.89
1	3959	11.56	3349	9.78	6348	18.53
2	6443	18.81	11677	34.09	4717	13.77
3	324	0.95	1626	4.75	2854	8.33
4	426	1.24	6271	18.31	1753	5.12
5+	70	0.20	3277	9.57	2178	6.36
Total	**34257**	**100.00**	**34257**	**100.00**	**34257**	**100.00**

For each trip purpose, a household-level- (or aggregate) and a person-level- (or disaggregate) model was estimated.

The household-level models (or the H-models) relate the total number of trips made by a household to the socio-economic characteristics of the household. Household size, vehicle ownership, and dwelling-unit type were taken as the explanatory variables as these are the most common variables used in FSUTMS. A segmented, non-linear, empirical structure is adopted (As a consequence, this regression model is essentially the same as the "cross-classification" trip-generation tables typically used in practice). The models are presented in Tables 5.3 (models for households in single-family dwelling units) and 5.4 (models for households in multi-family dwelling units). The models for HBNW and NHB trip purposes were estimated using the entire sample. However, the model for HBW trips was estimated using a subsample of households with at least one worker (as the number of work trips is necessarily zero for households without workers).

The person-level models (or the P-models) relate the number of trips made by a person to their socio-economic characteristics. These models include several explanatory variables at both the person-level (such as age, gender, and employment status) and the household level (such auto ownership, dwelling unit type, and tenure). For each trip purpose, separate models are estimated for adults (age >16 years) and children (age <= 15 years). These models are presented in Tables 5.5 (models for adults) and 5.6 (models for children). The models for HBNW and NHB trip purposes were estimated using the entire sample. The model for HBW trips for adults was estimated using a subsample of employed adults in the sample (as the number of work trips is necessarily zero for persons who are not employed). There is no model for the HBW trips of children as they are defined to be not employed.

As a general note on the models estimated, the linear-regression was chosen considering its simplicity and popularity in practical use. We recognize that this is not the best econometric structure to model integer variables within a finite range (i.e., the number of household/person trips in our context). An examination of the application of methods such as the Poisson-regression and ordered-probit is identified as an avenue for future research.

Overall, the H-models predict the number of trips made by a household while the P-models predict the number of trips made by a person. In addition, the number of explanatory variables used in the latter models is significantly greater than those used in the former. Thus, the H-models are treated as the "aggregate" models and the P-models are considered as "disaggregate" models.

Table 5.3 Household-level Trip Generation Models (H-models) for Households in Single-Family Dwelling units

Variables	HBW		HBNW		NHB	
	Coeff.	t stat.	Coeff.	t stat.	Coeff.	t stat.
Constant	1.160	25.802	2.266	26.566	1.542	20.447
hhsize = 1 & nveh = 0	-	-	-0.959	-3.488	-0.979	-3.940
hhsize = 1 & nveh = 1	-	-	-	-	-	-
hhsize = 1 & nveh = 2	-	-	-	-	-	-
hhsize = 1 & nveh = 3	-	-	-	-	-	-
hhsize = 1 & nveh = 4	-	-	-	-	-	-
hhsize = 1 & nveh = 5+	-	-	-	-	-	-
hhsize = 2 & nveh = 0	0.605	2.256	1.054	2.182	-	-
hhsize = 2 & nveh = 1	0.264	2.867	2.329	14.181	0.824	5.579
hhsize = 2 & nveh = 2	0.831	14.501	2.031	17.613	1.243	12.065
hhsize = 2 & nveh = 3	0.889	11.833	1.883	11.578	1.364	9.324
hhsize = 2 & nveh = 4	1.044	9.059	1.793	6.717	1.726	7.157
hhsize = 2 & nveh = 5+	0.819	5.033	1.800	4.784	1.690	4.965
hhsize = 3 & nveh = 0	0.324	2.704	3.384	3.737	-	-
hhsize = 3 & nveh = 1	0.324	2.704	4.972	17.558	2.530	9.889
hhsize = 3 & nveh = 2	0.760	10.780	4.213	25.981	2.454	16.823
hhsize = 3 & nveh = 3	1.279	16.099	3.816	20.226	2.425	14.265
hhsize = 3 & nveh = 4	1.364	11.341	4.272	14.257	3.113	11.495
hhsize = 3 & nveh = 5+	1.650	10.508	3.340	8.444	2.532	7.077
hhsize = 4 & nveh = 0	0.351	2.492	3.845	4.030	-	-
hhsize = 4 & nveh = 1	0.351	2.492	8.055	22.222	3.581	10.924
hhsize = 4 & nveh = 2	0.687	9.987	8.140	51.194	3.524	24.643
hhsize = 4 & nveh = 3	1.214	14.014	8.076	38.366	3.898	20.531
hhsize = 4 & nveh = 4	1.658	14.152	7.185	24.364	3.669	13.768
hhsize = 4 & nveh = 5+	2.088	13.755	6.386	16.564	4.328	12.412
hhsize = 5 & nveh <= 1	0.444	2.051	8.718	16.661	3.639	7.685
hhsize = 5 & nveh = 2	0.794	8.018	11.558	47.429	4.667	21.211
hhsize = 5 & nveh = 3	1.087	8.771	9.951	31.852	4.275	15.139
hhsize = 5 & nveh = 4	1.876	10.777	9.544	21.297	4.816	11.878
hhsize = 5 & nveh = 5+	2.231	9.633	10.560	17.583	5.589	10.282
hhsize = 6+ & nveh <= 1	0.446	1.640	13.629	20.661	5.090	8.525
hhsize = 6+ & nveh = 2	0.565	3.660	13.913	35.634	5.771	16.342
hhsize = 6+ & nveh = 3	1.380	6.923	14.004	27.186	5.585	11.983
hhsize = 6+ & nveh = 4	2.712	10.811	14.159	22.013	6.883	11.824
hhsize = 6+ & nveh = 5+	2.143	7.878	12.852	18.446	5.576	8.842
No. of Observations	8962		12052		12052	
R^2	0.089		0.425		0.152	
Adjusted R^2	0.087		0.423		0.15	

Table 5.4 Household-level Trip Generation Models (H-models) for Households in Multi-Family Dwelling units

Variables	HBW		HBNW		NHB	
	Coeff.	t stat.	Coeff.	t stat.	Coeff.	t stat.
Constant	1.271	27.366	1.982	19.945	1.604	17.201
hhsize = 1 & nveh = 0	-	-	-0.317	-1.605	-0.757	-4.129
hhsize = 1 & nveh = 1	-	-	-	-	-	-
hhsize = 1 & nveh = 2+	-	-	-	-	0.565	1.772
hhsize = 2 & nveh = 0	-	-	1.928	4.840	-	-
hhsize = 2 & nveh = 1	0.239	2.071	2.919	14.338	0.752	3.975
hhsize = 2 & nveh = 2	0.988	9.841	2.094	10.269	0.974	5.141
hhsize = 2 & nveh = 3+	0.702	3.162	1.927	4.316	1.669	4.032
hhsize = 3 & nveh = 0	-	-	4.059	6.792	1.271	2.295
hhsize = 3 & nveh = 1	-	-	5.423	16.415	1.241	4.052
hhsize = 3 & nveh = 2+	0.844	5.665	3.807	11.890	1.463	4.926
hhsize = 4+ & nveh = 0	0.774	2.715	7.172	12.474	2.358	4.425
hhsize = 4+ & nveh = 1	-	-	7.384	20.696	2.283	6.901
hhsize = 4+ & nveh = 2+	0.792	4.589	8.556	23.021	3.781	10.970
No. of Observations	1394		2118		2118	
R^2	0.087		0.403		0.115	
Adjusted R^2	0.083		0.4		0.11	

Table 5.5 Person-level Trip Generation Models (P-models) for Adults

Variables	HBW		HBNW		NHB	
	Coeff.	t stat.	Coeff.	t stat.	Coeff.	t stat.
Constant	0.843	32.302	2.120	23.303	0.636	8.250
Female	-	-	-	-	-	-
Male	-	-	-	-	-	-
18 yr <= Age<= 24 yr	-	-	-	-	-	-
25 yr <= Age<= 55 yr	-	-	-0.123	-2.352	-	-
56 yr <= Age<= 75 yr	-	-	-0.072	-1.273	-	-
Age>= 76 yr	-	-	-0.572	-8.220	-0.277	-5.935
Non-US citizen	-	-	-	-	-	-
US citizen	-	-	-	-	0.201	4.429
Not Employed	-	-	-	-	-	-
Part-time Employed	-	-	-0.365	-7.228	0.297	7.332
Full-time Employed	0.372	15.901	-1.237	-32.530	0.321	9.460
Low Income	-	-	-	-	-	-
Medium Income	-	-	0.182	5.365	0.170	5.315
High Income	-0.054	-3.226	0.284	7.898	0.326	9.621
White	-	-	-	-	-	-
Black	-0.118	-3.198	-0.130	-2.285	-	-
Others	-	-	-0.162	-3.371	-0.162	-3.390
Single Family	-	-	-	-	-	-
Multi Family	-	-	0.084	1.761	-	-
Own	-	-	-	-	-	-
Rented	0.062	2.824	-0.089	-2.144	-	-
Nveh/Nadult = 0	-	-	-	-	-	-
0 < Nveh/Nadult < 1	-	-	0.716	9.569	0.438	6.191
Nveh/Nadult = 1	-	-	0.739	10.407	0.536	8.193
Nveh/Nadult > 1	-	-	0.745	9.942	0.591	8.594
Single Person	-	-	-	-	-	-
Single Parent	-0.165	-3.088	0.511	5.561	0.396	4.187
Couple	0.099	4.350	-0.131	-3.807	-0.279	-7.034
Nuclear Family	-	-	-	-	-0.218	-3.991
Roommates	-	-	-	-	-0.411	-5.069
Others	0.104	4.822	-0.327	-8.977	-0.443	-9.697
No. of Chilrdren in HH	-0.085	-6.791	0.293	15.668	0.140	6.826
Male * No. of Children in HH	0.114	7.616	-0.223	-9.239	-0.117	-5.226
Male * Part-Time Employed	-	-	0.225	2.788	-	-
Male * Full-Time Employed	0.082	3.816	-0.064	-1.598	-0.111	-2.964
No. of Observations	16142		25945		25945	
R^2	0.042		0.099		0.037	
Adjusted R^2	0.042		0.098		0.037	

Table 5.6 Person-level Trip Generation Models (P-models) for Children

Variables	HBNW		NHB	
	Coeff.	t stat.	Coeff.	t stat.
Constant	2.137	28.117	0.865	11.250
Female	-	-	-	-
Male	-	-	-0.097	-3.192
Age <= 4 yrs	-	-	-	-
5 yr <= Age<= 7 yr	0.375	6.780	-	-
8 yr <= Age<= 10 yr	0.438	8.021	-	-
11 yr <= Age<= 13 yr	0.452	8.422	-0.115	-2.863
14 yr <= Age<= 17 yr	0.367	6.890	0.066	1.649
Non-US Citizen	-	-	-	-
US Citizen	-	-	-	-
Low Income	-	-	-	-
Medium Income	-	-	0.073	1.610
High Income	0.156	4.130	0.073	1.573
White	-	-	-	-
Black	-	-	-0.185	-2.871
Others	-0.105	-1.826	-0.224	-4.460
Single Family	-	-	-	-
Multi Family	-	-	-	-
Own	-	-	-	-
Rented	-0.112	-2.348	-	-
Nveh/Nadult = 0	-	-	-	-
0 < Nveh/Nadult < 1	-	-	-	-
Nveh/Nadult = 1	0.173	3.124	0.141	2.846
Nveh/Nadult > 1	0.155	2.517	0.158	2.936
Single Parent	-	-	-	-
Nuclear Family	0.172	4.064	-0.126	-2.302
Others	-	-	-0.197	-3.080
No. of Children in HH	-0.052	-3.148	0.020	1.433
No. of Observations	8312		8312	
R^2	0.026		0.011	
Adjusted R^2	0.024		0.01	

5.2 Population Synthesis

One of the objectives of this research is to examine the predictions of a trip-generation model when applied to a true population with the predictions when the same model is applied to a synthesized version of the same population. The Florida sample (1723 persons from 774 households) is extracted from the National, weekday sample of the NHTS 2001. This sample is treated as the true population as all the socio-economic characteristics are known for each person and household. The same population was also synthesized using the procedure described in Chapter 3. The control tables were generated by aggregating the household and population characteristics of the Florida sample of the NHTS (instead of using the US Census SF tables). The weekday, national sample of the NHTS was used as the seed data (instead of the PUMS data as described in Chapter 3). It is useful to note that the NHTS does not include institutionalized population and hence the population in group-quarters is not a part of this analysis.

Two populations were synthesized. The first population was synthesized using eleven control tables (Table 5.7). Eight of these tables are two dimensional and three are one dimensional. Seven of these are household-level tables (i.e., the universe is all households) and four are person-level tables. Overall, the structures of these tables are largely consistent with those presented in Chapter 3. The second population was synthesized using just four one-dimensional control tables (household size, age, gender, and employment status).

Table 5.7 Control Tables Used in the Synthesis of Population 1

		Dimension 1		Dimension 2	
S. No	Universe	Attribute	Categories	Attribute	Categories
1	Households	TENURE	Own, Rent	HHSIZE	1,2,3,4,5,6,7+
2	Households	TENURE	Own, Rent	DUTYPE	Single Family, Multi-Family
3	Households	TENURE	Own, Rent	NUMAUTO	0,1,2,3,4,5+
4	Households	HHSTRUCT	Family, Non-Family	HHSIZE	1,2,3,4,5,6,7+
5	Households	HHSTRUCT	Married couple, other family	CHAGE	none, only <6, only >=6, both <6 and >= 6
6	Households	HHSTRUCT	Married couple, other family	NUMWORK	0,1,2, 3+
7	Households	INCOME	< 30K, 30-50K, 50-75K, more than 75K		*NA*
8	Total Population	ETHNICITY	White, Black, Other		*NA*
9	Total Population	GENDER	Male, Female	AGE	0-5, 6-15, 16-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, over 75
10	Total Population	CITIZEN	Citizen, Non Citizen		*NA*
11	Total Population	GENDER	Male, Female	EMPSTAT	Parttime, Fulltime, Unemployed

Table 5.8 compares the characteristics of the true population (1723 persons and 774 households from the weekday, Florida sample of the NHTS 2001) with the corresponding characteristics of the two synthesized populations. In the case of Synthetic Population 1, the distributions of all the attributes match closely with the true distributions. This is because all these attributes are explicitly controlled-for in the synthesis procedure (see the Control Tables in Table 5.7). In the case of Synthetic Population 2, the distributions of household size, age, gender, and employment status match very well with the true values as these are controlled. However, the distributions of other attributes do not match as well.

Table 5.8 Characteristics of the True and Synthesized Populations

	True Population	Synthetic Population 1	Synthetic Population 2			True Population	Synthetic Population 1	Synthetic Population 2
HH Size					Number of Autos			
1	28.811	28.811	28.774		0	3.488	3.488	3.097
2	41.860	41.860	41.935		1	39.535	39.535	44.000
3	14.599	14.599	14.581		2	39.018	39.018	30.839
4	10.078	10.078	10.065		3	11.370	11.370	11.484
5	2.584	2.584	2.581		4	4.910	4.910	8.645
6	1.680	1.680	1.677		5+	1.680	1.680	1.935
7+	0.388	0.388	0.387		Gender			
HH Structure					Male	46.721	46.775	46.721
Married Couple	49.742	49.742	17.935		Female	53.279	53.225	53.279
Other Family	3.230	3.230	12.000		Age			
Non Family	47.028	47.028	70.065		0 to 5	5.223	5.230	5.223
Children					6 to 15	11.724	11.737	11.724
None	76.744	75.581	65.032		16 to 17	1.741	1.743	1.741
Only <= 6 years	4.393	5.297	9.161		18 to 24	4.527	4.532	4.527
Only > 6 years	13.824	13.953	22.581		25 to 34	9.402	9.413	9.402
Both	5.039	5.168	3.226		35 to 44	14.452	14.410	14.452
Number of Workers					45 to 54	15.032	15.049	15.032
0	35.401	35.401	23.613		55 to 64	13.929	13.887	13.929
1	34.496	34.496	56.774		65 to 74	13.465	13.481	13.465
2	26.357	26.357	17.290		>= 75	8.067	8.774	8.590
3+	3.747	3.747	2.323		Ethnicity			
Tenure					White	85.548	85.532	86.767
Own	83.463	83.46	80.258		Black	7.487	7.496	6.732
Rent	16.537	16.54	19.742		Other	6.965	6.973	6.500
Dwelling Unit Type					Employment Status			
Single Family	83.204	83.20	87.097		Not Employed	55.543	55.665	55.543
Multi Family	16.796	16.80	12.903		Part Time	8.474	8.483	8.474
Income					Full Time	35.984	35.851	35.984
< 30	37.468	37.468	38.323		Citizenship			
30 - 50	23.773	23.773	25.935		Native	86.767	86.752	92.571
50 -75	18.863	18.863	16.645		Non Citizen	13.233	13.248	7.429
> 75	19.897	19.897	19.097					

5.3 Accuracy of Model Predictions

The aggregate (or H-models) and disaggregate (or P-models) models developed for each of the three trip purposes (HBW, HBNW, and NHB) were presented in Section 5.1. Section 5.2 described the three populations to which these models were to be applied. One of these is the "true" population (1723 persons and 774 households extracted from the weekday, Florida sample of the NHTS 2001) and the other two are "synthesized". This chapter presents the results of an analysis of the predictive accuracy of each of the models when applied to each of the populations. The results are examined separately for each of the trip purposes. The following is an outline of the analysis procedure for each of the three trip purpose:

- Apply each of the aggregate (or H-models) and disaggregate (or P-models) models to each of the three populations to predict the number of trips for each household and person respectively in the corresponding population.

- Aggregate the predictions from the P-model to the household level to facilitate comparisons with the predictions from the H-model.
- The six predictions of travel volumes obtained (two models * three populations) are compared with the actual travel volumes as reported in the travel surveys. The 1723 persons and 774 households extracted from the weekday, Florida sample of the NHTS 2001 reported a total of 887 HBW trips, 4025 HBNW trips, and 2184 NHB trips.

The results are presented in Tables 5.9 (HBW Trips) 5.10 (HBNW Trips) and 5.11 (NHB Trips). All three tables have the same structure which is described prior to the discussion of the results. Each table has two major blocks of rows. The upper block represents the results from applying the aggregate (H) models and the lower block represents the results from applying the disaggregate (P) models. Each table also has four major columns. In the first column (labeled “Actual”), the true values from the surveys are reported. For example, from Table 5.9 we observe that 887 HBW trips were actually reported in the surveys with an average trip rate of 1.15 per household. The next three major columns represent the three populations to which the model is applied (“True Population”, Synthetic Population 1” and “Synthetic Population 2”). Within each of these three major columns, there are two more columns: “Predicted” and “% Error”. The column “Predicted” provides the value obtained by applying the model to the corresponding population and the column “% Error” compares the predicted value with the actual value in the first column and is calculated as the difference between the predicted and actual as the percentage of the actual.

The results for the HBW trips are presented in Table 5.9. Two key observations can be made. *First,* the error obtained by applying any model (H or P) to Synthetic Population 1 is rather close to the error obtained by applying the same model to the true population. However the error when the model is applied to Synthetic Population 2 is significantly larger. As discussed in Section 5.1, the models for HBW were estimated only for a subset of the population. Therefore, the model application results are strongly related to the number of workers in the population. The distribution of the number of workers per household for the Synthetic Population 2 does not closely match the true distribution (See Table 5.8) as this variable is not explicitly controlled for (the employment status is controlled for at the individual level). In fact, there are fewer zero-worker households in Synthetic Population 2 compared to the true population and this could be a reason for the substantial over-prediction of HBW trips in this population. This result underscores the need for using the right control variables in population synthesis. *Second*, for any population, the errors by applying the P model are greater than those by applying the H model. This is contrary to expectations as disaggregate models are generally expected to perform better than aggregate models. However, it is important to note that that the linear-regression structure may be inappropriate to model person-level HBW trips (people make at most two HBW trips per day with only about 2.5% of the population reporting more than 2 HBW trips – See Table 5.2). However, there is more variability in the number of household-level HBW trips. This limited variability of the dependent variable in the sample could severely limit the ability of the linear-regression model to explain the HBW trip-making behavior at the person-level.

Table 5.9 Predictive Accuracy of the Models for HBW Trips

	Actual	H Model					
		True Population		Synthetic Population 1		Synthetic Population 2	
		Predicted	% Error	Predicted	% Error	Predicted	% Error
Total # Trips	887.00	934.44	5.35	868.94	-2.04	1076.55	21.37
Trips Per HH							
Mean	1.15	1.21	5.35	1.12	-2.04	1.39	21.21
Std. Dev.	1.48	0.98	-33.98	0.91	-38.61	0.91	-38.43
Min.	0.00	0.00	NA	0.00	NA	0.00	NA
Max.	9.00	3.87	-56.98	3.87	-56.98	3.87	-56.98

	Actual	P Model					
		True Population		Synthetic Poulation 1		Synthetic Poulation 2	
		Predicted	% Error	Predicted	% Error	Predicted	% Error
Total # Trips	887.00	1089.85	22.87	1082.52	22.04	1286.97	45.09
Trips Per HH							
Mean	1.15	1.41	23.03	1.40	22.04	1.66	44.91
Std. Dev.	1.48	1.24	-16.25	1.25	-15.64	1.28	-13.46
Min.	0.00	0.00	NA	0.00	NA	0.00	NA
Max.	9.00	6.32	-29.82	5.25	-41.69	6.41	-28.83

The results for the HBNW trips are presented in Table 5.10. Two key observations can be made. *First*, for any model (H or P), the errors obtained by application to each of the three populations are comparable. Thus, the aggregate prediction of the total volume of the HBNW trips in the population is approximately the same, irrespective of whether it is applied to the true population or the rather approximate Synthetic Population 2. *Second*, for any population, the errors by applying the P model are comparable (albeit marginally greater) to those obtained by applying the H models. Together these results indicate that a disaggregate model applied to a synthetic population with few controls (i.e., the Synthetic Population 2) can predicts approximately the same aggregate volume of HBNW trips compared to an aggregate model applied to the "true" population.

Table 5.10 Predictive Accuracy of the Models for HBNW Trips

	Actual	H Model					
		True Population		Synthetic Population 1		Synthetic Population 2	
		Predicted	% Error	Predicted	% Error	Predicted	% Error
Total # Trips	4025.00	3953.97	-1.76	3932.92	-2.29	3951.75	-1.82
Trips Per HH							
Mean	5.20	5.11	-1.76	5.08	-2.29	5.10	-1.95
Std. Dev.	4.83	3.16	-34.67	3.09	-36.14	3.04	-37.07
Min.	0.00	1.31	NA	1.31	NA	1.31	NA
Max.	38.00	16.43	-56.78	16.43	-56.78	16.43	-56.78

	Actual	P Model					
		True Population		Synthetic Poulation 1		Synthetic Poulation 2	
		Predicted	% Error	Predicted	% Error	Predicted	% Error
Total # Trips	4025.00	3930.78	-2.34	3895.08	-3.23	3881.38	-3.57
Trips Per HH							
Mean	5.20	5.09	-2.24	5.03	-3.26	5.01	-3.72
Std. Dev.	4.83	3.16	-34.53	3.16	-34.61	2.79	-42.34
Min.	0.00	0.68	NA	0.88	NA	0.99	NA
Max.	38.00	21.21	-44.18	18.57	-51.14	17.58	-53.75

The results presented in Table 5.11 clearly indicate the superiority of disaggregate models over aggregate models for NHB trips. As in the case of HBNW trips, the aggregate prediction of the total volume of the NHB trips in the population is found to be approximately the same, irrespective of whether it is applied to the true population or either of the two synthetic populations. Further, the errors from the disaggregate (P) models are significantly lesser than those from the aggregate (H) models. In fact, the application of the P model to the least-accurate population (SP2) produces a more accurate estimate than even applying the aggregate model to the true population. It is possible that the decisions of NHB trips are inherently made at the individual-level as opposed the household level leading to the better performance of the P models. It is also possible that the large variations in the number of NHB trips (See Tables 5.1 and 5.2) at the person level makes linear regression still acceptable for the P models.

Table 5.11 Predictive Accuracy of the Models for NHB Trips

	Actual	H Model					
		True Population		Synthetic Population 1		Synthetic Population 2	
		Predicted	% Error	Predicted	% Error	Predicted	% Error
Total # Trips	2184.00	2286.17	4.68	2271.99	4.03	2280.34	4.41
Trips Per HH							
Mean	2.82	2.95	4.68	2.94	4.03	2.94	4.28
Std. Dev.	3.19	1.45	-54.66	1.39	-56.47	1.47	-54.06
Min.	0.00	0.56	NA	0.56	NA	0.56	NA
Max.	22.00	8.43	-61.70	8.43	-61.70	8.43	-61.70

	Actual	P Model					
		True Population		Synthetic Poulation 1		Synthetic Poulation 2	
		Predicted	% Error	Predicted	% Error	Predicted	% Error
Total # Trips	2184.00	2192.57	0.39	2169.55	-0.66	2187.68	0.17
Trips Per HH							
Mean	2.82	2.84	0.39	2.80	-0.79	2.82	-0.09
Std. Dev.	3.19	1.48	-53.71	1.45	-54.44	1.38	-56.69
Min.	0.00	0.56	NA	0.36	NA	0.84	NA
Max.	22.00	8.52	-61.29	8.96	-59.28	9.98	-54.64

The above analyses provide some evidence in favor of disaggregate models. Specifically, in two of the three trip purposes (HBNW and NHB), we find that the P models can perform just as good (if not better) as the H models. For the same trip purposes, we also find that the models when applied to the synthetic population produce just as accurate results as applied to the true population. Therefore, the inaccuracies in the synthetic populations may not be large relative to the model errors themselves. The results for the HBW trip purpose highlights the need for choosing more appropriate econometric structure when modeling entities at the disaggregate level. Further, the result also shows the need for choosing the right control variables for the population synthesis.

Finally, it is important to indicate here that this analysis only examined the prediction accuracy at the aggregate level (i.e., the total number of trips for the entire population). One of the strengths of disaggregate models is its ability to capture the differences in the trip patterns across different segments of the population. Given the relatively small size (774 households) of the total population used in this analysis, the evaluation of predictive accuracy was not extended to subsections of the population (market segments). This is identified as an important area for future research.

CHAPTER 6 SUMMARY

Over the past several years, there has been a growing interest in the development of disaggregate (individual- or household-level) travel-demand models. This interest is motivated by several factors such as (1) reduction of aggregation errors, (2) ensure sensitivity to demographic shifts like the ageing of the population, (3) capture differential sensitivity and response of travelers to policy actions, and (4) address special travel-needs of certain population groups.

The recognition of the above-described issues by Florida Department of Transportation is evident from their efforts to incorporate socio-demographic variables (*i.e.,* household characteristics) within the Florida Standard Urban Transportation Model Structure (FSUTMS). Specifically, the Tampa Bay and the South-East Florida regions have developed "lifestyle" trip production models. However, the lack of a systematic procedure to forecast the household characteristics (*i.e.,* the lifestyle variables) required by such disaggregate travel-demand models has been recognized as an important impediment to furthering these efforts for state-wide adoption.

In this context, the broad focus of this research is to contribute towards the development of methodology for comprehensively forecasting all traveler characteristics required as inputs to travel-demand forecasting models. This procedure is also referred to as synthetic population generation (SPG) in the literature.

The state-of-the-practice approach to population synthesis involves the use of the Iterative Proportional Fitting (IPF) method. While there have been several applications of this approach, the following issues still remain. First, the number of controls used in the synthesis of the population has been limited. In particular, most practical applications do not control for person-level attributes such as age and gender. Second, documentation of the validation of the procedure, especially in the context of a target year population is limited. Third, there does not seem to be any comparison of the travel patterns predicted using true populations with those predicted using synthetic populations.

This research addresses the issues identified above. A new greedy-heuristic data-fitting algorithm is developed that can be used to synthesize population with a large number of control tables both at household- and person-levels. The procedure is implemented in GAUSS, a matrix programming language. The code was used to synthesize the year 2000 population for 13 census tracts of varying populations and areas in Florida. Two sets of populations were estimated – the first with only household-level controls and the second with both household- and person-level controls. Validation analysis indicates that the second synthesized population matches the true distributions better. In fact, the extent of mismatch with the (uncontrolled) person-level tables is significant with the first population (i.e., synthesized with only household controls).

As a second step, the populations of 1990 were synthesized for the same 13 census tracts. Once again, two sets of populations were synthesized. One used the year-2000 population synthesized with only household-level controls as the seed data whereas the second used the year-2000 population synthesized with both household- and person-level controls as the seed data. The aggregate characteristics of the synthesized populations were compared with several control tables from the 1990 US Census. Once again, the results indicate that the use of both person- and household- controls in the base year synthesis leads to more accurate population estimates for the target year. Overall, the analysis highlights the value of a methodology that incorporates both controls in population synthesis.

Finally, travel estimates obtained by applying trip-generation models to the true population were compared with those obtained by applying the same models to a synthetic population. Trip generation models (household-level and person-level) were estimates using the weekday, national sample from the National Household Travel Survey of 2000. Subsequently, the estimated models were applied to the Florida sample of the survey data (i.e., the true population) to predict the travel estimates. The population characteristics of the Florida sample were also synthesized and the models were applied to these synthesized populations. The analyses provide some evidence in favor of disaggregate models. Specifically, for two trip purposes (home-based other and non-home-based), we find that the disaggregate models can perform just as good (if not better) as the aggregate models. For the same trip purposes, we also find that the travel estimates obtained by applying the models to the synthetic population are as accurate as the ones obtained by applying the same model to the true population. Thus, the need to synthesize the population characteristics does not necessarily deteriorate the trip-generation predictions (from linear-regression models) substantially. The results for the home-based work trip purpose highlights the need for choosing the appropriate econometric structure when developing disaggregate models and the right control variables for the population synthesis.

REFERENCES

Beckman, R.J., Baggerly, K.A., McKay, M.D. (1996) “Creating Synthetic Baseline Populations”, *Transportation Research Part A*, Vol. 30, No 6, pp. 415-429.

Bradley, M. and Bowman, J. (2006) “A Summary of the Design Features of Activity-Based Microsimulation models for US MPOs”, presented at the *Conference on Innovations in Travel-Demand Modeling,* Austin, TX. Available from: http://www.trb.org/Conferences/TDM/papers/BS1A%20-%20Austin_paper_bradley.pdf

Bowman, J.L. (2004) “A Comparison of Population Synthesizers Used in Microsimulation Models of Activity and Travel Demand”, draft paper available from http://jbowman.net/papers/B04.pdf

Bowman, J.L. and Bradley, M (2006) “Activity-Based Travel Forecasting Model for SACOG: Population Synthesis” *Technical Memo Number 2*, prepared for Sacramento Area Council of Governments, available from http://jbowman.net/ProjectDocuments/SacSim/SACOG%20tech%20memo%202--Pop%20Synth.20060731.pdf

Bowman, J.L. and Bradley, M (2007) “Activity-Based Travel Forecasting Model for SACOG: Household Auto Availability Model” *Technical Memo Number 9*, prepared for Sacramento Area Council of Governments, available from http://jbowman.net/ProjectDocuments/SacSim/SACOG%20tech%20memo%209--Auto%20availability.20060914.pdf

Bowman, J.L. and Rousseau, G (2006) “Validation of the Atlanta (ARC) Population Synthesizer”, presented at the *TRB Conference on Innovations in Travel Modeling,* Austin, Texas, USA. Available from http://jbowman.net/papers/20060425%20Bowman%20and%20Rousseau%20ARC%20PopSyn.pdf

Eluru, N., Pinjari, A., Guo, J.Y., Sener, I., Srinivasan, S., Copperman., R., and Bhat., C.R. (2008) “Population Updating System Structures and Models Embedded within the Comprehensive Econometric Microsimulator for Urban Systems”, *Transportation Research Record*, Vol. 2076, pp. 171-182.

Goulias, K. G., and Kitamura. R. (1996) A Dynamic Model System for Regional Travel Demand Forecasting. In *Panels for Transportation Planning: Methods and Applications*, Eds. Golob, T., R. Kitamura, and L. Long, Kluwer Academic Publishers, Boston, Ch. 13, pp. 321-348.

Guo, J.Y. and Bhat., C.R. (2007) “Population Synthesis for Microsimulating Travel Behavior”, *Transportation Research Record,* Vol. 2014, pp. 92-101.

Hunt, J.D., Abraham, J.E., and Weidner, T. (2004) “The Household Allocation (HA) Module of the Oregon2 Model", *Transportation Research Record* Vol. 1898, pp. 98-107.

Mackett, R. L. (1990) *MASTER Mode*. Report SR 237, Transport and Road Research Laboratory, Crowthorne, England.

Sundararajan, A., and Goulias, K. G. (2003) Demographic Microsimulation with DEMOS 2000: Design, Validation, and Forecasting. In *Transportation Systems Planning: Methods and Applications*, Eds. K.G. Goulias, CRC Press, Boca Raton, Ch. 14.

APPENDIX A ADJUSTMENT OF SF3 TABLES TO MATCH SF1 VALUES

The following is table “H15” obtained from SF1 files for a census tract. This table presents the joint distribution of household size against tenure. One can see that, 2638 households own their home while 341 rent their home in this tract.

		HH SIZE							
		1	2	3	4	5	6	7+	Total
TENURE	Own	644	1089	377	357	124	36	11	2638
	Rent	141	107	46	31	8	8	0	341
	Total	785	1196	423	388	132	44	11	2979

H15

The following is table “H32” obtained from SF3 files for the same census tract as above. This table presents the joint distribution of dwelling-unit type against tenure. One can see that, 2640 households own their home while 339 rent their home in this tract.

		DUTYPE		
		Single-Family	Multi-Family	Total
TENURE	Own	2279	361	2640
	Rent	164	175	339
	Total	2443	536	2979

H32

The intent of the adjustment procedure is to modify the above table such that the number of own and rented households are consistent with the values from the “H15” SF1 table. To do this, scaling factors are calculated for each of the “own” and “rent” categories. The factor for the “own” category is calculated as 2638/2640 = 0.999 (i.e., the number of own households in the SF1 table divided by the number of own households in the SF3 table). The corresponding value for the “rent” category is 341/339 = 1.006. Now, the values in each row of the table “H32” are multiplied by the corresponding scaling factors and rounded to an integer. For instance, the “adjusted” number of “own, single-family” households is 2279*0.999 = 2276.72 = 2277. Similarly, the adjusted number of “rent, single-family” households = 164*1.006 = 164.98 = 165.

The cells corresponding to the “multi-family” column are also similarly adjusted. The final adjusted table is presented below.

DUTYPE

		Single-Family	Multi-Family	Total
TENURE	Own	2277	361	2638
	Rent	165	176	341
	Total	2442	537	2979

H32

APPENDIX B NUMERICAL ILLUSTRATION OF THE POPULATION SYNTHESIS PROCEDURE

Section 3.2 outlined the procedure for population synthesis which involves selecting a set of households from the PUMS data in such a way that the tract-level controls are satisfied. One household is selected in each iteration of the procedure. A numerical illustration of this procedure is presented here.

For simplicity, we assume that there are two control tables (Figure B-1) for a hypothetical census tract (synthesis area). The first table (T_{1k}) is a two-dimensional household-level table joint distribution of household-size (household size is limited to being either 1 or 2 persons, again for simplicity) and tenure. The second table (T_{2k}) is a one-dimensional, person-level table representing the distribution of gender. The intent of the population-synthesis procedure is to generate households and persons that satisfy the distributions present in these control tables.

T_{1k}	HHSize = 1	HHSize = 2	Total
Own	1	5	6
Rent	2	2	4
Total	3	7	10

T_{2k}	Total
Male	11
Female	6
Total	17

Figure B-1 Control Tables

The seed data are presented in Figure B-2. There are two tables. The household-level table presents the tenure and household size of each household. The person-level table presents the gender of each person present in each of the households in the seed data. We see that there are five households and eight persons in this dataset. The reader will note that the seed-data has at least one household of each of the four types (i.e., Own, 1-person; Own, 2-person; Rent, 1-person; and Rent, 2-Person) represented in the control table. Similarly, there are both males and females in the seed data. Thus, this dataset has already been pre-treated.

HH ID (*i*)	Tenure	HH Size
1	Rent	1
2	Own	1
3	Rent	2
4	Own	2
5	Own	2

HH ID (*i*)	Person ID	Gender
1	1	Female
2	1	Male
3	1	Male
3	2	Male
4	1	Male
4	2	Female
5	1	Male
5	2	Male

Figure B-2 Seed Data

The population synthesis is an iterative procedure and one household is selected in each step. Count tables (CT) are used to track the number of households/persons of different types that have been selected until any point in the iteration. There are two count tables corresponding to the two control tables and the cell values of these tables are initialized to zero at the beginning of the population synthesis procedure.

CT^{0}_{1k}	HHSize = 1	HHSize = 2	Total
Own	0	0	0
Rent	0	0	0
Total	0	0	0

CT^{0}_{2k}	Total
Male	0
Female	0
Total	0

Figure B-3 Count Tables at the Start of the Population Synthesis.

The selection of households is based on a fitness function. This fitness function is calculated for each household in the seed data and for each iteration based on the values of the control tables (T), the count tables (CT) and the contribution of the corresponding household towards satisfying the controls (captured in HT-tables). Figure B-4 presents the values in the two HT-tables for the five households in the seed data. As already defined, the HT-tables define the contribution of each household towards satisfying the different controls. Household 1 (HH ID = 1) comprised a single female living in a rental house (Figure B2). Thus, “selecting” this household would contribute one single-person, rental household (as indicated in the first HT-table in Figure B-4) and one female (as indicated in the second HT-table) to the population. Similarly, selecting the third household from the seed data will contribute one two-person, rental household (as indicated in the first HT-table for the third household) and two males (as indicated in the second HT-table of the third household) to the population.

HH ID (i)= 1

HT^i_{1k}	HHSize = 1	HHSize = 2
Own	0	0
Rent	1	0

HT^i_{2k}	Total
Male	0
Female	1

HH ID (i)= 2

HT^i_{1k}	HHSize = 1	HHSize = 2
Own	1	0
Rent	0	0

HT^i_{2k}	Total
Male	1
Female	0

HH ID (i) = 3

HT^i_{1k}	HHSize = 1	HHSize = 2
Own	0	0
Rent	0	1

HT^i_{2k}	Total
Male	2
Female	0

HH ID (i) = 4

HT^i_{1k}	HHSize = 1	HHSize = 2
Own	0	1
Rent	0	0

HT^i_{2k}	Total
Male	1
Female	1

HH ID (i)= 5

HT^i_{1k}	HHSize = 1	HHSize = 2
Own	0	1
Rent	0	0

HT^i_{2k}	Total
Male	2
Female	0

Figure B-4 HT-tables for each of the Households in the Seed Data

Once all the tables have been defined, the “fitness” can be calculated for each of the households in the seed data. The fitness of a household *i* in iteration *n* $\left(F^{in}\right)$ is calculated using the following formula:

$$F^{in} = \sum_{j=1}^{J}\left[\frac{1}{e_j^i}\sum_{k=1}^{K_j}\left[\frac{\left(R_{jk}^{n-1}\right)^2}{T_{jk}} - \frac{\left(R_{jk}^{n-1} - HT_{jk}^i\right)^2}{T_{jk}}\right]\right]$$

$$\text{Where: } R_{jk}^{n-1} = T_{jk} - CT_{jk}^{n-1}$$

In the current example, *J* = 2 (there are two control tables) and K_1 = 4 (four cells in the first control table) and K_2 = 2 (two cells in the second control table). e_1^i takes the value of 1 (first control table is a household-level table) and e_2^i is the size of household *i* (second control table is a person-level table and hence $e_2^1 = e_2^2 = 1$ and $e_2^3 = e_2^4 = e_2^5 = 2$). Thus, the formula can be rewritten for the example under consideration as follows:

$$F^{in} = \left\{\sum_{k=1}^{4}\left[\frac{\left(R_{1k}^{n-1}\right)^2}{T_{1k}}\right] - \sum_{k=1}^{4}\left[\frac{\left(R_{1k}^{n-1} - HT_{1k}^i\right)^2}{T_{1k}}\right]\right\} + \frac{1}{\left(HHSize^i\right)}\left\{\sum_{k=1}^{2}\left[\frac{\left(R_{2k}^{n-1}\right)^2}{T_{2k}}\right] - \sum_{k=1}^{2}\left[\frac{\left(R_{2k}^{n-1} - HT_{2k}^i\right)^2}{T_{2k}}\right]\right\}$$

The iterative algorithm begins with the calculation of the fitness values for each of the five households. The household with the highest value of fitness is selected into the synthetic population of the census tract. The count tables are then updated to reflect the household added into the population. The fitness values are then recalculated and the iterations continue. A household from the seed data can be selected into the population of the census tract multiple times. The algorithm terminates when the fitness values of all households are negative.

The numerical calculations corresponding to the application of the algorithm to the example problem is presented in Figures B-5 through B-15. Each figure represents the calculations corresponding to one of the iterations. In the first iteration (Figure B-5), the values in the count tables (CT) are zero as no household has been selected into the population yet. On calculating the fitness values, we find that household 4 (HH ID = 4) has the highest value. Hence this household is added into the population of the census tract. The count tables are then suitably updates (Figure B-6). Note that household 4 is a two-person own-home household with one male and one female. Thus, the corresponding cells of the count tables (CT) are updated by 1. The fitness values are then re-calculated. Now, household 3 is found to have the highest fitness (Figure B-6) and hence this is added into the synthetic population. The count tables are then updated once again (Figure B-7). Note that the cell corresponding to the two-person, rental households is increased by 1 and the number of males is increased by 2 reflecting the household added to the population. The iterations continue in the same way. In iteration 6 (Figure B-10), household 2 (single female living in a rental house) in the seed data is added to the synthetic population. In the subsequent iterations, the fitness value for this household is negative. This is because, there is only one 1-person rental household required for the census tract (see the control tables, T) and this has already been achieved in iteration 6 by adding household 2 to the population. Thus, adding any more of household 2 would violate the control target and this is reflected by the negative value of the fitness for this household in the subsequent iterations. In iteration 11 (Figure B-15), we find that the fitness values for all the households are negative and all the control targets have been achieved. In fact, in this case, the count tables have exactly the same values as the control tables (See T and CT tables in Figure B-15) indicating a perfect fit of the synthesized population to the controls. When the number of households is large and there are several control tables, such a perfect fit may not be possible. The algorithm terminates when all households in the seed data have negative fitness values. The synthetic population (17 persons from 10 households) generated is presented in Figure B-16.

Iteration 1 (n=1)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	0
Rent	0	0

CT^{n-1}_{2k}

	Total
Male	0
Female	0

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

R^{n-1}_{2k}

	Total
Male	11
Female	6

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	1.000	5.000
Rent	2.000	2.000

Sum: 10.000

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	11.000
Female	6.000

Sum: 17.000

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	1.000	5.000
Rent	0.500	2.000

Sum: 8.500

i=2	Size = 1	Size = 2
Own	0.000	5.000
Rent	2.000	2.000

Sum: 9.000

i=3	Size = 1	Size = 2
Own	1.000	5.000
Rent	2.000	0.500

Sum: 8.500

i=4	Size = 1	Size = 2
Own	1.000	3.200
Rent	2.000	2.000

Sum: 8.200

i=5	Size = 1	Size = 2
Own	1.000	3.200
Rent	2.000	2.000

Sum: 8.200

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	11.000
Female	4.167

Sum: 15.167

i=2	Total
Male	9.091
Female	6.000

Sum: 15.091

i=3	Total
Male	7.364
Female	6.000

Sum: 13.364

i=4	Total
Male	9.091
Female	4.167

Sum: 13.258

i=5	Total
Male	7.364
Female	6.000

Sum: 13.364

F^{in}

3.333 = (10 - 8.5) + (17 - 15.167)

2.909 = (10 - 9) + (17 - 15.091)

3.318 = (10 - 8.5) + (1/2*(17 - 13.364))

3.671 = (10 - 8.2) + (1/2*(17 - 13.258))

3.618 = (10 - 8.2) + (1/2*(17 - 13.364))

HH ID (i) = 4 has the highest fitness and is slectected

Figure B-5 Iteration 1 of the Population Synthesis Procedure

Iteration = 2 (n=2)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	1
Rent	0	0

CT^{n-1}_{2k}

	Total
Male	1
Female	1

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	4
Rent	2	2

R^{n-1}_{2k}

	Total
Male	10
Female	5

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	1.000	3.200
Rent	2.000	2.000

Sum: 8.200

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	9.091
Female	4.167

Sum: 13.258

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	1.000	3.200
Rent	0.500	2.000

Sum: 6.700

i=2	Size = 1	Size = 2
Own	0.000	3.200
Rent	2.000	2.000

Sum: 7.200

i=3	Size = 1	Size = 2
Own	1.000	3.200
Rent	2.000	0.500

Sum: 6.700

i=4	Size = 1	Size = 2
Own	1.000	1.800
Rent	2.000	2.000

Sum: 6.800

i=5	Size = 1	Size = 2
Own	1.000	1.800
Rent	2.000	2.000

Sum: 6.800

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	9.091
Female	2.667

Sum: 11.758

i=2	Total
Male	7.364
Female	4.167

Sum: 11.530

i=3	Total
Male	5.818
Female	4.167

Sum: 9.985

i=4	Total
Male	7.364
Female	2.667

Sum: 10.030

i=5	Total
Male	5.818
Female	4.167

Sum: 9.985

F^{in}

F^{in}
3.000
2.727
3.136
3.014
3.036

HH ID (i) = 3 has the highest fitness and is slectected

Figure B-6 Iteration 2 of the Population Synthesis Procedure

Iteration = 3 (n=3)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	1
Rent	0	1

CT^{n-1}_{2k}

	Total
Male	3
Female	1

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	4
Rent	2	1

R^{n-1}_{2k}

	Total
Male	8
Female	5

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	1.000	3.200
Rent	2.000	0.500

Sum: 6.700

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	5.818
Female	4.167

Sum: 9.985

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	1.000	3.200
Rent	0.500	0.500

Sum: 5.200

i=2	Size = 1	Size = 2
Own	0.000	3.200
Rent	2.000	0.500

Sum: 5.700

i=3	Size = 1	Size = 2
Own	1.000	3.200
Rent	2.000	0.000

Sum: 6.200

i=4	Size = 1	Size = 2
Own	1.000	1.800
Rent	2.000	0.500

Sum: 5.300

i=5	Size = 1	Size = 2
Own	1.000	1.800
Rent	2.000	0.500

Sum: 5.300

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	5.818
Female	2.667

Sum: 8.485

i=2	Total
Male	4.455
Female	4.167

Sum: 8.621

i=3	Total
Male	3.273
Female	4.167

Sum: 7.439

i=4	Total
Male	4.455
Female	2.667

Sum: 7.121

i=5	Total
Male	3.273
Female	4.167

Sum: 7.439

F^{in}

3.000

2.364

1.773

2.832

2.673

HH ID (i) = 1 has the highest fitness and is slectected

Figure B-7 Iteration 3 of the Population Synthesis Procedure

Iteration = 4 (n=4)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	1
Rent	1	1

CT^{n-1}_{2k}

	Total
Male	3
Female	2

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	4
Rent	1	1

R^{n-1}_{2k}

	Total
Male	8
Female	4

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2	
Own	1.000	3.200	Sum:
Rent	0.500	0.500	5.200

$(R^{n-1}_{2k})^2/T_{2k}$

	Total	
Male	5.818	Sum:
Female	2.667	8.485

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2	
Own	1.000	3.200	Sum:
Rent	0.000	0.500	4.700

i=2	Size = 1	Size = 2	
Own	0.000	3.200	Sum:
Rent	0.500	0.500	4.200

i=3	Size = 1	Size = 2	
Own	1.000	3.200	Sum:
Rent	0.500	0.000	4.700

i=4	Size = 1	Size = 2	
Own	1.000	1.800	Sum:
Rent	0.500	0.500	3.800

i=5	Size = 1	Size = 2	
Own	1.000	1.800	Sum:
Rent	0.500	0.500	3.800

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total		F^{in}
Male	5.818	Sum:	
Female	1.500	7.318	1.667

i=2	Total		
Male	4.455	Sum:	
Female	2.667	7.121	2.364

i=3	Total		
Male	3.273	Sum:	
Female	2.667	5.939	1.773

i=4	Total		
Male	4.455	Sum:	
Female	1.500	5.955	2.665

i=5	Total		
Male	3.273	Sum:	
Female	2.667	5.939	2.673

HH ID (i) = 5 has the highest fitness and is slectected

Figure B-8 Iteration 4 of the Population Synthesis Procedure

Iteration = 5 (n=5)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	2
Rent	1	1

CT^{n-1}_{2k}

	Total
Male	5
Female	2

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	3
Rent	1	1

R^{n-1}_{2k}

	Total
Male	6
Female	4

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	1.000	1.800
Rent	0.500	0.500

Sum: 3.800

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	3.273
Female	2.667

Sum: 5.939

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	1.000	1.800
Rent	0.000	0.500

Sum: 3.300

i=2	Size = 1	Size = 2
Own	0.000	1.800
Rent	0.500	0.500

Sum: 2.800

i=3	Size = 1	Size = 2
Own	1.000	1.800
Rent	0.500	0.000

Sum: 3.300

i=4	Size = 1	Size = 2
Own	1.000	0.800
Rent	0.500	0.500

Sum: 2.800

i=5	Size = 1	Size = 2
Own	1.000	0.800
Rent	0.500	0.500

Sum: 2.800

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	3.273
Female	1.500

Sum: 4.773

i=2	Total
Male	2.273
Female	2.667

Sum: 4.939

i=3	Total
Male	1.455
Female	2.667

Sum: 4.121

i=4	Total
Male	2.273
Female	1.500

Sum: 3.773

i=5	Total
Male	1.455
Female	2.667

Sum: 4.121

F^{in}

1.667

2.000

1.409

2.083

1.909

HH ID (i) = 4 has the highest fitness and is slectected

Figure B-9 Iteration 5 of the Population Synthesis Procedure

Iteration = 6 (n=6)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	3
Rent	1	1

CT^{n-1}_{2k}

	Total
Male	6
Female	3

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	2
Rent	1	1

R^{n-1}_{2k}

	Total
Male	5
Female	3

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	1.000	0.800
Rent	0.500	0.500

Sum: 2.800

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	2.273
Female	1.500

Sum: 3.773

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	1.000	0.800
Rent	0.000	0.500

Sum: 2.300

i=2	Size = 1	Size = 2
Own	0.000	0.800
Rent	0.500	0.500

Sum: 1.800

i=3	Size = 1	Size = 2
Own	1.000	0.800
Rent	0.500	0.000

Sum: 2.300

i=4	Size = 1	Size = 2
Own	1.000	0.200
Rent	0.500	0.500

Sum: 2.200

i=5	Size = 1	Size = 2
Own	1.000	0.200
Rent	0.500	0.500

Sum: 2.200

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	2.273
Female	0.667

Sum: 2.939

i=2	Total
Male	1.455
Female	1.500

Sum: 2.955

i=3	Total
Male	0.818
Female	1.500

Sum: 2.318

i=4	Total
Male	1.455
Female	0.667

Sum: 2.121

i=5	Total
Male	0.818
Female	1.500

Sum: 2.318

F^{in}
1.333
1.818
1.227
1.426
1.327

HH ID (i) = 2 has the highest fitness and is slectected

Figure B-10 Iteration 6 of the Population Synthesis Procedure

Iteration = 7 (n=7)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	3
Rent	1	1

CT^{n-1}_{2k}

	Total
Male	7
Female	3

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	2
Rent	1	1

R^{n-1}_{2k}

	Total
Male	4
Female	3

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	0.000	0.800
Rent	0.500	0.500

Sum: 1.800

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	1.455
Female	1.500

Sum: 2.955

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	0.000	0.800
Rent	0.000	0.500

Sum: 1.300

i=2	Size = 1	Size = 2
Own	1.000	0.800
Rent	0.500	0.500

Sum: 2.800

i=3	Size = 1	Size = 2
Own	0.000	0.800
Rent	0.500	0.000

Sum: 1.300

i=4	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.500	0.500

Sum: 1.200

i=5	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.500	0.500

Sum: 1.200

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	1.455
Female	0.667

Sum: 2.121

i=2	Total
Male	0.818
Female	1.500

Sum: 2.318

i=3	Total
Male	0.364
Female	1.500

Sum: 1.864

i=4	Total
Male	0.818
Female	0.667

Sum: 1.485

i=5	Total
Male	0.364
Female	1.500

Sum: 1.864

	F^{in}
i=1	1.333
i=2	-0.364
i=3	1.045
i=4	1.335
i=5	1.145

HH ID (i) = 4 has the highest fitness and is slectected

Figure B-11 Iteration 7 of the Population Synthesis Procedure

Iteration = 8 (n=8)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	4
Rent	1	1

CT^{n-1}_{2k}

	Total
Male	8
Female	4

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	1
Rent	1	1

R^{n-1}_{2k}

	Total
Male	3
Female	2

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.500	0.500

Sum: 1.200

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	0.818
Female	0.667

Sum: 1.485

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.000	0.500

Sum: 0.700

i=2	Size = 1	Size = 2
Own	1.000	0.200
Rent	0.500	0.500

Sum: 2.200

i=3	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.500	0.000

Sum: 0.700

i=4	Size = 1	Size = 2
Own	0.000	0.000
Rent	0.500	0.500

Sum: 1.000

i=5	Size = 1	Size = 2
Own	0.000	0.000
Rent	0.500	0.500

Sum: 1.000

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	0.818
Female	0.167

Sum: 0.985

i=2	Total
Male	0.364
Female	0.667

Sum: 1.030

i=3	Total
Male	0.091
Female	0.667

Sum: 0.758

i=4	Total
Male	0.364
Female	0.167

Sum: 0.530

i=5	Total
Male	0.091
Female	0.667

Sum: 0.758

F^{in}

1.000

-0.545

0.864

0.677

0.564

HH ID (i) = 1 has the highest fitness and is slectected

Figure B-12 Iteration 8 of the Population Synthesis Procedure

Iteration = 9 (n=9)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	4
Rent	2	1

CT^{n-1}_{2k}

	Total
Male	8
Female	5

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	1
Rent	0	1

R^{n-1}_{2k}

	Total
Male	3
Female	1

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.000	0.500

Sum: 0.700

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	0.818
Female	0.167

Sum: 0.985

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.500	0.500

Sum: 1.200

i=2	Size = 1	Size = 2
Own	1.000	0.200
Rent	0.000	0.500

Sum: 1.700

i=3	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.000	0.000

Sum: 0.200

i=4	Size = 1	Size = 2
Own	0.000	0.000
Rent	0.000	0.500

Sum: 0.500

i=5	Size = 1	Size = 2
Own	0.000	0.000
Rent	0.000	0.500

Sum: 0.500

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	0.818
Female	0.000

Sum: 0.818

i=2	Total
Male	0.364
Female	0.167

Sum: 0.530

i=3	Total
Male	0.091
Female	0.167

Sum: 0.258

i=4	Total
Male	0.364
Female	0.000

Sum: 0.364

i=5	Total
Male	0.091
Female	0.167

Sum: 0.258

F^{in}

F^{in}
-0.333
-0.545
0.864
0.511
0.564

HH ID (i) = 3 has the highest fitness and is slectected

Figure B-13 Iteration 9 of the Population Synthesis Procedure

Iteration = 10 (n=10)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	4
Rent	2	2

CT^{n-1}_{2k}

	Total
Male	10
Female	5

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	1
Rent	0	0

R^{n-1}_{2k}

	Total
Male	1
Female	1

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.000	0.000

Sum: 0.200

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	0.091
Female	0.167

Sum: 0.258

HT^i_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^i_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.500	0.000

Sum: 0.700

i=2	Size = 1	Size = 2
Own	1.000	0.200
Rent	0.000	0.000

Sum: 1.200

i=3	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.000	0.500

Sum: 0.700

i=4	Size = 1	Size = 2
Own	0.000	0.000
Rent	0.000	0.000

Sum: 0.000

i=5	Size = 1	Size = 2
Own	0.000	0.000
Rent	0.000	0.000

Sum: 0.000

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	0.091
Female	0.000

Sum: 0.091

i=2	Total
Male	0.000
Female	0.167

Sum: 0.167

i=3	Total
Male	0.091
Female	0.167

Sum: 0.258

i=4	Total
Male	0.000
Female	0.000

Sum: 0.000

i=5	Total
Male	0.091
Female	0.167

Sum: 0.258

F^{in}
-0.333
-0.909
-0.500
0.329
0.200

HH ID (i) = 4 has the highest fitness and is slectected

Figure B-14 Iteration 10 of the Population Synthesis Procedure

Iteration = 11 (n=11)

T_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

T_{2k}

	Total
Male	11
Female	6

CT^{n-1}_{1k}

	Size = 1	Size = 2
Own	1	5
Rent	2	2

CT^{n-1}_{2k}

	Total
Male	11
Female	6

R^{n-1}_{1k}

	Size = 1	Size = 2
Own	0	0
Rent	0	0

R^{n-1}_{2k}

	Total
Male	0
Female	0

$(R^{n-1}_{1k})^2/T_{1k}$

	Size = 1	Size = 2
Own	0.000	0.000
Rent	0.000	0.000

Sum: 0.000

$(R^{n-1}_{2k})^2/T_{2k}$

	Total
Male	0.000
Female	0.000

Sum: 0.000

HT^{i}_{1k}

i=1	Size = 1	Size = 2
Own	0	0
Rent	1	0

i=2	Size = 1	Size = 2
Own	1	0
Rent	0	0

i=3	Size = 1	Size = 2
Own	0	0
Rent	0	1

i=4	Size = 1	Size = 2
Own	0	1
Rent	0	0

i=5	Size = 1	Size = 2
Own	0	1
Rent	0	0

HT^{i}_{2k}

i=1	Total
Male	0
Female	1

i=2	Total
Male	1
Female	0

i=3	Total
Male	2
Female	0

i=4	Total
Male	1
Female	1

i=5	Total
Male	2
Female	0

$(R^{n-1}_{1k}-HT^{n-1}_{1k})^2/T_{1k}$

i=1	Size = 1	Size = 2
Own	0.000	0.000
Rent	0.500	0.000

Sum: 0.500

i=2	Size = 1	Size = 2
Own	1.000	0.000
Rent	0.000	0.000

Sum: 1.000

i=3	Size = 1	Size = 2
Own	0.000	0.000
Rent	0.000	0.500

Sum: 0.500

i=4	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.000	0.000

Sum: 0.200

i=5	Size = 1	Size = 2
Own	0.000	0.200
Rent	0.000	0.000

Sum: 0.200

$(R^{n-1}_{2k}-HT^{n-1}_{2k})^2/T_{2k}$

i=1	Total
Male	0.000
Female	0.167

Sum: 0.167

i=2	Total
Male	0.091
Female	0.000

Sum: 0.091

i=3	Total
Male	0.364
Female	0.000

Sum: 0.364

i=4	Total
Male	0.091
Female	0.167

Sum: 0.258

i=5	Total
Male	0.364
Female	0.000

Sum: 0.364

F^{in}
-0.667
-1.091
-0.682
-0.329
-0.382

All households have neative fitness. Iterations complete

Figure B-15 Iteration 11 of the Population Synthesis Procedure

Syn. Pop HH ID	Seed Data HH ID (i)	Tenure	HH Size
1	4	Own	2
2	3	Rent	2
3	1	Rent	1
4	5	Own	2
5	4	Own	2
6	2	Own	1
7	4	Own	2
8	1	Rent	1
9	3	Rent	2
10	4	Own	2

Syn. Pop HH ID	Seed Data HH ID (i)	Person ID	Gender
1	4	1	Male
1	4	2	Female
2	3	1	Male
2	3	2	Male
3	1	1	Female
4	5	1	Male
4	5	2	Male
5	4	1	Male
5	4	2	Female
6	2	1	Male
7	4	1	Male
7	4	2	Female
8	1	1	Female
9	3	1	Male
9	3	2	Male
10	4	1	Male
10	4	2	Female

Figure B-16 Synthetic Population for the Census Tract