A Framework for Travel Demand Model Related Household and Person Level Control

[Draft Report]

Prepared by National Center for Smart Growth Research and Education

University of Maryland College Park

Authors

Sabyasachee Mishra

Xiaoyu Zhou

Timothy Welch

Frederick Ducca

July 2012

1. Introduction

Over the last few decades many parts of the world have seen rapid urbanization, growing urban boundaries and increasing congestion in cities. Answering various questions raised by urbanization with added demand poses a challenge to policymakers, planners and researchers. Adequate understanding of travel behavior and traveler demographics is a critical component in devising policies to tackle this problem. Currently, there is a trend of focus from macro-level to micro, disaggregate and activity-oriented level in travel behavior and travel demand modeling. The application of these studies in travel forecasting and land use policy requires more detailed population information on socioeconomic and demographic. Currently, synthesis methods, such as Iterative Proportional Fitting (IPF), are greatly used to generate a detailed socio-demographic characteristic of every resident household in the study area. There are limitations of controlled attributes used as input to these synthesis models: (1) The population projection models (e.g., Cohort-component method) to derive control attributes are commonly used for larger level of geography (county, state, national); (2) Limited set of variables, such as household size, income, race, number of workers, are currently predicted but not enough.

Meanwhile, there is a growing concern in small area (TAZ, community) population projections because it is highly related to community service, transportation level of service and other social wellness. The population size by socio-demographic in each TAZ or community is an important indicator to predict the trip generation and distribution, intra-zonal linkage and housing growth. Because of the limitations in current projection methods, there are several attempts to build a framework for small area population projection. Issues of lacking historical and current trend and developing reasonable migration assumptions are the critical problems.

In this project, we are facing the issue providing the supplemental input to PopGen-BMC, which is greatly used to generate a detailed socio-demographic characteristic of every resident household in the model area. PopGen-BMC synthesizes future year population based on observed base year data. Therefore, estimations of Socio-demographic characteristics that change over the time such as aging of population are less dependable. At present, limited set of variables (Number of Household by Size; Number of Household by Income; Number of Household by Worker and Total Population; Group Quarters Population) are used as controlled inputs to PopGen-BMC to generate other detail variables of interest.

Within this context, BMC desires to establish an aggregated sub-model that will allow estimating supplemental control variables required for PopGen-BMC such as housing type, householder age group, person by age group, worker type, and worker by occupation at Transportation Analysis Zone level (TAZ). Among the variables of interest, county level control estimates for population by age, sex and race and age of householder are available through Maryland Department of Planning. These county totals need to be allocated at TAZ level for PopGen-BMC input. These changing socio-demographic trends can be confirmed in the synthetic population estimates only if these changes are controlled as inputs to the PopGen-BMC.

Therefore, we seek for a population projection approach applicable to small areas (TAZ) considering historical and current trend. Population distribution by various household and personal socio-demographic characteristics will be estimated and forecasted, such as housing type, householder age group, person by age group, employment type, and worker by occupation.

2. Literature Review

The demographic and socioeconomic updating methods within the travel demand forecasting community and quantitative household and person level analysis and forecast are relatively limited (Miller, 2003). Traditional four-step modeling technique has been used by most of the planning agencies to forecast travel demand. Transition to a disaggregate model requires much more intensive data processes and faster computing abilities. For example, simulating the evolution of households and firms requires disaggregate data to estimate various life-cycle transition models. In the absence of disaggregate data, many practices have used growth factors or past experiences to forecast socio-economic data. In this section, different socio-economic and demographic evolution processes are outlined.

The popular approaches to forecast the demographic characteristics of future population are mostly used for the larger levels of geography, e.g., US Census Bureau uses the cohort-component method to produce the national and state population projections. Information of birth, death and migration are necessary in the forecast and the accuracy is relatively high. As the growing need in small area studies, researchers from various fields (social science, statistics, urban planning) have adapted various methods in small areas. Rees et al. (2004) discussed a framework for small area population estimation, which is constructed by four stages. Estimation methods, such as apportionment, ratio, IPF, Cohort-component and enhancements (hybrid method, district level constraints) are compared. Kanaroglou et al. (2007)

studied the spatial distribution of population at the census tract level using Cohort-component and aggregate spatial multinomial logit (ASMNL) model. A recent application of multinomial logistic model for Transportation Analysis Zone (TAZ) level population projection is proposed by Choi and Ryu (2011). Beyond the traditional methods, this is a new approach to forecast demographic distribution by capturing the historical and current trend.

Over the last few decades, a number of demographic and socioeconomic updating modules have been developed over multiple disciplines including DYNAMOD (King et al., 1999), DYNACAN (Morrison, 1998/Dussault, 2000), NEDYMAS (Nelissen, 1995), and LIFEPATHS (Gribble, 2000). These modules explicitly model demographic processes at a high level of detail. However, they are not well suited for applications in the context of an activity-based travel microsimulation system because generating the necessary land-use and transportation system characteristics with these models is not straightforward. Sundararajan and Goulias (2003) studied simulation of demographic evolution for the purposes of travel forecasting in a tool called as DEMOgraphic (Micro) Simulation (DEMOS) system. Other population updating systems have been developed in the travel demand forecasting community with varying levels of detail and sophistication, including the Micro-analytic Integrated Demographic Accounting System (MIDAS) proposed by Goulias and Kitamura (1996) and the Micro-Analytical Simulation of Transport Employment and Residences (MASTER) recommended by Mackett (1990). Certain aspects of the population evolution processes, such as residential relocations and automobile ownership are focused by land-use transportation modeling systems, including TRANUS (de la Barra, 1989), MEPLAN (Hunt, 1993), URBANSIM (Waddell, 2002), STEP2 (Caliper Corporation, 2003), ILUTE (Miller et al., 2004), PECAS (Hunt et al., 2010), and POPGEN (Pendyala et al., 2011).

Models of life-cycle transitions require special panel surveys to track changes in the demographics of a household. Since such surveys are rare, there have been very few models which track household evolution in great detail. MIDAS by Goulias and Kitamura's (1996) is one of such models, which combines models of travel behavior with a microsimulation model of household demographics. MIDAS was calibrated using the Dutch National Mobility Panel dataset. Another study of interest is STEP2 model for Nevada's Clark County (Caliper Corporation, 2003), which is closely mimicked by this studys' rules of household evolution.

In this study, the supplemental data needed for POPGEN is studied. IPF procedure used in POPFGEN only matches the control totals in the disaggregation process, but is blind to the temporal evolution. The

Table 1: Literature on Socioeconomic and Demographic Projection and Evolution Methods

Author	Year	Objective	Method	Data
Rees et al.	2004	A framework for small area population estimation	Framework: IPF, Cohort, combined	1991-1998 UK Yorkshire and Humber
Kanaroglou et al.	2007	Project the spatial distribution of population age and sex census tract	Rogers multi-regional population projection model (cohort) aggregate spatial multinomial logit (ASMNL) model	Hamilton Census Metroplitan Area (CMA), Canada 1996 2001
Choi & Ryu	2011	Disaggregate small area population in to demographic characteristics age, race TAZ level	Multinomial logistic	1990 2000 SF3 CTPP
Gouillas and Kitamura	1991	MIDAS	Micro-simulation	Dutch National Mobility Data
Mackett	1990	MASTER	Micro-simulation	Leeds
de la Barra	1989	TRANUS	Micro-simulation	Charlotte
Hunt	1993	MEPLAN	Micro-simulation	Sacremanto
Waddel	2002	Urbansim	Micro-simulation	Puget Sound
Caliper Corporation	2003	STEP2	Household microsimulation	Nevada’s Clark County
Miller	2004	ILUTE	Microsimulation	Toronto
Hunt	2010	PECAS	Spatial Economic Simulation	California
Pendyala	2011	POPGEN	Iterative Proportional Updating	SCAG and Baltimore

disadvantages of IPF are (1) only controls for household attributes but not personal attributes, (2) fails to synthesize populations to match distributions of target person characteristics, and (3) ignores differences in household composition among households within a TAZ (Pendyala and Konduri 2011). In the next section, methodology framework used to prepare supplemental data is discussed.

3. Methodology Framework and Forecasting Process

The modeling framework in this research is shown in the following flow chart. Threes steps: estimation, forecast and validation are proposed for this project. The methodology in each process is discussed in this section.

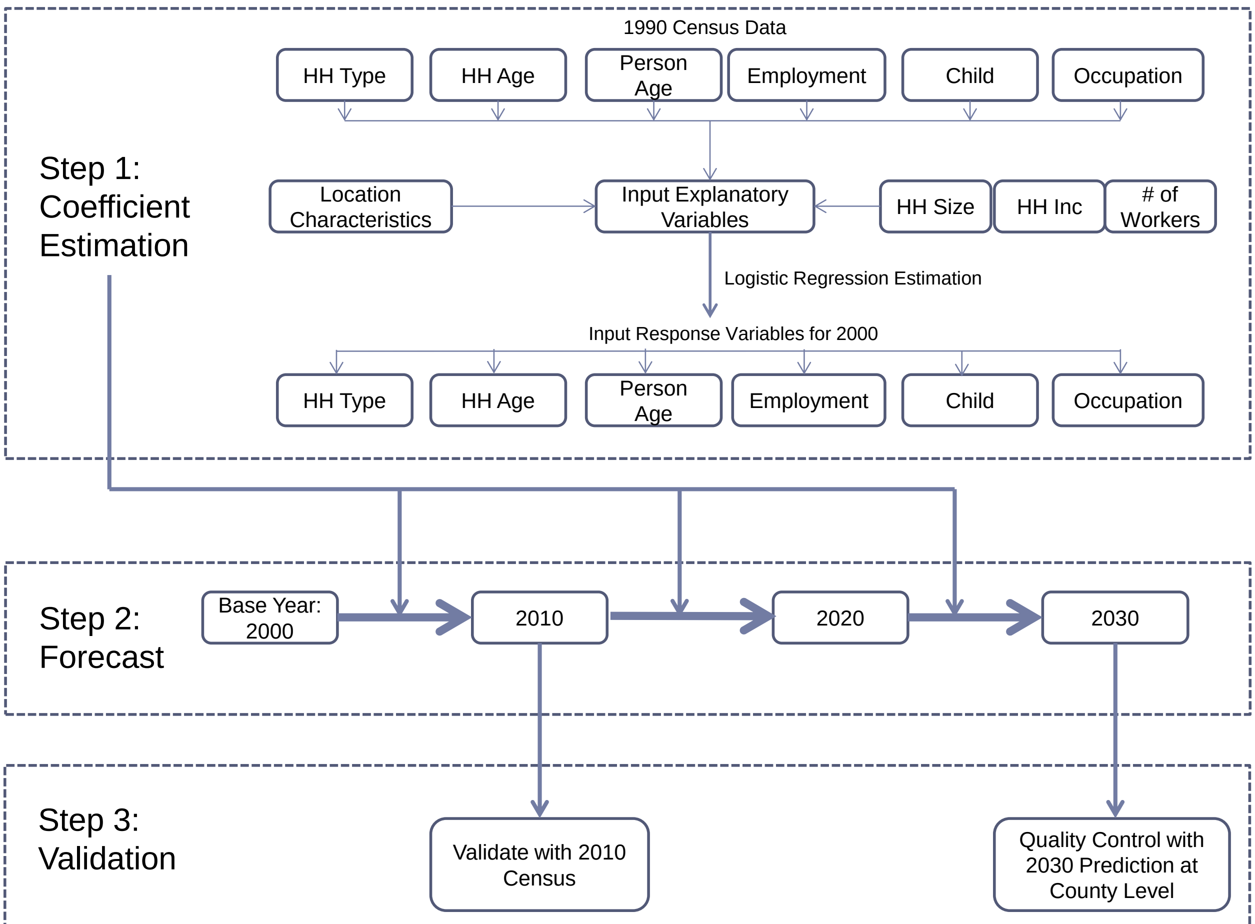

Figure 1. Flowchart of Project Process

3.1 Coefficient Estimation

In this step of coefficient estimation, we have 6 targets: Household type, householder's age, personal age, employment type, school child year and worker by occupation. Corresponding to each target, variables for these six targets can be grouped as major variables. All the other variables are secondary variables with the given major variable estimate, such as household size, income, workers, and zone characteristics. The methodology in this process is baseline-category logits model, one of the logistic regression models. It is also recognized as multinomial logistic regression in the study by Choi and Ryu (2011) to disaggregate the small area population in to demographic characteristics in Southern California Association of Governments (SCAG) region. To predict the future population distribution by various socio-economic and demographic in each zone, the population distribution data for two base years, say 1990 and 2000, in these zones are required. The impact of historical population (1990) to on the population ten years later (2000) is examined and the evolution trend is captured skipping the detailed birth, death and migration. The formulation is explained taking person by age group as an example.

Let probability of population in each age group defined as $\pi_j = P(Y = j), j = 1,2, \dots 8$. $j = 1$ for age 0-4; $j = 2$ for 5-14; $j = 3$ for 15-37; $j = 4$ for 18-24; $j = 5$ for 25-34; $j = 6$ for 35-44; $j = 7$ for 45-64; $j = 8$ for over 64. The age group $j = 8$ is chosen to be the baseline category.

$$\ln\left(\frac{\pi_{j\_00}}{\pi_{8\_00}}\right) = X\beta_j, \qquad j = 1, 2, \dots, 7.$$

$$\frac{\pi_{j\_00}}{\pi_{8\_00}} = \exp(X\beta_j), \qquad j = 1, 2, \dots, 7.$$

Where, π_{j_00} is the $(n \times 1)$ vector of probability of age group j in year 2000. n is the number of TAZs. X is the input explanatory variables, which contain the major variables 1990 population by age group, and secondary variables like total population. $\beta_j, j = 1, 2,3,7$ will be estimated in R. $\frac{\pi_{j\_00}}{\pi_{8\_00}}$ is the odds ratio of group j to group 8.

3.2 Forecast

The second step is using the estimation result $\widehat{\beta_j}, j = 1, 2,3,7$ from step 1 as the growth trend and 2000 census data as base year input X_{00} to forecast the population in 2030. The forecast is conducted as the following process by each decade. First, probability of 2010 population by each age group π_{j_10} will be calculated using 2000 as base year.

$$\pi_{j\_10} = \frac{\exp(X_{00}\widehat{\beta_J})}{1+\sum_{i=1}^{7}\exp(X_{00}\widehat{\beta_\iota})}, j = 1\ 2, \dots, 7$$

$$\pi_{8\_10} = \frac{1}{1+\sum_{i=1}^{7}\exp(X_{00}\widehat{\beta_i})}$$

Then the population by each group X_{10} could be calculated based on the estimated 2010 total population Pop_{10}in each TAZ by formulation $X_{10} = Pop_{10} \times \Pi_{10}, \ \Pi_{10} = [\pi_{1_{10}}, \pi_{2_{10}}, \dots, \pi_{8\_10}]$. Similarly to the above step, we will calculate the probability of 2020 population by each age group π_{j_20} using X_{10} as input.

$$\pi_{j\_20} = \frac{\exp(X_{10}\widehat{\beta_J})}{1+\sum_{i=1}^{7}\exp(X_{10}\widehat{\beta_\iota})}, j = 1\ 2, \dots, 7$$

$$\pi_{8\_20} = \frac{1}{1+\sum_{i=1}^{7}\exp(X_{10}\widehat{\beta_i})}$$

Repeatedly, $\pi_{j\_30}, j = 1\ 2, \dots, 8$ could be calculated and the target population by each age group X_{30} will be achieved.

3.3 Validation

The validation is designed at two stages. First, with the 2010 census at county and TAZ level, the 2010 forecast could be compared with the actual census outcome.

$$\pi_{j\_00} = \frac{\exp(X_{90}\widehat{\beta_J})}{1+\sum_{i=1}^{7}\exp(X_{90}\widehat{\beta_\iota})}, j = 1\ 2, \dots, 7$$

The second step is validating the final forecast of 2030, by aggregating the TAZ forecast to county and comparing with the county control for the demographic distribution. Mean absolute percentage error (MAPE) will be computed to test the fitness of prediction.

4. Data

There are four datasets retrieved for the study. The first group is for 1990 and the second for 2000. The 1990 data was collected from the census ftp site and included summary file 1 (SF1), which is 100% data from the short form census and summary file 3 (SF3), which is sample data from the long form census. The year 2000 data was collected from the same ftp site and consisted of summary files 1 and 3. The entire collection, allocation and aggregation process is shown in figure 2, with the retrieved data at the top of the figure for each census year and summary file. The mid-section of the figure describes the data formatting and the bottom of the figure shows how the data was either allocated or aggregated to TAZs depending on the type of summary file.

Limitations

The most disaggregate publically available 1990 long form census data in SF3 is at the block group level. The rest of the data (1990 SF1 and 2000 SF1 and SF3) is available at the block level. Further, the 1990 SF3 data has many fewer variables available than is available in 2000. The methodology used to handle data at the block group and block level is described in the following sections.

Formatting

The raw census data comes in flat ASCII files. The files are horizontal, that is, each row represents a single record at a desired geography and each column is a single variable. The raw data file is not comma delimited, has no geocoded data and does not have variable headings. Each of these attributes must be added to the formatted tables to make the census data useful.

Each summary file is broken down into several ASCII files.

The year 2000 census data consists of 39 tables for SF1 and 76 tables for SF3. The census provides a description of each table and variable in the raw data and a geocoded file that uses a common variable to merge the data with the geographic record.

To format each table and produce a final complete census record, a structure file is required, which allows the raw ASCII to be imported into Microsoft Access to be properly formatted into columns for each variable. The structure file was provided for year 2000 SF1 and SF3 files but was not readily available for 1990. A structure file was created for the 1990 files based on a data dictionary provided by census that gave the fixed variable lengths for each record. The formatted files were exported to DBF format so that the files could be merged with the graphic data.

Allocation

The SF1 files for both 1990 and 2000 are available at the block level, which nests very well into the BMC TAZ geography. However, the data is only available at the block group level for 1990, which does not always nest within TAZs. To make the SF3 data useful, each block group record had to be allocated to TAZs which in some cases were larger than block groups and in other cases smaller. To maintain consistency the block group data for both 1990 and 2000 was used.

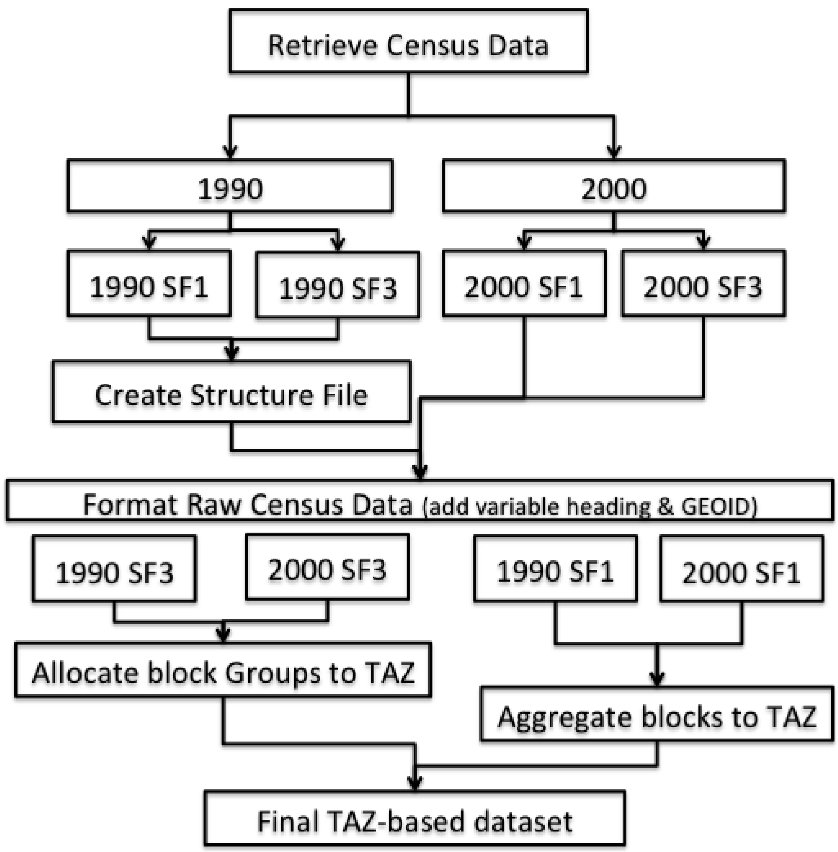

Figure 2. Census data collection and TAZ allocation Process

To properly allocate the block group data to TAZs, each census block group boundary file was imported into ARCGIS. The block group boundaries were overlaid on a 2010 TAZ shapefile. Each of the shapefiles was clipped to remove water and other non-developable features where census data likely did not exist. For the remaining area, in the absence of more detailed spatial data, it was assumed that population and households are evenly distributed across each block group. The ARCGIS intersect tool was used to calculate the area of each block group that overlapped a TAZ. The results were used to create a ratio for each block group to proportionately re-allocate each record to the TAZ. Once the ratios were established the 1990 and 2000 formatted census data was merged with the block group geographic data, with the ratios dividing the results by TAZ.

Aggregation

For the 1990 and 2000 SF1 data, the level of disaggregation was already at the block level, which nests well into TAZs. Each census block boundary file was imported into ARCGIS. The block boundaries were overlaid on a 2010 TAZ shapefile. The ARCGIS spatial join tool was used to attach the TAZ number that each block fit into. Once this relationship was established, the final block data was summed by TAZ.

5. Estimation Results

In this section, we present the model framework and result using population age group as an example. The estimation result is shown in Table 2 and the data description for the variables is displayed in Table 3.

Estimation sample is selected based on variable distributions, such as removing outliers and missing values. Also we firstly worked on the TAZs in six counties, but the validation did not fit well because the Baltimore City is quite various from others. Finally, the model is applied on five counties: Anna Arundel, Baltimore County, Carroll, Harford, and Howard. The sample size is 763 TAZs, with dependent variable by 8 age groups. The variables examined in this model correlated with age distribution in TAZs include the historical age distribution ten years ago, current median income, population density, employment density and group quarter density. We also examined variables, such as distribution of household size, income and number of workers. But these variables are proved to be not highly correlated with age distribution.

As in Table 2, most the coefficients are over 99% significant (shown in black) by examine the p-value and insignificant coefficients are shown in gray. Positive sign means the larger value in this row category is positively correlated with more population in the category by column, vice versa. We explain the result table using coefficient of independent variable “P_Age25_34” and dependent variable ”Age35_44_00”, which equals to 3.626 (highlight in grey) as an example. This coefficient is interpreted that if there is 1 percent increase in 25-34 age group in 1990, there would be a multiplicative effect by $\exp(3.626\ \times 1\%) =\ 1.037$ on odds of Age35_44 rather than odds of Age45_64 in 2000. This means a likely increasing in probability of population distributed in age 35-44 in 2000. Another example is the coefficient of -0.127 in row “HHDEN00” and col “Age25_34_00”. Odds ratio of Age25_34 to Age45_64 would decrease with higher household density. This indicates that comparing with 45-64 age group, younger (25-34) are less likely to leave in high density area. While positive or negative sign does not definitely imply the increase of decease in probability for a particular age group. The impact of one parameter on the probability of any age group is finally decided by all the coefficients in the same row with this parameter (refer from equation xx).

The following step before prediction and validation is the model evaluation, comparing the fitted value of model with the observed data in 2000.The observed population is plotted against the fitted population in 763 TAZs for each age group. The validation result is displayed in Figure 3. Most of the points are along the diagonal line with few unmatched points. The validation proves the model fits well and error is moderate. We also evaluate the model with a mean absolute percentage error (MAPE) of 15% and median absolute percentage error (MedAPE) of 10%.

Table 2. Estimation results for age group

	Age0_4	Age5_14	Age15_17	Age18_24	Age25_34	Age35_44	Age65
intersection	**-3.200**	**-1.769**	**-3.232**	**-4.082**	**-2.432**	**-1.494**	**1.533**
PAge0_4	**7.288**	**5.844**	**4.530**	**3.226**	**3.545**	**3.260**	**-2.764**
PAge5_14	**4.295**	**5.093**	**5.713**	**4.873**	**1.238**	**2.048**	**-1.531**
PAge15_17	-0.384	-0.389	**2.036**	**3.530**	**5.288**	**3.091**	**-5.831**
PAge18_24	**3.752**	**1.220**	**2.471**	**11.221**	**4.099**	**0.629**	**-3.686**
PAge25_34	**3.608**	**2.249**	**1.442**	**3.345**	**5.699**	**3.626**	**-1.945**
PAge35_44	**-3.761**	**-3.414**	**-1.869**	**0.976**	**-1.314**	**-2.351**	**-4.115**
PAge65	**1.838**	**1.206**	**2.058**	**4.060**	**2.171**	**1.573**	**1.433**
medinc (10K)	**0.050**	**0.039**	**0.014**	**-0.060**	**-0.038**	**0.031**	**-0.066**
HHDEN00	**-0.019**	-0.005	**-0.082**	**-0.059**	**-0.127**	**-0.067**	**0.096**
EMPDEN00	**0.030**	**0.014**	**0.063**	**0.070**	**0.132**	**0.052**	**-0.076**
GQDEN00	**-0.192**	**-0.158**	**-0.091**	-0.004	**0.065**	0.017	**0.339**

Table 3 Independent variable description in the age sample

	Label	mean	min	max
PAge0_4	Percentage of Age 0-4 in 1990	0.0742	0.0000	0.1667
PAge5_14	Percentage of Age 5-14 in 1990	0.1335	0.0000	0.2445
PAge15_17	Percentage of Age 15-17 in 1990	0.0375	0.0000	0.1962
PAge18_24	Percentage of Age 18-24 in 1990	0.0897	0.0000	0.2915
PAge25_34	Percentage of Age 25-34 in 1990	0.1793	0.0364	0.5000
PAge35_44	Percentage of Age 35-44 in 1990	0.1723	0.0000	0.3654
Page45_64	Percentage of Age 45-64 in 1990	0.2090	0.0523	0.4167
PAge65	Percentage of Age over 65 in 1990	0.1045	0.0000	0.5117
medinc (10K)	Median income in TAZ (in unit 10,000)	6.3966	1.1035	13.5460
HHDEN00	Household density in TAZ (per acre)	1.7364	0.0357	9.5340
EMPDEN00	Employment density (per acre)	2.2654	0.0464	9.7926
GQDEN00	Group quarter density (per acre)	0.0503	0.0000	2.4783

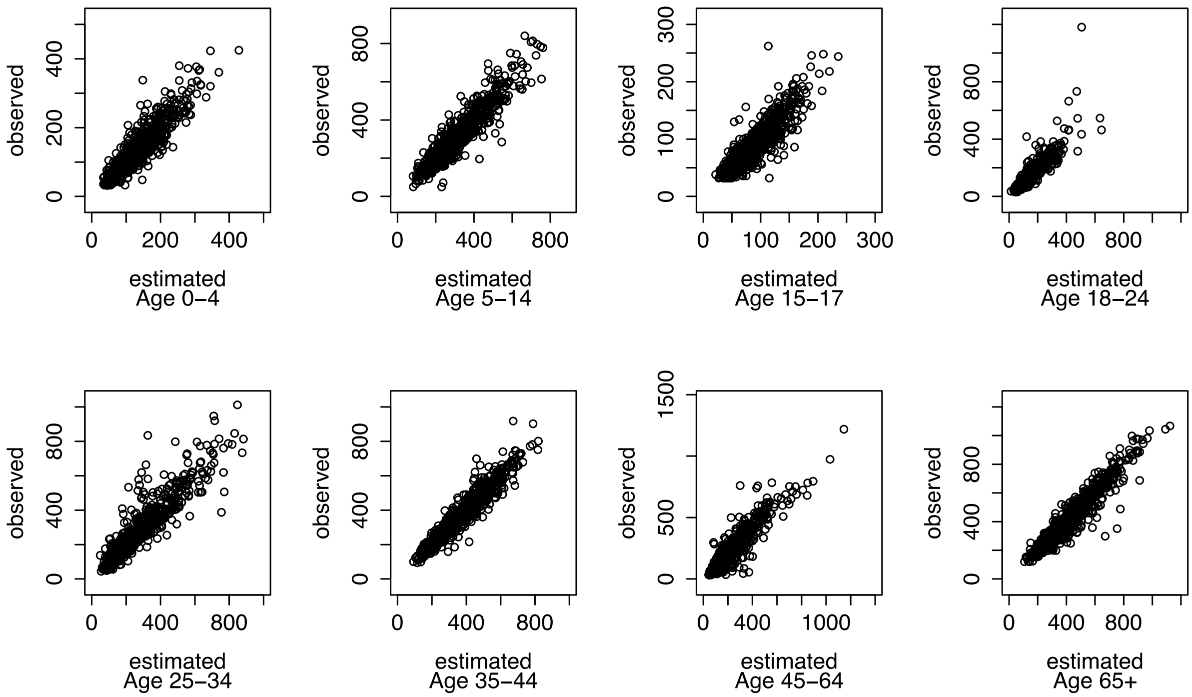

Figure 3. Validation plot of observed population against fitted population in 2000

Then we start the prediction and validation step. First is the prediction and the validation for 2010 before we predict the final age distribution in 2030. With the estimated coefficient, we could calculate the probability of population distribution by age group in 2010, using the observed population by age group in 2000. With approximated total population in each TAZ in 2010, we could calculate the number of population by age category in these TAZs. The prediction procedure is conducted on 1047 zones in 5 counties. The validation is at county level because the observed age distribution in 2010 at county level is easily to achieve and convenient to compare with the prediction at TAZ level. To validate the 2010 forecast, we aggregate the forecast by county and the result is shown in Table 4. The county level population by age group in 2010 is achieved from Maryland State Data Center and is used to examine the accuracy. The percentage error of the validation is shown in Table 5. The average error (MAPE) at county level is 10.2% and median error (MedAPE) is 6.2%. We observe a larger error in Age 0-4, 25-44, and over 65. Overall, the validation result is acceptable. We also display the error by age group in Figure 4. Predictions for Age 0-4 and35-44 are matched with observation quite well, with the points along the diagonal line. The percentage error for Age 0-4 in Carroll and Harford are above 10% because the population in this group is small. From the figure, we also observed an underestimation in age group 15-24, 25-34, and 45-64. In addition, age 5-14 and over 65 are overestimated.

Table 4. Estimated County level population by age group in 2010

County	Age0_4	Age5_14	Age15_17	Age18_24	Age25_34	Age35_44	Age45_64	65plus
Anna Arundel	32925	87364	25929	37356	58769	95090	66390	128765
Baltimore County	46822	122218	37365	65990	93122	135910	129054	185999
Carroll	10317	29839	9571	11991	16644	31281	21719	44161
Harford	15207	42376	13088	15372	25269	44054	31508	61827
Howard	18976	52662	15204	15024	29474	53870	28155	70195

Table 5. Percentage Error of estimated 2010 with Estimates by Gender and Age Group

County	Age0_4	Age5_14	Age15_24	Age25_34	Age35_44	Age45_64	65plus
Anna Arundel	8.5%	13.7%	2.4%	25.2%	1.0%	1.2%	25.0%
Baltimore County	4.4%	9.7%	0.9%	15.1%	1.7%	2.5%	8.1%
Carroll	12.2%	3.6%	6.2%	20.0%	4.9%	5.1%	14.8%
Harford	15.2%	5.4%	1.6%	22.5%	3.9%	5.0%	25.0%
Howard	9.1%	13.5%	0.5%	29.5%	3.2%	3.9%	33.2%

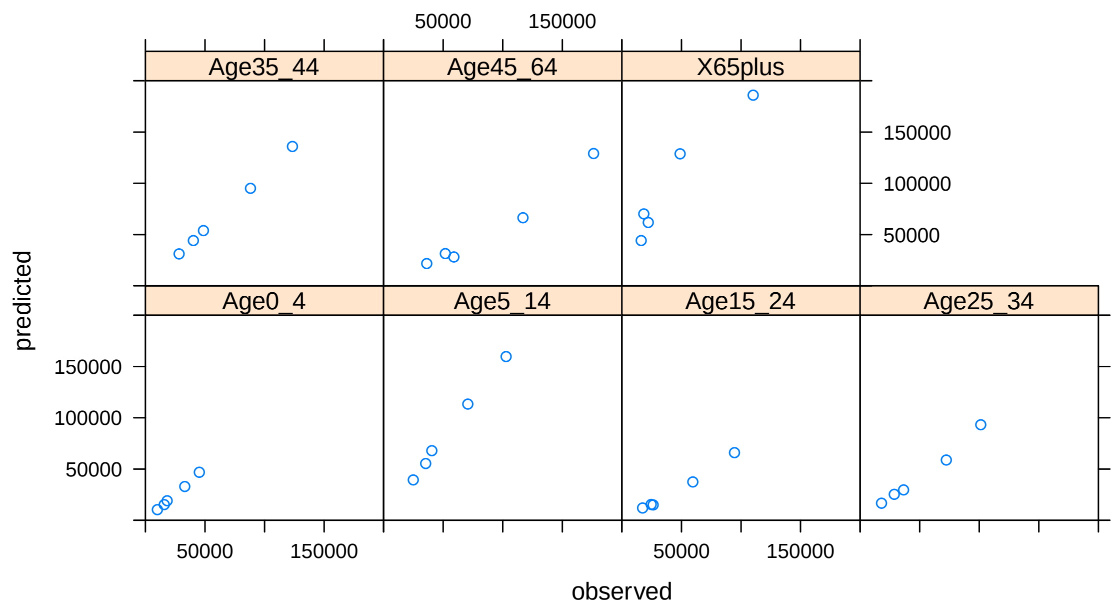

Figure 4. Validation plot of predicted population and observed population at county level by age group

After the above validation, we continue the forecast step shown in the framework and achieve the forecasting result in 2030. The approximate total populations for each TAZ in 2030 are provided by BMC. The aggregated county level population by age group is presented in Table 6. Table 7 is a comparison of county level prediction of 2030 with demographic projection provided by Maryland State Data Center. The age category provided by the projection in 0-4, 5-19, 20-44, 45-64 and over 65. We could not compare exactly using our prediction at age category shown in Table 6. The second column result in Table 7 is comparing with predicted age 5-17 with the observed age 5-19. The prediction in our model is more than the current projection. Generally, we obtained that our model has an underestimation comparing with the projection data.

Table 6. Estimated County level population by age group in 2030

County	Age0_4	Age5_14	Age15_17	Age18_24	Age25_34	Age35_44	Age45_64	65plus
Anna Arundel	37041	97821	33488	49993	56169	104099	125462	69953
Baltimore County	51171	131340	44414	83906	92933	144533	181729	132110
Carroll	12804	34624	12492	17762	19407	38415	47133	24674
Harford	17847	47749	16901	22962	28070	52122	64761	37420
Howard	23392	64162	21472	22951	28924	62553	70429	33573

Table 7. Compare with 2030 projection

County	Age0_4	Age5-19	Age20-44	Age5-44	Age45-64	65+
Anna Arundel	8%	29%	12%	18%	-8%	-39%
Baltimore County	5%	10%	17%	14%	-7%	-28%
Carroll	8%	21%	21%	21%	4%	-49%
Harford	1%	20%	13%	16%	0%	-38%
Howard	17%	41%	11%	22%	-8%	-50%

6. Conclusions

In conclusion, this paper provides a framework of forecasting future demographic and social economic distribution in small area (TAZ level). The framework is examined in forecasting age group and occupation. We present the modeling results, model evaluation, forecasting and validation of age group in this paper. The model evaluation and validation of prediction proves that the method is reasonable and the prediction is acceptable.

In this study, we also encounter many obstacles. The major problem is the accuracy of the data for estimation and prediction. For example, the TAZ changes from 1990 to 2010. To maintain consistency in estimation and prediction, we divide the population in 1990 TAZ to 2010 using the area size approximately. Also, we use the values such as population, income, household density of each TAZ in 2020 and 2030 in the prediction procedure, which could not be evaluated how accurate they are. Also we compare the final forecast in 2030 with projection in 2030 provided by MDP approximately. Besides, we wish to include variables corresponding to each TAZ, such as number of schools, university, recreation center, shopping center, which may relate to the population residential location choice. These variables are useful for scenario test, e.g., an increasing TAZ with more schools or business area, but not available currently.

In the future, we are going to work on other demographic focus except age and occupation, such as household type. Meanwhile, we will fulfill this framework and build up an applicable and deliverable production in R.

References

Choi, S., Ryu, S. Linking the Regional Demographic Process and the Small Area Housing Growth: Implications for the Small Area Demographic Projections. Presented at the 52nd Association of Collegiate Schools of Planning Conference, October 13-15, 2011.

Kanaroglou, P.S., Maoh, H.F., Newbold, K. B., Scott, D. M., Paez, A. A Demographic Model for Small Area Population Projections: Am Application to the Census Metroplitan Area (CMA) of Hamilton in Ontario, Canada. Working paper, 2007.

Rees, P., Norman, P., Brown, D. A framework for progressively improving small area population estimates. *Journal of The Royal Statistical Society*. Series A (Statistics In Society), 167 1: 5-36, 2004.

Sundararajan, A., and K. G. Goulias. 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, 2003.

Waddell, P. UrbanSim, Modeling Urban Development for Land Use, Transportation, and Environmental Planning. *Journal of the American Planning Association*, Vol. 68, 2002, pp. 297-314.

Miller, E. J. Microsimulation. In *Transportation Systems Planning: Methods and Applications*, Eds. K. G. Goulias, CRC Press, Boca Raton, Ch. 12, 2003.

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

Goulias, K. G., and R. Kitamura. 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, 1996, pp. 321-348.

Hunt, J. D. A Description of the MEPLAN Framework for Land Use and Transport Interaction Modeling. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1993.

Hunt, J.D. PECAS, University of California Land Use and Transportation Center, University of California, Davis.

Gabriel, S. A., and S. S. Rosenthal. 1989. Household Location and Race: Estimates of a Multinomial Logit Model. *The Review of Economics and Statistics* 71(2): 240-249.

Pendyala, R.M., K.P. Christian, and K.C. Konduri (2011) *PopGen 1.1 User's Guide*. Lulu Publishers, Raleigh, North Carolina.