ENHANCED SYNTHETIC POPULATION GENERATOR THAT ACCOMMODATES CONTROL VARIABLES AT MULTIPLE GEOGRAPHIC RESOLUTIONS

Karthik C. Konduri
University of Connecticut
Department of Civil and Environmental Engineering
261 Glenbrook Road, Unit 3037, Storrs, CT 06269-3037
Phone: 860-486-2733, Fax: 860-486-2298; Email: kkonduri@engr.uconn.edu

Daehyun You (corresponding author)
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 480-626-3716; Fax: 404-894-2278; Email: daehyun.you@gatech.edu

Venu M. Garikapati
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 480-522-8067; Fax: 404-894-2278; Email: venu.garikapati@gatech.edu

Ram M. Pendyala
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Tel: 404-894-2201; Fax: 404-894-2278; Email: ram.pendyala@ce.gatech.edu

Submitted for Final Publication

Paper Number 16-6639
Abstract: 241 words
Word count: 6,075 text + 6 tables/figures x 250 = 7,575 words

95th Annual Meeting of the Transportation Research Board
Committee on Transportation Planning Applications (ADB50)

Revised March 2016

ABSTRACT

Microsimulation models that simulate travel demand at the level of individual travelers have been gaining increasing interest among practitioners. Transportation planning agencies across the country are steadily migrating to activity-based microsimulation models which provide considerable flexibility in testing policy scenarios. Generating a synthetic population is the first step in the application of any activity-based model system and hence has been a topic of extensive research in the activity-based modeling arena. Several researchers have developed population synthesizers that are able to generate synthetic populations while matching household- and person-level constraints at a specified geographical resolution, e.g., census block group. However, information regarding control variables may not always be available at the specified spatial resolution. While information for some control variables may be available at the specified resolution, information on other control variables may be available only at a more aggregate spatial resolution. Ignoring control variables at different levels of spatial resolution could result in the generation of a synthetic population that is not representative of the underlying population. However, there has been limited progress on the development of synthetic population generators that are capable of accommodating control variables at multiple spatial resolutions. This paper proposes a robust approach to control for constraints at multiple geographic resolutions in generating a synthetic population. The methodology is an extension of the Iterative Proportional Updating (IPU) algorithm previously proposed and implemented by the authors. A case study demonstrating the efficacy of the enhanced algorithm is presented.

Keywords: activity-based model, synthetic population generator, population synthesis, iterative proportional fitting, iterative proportional updating, sample weight

1. INTRODUCTION

Activity-based microsimulation model systems (ABMs) are increasingly emerging as the next generation of travel forecasting models. The limitations of the four-step travel demand modeling process are well-documented (*1*). For the past two decades, many researches have been pursuing the development of ABMs that are built on sound behavioral foundations and represent travel in a realistic fashion. ABMs built on varying modeling paradigms have been introduced in the recent past. These include but are not limited to: DaySim (*2*); FAMOS (*3*); CEMDAP (*4*); ALBATROSS (*5*); TASHA (*6*); and CT-RAMP (*7*).

ABMs require household (e.g., household size, number of vehicles, and number of workers) and person-level (e.g., age, gender, and race) attributes for the entire population in a model region to be applied in a forecasting or application mode. Such disaggregate information for the entire population of a region is not publicly available and is very difficult to obtain. However, disaggregate data may be available for a random sample of households in the region, along with aggregate household and person-level frequency distributions of several attributes of interest (at various spatial resolutions) through census databases. These databases provide a basis to generate a synthetic population with attributes of interest; the synthetic population serves as input to an ABM capable of predicting travel demand at the level of the individual traveler.

Since the inception of ABMs, a number of synthetic population generators have been developed. The iterative Proportional Fitting (IPF) procedure (*8*-*10*) is at the heart of many of these synthesizers. Synthetic population generators generally employ the IPF procedure to estimate joint distributions of household and/or person attributes based on known univariate control distributions (*11*). Guo and Bhat (*12*) extended the IPF procedure to enhance the fit to person-level attributes. Arentze et al. (*13*) proposed an algorithm in which person-level marginal constraints are converted into household-level constraints using relational matrices. The standard IPF-based procedure is then applied to estimate household joint distributions. Pritchard and Miller (*14*) implemented the IPF procedure with a sparse list-based data structure that can accommodate a large number of control variables. Auld and Mohammadian (*15*) proposed a novel approach that accounts for person-level constraints to compute the selection probability for a sample household after applying the IPF procedure to estimate the household joint distribution. The synthesizer developed by Barthelemy and Toint (*16*) does not use the IPF procedure. Instead, a pool of individuals is generated and households are formed using individuals from the pool to satisfy control distributions. Müller and Axhausen (*17*, *18*) presented a novel algorithm labeled the 'Hierarchical IPF' to compute sample expansion factors at the person level and compare its performance with two similar algorithms (*19*-*20*). Vovsha et al. (*21*) presented an enhanced entropy maximization approach that can accommodate household/person-level controls at multiple spatial resolutions. Lee and Fu (*22*) applied a cross-entropy optimization model to generate synthetic populations by generalizing constraints of variables of interest at different levels (household/person). Abraham et al (*23*) employed a combinatorial optimization algorithm to match controls at both household and person-levels while accounting for constraints at multiple spatial resolutions. Farooq et al (*24*) adopted Markov Chain Monte Carlo processes to retrieve the underlying joint distribution of the population, from which a realization of the synthetic population can be generated. Ma and Srinivasan (*25*) introduced the Fitness-Based-Synthesis (FBS) methodology in which households are iteratively selected with replacement from sample data until the control totals are matched. Casati et al. (*26*) developed the Hierarchical Markov Chain Monte Carlo (MCMC) procedure in order to perfectly match a synthetic population against known marginal control totals at the household and person levels.

This paper extends the work of Ye et al. (*19*), who introduced a heuristic iterative procedure called the Iterative Proportional Updating (IPU) algorithm that can generate a synthetic population while matching both household and person-level distributions closely. Sample weights are estimated such that both household and person constraints are matched at a specified geographical resolution. Despite some recent progress (e.g., *21*, *23*), the IPU algorithm and many of the other algorithms noted previously have not been able to accommodate household and person-level attributes of interest at multiple geographical resolutions simultaneously. This shortcoming may lead to a potential mismatch between the synthetic population and true population on known characteristics of interest. The primary objective of this paper is to present an enhanced IPU algorithm that is able to accommodate constraints at multiple spatial resolutions.

The remainder of the paper is organized as follows. The next section presents an overview of the synthetic population generator called PopGen, which is based on the algorithm described in Ye et al. (*19*). This is followed by a description of an enhanced approach to accommodate controls for attributes of interest simultaneously at multiple geographic resolutions. The fourth section presents a case study in which the enhanced algorithm is employed to generate a synthetic population for a model region. This section also presents a scenario analysis to illustrate the efficacy of the approach. The final section presents some concluding thoughts and directions for future research.

2. OVERVIEW OF POPGEN METHODOLOGY

The traditional IPF procedure is adopted by most synthetic population generators to estimate household and person-level joint distributions that satisfy the given marginal distributions of interest. This facilitates the estimation of sample weights and selection probabilities for drawing households and composing a synthetic population that closely matches the true population with respect to household-level marginal distributions. The synthetic population is essentially comprised of all persons in the chosen households. A critical shortcoming of this procedure is that attributes of interest are not controlled and matched at the person-level. The resulting synthetic population fails to closely match both household and person-level marginal distributions, except under extremely unrealistic conditions.

Ye et al. (*19*) introduced a heuristic iterative procedure known as the IPU (iterative proportional updating) algorithm to overcome the shortcoming of using only the household-level IPF procedure in population synthesis. This algorithm has been implemented in PopGen, an open source synthetic population generator. In PopGen, the IPF procedure is applied first to both household- and person-level control variables of interest to obtain the number of households and persons in each cell of the respective joint distributions. Appropriate rounding procedures are applied to obtain cell “constraints” that must be matched through the population synthesis process. The IPU algorithm computes weights for sample households such that household-level as well as person-level marginal distributions are matched as closely as possible. An illustration of the IPU procedure is presented in Ye et al. (*19*) and replicated here briefly for the sake of completeness (Table 1). In the IPU procedure, unit weights are initially assigned to all sample households depicted in a sparse matrix format (see the column labeled “Weights” in Table 1). The weighted sum is computed using the initial set of weights. Next, an adjustment factor for a household-type or person-type is computed by dividing the constraint by the corresponding weighted sum (e.g., 35/3=11.67 for household type 1). The first three households that belong to household type 1 receive a weight of 11.67 (see Table 1). The algorithm proceeds by continuously adjusting household weights to account for subsequent constraints. After adjusting sample household weights, the weighted sums are updated for all household-types and person-types. When all constraints have been considered once, a full iteration is said to have been completed. In Table 1, the column labeled "Weights 5" depicts the weights obtained at the end of the first full iteration.

The completion of each iteration is followed by a check of the goodness-of-fit. If the goodness-of-fit satisfies a user-specified tolerance, the IPU procedure is terminated. A deviation measure (δ_j) for each household-type or person-type is computed as:

$$\delta_j = \frac{|d_j - c_j|}{c_j} \tag{1}$$

where j denotes the constraint or population characteristic of interest (j = 1, 2,…, 5)

d_j represents the weighted sum of households for population characteristic j

c_j is the actual number of households or persons in the true population for characteristic j.

The average deviation value across all household/person type constraints is compared between successive iterations. If the absolute difference of average deviation values between two full iterations falls below a threshold value set by the analyst, the IPU procedure is terminated. For example, in Table 1, the absolute difference between average deviation measure values is 0.8173 after the first full iteration (average δ_b= 0.9127 and average δ_a= 0.0954).

After the IPU procedure, selection probabilities are computed for sample households based on the IPU-computed weights and Monte Carlo drawing procedures are employed to construct the synthetic population. Since the IPU procedure accounts for both household and person-level joint distributions, households of the same type (cell in the joint distribution) may have different selection probabilities. The number of households in the synthetic population should match the frequencies of households in the rounded joint distribution table for all household types. The unique aspect of the PopGen methodology is that it facilitates drawing households such that the number of persons of various types (in the synthetic population) closely matches the frequencies of persons in the rounded joint distribution table for all person types.

3. THE ENHANCED MULTI-RESOLUTION METHODOLOGY

Although the IPU procedure proposed by Ye et al. (*19*) effectively controls for both household and person-level attributes of interest, it is still constrained by its applicability to only one geographical resolution at a time. For example, if control distributions for a few variables of interest are available at the Traffic Analysis Zone (TAZ) level, and distributions of others are available only at the census tract level, the existing algorithm cannot be used to control variables of interest at both geographical resolutions simultaneously. The resulting synthetic population may not be as representative of the true population as it might have been had information available at both geographic resolutions been used. Inaccuracies in population representativeness will inevitably have adverse downstream impacts on forecasts obtained from activity-based microsimulation models that take the synthetic population as input. To address this issue, this paper proposes an enhanced IPU procedure that can control for variables of interest at multiple geographical resolutions simultaneously. The enhanced algorithm is explained in detail in this section. An illustrative example is provided for a region with two geographic units. Control distributions are available at both the region and geographic unit levels, where each unit has its own set of household and person-level marginal distributions to be matched.

3.1 Initialize Household Sample Weight

The enhanced algorithm begins by assigning an initial set of weights to all sample households in all geographic units. Unit weights are assigned to each sample household to start the sample weight estimation process, as shown in Table 2. The illustrative example corresponds to 8 sample households with control distributions for two geographical units. Separate marginal distributions are available at the region level and at the level of two geographic units. Person-level marginal distributions are assumed to be available only at the level of geographic units (see row labeled 'Constraint' in Table 2). Weighted sums for each household type (and person type) are populated in the row labeled 'weighted sum' and the goodness-of-fit deviation measure explained previously is populated in the row labeled "δ". The IPF procedure is run for the region as a whole and for individual geographic units to obtain constraints that need to be matched at various spatial resolutions.

Sample data for each geographic unit is furnished in the form of a frequency matrix. In Table 2, each row corresponds to a single household record and provides data describing the composition of the household. The entries in the household type columns of the frequency matrix include either 0 or 1, indicating whether the household belongs to the category in question. Columns pertaining to person type in the frequency matrix include entries to indicate the number of persons of each type in the household. There are two household types and three person types considered in this example with marginal distributions for variables provided at two levels (Region and Geo). Household types may be defined by variables such as household size, household income, or car ownership, and person types may be defined by variables such as age and gender. For ease of interpretation, 'Region' is considered a more aggregate 'upper' level spatial resolution (e.g., county), and Geo is the more disaggregate 'lower' level spatial resolution (e.g., census tract).

The deviation measure "δ" helps assess the match between the weighted sums and the constraints at the end of each iteration. The deviation value can be computed at both Region and Geo levels using Equation 1. At the Region level, the deviation measured should be computed considering all geographic units together. For example, in Table 2, household type 1 has a weighted sum of 4 at the Region level, as there are two households of that type in Geo 1 and two households in Geo 2. The Region level constraint for household type 1 is 86, resulting in an initial deviation measure of $|4 - 86| \div 86 = 0.953$. The weighted sums are said to perfectly satisfy the constraints when all of the δ values simultaneously approach zero. In the first iteration (Table 2, Panel A), the δ values are usually quite large as the initial weights are set arbitrarily to unity.

3.2 Adjust Household Sample Weights to Match Region Level Constraints

In this step, sample household weights for all geographic units in a Region are adjusted to match the marginal distributions at the Region level. The procedure consists of three sub-steps.

1. An adjustment factor for the first household type is computed as the Region level constraint divided by the corresponding weighted sum in all geographic units taken together. In the example shown in Table 2, the adjustment factor is $86 \div 4 = 21.5$ for households of type 1.
2. Weight values for the sample households that correspond to the household type under consideration are multiplied by the adjustment factor. Thus, the second and fourth household records in the sample now have weights of 21.5 (see the 'weight' column of Panel B in Table 2).
3. All weighted sum and deviation values are updated based on the new weights for all household and person types at both Region and Geo levels. In Table 2 (Panel B),

multiplying the column ‘weight’ with the column corresponding to household type 1 yields a weighted sum of 86 (matching the constraint perfectly). The corresponding weighted sum for household type 1 is 45.33 at the level of Geo 1 and Geo 2 (resulting in large deviations from constraint values in the respective geographic units).

Steps 2 and 3 are repeated for each household type column at the Region level. Panel B of Table 2 shows the results at the end of the first full iteration at the Region level (the weight computation procedure is run thrice within the first full iteration, once for each of the three distinct household types at the Region level). It can be observed that the δ values for all household types at the region level are zero as the weighted sums match the Region level constraints perfectly. However, the δ values at the disaggregate Geo level are not close to zero.

3.3 Adjust Household Sample Weights to Match Constraints for Each Geographic Unit

The objective of this step is to satisfy the household type and person type constraints at a finer spatial resolution by adjusting sample household weights within each geographic unit (Geo). To achieve this, the sample weighting process is applied separately to each geographic unit. First, an adjustment factor for the first household type in a geographic unit (say Geo 1) is computed as the corresponding constraint divided by the weighted sum. For example, the adjustment factor for household type 1 in Geo 1 is $46 \div 45.33 = 1.0147$ (Table 2, Panel B). Second, weight values for the sample households that belong to household type 1 are adjusted by multiplying the current weight with the adjustment factor. This is shown in Table 3 (Panel A, first row) where the weight for the first sample household is adjusted as $13.67 \times 1.0147 = 13.87$. This process is repeated for all household and person types in the geographic unit. Weighted sums and corresponding deviation values are updated (based on the new weights) for the geographic unit under consideration. This procedure is carried out for all geographic units within a Region to complete one full iteration of the enhanced algorithm. The weighted sum and deviation values at the Region level are also updated at the end of each adjustment (last three rows in Panel A of Table 3).

One complete set of adjustments of weights at the Region and Geo levels comprises an iteration of the enhanced IPU procedure. After the first iteration, there is an improvement in the match between weighted sums and constraints, but some differences persist. The entire Region and Geo level adjustment process is repeated and the weights are iteratively adjusted until there is no further improvement in the match with respect to the different constraints. As iterations progress, the average δ value approaches zero indicating that the sample weights are converging, with weighted sums for all household and person types matching the geographic unit level constraints.

The enhanced IPU algorithm is an iterative procedure that is terminated when the improvement in the average δ value drops below a user-specified threshold. If all of the constraints are consistent across geographic levels, then the solution should result in a perfect match between weighted sums and constraints. On the other hand, if there are inconsistencies in marginal distributions across the geographic levels, then the solution is likely to result in a perfect match for some constraints and only a close match for others. Thus consistency of input data (across geographic levels) is of considerable importance in population synthesis that accommodates control variables at multiple spatial resolutions.

The solution after 1000 complete iterations are shown in Panel B of Table 3 for the illustrative example. It can be seen that household type constraints are perfectly matched at the Region level. At the disaggregate geographical unit level (Geo), it can be seen that the algorithm matches both household and person type constraints quite closely (with δ values close to zero).

4. CASE STUDY

To test the efficacy of the enhanced IPU algorithm, a case study is carried out where a synthetic population is generated for a model area while controlling for household and person type marginal distributions at both the county (Region) and TAZ (traffic analysis zone serving as Geo) levels. In addition, the case study demonstrates the value of using additional controls at more aggregate spatial levels in generating a synthetic population. First, the model area, the input data, and the population synthesis setup is described. This is followed by comparisons of goodness-of-fit of the synthetic population and performance metrics for the estimated sample weights.

4.1 Description of the Model Region and Input Data

The model area for the case study is the planning region of Baltimore Metropolitan Council (BMC) that consists of ten counties (District of Columbia, Anne Arundel, Baltimore, Carroll, Frederick, Harford, Howard, Montgomery, Prince George's, and Baltimore City) across Maryland and the District of Columbia. Household marginal distributions were provided by BMC at the county and TAZ levels for household size, household income, and number of workers in the household. Marginal distributions for variables that were not available at the county level were derived by aggregating TAZ-level marginal distributions. The marginal distribution for 'age of household head' was available only at the county level. For groupquarters, marginal distributions were available for type of groupquarter at the TAZ level, and the total number of groupquarter units was available at the county level. Among person-level variables, the marginal distribution for employment status was available at the TAZ level, while the marginal distribution for person age was available at the county level. The total population for the model area is 5,416,563 persons (based on the employment status distribution) residing in 2,076,236 households (derived from the distribution of number of workers in the household), and 145,718 groupquarters (from the type of groupquarter unit distribution).

The 2008-2012 five-year American Community Survey (ACS) Public Use Microdata Sample (PUMS) data for Maryland and District of Columbia served as the sample data. The data included 123,027 household records, 8,912 groupquarter records, and 310,252 person records. The household records in the PUMS data are geocoded to a Public Use Microdata Area (PUMA). Because the population is being synthesized based on county and TAZ level control distributions, a geographic correspondence file mapping the three geographical entities (County ⇔ PUMA ⇔ TAZ) was developed. This case study considered four different scenarios to test the enhanced algorithm and the benefits gained through the inclusion of control variable distributions at multiple geographic resolutions. They are:

- *Scenario 1 – Only TAZ Level Controls*: This scenario is consistent with general practice where synthetic populations are generated at the TAZ level based on controls that are available at this level. In this case study, household-level TAZ controls include household size, household income, and number of workers and person-level TAZ controls include employment status.
- *Scenario 2 – All TAZ Level Controls + Householder Age Control at County Level*: This scenario is the same as Scenario 1, but includes an additional control at the county level for age of householder. Thus, this scenario entails accommodating controls at multiple geographic resolutions.
- *Scenario 3 – All TAZ Level Controls + Person Age Control at County Level*: This scenario is similar to Scenario 2, except that the additional control at the county level is 'person age'.

- *Scenario 4 – All TAZ Level Controls + Householder Age and Person Age Controls at County Level*: This is the comprehensive scenario where all available information on controls at both geographic levels is utilized.

4.2 Results of Synthetic Population Generation

Table 4 presents an aggregate comparison of the distributions of various household attributes in the synthetic population against corresponding distributions in the true population. The regional comparison of distributions of various attributes serves as an overall assessment of the synthetic population generation process and the efficacy of the multilevel enhanced IPU algorithm. Table 5 presents similar statistics for the person-level attributes. An examination of Table 4 shows that the fit of the synthetic population is excellent for household size, household income, and household worker count in Scenario 1. This is consistent with expectations because these three variables were controlled in the synthesis process in Scenario 1. Because householder age (control available at county level only) was not controlled, the deviation is considerably larger for this variable. In Scenario 2, where the householder age variable is controlled (at the county level) through the enhanced IPU algorithm, the fit is considerably improved without any compromise with respect to fit to TAZ level controls. In the third scenario, the fit to TAZ level control variables is excellent as expected, but the fit to householder age is rather poor – once again reflecting the difficulty in matching distributions of uncontrolled variables. In this instance, the fit to householder age is poor even when controlling for person age at the county level (in the third scenario). In other words, person age is not a sufficient substitute for householder age. In the fourth scenario, where all variables are controlled, the match between synthetic and actual population distributions is excellent for all variables. There is a slight compromise in Scenario 4 with respect to the fit to householder age (relative to Scenario 2), but this compromise must be viewed in the context of the vast improvement of fit obtained in matching the person age distribution in Scenario 4. In Table 5, it can be seen that the person age distribution matches the true population age distribution quite closely, suggesting that the addition of the person age control variable to the synthesis process provides a more representative synthetic population overall (despite the modest compromise with respect to householder age). The fit to employment status is quite good in all scenarios, consistent with the fact that employment status, which is a TAZ level control variable, is controlled in all scenarios. The fit to person age distribution is best in Scenarios 3 and 4 because it is included as a control variable in these scenarios. Scenario 4 offers a slightly worse fit relative to Scenario 3 for person age distribution, but the vastly improved fit to householder age (seen in Table 4) more than makes up for this modest compromise.

Table 6 presents a comparison of the performance of the synthetic population generation process across the four scenarios for the two variables for which controls are available solely at the county level. The comparison is performed for each county to obtain more disaggregate insights into the quality of the synthetic population generated in each scenario. The population synthesis was performed such that the household worker count was controlled last and hence controlled perfectly. The TAZ level controls showed a total household count of 1,801,191 for the region based on this control variable. Thus, the synthetic population generation process consistently generates 1,801,191 households in every scenario (because this control variable is used in every scenario). The total deviation and percent deviation values for householder age are therefore consistent across all scenarios, as the synthesis process generates the same number of households in each county regardless of scenario. However, the percent deviation across householder age categories differ substantially depending on whether or not householder age is controlled. In Scenarios 2 and 4, where householder age is controlled, the percent deviation across categories shows a much smaller range than in Scenarios 1 and 3. Controlling for householder age substantially improved the fit of the synthetic population with respect to this variable. Similarly, a comparison was also performed for person age. As the synthesis process yields a slightly different population count depending on the control variables used in each scenario, the total deviation and percent deviation will vary by county across scenarios. The percent deviation across person age categories shows a much smaller range in Scenarios 3 and 4, the two scenarios where this variable is controlled. The range of error in replicating the person age distribution is quite large in Scenarios 1 and 2 where person age is not controlled. It is clear that the synthetic population generated in Scenario 4 offers the best fit with respect to all of the available control variables, including those at the TAZ level and County level. Not including county level controls severely compromises the representativeness of the synthetic population; including such controls, on the other hand, greatly improves the representativeness of the synthetic population with respect to all variables at all levels with a very modest and virtually negligible compromise in goodness of fit that often comes with increasing the number of constraints that must be matched, and the consequent greater likelihood of inconsistency in input control distributions across variables both within and between geographic levels.

5. CONCLUSION

Activity-based microsimulation model (ABM) systems are being increasingly adopted to simulate activity-travel choices at the disaggregate level of individual travelers. ABMs require information at the level of the individual household and person for the entire population of a model region so that traveler behavior can be modeled at the level of the individual agent as a function of socio-economic, demographic, and built environment variables. However, such information is neither readily available nor easy to obtain. Synthetic population generators are therefore used to create a synthetic population that closely mirrors the actual population of a region with respect to known distributions on variables of interest. Synthetic population generators that use readily available sample data and marginal distributions (provided by the Census Bureau) for household-level and person-level attributes of interest have been developed and deployed in the past decade to generate synthetic populations.

With the exception of a few recent developments, virtually all of the synthetic population generators are able to control for variables at a single geographical resolution (say, traffic analysis zone or census block group). However, it is often difficult to obtain marginal distributions on all variables of interest at a single geographical resolution. Moreover, in some jurisdictions, different entities may be responsible for producing aggregate population forecasts at various geographic levels. For example, a metropolitan planning organization may produce population forecasts at the TAZ level, but a county government may produce forecasts at the county level (for the same or different set of variables). It may be desirable to ensure that the synthetic population adheres to population forecasts produced by multiple entities at different geographic levels. In addition, the ability to consider control variable distributions at multiple geographic levels would lead to a population synthesis process that utilizes full information (as opposed to information available solely at a single spatial resolution).

To overcome this shortcoming, this paper presents an enhanced methodology for population synthesis that extends the original iterative proportional updating (IPU) algorithm proposed by Ye et al. (*19*). The enhanced algorithm is able to accommodate controls at multiple geographic levels through an iterative process that alternates between adjusting sample household weights to match constraints at different levels. Through such an iterative alternating process, the extended IPU algorithm is able to control for different attributes of interest at multiple spatial resolutions simultaneously. Using this algorithm, it is possible to control for variables of interest for which marginal control distributions may only be available at a more aggregate geographic scale. Results from a case study in which the extended algorithm was applied to the model region of the Baltimore Metropolitan Council demonstrate the efficacy of the proposed approach. The representativeness of the synthetic population was substantially improved through the inclusion of additional control variables at multiple geographic levels. This research effort helps advance the development of synthetic population generators that can control for attributes of interest at multiple spatial resolutions simultaneously, and shows that it is better to control for variables at the resolution for which data is available than not controlling for them at all.

REFERENCES

1. Kitamura, R., C. Chen, R. M. Pendyala, and R. Narayanan. Micro-Simulation of Daily Activity-Travel Patterns for Travel Demand Forecasting. *Transportation*, Vol. 27, No. 1, 2000, pp. 25-51.
2. Bradley, M., J.L. Bowman, and B. Griesenbeck. SACSIM: An Applied Activity-Based Model System with Fine Level Spatial and Temporal Resolution. *Journal of Choice Modeling*, Vol. 3, No. 1, 2010, pp. 5-31.
3. Pendyala, R. M., R. Kitamura, A. Kikuchi, T. Yamamoto, and S. Fujii. Florida Activity Mobility Simulator: Overview and Preliminary Validation Results. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1921*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 123-130.
4. Bhat, C.R., J. Y. Guo, S. Srinivasan, and A. Sivakumar. Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1894*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 57-66.
5. Arentze, T.A., and H.J.P. Timmermans. A Learning-Based Transportation Oriented Simulation System, *Transportation Research Part B: Methodological*, Vol. 38, No. 7, 2004, pp. 613-633.
6. Miller, E. J., and J. Roorda. A Prototype Model of Household Activity/ Travel Scheduling. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1831*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 114-121.
7. Vovsha, P., J. Freedman, V. Livshits, and W. Sun. Design Features of Activity-Based Models in Practice: Coordinated Travel-Regional Activity Modeling Platform. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2254*, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 19-27.
8. Deming, W.E., and F.F. Stephan. On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known. *Annals of Mathematical Statistics*, Vol. 11, 1940, pp. 427-444.
9. Fienberg, S.E. An Iterative Procedure for Estimation in Contingency Tables. *Annals of Mathematical Statistics*, Vol. 41, 1970, pp. 907-917.
10. Ireland, C.T., and S. Kullback. Contingency Tables with Given Marginals. *Biometrika*, Vol. 55, No. 1, 1968, pp. 179-188.

11. Beckman, R.J., K.A. Baggerly, and M.D. McKay. Creating Synthetic Baseline Populations. *Transportation Research Part A: Policy and Practice*, Vol. 30, No. 6, 1996, pp. 415-429.
12. Guo, J. Y., and C.R. Bhat. Population Synthesis for Microsimulating Travel Behavior. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2014*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 92-101.
13. Arentze T., H.J.P. Timmermans, and F. Hofman. Creating Synthetic Household Populations: Problem and Approach. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2014*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 85-91.
14. Pritchard, D. R., and E. J. Miller. Advances in Agent Population Synthesis and Application in an Integrated Land Use and Transportation Model. Presented at 88th Annual Meeting of Transportation Research Board, Washington, D.C., 2009.
15. Auld, J., and A. Mohammadian. Efficient Methodology for Generating Synthetic Populations with Multiple Control Levels. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2175*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 138-147.
16. Barthelemy, J., and P. L. Toint. Synthetic Population Generation without a Sample. *Transportation Science*, Vol. 47, No. 0, 2013, pp. 266-279.
17. Müller, K., and K. W. Axhausen. Population Synthesis for Microsimulation: State of the Art. ETH Zürich, Institut für Verkehrsplanung, Transporttechnik, Strassen-und Eisenbahnbau (IVT), 2010.
18. Müller, K., and K. W. Axhausen. Hierarchical IPF: Generating a Synthetic Population for Switzerland. Eidgenössische Technische Hochschule Zürich, IVT, 2011.
19. Ye, X., K. Konduri, R. M. Pendyala, B. Sana, and P. Waddell. A Methodology to Match Distributions of Both Household and Person Attributes in the Generation of Synthetic Populations. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
20. Bar-Gera, H., K. Konduri, B. Sana, X. Ye, and R. M. Pendyala. Estimating Survey Weights with Multiple Constraints Using Entropy Optimization Methods. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
21. Vovsha, P., J. E. Hicks, B. M. Paul, V. Livshits, P. Maneva, and, K. Jeon. New Features of Population Synthesis. 2014, Available at: http://docs.trb.org/prp/15-5343.pdf, Accessed July 29, 2015.
22. Lee, D. H., and Y. Fu. Cross-Entropy Optimization Model for Population Synthesis in Activity-based Microsimulation Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2255*, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 20-27.
23. Abraham, J. E., K. J. Stefan, and J. D. Hunt. Population Synthesis Using Combinatorial Optimization at Multiple Levels. Presented at the 91st Annual Meeting of Transportation Research Board, Washington, D.C., 2012.
24. Farooq, B., M. Bierlaire, R. Hurtubia, and G. Flötteröd. Simulation Based Population Synthesis. *Transportation Research Part B: Methodological*, Vol. 58, 2013, pp. 243-263.
25. Ma, L., and S. Srinivasan. Synthetic Population Generation with Multilevel Controls: A Fitness-Based Synthesis Approach and Validations. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 30, No. 2, 2015, 135-150.

26. Casati, D., K. Müller, P. J. Fourie, A. Erath, and K. W. Axhausen. Synthetic Population Generation by Combining a Hierarchical, Simulation-Based Approach with Reweighting by Generalized Raking. Presented at 94th Annual Meeting of Transportation Research Board, Washington, D.C., 2015.

Table 1 Illustration of Iterative Proportional Updating (IPU) Algorithm

Household ID	Weights	Household Type 1	Household Type 2	Person Type 1	Person Type 2	Person Type 3	Weights 1	Weights 2	Weights 3	Weights 4	Weights 5	Final Weights
1	1	1	0	1	1	1	11.67	11.67	9.51	8.05	12.37	1.36
2	1	1	0	1	0	1	11.67	11.67	9.51	9.51	14.61	25.66
3	1	1	0	2	1	0	11.67	11.67	9.51	8.05	8.05	7.98
4	1	0	1	1	0	2	1.00	13.00	10.59	10.59	16.28	27.79
5	1	0	1	0	2	1	1.00	13.00	13.00	11.00	16.91	18.45
6	1	0	1	1	1	0	1.00	13.00	10.59	8.97	8.97	8.64
7	1	0	1	2	1	2	1.00	13.00	10.59	8.97	13.78	1.47
8	1	0	1	1	1	0	1.00	13.00	10.59	8.97	8.97	8.64
Weighted Sum		3.00	5.00	9.00	7.00	7.00						
Constraints		35.00	65.00	91.00	65.00	104.00						
δ_b		0.9143	0.9231	0.9011	0.8923	0.9327						
Weighted Sum 1		***35.00***	5.00	51.67	28.33	28.33						
Weighted Sum 2		35.00	***65.00***	111.67	88.33	88.33						
Weighted Sum 3		28.52	55.38	***91.00***	76.80	74.39						
Weighted Sum 4		25.60	48.50	80.11	***65.00***	67.68						
Weighted Sum 5		35.02	64.90	104.84	85.94	***104.00***						
δ_a		0.0006	0.0015	0.1521	0.3222	0.0000						
Final Weighted Sum		35.00	65.00	91.00	65.00	104.00						

Table 2 Initial Sample Household Weights and Weights After Adjustment for Region Level Constraints

		Panel A. Initial Household Sample Weight									Panel B. Adjusted Weights after Full Iteration at Region Level									
			Region HH Type			**HH Type**		**Person Type**				**Region HH Type**			**HH Type**		**Person Type**			
	hid	**weight**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**3**	**weight**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**3**	
First geographic unit (Geo 1)	1	1	0	0	1	1	0	1	1	1	13.67	0	0	1	1	0	1	1	1	
	2	1	1	0	0	1	0	1	0	1	21.50	1	0	0	1	0	1	0	1	
	3	1	0	1	0	1	0	2	1	0	10.17	0	1	0	1	0	2	1	0	
	4	1	1	0	0	0	1	1	0	2	21.50	1	0	0	0	1	1	0	2	
	5	1	0	1	0	0	1	0	2	1	10.17	0	1	0	0	1	0	2	1	
	6	1	0	0	1	0	1	1	1	0	13.67	0	0	1	0	1	1	1	0	
	7	1	0	1	0	0	1	2	1	2	10.17	0	1	0	0	1	2	1	2	
	8	1	0	0	1	0	1	1	2	0	13.67	0	0	1	0	1	1	2	0	
		Geo 1		Weighted Sum		3	5	9	8	7	Geo 1		Weighted Sum		45.33	69.17	124.67	95.33	108.67	
				Constraint		46	51	92	88	84			Constraint		46	51	92	88	84	
				δ		0.94	0.90	0.90	0.91	0.92			δ		0.01	0.36	0.36	0.08	0.29	
			Region HH Type			**HH Type**		**Person Type**				**Region HH Type**			**HH Type**		**Person Type**			
	hid	**weight**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**3**	**weight**	**1**	**2**	**3**	**1**	**2**	**1**	**2**	**3**	
Second geographic unit (Geo 2)	1	1	0	0	1	1	0	1	1	1	13.67	0	0	1	1	0	1	1	1	
	2	1	1	0	0	1	0	1	0	1	21.50	1	0	0	1	0	1	0	1	
	3	1	0	1	0	1	0	2	1	0	10.17	0	1	0	1	0	2	1	0	
	4	1	1	0	0	0	1	1	0	2	21.50	1	0	0	0	1	1	0	2	
	5	1	0	1	0	0	1	0	2	1	10.17	0	1	0	0	1	0	2	1	
	6	1	0	0	1	0	1	1	1	0	13.67	0	0	1	0	1	1	1	0	
	7	1	0	1	0	0	1	2	1	2	10.17	0	1	0	0	1	2	1	2	
	8	1	0	0	1	0	1	1	2	0	13.67	0	0	1	0	1	1	2	0	
		Geo 2		Weighted Sum		3	5	9	8	7	Geo 2		Weighted Sum		45.33	69.17	124.67	95.33	108.67	
				Constraint		33	99	138	122	104			Constraint		33	99	138	122	104	
				δ		0.91	0.95	0.94	0.93	0.93			δ		0.37	0.30	0.10	0.22	0.05	
Region		Weighted Sum	4.0	6.0	6.0							86.0	61.0	82.0						
		Constraint	86.0	61.0	82.0							86.0	61.0	82.0						
		δ	0.953	0.902	0.927							0.000	0.000	0.000						

Table 3 Sample Household Weights After Adjusting for One Control at the Geographic Unit Level (Geo 1) and After 1000 Complete Iterations of the Enhanced IPU Algorithm

Panel A. Result after Controlling for HH Type 1 in Geo Iteration 1

	hid	**weight**	**Region HH Type 1**	**Region HH Type 2**	**Region HH Type 3**	**HH Type 1**	**HH Type 2**	**Person Type 1**	**Person Type 2**	**Person Type 3**
First geographic unit (Geo 1)	1	13.87	0	0	1	1	0	1	1	1
	2	21.82	1	0	0	1	0	1	0	1
	3	10.32	0	1	0	1	0	2	1	0
	4	21.50	1	0	0	0	1	1	0	2
	5	10.17	0	1	0	0	1	0	2	1
	6	13.67	0	0	1	0	1	1	1	0
	7	10.17	0	1	0	0	1	2	1	2
	8	13.67	0	0	1	0	1	1	2	0
		Geo 1			Weighted Sum	46.00	69.17	125.48	95.68	109.18
					Constraint	46	51	92	88	84
					δ	0.000	0.356	0.364	0.087	0.300
Second geographic unit (Geo 2)	1	13.67	0	0	1	1	0	1	1	1
	2	21.50	1	0	0	1	0	1	0	1
	3	10.17	0	1	0	1	0	2	1	0
	4	21.50	1	0	0	0	1	1	0	2
	5	10.17	0	1	0	0	1	0	2	1
	6	13.67	0	0	1	0	1	1	1	0
	7	10.17	0	1	0	0	1	2	1	2
	8	13.67	0	0	1	0	1	1	2	0
		Geo 2			Weighted Sum	45.33	69.17	125.67	95.33	108.67
					Constraint	33	99	138	122	104
					δ	0.374	0.301	0.097	0.219	0.045
Region		Weighted Sum	86.3	61.2	82.2					
		Constraint	86.0	61.0	82.0					
		δ	0.004	0.002	0.002					

Panel B. Result after 1000 Full Iterations

	hid	**weight**	**Region HH Type 1**	**Region HH Type 2**	**Region HH Type 3**	**HH Type 1**	**HH Type 2**	**Person Type 1**	**Person Type 2**	**Person Type 3**
First geographic unit (Geo 1)	1	8.33	0	0	1	1	0	1	1	1
	2	25.71	1	0	0	1	0	1	0	1
	3	12.19	0	1	0	1	0	2	1	0
	4	12.19	1	0	0	0	1	1	0	2
	5	20.02	0	1	0	0	1	0	2	1
	6	8.22	0	0	1	0	1	1	1	0
	7	2.78	0	1	0	0	1	2	1	2
	8	8.22	0	0	1	0	1	1	2	0
		Geo 1			Weighted Sum	46.23	51.43	92.60	88.00	84.00
					Constraint	46	51	92	88	84
					δ	0.005	0.009	0.007	0.000	0.000
Second geographic unit (Geo 2)	1	4.46	0	0	1	1	0	1	1	1
	2	17.71	1	0	0	1	0	1	0	1
	3	11.00	0	1	0	1	0	2	1	0
	4	30.39	1	0	0	0	1	1	0	2
	5	10.31	0	1	0	0	1	0	2	1
	6	26.85	0	0	1	0	1	1	1	0
	7	5.38	0	1	0	0	1	2	1	2
	8	26.85	0	0	1	0	1	1	2	0
		Geo 2			Weighted Sum	33.17	99.77	139.00	122.00	104.00
					Constraint	33	99	138	122	104
					δ	0.005	0.008	0.007	0.000	0.000
Region			86.0	61.7	82.9					
			86.0	61.0	82.0					
			0.000	0.011	0.011					

Table 4 Comparison of Household-Level Marginal Distributions for Entire Model Region

Variable Name	Category	Actual	% Difference: (Synthetic - Actual)/Actual			
			Scenario 1	Scenario 2	Scenario 3	Scenario 4
Household Size	1	476509 (26.46)	-0.29	-0.29	-0.29	-0.29
	2	539006 (29.93)	0.12	0.12	0.12	0.12
	3	322968 (17.93)	0.11	0.11	0.11	0.11
	4	259973 (14.43)	0.09	0.09	0.09	0.09
	5	202706 (11.25)	0.08	0.08	0.08	0.08
	Total	**1801162**	**1801191**	**1801191**	**1801191**	**1801191**
Household Income	<$15K	150344 (8.35)	-0.18	-0.18	-0.18	-0.18
	$15K to <$30K	165993 (9.22)	-0.10	-0.10	-0.10	-0.10
	$30K to <$50K	260220 (14.45)	-0.01	-0.01	-0.01	-0.01
	$50K to <$100K	610697 (33.91	0.01	0.01	0.01	0.01
	$100K or over	613879 (34.1)	0.08	0.08	0.08	0.08
	Total	**1801133**	**1801191**	**1801191**	**1801191**	**1801191**
Household Worker Count	0	348324 (19.34)	-0.07	-0.07	-0.07	-0.07
	1	647931 (35.97)	0.03	0.03	0.03	0.03
	2	649985 (36.09)	0.01	0.01	0.01	0.01
	3	154951 (8.60)	-0.01	-0.01	-0.01	-0.01
	Total	**1801191**	**1801191**	**1801191**	**1801191**	**1801191**
Householder Age (years)	<25	62960 (3.43)	-4.32	-1.84	42.76	-2.77
	25-34	289882 (15.77)	-10.41	-2.06	0.59	-2.29
	35-44	330245 (17.97)	8.61	-2.02	0.83	-2.36
	45-54	396840 (21.59)	5.36	-2.07	-1.53	-2.13
	55-64	360068 (19.59)	-5.90	-1.77	-7.01	-1.85
	65 or above	398121 (21.66)	-8.17	-2.14	-9.29	-1.42
	Total	**1838116**	**1801191**	**1801191**	**1801191**	**1801191**

Scenario 1: Only TAZ level controls
Scenario 2: TAZ level controls + 'Householder Age' control at the county level
Scenario 3: TAZ level controls + 'Person Age' control at the county level
Scenario 4: TAZ level controls + 'Householder Age' and 'Person Age' controls at the county level
Numbers in parentheses are percentage values

Table 5 Comparison of Person- and Groupquarter-level Marginal Distributions for Entire Model Region

Variable Name	Category	Given	% Difference: (Synthetic - Given)/Given			
			Scenario 1	Scenario 2	Scenario 3	Scenario 4
Employment Category	Employed	2451494 (51.14)	2.42	1.85	0.33	0.43
	Unemployed	2342486 (48.86)	-2.69	-1.87	-3.38	-3.54
	Total	**4793980**	**4790075**	**4795380**	**4722872**	**4721624**
Person Age	0-4	300100 (6.24)	-0.09	2.51	-2.74	-3.51
	5-9	300681 (6.25)	3.03	2.61	-2.80	-4.03
	10-14	310730 (6.46)	3.18	2.18	-2.73	-3.84
	15-19	325957 (6.78)	8.74	6.42	-2.31	-3.19
	20-24	333563 (6.93)	-5.89	-6.86	-1.80	-2.28
	25-29	345548 (7.18)	-11.70	-7.63	-2.12	-2.68
	30-34	324888 (6.75)	-11.57	-6.01	-1.60	-1.39
	35-39	314668 (6.54)	3.45	-2.83	-1.54	-2.72
	40-44	330160 (6.86)	9.60	2.54	-1.60	-2.17
	45-49	360197 (7.49)	7.92	3.15	-1.33	-2.30
	50-54	363570 (7.56)	-0.10	-4.36	-1.53	-2.09
	55-59	324387 (6.74)	-0.49	3.71	-1.17	0.51
	60-64	270434 (5.62)	-5.48	-1.60	-1.44	0.54
	65-69	201639 (4.19)	-10.47	-3.72	-1.74	1.01
	70-74	139244 (2.89)	-3.69	2.01	-1.09	1.19
	75-79	101765 (2.12)	6.38	11.40	-1.17	1.27
	80-84	77964 (1.62)	7.14	11.12	-0.81	1.55
	85 or above	84545 (1.76)	-11.33	-8.52	-1.21	0.26
	Total	**4810040**	**4790075**	**4795380**	**4722872**	**4721624**
Groupquarter Type	Institutional	43895 (42.00)	0.00	0.00	0.00	0.00
	Non-Institutional	60627 (58.00)	0.00	0.00	0.00	0.00
	N	**104522**	**104522**	**104522**	**104522**	**104522**

Scenario 1: Only TAZ level controls
Scenario 2: TAZ level controls + 'Householder Age' control at the county level
Scenario 3: TAZ level controls + 'Person Age' control at the county level
Scenario 4: TAZ level controls + 'Householder Age' and 'Person Age' controls at the county level
Numbers in parentheses are percentage values

Table 6 Comparison of Scenarios for County-level Control Variables

Scenario	County ID	Householder Age: Total Deviation	Householder Age: % Total Deviation	Householder Age: % Deviation Across Categories Min	Householder Age: % Deviation Across Categories Max	Person Age: Total Deviation	Person Age: % Total Deviation	Person Age: % Deviation Across Categories Min	Person Age: % Deviation Across Categories Max
1	3	-4745	-2.29	-10.08	9.50	3889	0.71	-9.89	17.05
	5	-6878	-2.11	-10.03	9.11	1051	0.13	-32.11	13.10
	13	1517	2.46	-12.68	58.81	1562	0.93	-14.38	30.04
	21	-3379	-3.79	-9.72	6.82	-4316	-1.81	-18.20	12.05
	25	-1831	-1.96	-11.14	20.36	1428	0.58	-10.62	10.35
	27	-3725	-3.34	-8.47	7.48	2069	0.70	-10.78	17.45
	31	-10492	-2.79	-16.94	9.71	-23834	-2.39	-31.51	11.74
	33	-4644	-1.46	-11.83	7.56	-9574	-1.09	-18.39	21.97
	510	-2748	-1.08	-31.05	15.87	7760	1.25	-30.50	41.08
2	3	-4745	-2.29	-3.06	-2.01	4308	0.79	-4.94	24.21
	5	-6878	-2.11	-2.31	-1.14	1935	0.24	-28.18	12.31
	13	1517	2.46	0.13	4.32	2058	1.23	-8.09	19.75
	21	-3379	-3.79	-4.41	-2.65	-4248	-1.78	-16.45	15.50
	25	-1831	-1.96	-2.48	-1.20	1807	0.73	-5.77	14.59
	27	-3725	-3.34	-4.28	-2.10	2166	0.73	-11.67	25.51
	31	-10492	-2.79	-3.70	-2.45	-23130	-2.32	-25.95	9.62
	33	-4644	-1.46	-1.77	-0.71	-8623	-0.98	-17.25	28.31
	510	-2748	-1.08	-1.85	-0.67	9067	1.46	-13.22	25.56
3	3	-4745	-2.29	-14.56	68.53	-2211	-0.40	-1.77	1.66
	5	-6878	-2.11	-8.67	43.22	-6943	-0.85	-1.74	0.14
	13	1517	2.46	-9.19	161.83	-161	-0.10	-3.08	1.72
	21	-3379	-3.79	-11.37	49.43	-7232	-3.04	-6.42	-1.33
	25	-1831	-1.96	-9.71	81.13	-1511	-0.61	-1.82	1.97
	27	-3725	-3.34	-12.39	76.92	-1204	-0.41	-3.13	1.37
	31	-10492	-2.79	-6.18	30.49	-40536	-4.06	-6.47	-1.64
	33	-4644	-1.46	-9.36	41.42	-26250	-2.99	-5.14	-1.11
	510	-2748	-1.08	-8.87	17.17	-1120	-0.18	-2.00	0.74
4	3	-4745	-2.29	-3.52	-1.37	-2512	-0.46	-3.38	5.83
	5	-6878	-2.11	-3.37	-1.32	-6951	-0.85	-3.34	2.47
	13	1517	2.46	1.33	4.37	-201	-0.12	-5.25	6.97
	21	-3379	-3.79	-4.91	-3.07	-7308	-3.07	-5.77	3.50
	25	-1831	-1.96	-4.58	-1.55	-1696	-0.68	-3.54	5.94
	27	-3725	-3.34	-5.04	-2.58	-1353	-0.46	-3.56	5.06
	31	-10492	-2.79	-3.64	-1.58	-40663	-4.08	-7.16	-0.86
	33	-4644	-1.46	-2.49	-0.12	-26418	-3.01	-6.01	2.12
	510	-2748	-1.08	-1.61	-0.48	-1314	-0.21	-1.45	3.28

Scenario 1: Only TAZ level controls
Scenario 2: TAZ level controls + 'Householder Age' control at the county level
Scenario 3: TAZ level controls + 'Person Age' control at the county level
Scenario 4: TAZ level controls + 'Householder Age' and 'Person Age' controls at the county level