Using survey calibration and statistical matching to reweight and distribute activity schedules

Journal Article

Author(s):
Müller, Kirill; Axhausen, Kay W.

Publication date:
2014

Permanent link:
https://doi.org/10.3929/ethz-b-000074719

Originally published in:
Transportation Research Record 2429, https://doi.org/10.3141/2429-17

Using Survey Calibration and Statistical Matching to Reweight and Distribute Activity Schedules

Kirill Müller and Kay W. Axhausen

The processes used to generate a 10^6-member agent population for a long-range study of 2030 travel in Switzerland are presented in this paper. This study was part of an effort to assess the effects of electric vehicles on the energy production and stability of the electric supply network. The process used well-established statistical methods—survey calibration and statistical matching. Both methods are described, and consistency with known approaches in transportation planning is shown. The paper introduces a new approach that allows exogenous specification of shares of activity types while maintaining the representativeness of the population: survey calibration is applied to satisfy these constraints, and statistical matching allows the joining of data sets with common variables. The discussion of the results for Switzerland focuses on the quality metrics available and highlights the links between the activity schedules and total shares of the activity types. Furthermore, the error introduced by the calibration and matching stages was analyzed and quantified, with special emphasis placed on uncontrolled sociodemographic and travel variables. The specified activity shares could be replicated almost perfectly; the resultant mean error in the uncontrolled variables was within the range of a few percentage points. Therefore, this approach is a viable alternative to a complicated estimation of an activity schedule model that would be necessary otherwise. Finally, two practical issues—data sets describing different populations and biased surveys—and their effect on the outcome were tested by suitably adjusting the input data.

Agent-based transportation models simulate activities and trips over time at the level of individual agents. In contrast to traditional four-step models, the simulation results provide detailed activity and trip information for each agent. This allows incorporating features such as emissions (*1*) or the loading on the electric supply network incurred by electric vehicles (*2*). However, agent-based models also require a detailed synthetic agent population as input [see Müller and Axhausen for a review (*3*)], including a distribution of daily activity schedules.

The agent-based models in the tradition of Bowman et al. (*4*) use multilevel nested logit models to predict the likelihood of different types of daily activity schedules (*5–12*). To keep the logit model estimable, the number of allowed patterns is generally restricted to a reasonable subset of all observed daily schedules. Decision tree–based models, such as ALBATROSS, equally restrict the structures available (*13*). Still, these models capture some of the trade-offs made by travelers and can therefore assess the impact of certain transport policies that affect the generalized costs of travel, as captured by the logsum terms of the different nests feeding into the main scheduling choice models at the top of the overall model structure. This ability comes at the cost of careful and time-consuming model estimation, which is not always available to a project. In addition, the models are by definition constrained to transport investment- and policy-related processes. In some cases, the analyst is asked to assess the impact of exogenous targets on the activity schedules or policies or both that a transport model cannot capture, for example, the arrival of reliable and affordable Internet-based daily and nondaily shopping or work-from-home initiatives.

Balmer describes the extraction of observed activity schedules from the Swiss transportation microcensus (*14*). The extracted activity schedules are stochastically attached to synthetic agents to create the initial demand for a MATSim model (*15*). The distribution of activity schedules can be influenced only by using some of the extracted activity schedules more often than others. Hettinger proposes a custom weighting procedure, the results of which could be used to affect the selection of activity schedules during distribution (*16*).

A novel approach is presented here to estimate weights for a sample of persons with observed activity schedules to reproduce given targets for the future shares or total durations of different activity types, while the sociodemographic distribution is held fixed. In contrast to Balmer (*14*) and Hettinger (*16*), the activity schedule remains connected to the sociodemographic data. While the approach cannot give insights into the individual trade-offs made by travelers, it produces a distribution consistent with the target values at low computational costs and without the model estimation effort. Special emphasis is placed on using methods that are well established in the statistics community but perhaps not so widely recognized in the transportation community:

- Survey calibration to reweight data sets to satisfy exogenous constraints (*17, 18*) and
- Statistical matching to combine multiple data sets (*19*).

These methods are presented in detail, along with results for an application for all of Switzerland in which a synthetic population of about 808,000 agents (10% of the population) with corresponding activity schedules is created. The case study is part of an effort to assess the effects of electric vehicles on energy production and the stability of the electric supply network.

The rest of this paper is organized as follows. After the data are outlined, the methodology is presented in detail. Subsequently, simulation results are shown and discussed; the paper then concludes with a summary.

IVT, ETH Zurich, Wolfgang-Pauli-Strasse 15, CH-8093 Zurich, Switzerland. Corresponding author: K. Müller, kirill.mueller@ivt.baug.ethz.ch.

Transportation Research Record: Journal of the Transportation Research Board, No. 2429, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 157–167.
DOI: 10.3141/2429-17

DATA

Two major data sets and a classification of communes have been used for generating the synthetic population. This section presents the data sources and shows which parts of the final population are derived from which data source.

Register Survey

As of 2010, the full population census of Switzerland has been replaced by the combination of a full register survey and a detailed 2.5% population survey. The data are collected every year, as opposed to the census that was collected every 10 years. The register survey describes the full population of Switzerland on a certain day of 2010. In the present study, detailed data for all persons have been used (8.08 million observations); however, an aggregated version that lists only person counts per hectare is freely available and could be used as well.

The data set contains the de jure spatial location at the hectare level, in addition to basic sociodemographics available from the civil registry. A 10% random sample without replacement has been drawn, as this was the target sampling fraction for the transport model. Only the age, sex, and location attributes were used. The age was binned in three classes: younger than 25 years, 25 to 64 years, and 65 years or older.

The population survey has not been used for this study, as the transportation microcensus (see below) already contains all necessary information.

Transportation Microcensus

A nationwide representative survey on mobility behavior, the transportation microcensus, is collected every 5 years in Switzerland, the last time in 2010. It contains, among other information, extended sociodemographics and information on mobility behavior (activities and detailed trips) for 1 day for about 62,900 persons (0.78%). Only persons 6 years old and above are included in the sample.

The survey days are distributed uniformly over the year for the entire sample. To generate a population that is representative of a typical working day, only roughly 24,300 (0.3%) midweek observations (Tuesday till Thursday) were used. Weights are provided to make the sample representative of the full population and of all weekdays. The data set also contains location information at the hectare level, but that information cannot be used directly because of the relative sparsity of the sample.

The microcensus contains a very detailed description of the activities. Each activity type has been mapped to one of home, work, education, leisure, or shopping. (For the present study, daily and long-term shopping is not distinguished.) The activity schedule consisting of the five activity types above, as well as age, sex, work status, and location, has been used from this data set. In addition, activity durations have been used for validation (see section on results).

Commune Classification

The Swiss communes are classified into 22 types according to commuter movement, occupation, housing conditions, wealth, tourism, population, and role in the central place theory as defined by Christaller (*20*). A coarser version of this classification with nine levels is provided. This nine-level classification has been applied to the transportation microcensus and the register survey data to keep the reweighted sample representative of the population (see section on survey calibration) and to determine matching classes (see section on statistical matching).

Target Population

The transport model requires the following attributes for each agent:

- Age,
- Sex,
- Precise home location,
- Precise workplace location,
- Education level,
- Income, and
- Activity schedule with durations for each activity.

The first three attributes are provided by the register survey. For imputing the workplace location, a calibrated commuter matrix and detailed data on businesses are matched; that procedure is beyond the scope of this paper, however. Education level and activity schedule are imputed to the register data by statistically matching them with the data taken from the transportation microcensus. The implementation of the population synthesis procedure is detailed in the following section.

METHODOLOGY

This section shows the process of generating a synthetic population as described above in regard to survey calibration and statistical matching. The process consists of two stages. In the first stage, survey calibration is used to reweight a person sample with activity schedules to reflect postulated changes in the frequency of certain activity types. The second stage combines this calibrated data set with the register survey data by means of statistical matching. The method used in each stage is described in detail.

Survey Calibration

Estimating a contingency table that satisfies known marginal constraints is a classical application of iterative proportional fitting (IPF) (*21*). In the field of transportation modeling, it is used to update a commuter matrix to known in- and outflows. Beckman et al. described its use for generating a synthetic population, a data set that is statistically consistent with microdata and aggregate controls (*22*). For Beckman et al., the rationale for using IPF was to maintain the odds ratios in the contingency table, as these define the correlation structure (*22*). From the perspective of population modeling, using IPF is equivalent to estimating a log-linear model in which the aggregate controls define the model structure and the survey data define the previous distribution (*23*). From an information-theoretic perspective, IPF minimizes the Kullback–Leibler divergence between the original and the generated data (*24*). As noted by Pritchard and Miller, the IPF algorithm can also be run without first creating a contingency table from the microdata; in that case, the result is a weighting of the microdata (*25*). This technique is also applied in various propositions to fit a hierarchical sample (e.g., persons in households) to constraints at both levels (*26–28*). See also Müller and Axhausen (*3*) for a detailed description and recent improvements of the original approach by Beckman et al. (*22*).

The problem of adapting a detailed data set to aggregate controls in a statistically consistent fashion also occurs in the domain of survey statistics. Sampling weights are adjusted to match known population totals, to make the (usually nonuniform) sample more representative of the population and to improve the precision of estimation (*29*). The simplest form of adjustment of weights is poststratification: the population is divided into strata for which the population totals are known, and each stratum is reweighted independently to match the given totals. If population totals are known for several overlapping stratifications, poststratification can be applied repeatedly for each stratum in a round-robin fashion until convergence. This procedure is called "raking" and is essentially the IPF algorithm. Both poststratification and raking operate on categorical variables only. Deville and Särndal (*17*) suggest a more general set of weighting schemes called "generalized raking," which is developed further in Deville et al. (*18*). Generalized raking allows direct estimation of the calibration weights, with classical raking as a special case. The procedure offers direct support for calibrating continuous variables, for calibrating against totals given for different categorizations of a categorical variable, and for calibrating nested structures such as persons in households. The calibration weights are optimized with respect to a distance function; some control over their distribution is possible by choosing one of the four alternatives suggested in Deville et al. (*18*).

The generation of a synthetic population can also be looked at as estimation of a population total (the total number of, e.g., persons in the population) for each combination of characteristics. The fact that the IPF method is used for both population synthesis and survey calibration further suggests that the problems are inherently similar. Consequently, the methods used for survey calibration are well applicable to generating synthetic populations. In particular, generalized raking is able to solve a superset of the problems that occur with population synthesis. The method is well understood and theoretically justified, and a free implementation (*30, 31*) is available for the R platform for statistical computing (*32*). In the following, a brief outline of the generalized raking algorithm is given; see Deville et al. for a detailed description (*18*). For simplicity, boldface lowercase letters denote row vectors.

The input is a real-valued matrix $\mathbf{X}$ with n rows $\mathbf{x}_k$ (one per survey record) and m columns (one per calibration variable). A vector of totals $\mathbf{t}$ is provided, and optionally a vector $\mathbf{d}$ of original weights (defaults to a vector of ones). For a given distance measure, the objective is to find a vector $\mathbf{w}$ of weights so that $\mathbf{w} \cdot \mathbf{X} = \mathbf{t}$ and the distance between $\mathbf{w}$ and $\mathbf{d}$ is minimal. Several distance measures are suggested, one of them (the multiplicative or raking ratio method) makes the algorithm equivalent to IPF. The underlying optimization problem is then solved numerically by using the method of Lagrange multipliers.

The generation of the input matrix $\mathbf{X}$ and the totals from survey data are described below.

Continuous Variables

As seen above, generalized raking fits the weighted sum over all attributes to a vector of totals; thus, continuous variables are calibrated in a straightforward fashion. Continuous variables and their totals are used directly, without transformation, as columns in $\mathbf{X}$ and $\mathbf{t}$. (Example: number of cars in a household versus a known number of privately owned cars.)

Categorical Variables

The categorical variable is the only kind of variable supported by IPF. In generalized raking, for a categorical variable with j categories, j binary indicator columns are created in $\mathbf{X}$, and the corresponding totals are appended to $\mathbf{t}$ (example: household type—single, family, single with child, . . .).

Nested Structures

When nested structures are calibrated, for example, persons in households, each row in $\mathbf{X}$ represents a household. For a continuous variable at the person level, the per-household sum is stored in one column in $\mathbf{X}$. Likewise, a categorical variable at the person level is converted to j count columns in $\mathbf{X}$, indicating the number of persons in the household with the respective characteristics. A similar transformation has been suggested by Bar-Gera et al. (*26*) (example: individual age group.)

Survey calibration can be used in a straightforward fashion for reweighting a sample of persons with activity schedules as described in the section on transportation microcensus. In fact, activity schedules are also a nested structure—different activities grouped together. Therefore, the $\mathbf{X}$ matrix contains four columns, one for each activity type (except home), and each cell contains the number of activities of the respective type in the given activity schedule. (Controlling home activities would unnecessarily restrict the optimization problem and lead to more extreme weights; this issue is discussed further in the section on deterministic results.) The $\mathbf{t}$ vector contains the new total number of trips for each type. Four scenarios are analyzed here, reflecting the following assumptions based on totals derived from the microcensus data (percentages correspond to those totals derived from the microcensus data):

- Baseline. Original frequency of activities, nothing changes.
- Home office. Working from home is encouraged; number of work trips decreases by 20%.
- Delivery. Home delivery of goods plays a larger role; people replace 20% of their shopping trips by leisure trips.
- Combined. The combination of the former two scenarios.

In addition, the sample is stratified by age, sex, and commune classification, and the population in each stratum is kept unchanged to ensure that the reweighted sample remains representative of the person population. The three age, two sex, and nine commune classes yield 54 additional columns in $\mathbf{X}$. This is essentially a poststratification applied on top of the actual update of activity schedule frequencies—weights are estimated to simultaneously satisfy both constraints. The weights supplied with the transportation microcensus are used naturally as previous weights, which are not changed by the baseline scenario.

The matrix

$$\mathbf{X} = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{1}$$

corresponds to the following hypothetical set of persons with activity schedules:

- Young male, HEHLH;
- Middle-aged female, HSLH;
- Middle-aged male, HWLH; and
- Elderly female, HLHLH.

The first four columns of the matrix represent the frequencies of the work (W), education (E), leisure (L), and shopping (S) activities, and the remaining six columns denote the sociodemographic attributes of the person. (For simplification, the commune classification is not considered here.) For the delivery scenario, $\mathbf{t} = (1, 1, 6, 0.8, 1, 0, 1, 1, 0, 1)$ would be used as the calibration totals. (For this small example, no solution exists; this is not a problem when the microcensus data are used, even in respect to the commune classification.)

All scenarios potentially change the total number of activities in the population compared with the baseline scenario. In particular the home office scenario leads to an unrealistic decrease in the total number of activities, contrary to evidence by, for example, Pendyala et al. (*33*). This issue can be alleviated by choosing appropriate relative changes for other activity types; however, for the present case study it is assumed that the suppressed work trips are not substituted.

Statistical Matching

The term "statistical matching" (or "data fusion") refers to a stochastic procedure for integration of nonoverlapping data sets (X, Y) and (X, Z) with a common variable X. Here, X, Y, and Z can be multivariate. For this procedure, two approaches are distinguished:

- Macro. The joint distribution $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ or its key characteristics are estimated directly from the input data sets.
- Micro. A complete synthetic data set $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is constructed.

The micro approach is of particular interest for generating a synthetic population. In the simplest case, conditional independence is assumed between Y and Z given X. The matching can be performed by choosing one data set as recipient and the other as donor and then drawing, for each recipient record, a compatible record from the donor data set. For categorical X, two records (x_y, y) and (x_z, z) can be treated as compatible if $x_y = x_z$ or if they are sufficiently close with respect to some distance measure. Weights in the donor data set can be used during the drawing of a compatible record. This particular kind of matching is also referred to as "hot deck imputation," and it seems to be a natural and plausible approach to the problem. A practice-ready R package offers an implementation of this algorithm (*34*).

Many more procedures and tools are available in the statistical matching framework, such as parametric methods, Bayesian approaches, replacement of the assumption of conditional independence by the assumption of pairwise independence or by auxiliary information, and evaluation of matching uncertainty by using multiple imputation and expectation maximization. These procedures and tools can be seen as extensions of the hot deck approach. See D'Orazio et al. for a detailed treatment of the subject (*19*).

For the case study presented here, a 10% sample of the register survey is combined with the calibrated transportation microcensus by using statistical matching with hot deck imputation. The first data set is a complete data set with detailed location information, and the second data set contains important attributes such as extended sociodemographics and activity schedules. As the interest here is in generating a complete population, the register survey is used as the recipient data set. The calibration weights derived in the section on survey calibration are used as matching weights. Common variables in both data sets are age, sex, and commune classification; records with identical values for all variables are considered compatible. As the transportation microcensus contains only persons 6 years old and above, these persons were excluded from the register survey before matching. (The effect of ignoring this is analyzed in the subsection on population mismatch of the section on stochastic results.)

D'Orazio et al. recommend using the larger data set as donor and the smaller data set as recipient (*19*). Violating this recommendation obviously leads to donor records used more than once and, therefore, to a modification of the variability of the imputed variables (in this case, extended sociodemographics and activity schedules). Potential sampling errors in the transportation microcensus will be amplified. However, both data sets can be considered reliable, despite their differences in size. While the generated data set is perhaps not optimal for statistical inference, it is usable as input to a transport microsimulation.

RESULTS

This section presents experimental results from generating 100 synthetic populations with different random seeds, with a fixed 10% sample of the register survey. The section on deterministic results presents the results after the calibration of the transportation microcensus to the four scenarios (see the section on survey calibration). Because this method is entirely deterministic, only one run per scenario is considered here. Matching the calibrated microcensus to the register survey sample yields the final synthetic population; this result is discussed in the subsequent section on stochastic results.

The experiments were conducted by using the R platform for statistical computing and the packages survey (*30*) and StatMatch (*34*), among others. The calibration took only a few seconds per run; creating the 10% population requires just under 15 min on a current computer server, including data input and output.

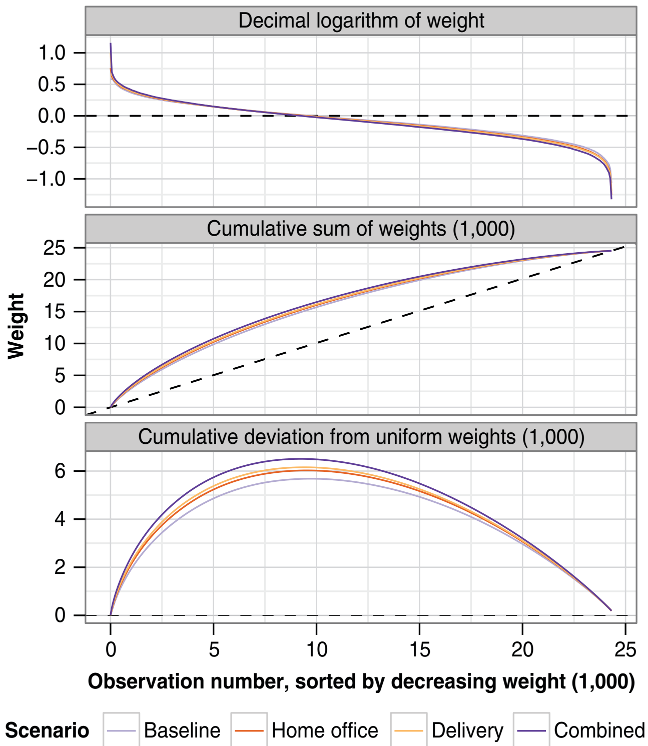
Deterministic Results

First, the weights that result from calibrating the transportation microcensus are considered. After that, average activity counts and durations and activity chain frequencies are analyzed.

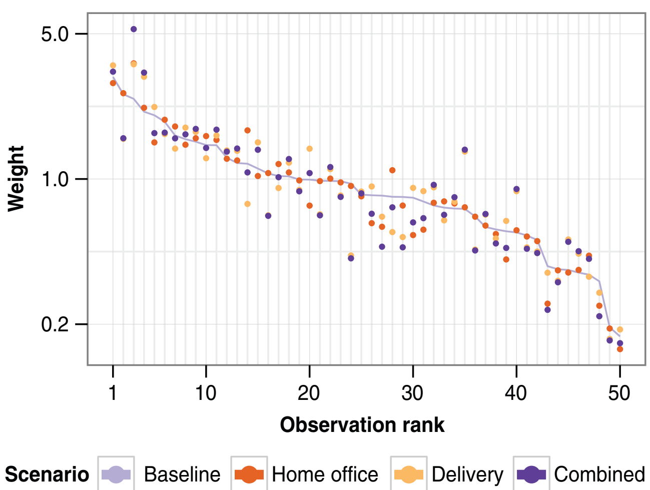
Weights

The original data are also weighted; these weights are used for the baseline scenario. The larger the relative difference of a weight from the average weight, the more (or less) emphasis is put on the corresponding observation. Extreme weights might be necessary to ensure representativeness (or, in this case, satisfaction of external controls); however, observations with extreme weights substantially contribute to the variance of population estimators (*17*).

Figure 1*a* shows three graphs; all of these graphs show the distribution of weights against the rank by decreasing weight. The top graph shows absolute values on a logarithmic scale (rotated cumulative distribution function). All curves are very similar, most weights are between 0.3 and 3, and the median weight is below 1 for all scenarios. The middle graph shows the cumulative sum; the deviation from the straight line is an indicator of nonuniformness (*35*). In turn, the graph at the bottom is a sheared version of the middle plot; uniform weights now correspond to a horizontal line. As expected, the combined scenario deviates most, followed by the delivery and the home office scenarios. However, it appears that the bias in the baseline scenario, which uses the original microcensus



(a)



(b)

FIGURE 1 Distribution of weights after survey calibration: (*a*) distribution against rank by decreasing weight and (*b*) plot of weights of sample of 50 observations for all scenarios.

weights, is much larger than the additional bias introduced by each of the scenarios.

Although the weights do not seem to differ much on average between the scenarios, they do differ significantly for individual observations. Figure 1*b* is a plot of the weights of a sample of 50 observations for all scenarios. For individual observations, when the scenarios are compared with the baseline scenario, the deviation of the combined scenario seems to equal the added deviation of the home office and delivery scenarios (on the logarithmic scale); this equality is true only in approximation.

Activity Count and Duration

Table 1 shows the relative difference compared with the baseline scenario of mean activity count and duration for the different activity types and scenarios. In the frequency side of the table the change in activity counts from the baseline scenario is shown; the frequency is exactly as specified by the configuration of the scenario, except for home activities, which have not been controlled for and for which the frequency is between –1.6% and 1.9%. The deviation of the average activity duration per activity type is in an acceptable range, between –2.4% and 8.1%. Notably, among all scenarios, the activity duration is altered most for the home office scenario. Activities of all types, especially work activities, tend to take longer on average. An interpretation is attempted in the following subsection.

Activity Schedules

The distribution of the 20 most frequent types of activity schedules in the different scenarios, sorted according to their frequency in the baseline scenario, is shown in Table 2. The table shows, for each activity schedule type, the relative frequency, the deviation from the baseline scenario, and the change in rank compared with the baseline. Perhaps the biggest surprise is the first row, activity schedule type HWH. For the home office scenario, the frequency of this activity schedule type increases, while it decreases for the other scenarios. A possible reason for this counterintuitive result might be a correlation between this activity schedule type and the person classification (age, sex, and type of residential commune) (see the section on survey calibration). This result is even worse when the analyst is also controlling for the average frequency of home activities during calibration. In the home office scenario the share of the HLH and HSH schedules increases to compensate for the unexpected increase of the share of HWH schedules. The HWHWH and HWLWH schedules drop by six positions each in the overall frequency ranking. That result might explain the increase in the average duration of work activities: the share of HWH schedules (with long work activities) increases slightly while the share of schedules with two (shorter) work activities decreases.

For the combined scenario, the absolute increase of share is largest for the HLH and HLHLH schedules. The relative differences for the home office and delivery scenarios sum up very closely to that of the combined scenario.

Stochastic Results

In the following, the results after matching are presented. In contrast to the previous section, the results are of a stochastic nature, and therefore the analysis will also cover mostly the distribution along with point estimates. First, the frequency and duration of activity types under the four scenarios are examined. After the effect on uncontrolled variables is analyzed, a sensitivity analysis to assess the impact of data sets describing different populations and biased survey data is performed.

Activity Count and Duration

The results after matching are summarized in Figure 2, *a* and *b*. The box plot in Figure 2, *a* and *b*, respectively, compares the distribution of the frequency and duration, for the supported activity types

TABLE 1 Frequency and Duration by Activity Type, After Calibration

Activity Type	Frequency, by Scenario: Baseline	Home Office	Delivery	Combined	Duration (h), by Scenario: Baseline	Home Office	Delivery	Combined
Home								
Absolute	2.44	2.40	2.49	2.46	2.08	2.20	2.06	2.17
Relative (%)	0.00	–1.62	1.93	0.63	0.00	5.77	–0.61	4.63
Work								
Absolute	0.66	0.53	0.66	0.53	5.79	6.26	5.68	6.14
Relative (%)	0.00	–20.00	0.00	–20.00	0.00	8.12	–1.85	5.97
Education								
Absolute	0.21	0.21	0.21	0.21	3.91	3.99	3.88	3.96
Relative (%)	0.00	0.00	0.00	0.00	0.00	1.89	–0.72	1.11
Leisure								
Absolute	0.89	0.89	1.07	1.07	1.58	1.66	1.54	1.62
Relative (%)	0.00	0.00	20.00	20.00	0.00	5.23	–2.35	2.44
Shopping								
Absolute	0.41	0.41	0.33	0.33	0.56	0.59	0.56	0.59
Relative (%)	0.00	0.00	–20.00	–20.00	0.00	5.25	0.06	4.91

in the final population with the corresponding value that resulted from calibration. The comparison was performed for all scenarios; each scenario was run 100 times to obtain a sample of these error distributions. There is some slight bias, mostly well within the range of ±1%, and about −3% for frequency of education activities (Figure 2*a*, middle panel). Because this bias is present even for the baseline scenario where the calibration does not alter the original weights, the bias can be attributed to the statistical matching. A possible reason for the bias might be that the data sets used for the matching describe slightly different populations. The most notable difference, the absence of persons under 6 years of age in the microcensus, has been accounted for; however, the microcensus weights are based on the de facto population while the register survey reports the de jure population. Fortunately, the relative deviation of the other scenarios from the baseline scenario is, on average, much smaller than the bias in the baseline scenario. On average, the external

TABLE 2 Most Frequent Activity Schedule Types, After Calibration

Schedule	Baseline: f (%)	Rank	Home Office: f (%)	Rank	Delivery: f (%)	Rank	Combined: f (%)	Rank
HWH	10.8	1	10.9 (0.1)	2↓1	9.7 (–1.1)	2↓1	9.8 (–1.0)	2↓1
HLH	9.9	2	11.6 (1.7)	1↑1	11.5 (1.6)	1↑1	13.4 (3.5)	1↑1
HSH	6.7	3	7.4 (0.7)	3↔	4.5 (–2.2)	4↓1	4.9 (–1.8)	4↓1
HWHWH	3.4	4	2.5 (–1.0)	10↓6	3.0 (–0.4)	6↓2	2.2(–1.3)	11↓7
HWHLH	3.3	5	3.3 (0.0)	6↓1	3.7 (0.4)	5↔	3.7 (0.4)	5↔
HLHLH	3.1	6	3.6 (0.5)	4↑2	4.5 (1.4)	3↑3	5.3 (2.2)	3↑3
H	3.0	7	3.4 (0.5)	5↑2	2.8 (–0.2)	7↔	3.2 (0.2)	6↑1
HSHLH	2.5	8	2.8 (0.3)	7↑1	2.1 (–0.4)	11↓3	2.3 (–0.2)	9↓1
HWLWH	2.5	9	1.8 (–0.7)	15↓6	2.7 (0.3)	8↑1	2.0 (–0.5)	13↓4
HEH	2.5	10	2.7 (0.2)	8↑2	2.2 (–0.3)	10↔	2.4 (–0.1)	8↑2
HLSH	2.4	11	2.7 (0.3)	9↑2	2.1 (–0.4)	12↓1	2.3 (–0.2)	10↑1
HWLH	2.3	12	2.3 (0.0)	11↑1	2.6 (0.3)	9↑3	2.6 (0.3)	7↑5
HEHEH	1.9	13	1.9 (0.0)	14↓1	1.6 (–0.2)	14↓1	1.6 (–0.3)	14↓1
HSLH	1.7	14	1.9 (0.2)	12↑2	1.5 (–0.3)	15↓1	1.6 (–0.1)	15↓1
HEHLH	1.7	15	1.9 (0.1)	13↑2	1.9 (0.2)	13↑2	2.0 (0.3)	12↑3
HWSH	1.7	16	1.7 (–0.1)	16↔	1.1 (–0.6)	17↓1	1.1 (–0.7)	20↓4
HEHEHLH	1.1	17	1.1 (0.0)	19↓2	1.2 (0.1)	16↑1	1.2 (0.1)	19↓2
HLSHLH	1.0	18	1.1 (0.2)	18↔	1.0 (0.1)	21↓3	1.2 (0.2)	18↔
HWHWHLH	1.0	19	0.7 (–0.3)	23↓4	1.1 (0.1)	19↔	0.7 (–0.2)	23↓4
HLHSH	0.9	20	1.0 (0.1)	20↔	0.8 (–0.2)	24↓4	0.9 (–0.1)	21↓1

NOTE: Figures in parentheses denote relative change to share in baseline scenario; arrows denote change in rank. Differences may be off by ±0.1 as a result of rounding.

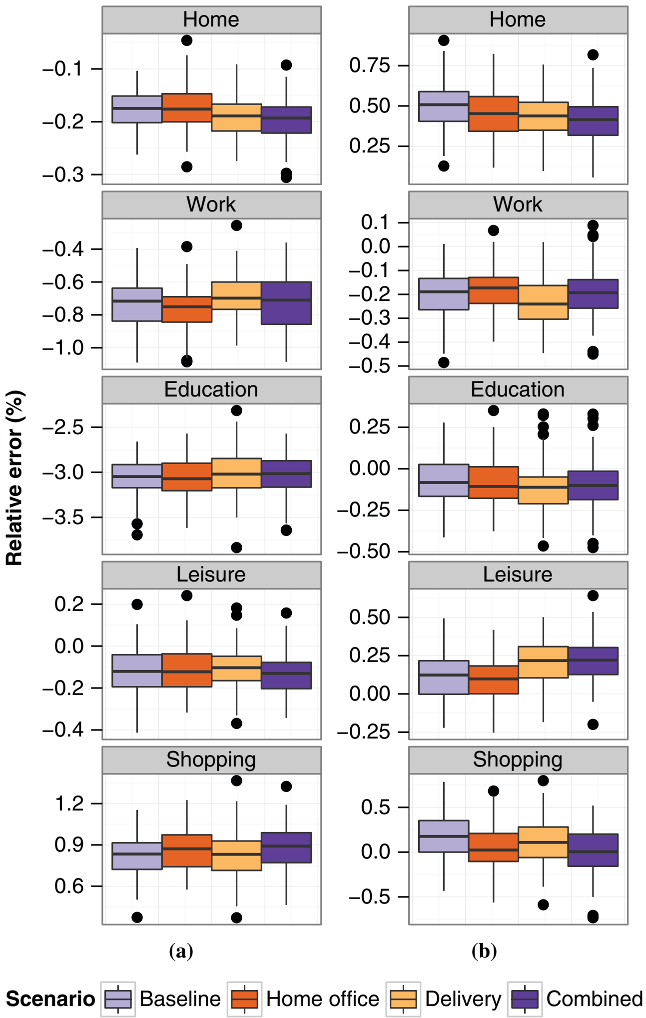


FIGURE 2 Relative error of (*a*) frequency and (*b*) duration per activity type.

specification of activity type frequency is not expected to contribute substantial bias to the resultant synthetic population.

Table 3 shows the relative error and the coefficient of variation (CV) of activity frequency and duration, compared with the baseline scenario. The relative error is very close to that shown in Table 1, an indication that little additional error is introduced with the statistical matching. The CVs can be considered negligible.

Uncontrolled Variables

Figures 3 and 4 show the effect of the statistical matching on uncontrolled sociodemographic and travel variables, respectively. Figure 3 shows, for each classification from a list of sociodemographic variables, its absolute share in the survey data and the distribution of the error under each scenario when scenario data are compared with the survey data. Figure 4 is a similar plot for travel variables. According to Figure 3, the error is mostly well within ±6% for sociodemographic variables. For marital status, separated is over- and single is underrepresented even in the baseline scenario. Education remains mostly unchanged; a tendency toward higher education can be seen in the delivery scenario. Generally, the resultant populations are biased toward lower income levels, fewer cars, fewer driving licenses, and less access to a car. This bias is reduced in the delivery scenario, aggravated in the home office scenario, and balanced again in the combined scenario. A conceivable correlation between leisure activities and an expensive lifestyle could explain the preference for the latter in the delivery scenario. However, the reduction of work trips in the home office scenario seems to put more weight on the nonworking part of the population.

For travel variables (Figure 4), almost unchanged distributions in the baseline scenario are observed; the error varies mostly between –6% and 10% for the other scenarios. For daily travel times, the home office scenario slightly prefers shorter times and distances, while the delivery and combined scenarios put a strong preference on longer trips—again, the result of the increased share of leisure trips. The situation is similar for the travel distance by car; however, the preference for longer distances is not as strong as with the daily travel time, and the combined scenario is almost unbiased with

TABLE 3 Frequency and Duration by Activity Type, After Matching

	Frequency (%), by Scenario				Duration (%), by Scenario			
Activity Type	Baseline	Home Office	Delivery	Combined	Baseline	Home Office	Delivery	Combined
Home								
Mean	0.00	–1.62	1.91	0.61	0.00	5.72	–0.67	4.54
CV	0.04	0.04	0.04	0.04	0.15	0.16	0.14	0.14
Work								
Mean	0.00	–20.04	0.03	–20.01	0.00	8.15	–1.88	5.98
CV	0.15	0.11	0.14	0.13	0.10	0.10	0.10	0.11
Education								
Mean	0.00	–0.01	0.05	0.02	0.00	1.90	–0.75	1.10
CV	0.22	0.23	0.25	0.24	0.15	0.14	0.15	0.15
Leisure								
Mean	0.00	0.00	20.01	19.98	0.00	5.20	–2.26	2.54
CV	0.11	0.11	0.12	0.13	0.15	0.15	0.13	0.15
Shopping								
Mean	0.00	0.05	–19.99	–19.95	0.00	5.12	0.00	4.73
CV	0.13	0.15	0.14	0.13	0.25	0.26	0.25	0.27

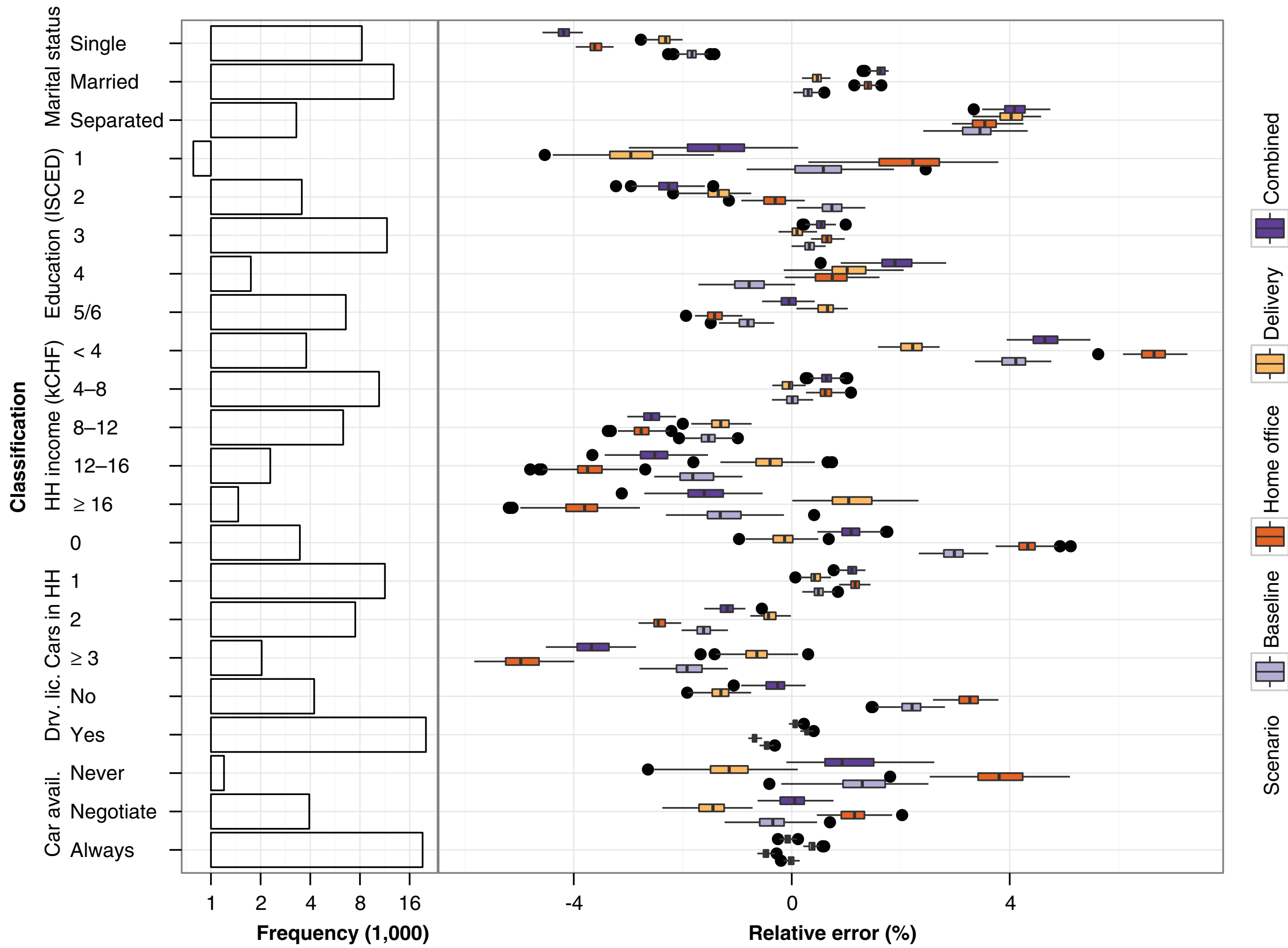


FIGURE 3 Error of sociodemographic indicators, after matching (ISCED = International Standard Classification of Education; kCHF = 1,000 Swiss francs; drv. lic. = driver's license; avail. = available).

respect to these indicators. In all scenarios there seems to be a disapproval of public transport, perhaps because leisure trips are mostly undertaken by car, and a reduction of work trips implicitly leads to a reduction of public transport trips.

Population Mismatch

In an earlier version of the simulation an attempt was made to match the calibrated microcensus with the full register survey, not accounting for the population mismatch resulting from the absence of small children younger than 6 years of age in the microcensus. That procedure has resulted in an extreme overrepresentation of education activities and a slight underrepresentation of other activity types. All records in the register survey corresponding to small children are matched to microcensus records that very likely include an education activity and prefer out-of-home leisure time to time at home.

The results presented earlier in the subsection on activity count and duration were obtained by ignoring small children for the statistical matching. In this subsection simulations in which the age threshold has been varied between zero (i.e., using all records) and 10 in steps of two are analyzed. For each scenario and each age threshold, only one simulation has been run. Figure 5, *a* and *b*, shows the relative error of the frequency and duration, respectively, of education activities. The error in frequency introduced by the population mismatch is much larger than the difference between the scenarios. However, although the error in duration shows a negative trend with increasing age threshold, the stochastic variation reverses the trend in some cases.

Biased Survey Data

For the simulation of a highly biased survey, the previous microcensus weights **d** are transformed. The weight of each record is substituted by its *k*th power with $k \in \{0, 1, 2\}$; this process leads to three scenarios: uniform weights, unaltered weights, and exaggerated weights. For each scenario and each weight transformation, 10 synthetic populations were generated; all of them satisfied the external constraints. Figure 6, *a* and *b*, shows the distribution of the errors compared with the original microcensus data.

The differences between the scenarios are often negligible compared with the differences between different initial weights. A large additional error introduced by the matching process might be an indicator for potential bias in the data.

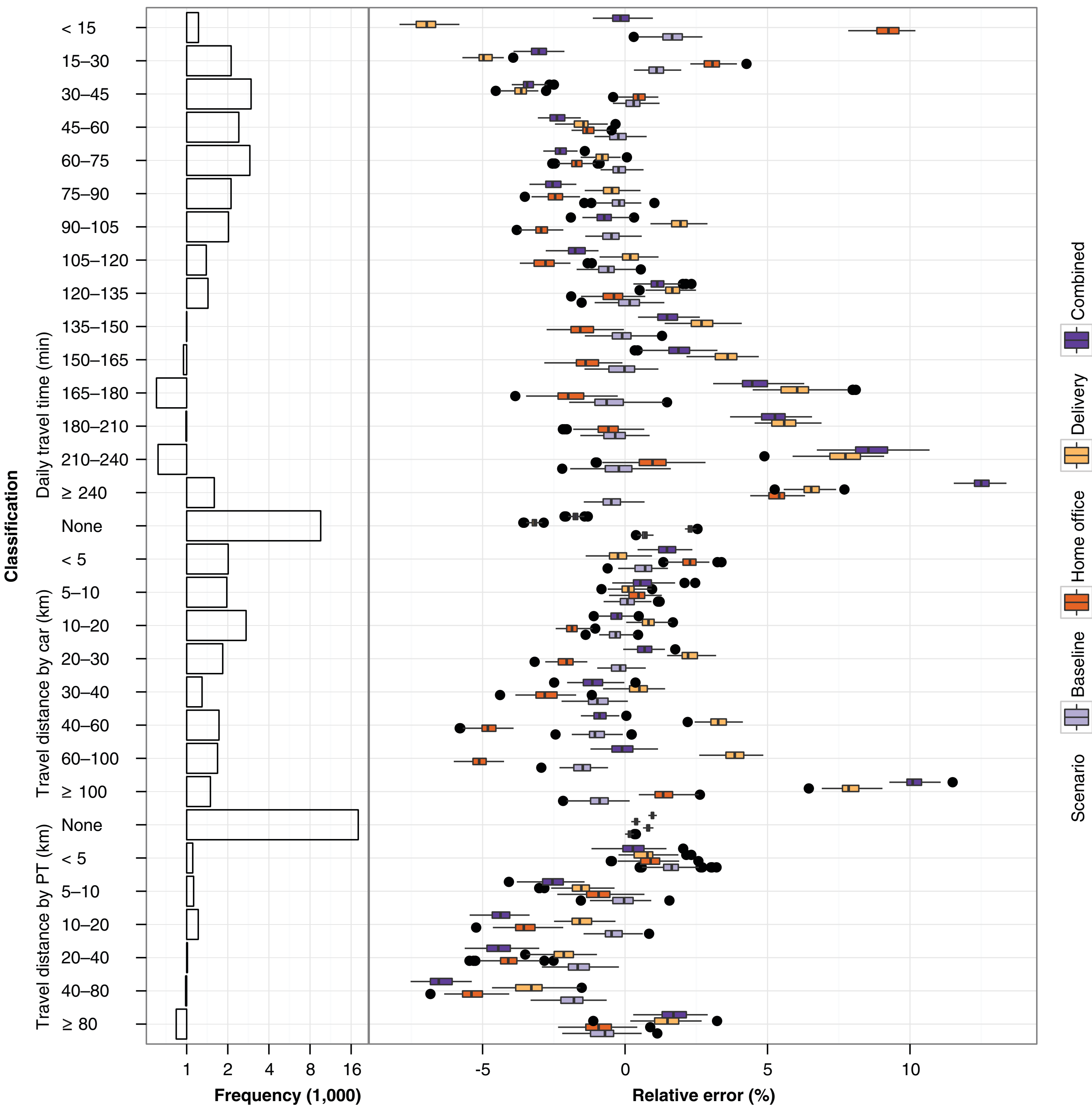


FIGURE 4 Error of travel indicators, after matching (PT = public transport).

FIGURE 5 Relative error of education activities by age threshold: (*a*) frequency and (*b*) duration.

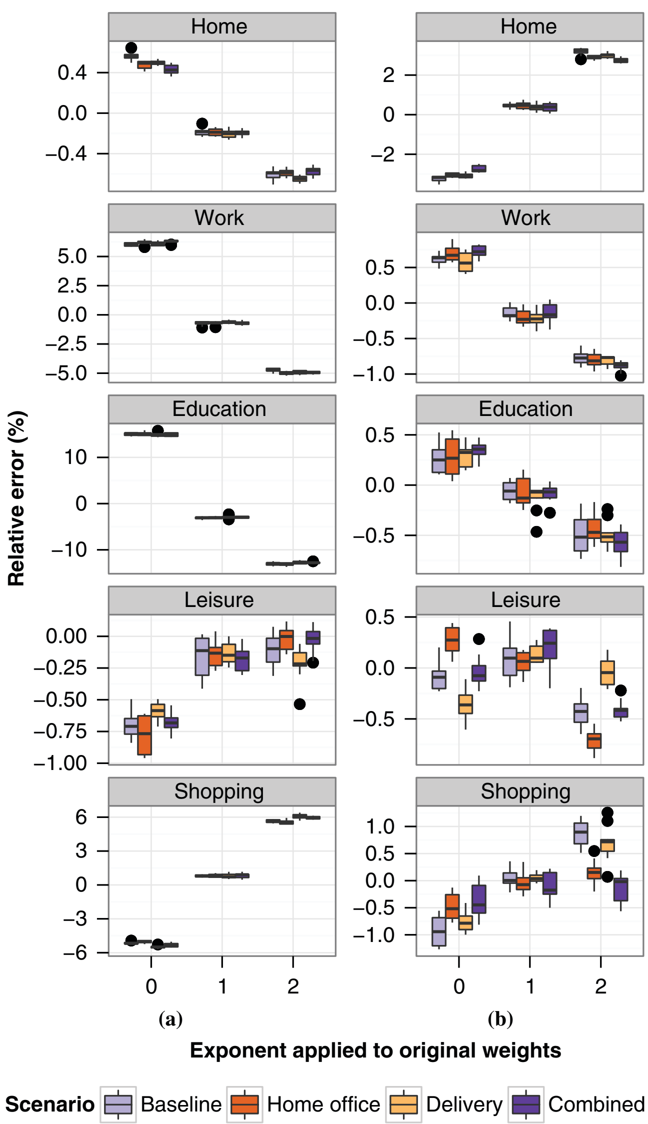


FIGURE 6 Relative error of (*a*) frequency and (*b*) duration for distorted weights.

CONCLUSION

This paper presents a novel approach to weighting a sample of persons with activity schedules so that given targets for shares of different activity types are satisfied. The weighting is then used to randomly distribute the persons over the study area. An application is presented in which a 10% synthetic population for all of Switzerland is created.

The methodology used for reweighting the synthetic population has been successfully applied in the field of survey statistics for more than 20 years. Stochastic distribution is considered as a special case of statistical matching, a relatively new area of research focusing on combining data sets according to common attributes. Both methods can be applied in a straightforward fashion to the problem at hand, with very little modeling and computational effort. The result is a synthetic population that matches the targets almost perfectly and introduces very little bias. Only a little programming effort was required thanks to the availability of free implementations for survey calibration and statistical matching for the R platform for statistical computing.

Assessing the error introduced by statistical matching is possible only if the matched data sets refer to the same population. Potential enhancements include the control for total duration per activity type and support for a finer categorization of activities (e.g., the distinction between daily and long-term shopping). Finally, the weighting approach could be compared with a model-based approach by using actual simulation results of a transportation model.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the sponsors of the Technology-Centered Electric Mobility Assessment project and of the Swiss National Science Foundation (*35*).

REFERENCES

1. Kickhöfer, B., F. Hülsmann, R. Gerike, and K. Nagel. Rising Car User Costs: Comparing Aggregated and Geo-Spatial Impacts on Travel Demand and Air Pollutant Emissions. In *Smart Transport Networks: Decision Making, Sustainability and Market Structure,* NECTAR Series, 11–16 (T. Vanoutrive and A. Verhetsel, eds.), Edward Elgar, Cheltenham, United Kingdom, 2012.
2. Waraich, R.A., M.D. Galus, C. Dobler, M. Balmer, G. Andersson, and K.W. Axhausen. Plug-in Hybrid Electric Vehicles and Smart Grids: Investigations Based on a Microsimulation. *Transportation Research Part C: Emerging Technologies,* Vol. 28, 2013, pp. 74–86.
3. Müller, K., and K.W. Axhausen. Population Synthesis for Microsimulation: State of the Art. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
4. Bowman, J.L., M.A. Bradley, and J. Gibb. The Sacramento Activity-Based Travel Demand Model: Estimation and Validation Results. Presented at European Transport Conference, Strasbourg, Germany, 2006.
5. Bowman, J.L. Historical Development of Activity Based Model Theory and Practice, Part 1. *Traffic Engineering and Control,* Vol. 50, No. 2, 2009, pp. 59–62.
6. Bowman, J.L. Historical Development of Activity Based Model Theory and Practice, Part 2. *Traffic Engineering and Control,* Vol. 50, No. 7, 2009, pp. 314–318.
7. Axhausen, K.W. Neue Modellansätze der Verkehrsnachfragesimulation: Entwicklungslinien, Stand der Forschung, Forschungsperspektiven. *Stadt Region Land,* Vol. 81, 2006, pp. 149–164.
8. Vovsha, P., R. Donnelly, and S. Gupta. Network Equilibrium with Activity-Based Microsimulation Models. In *Transportation Research*

Record: Journal of the Transportation Research Board, No. 2054, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 102–109.
9. Shiftan, Y., K.J. Button, and P. Nijkamp, eds. *Transportation Planning*. Classics in Planning, Vol. 7. Edward Elgar, Cheltenham, United Kingdom, 2007.
10. Pendyala, R.M., and C.R. Bhat. Validation and Assessment of Activity-Based Travel Demand Modeling Systems. Presented at Innovations in Travel Demand Modeling (ITM'06), Austin, Tex., 2006.
11. Bhat, C.R., and F.S. Koppelman. Activity-Based Modeling of Travel Demand. In *Handbook of Transportation Science* (R.W. Hall, ed.), Springer, New York, 2003, pp. 39–65.
12. Bhat, C.R., and R.M. Pendyala, eds. *Travel Behaviour Research for an Evolving World.* Emerald Publishing Group, Bingley, United Kingdom, 2012.
13. Arentze, T.A., and H.J.P. Timmermans. *ALBATROSS: A Learning-Based Transportation Oriented Simulation.* EIRASS, Eindhoven, Netherlands, 2000.
14. Balmer, M. *Travel Demand Modeling for Multi-Agent Traffic Simulations: Algorithms and Systems*. PhD dissertation. ETH Zurich, Switzerland, 2007.
15. Meister, K., M. Rieser, F. Ciari, A. Horni, M. Balmer, and K.W. Axhausen. Anwendung eines agentenbasierten Modells der Verkehrsnachfrage auf die Schweiz. *Straßenverkehrstechnik,* Vol. 53, No. 5, 2009, pp. 269–280.
16. Hettinger, T. *Anpassung von Aktivitätenketten mittels wiederholter proportionaler Anpassung*. Term paper. IVT, ETH Zurich, Switzerland, 2007.
17. Deville, J.-C., and C.-E. Särndal. Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association,* Vol. 87, No. 418, 1992, pp. 376–382.
18. Deville, J.-C., C.-E. Särndal, and O. Sautory. Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association,* Vol. 88, No. 423, 1993, pp. 1013–1020.
19. D'Orazio, M., M.D. Zio, and M. Scanu. *Statistical Matching: Theory and Practice.* John Wiley & Sons, New York, 2006.
20. Christaller, W. *Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen.* Wissenschaftliche Buchgesellschaft, Darmstadt, Germany, 1933.
21. Deming, W.E., and F.F. Stephan. On the Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known. *Annals of Mathematical Statistics,* Vol. 11, No. 4, 1940, pp. 427–444.
22. 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.
23. Bishop, Y.M., S.E. Fienberg, P.W. Holland, R.J. Light, and F. Mosteller. *Discrete Multivariate Analysis: Theory and Practice,* 2007 ed. Springer, 2008.
24. Kullback, S., and R.A. Leibler. On Information and Sufficiency. *Annals of Mathematical Statistics,* Vol. 22, No. 1, 1951, pp. 79–86.
25. Pritchard, D.R., and E.J. Miller. Advances in Population Synthesis: Fitting Many Attributes per Agent and Fitting to Household and Person Margins Simultaneously. *Transportation,* Vol. 39, No. 3, 2012, pp. 685–704.
26. Bar-Gera, H., K.C. 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.
27. Ye, X., K.C. Konduri, R.M. Pendyala, B. Sana, and P. Waddell. Methodology to Match Distributions of Both Household and Person Attributes in Generation of Synthetic Populations. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
28. Müller, K., and K.W. Axhausen. Hierarchical IPF: Generating a Synthetic Population for Switzerland. Presented at 51st Congress of the European Regional Science Association, Barcelona, Spain, 2011.
29. Lumley, T. *Complex Surveys: A Guide to Analysis Using R,* Vol. 565, John Wiley & Sons, Hoboken, N.J., 2010.
30. Lumley, T. *Survey: Analysis of Complex Survey Samples.* R Package Version 3.28-2, 2012.
31. Lumley, T. Analysis of Complex Survey Samples. R package version 2.2. *Journal of Statistical Software,* Vol. 9, No. 1, 2004, pp. 1–19.
32. R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, 2014. http://www.r-project.org.
33. Pendyala, R.M., K.G. Goulias, and R. Kitamura. Impact of Telecommuting on Spatial and Temporal Patterns of Household Travel. *Transportation,* Vol. 18, No. 4, 1991, pp. 383–409.
34. D'Orazio, M. *StatMatch: Statistical Matching,* R package version 1.2.0. 2012. http://CRAN.R-project.org/package=StatMatch.
35. THELMA. Technology-Centered Electric Mobility Assessment. Swiss Federal Institutes of Technology, Zurich, 2013. http://www.thelma-emobility.net/.

The Transportation Demand Forecasting Committee peer-reviewed this paper.