abm_illustration

Hayden Atchley　Gregory S. Macfarlane

14 April 2023

Table of contents

index **3**

1 Introduction **4**

2 Literature Review **5**

2.1 Comparison of Modeling Frameworks 6

2.2 Claimed Advantages 9

2.3 Research Gap 12

3 flowchart-prelim **13**

References **15**

index

1 Introduction

This is a book created from markdown and executable code.

2 Literature Review

In the early 1990s, research began in a new “activity-based” modeling framework as an alternative to the traditional “four-step” or “trip-based” approaches (Rasouli and Timmermans 2014) . This was part of a broader shift toward research in behavioral models over aggregate spatial interaction models in urban planning and regional science that began in the 1970s. Rasouli and Timmermans (2014) list several arguments for this shift: four-step models use several sub-models that are (implicitly or explicitly) assumed independent, and this can result in a lack of consistency or integrity between sub-models; four-step models are strongly aggregated in nature, which can cause significant aggregation bias; and these models lack “behavioral realism”—that is, they do not have a concept of individuals making decisions, which is what travel behavior actually is.

Activity-based models (ABMs) were proposed as a means of addressing the shortcomings of traditional four-step models. Unlike four-step models, ABMs place the focus on “activities” rather than “trips” as the basic unit of analysis. ABMs predict a sequence of activities for each individual and household, with information such as activity location, start time, and duration using a high level of temporal and spatial granularity. “Trips” are then the result of getting from one activity to the next. By adopting this activity-centric framework, ABMs provide a more consistent and comprehensive representation of travel behavior. They take into account complex dependencies and interactions within the model as a whole and at an individual level. ABMs acknowledge that travel choices are not made in isolation, but rather influenced by the preceding activities. This means that e.g. if an individual takes transit to work, they will not be able to drive home. ABMs therefore present a more conceptually accurate model of actual travel behavior than traditional four-step models.

Despite these advantages, however, many agencies have yet to adopt ABMs, instead continuing to use four-step models. While ABMs are superior in certain aspects, they also have disadvantages, such as requiring more detailed input data and greater computational resources. It is also not always clear if ABMs provide substantially better forecasts than their four-step counterparts, nor if this tradeoff is worth it for every agency. This literature review presents an overview of both modeling frameworks, and discusses the potential advantages and disadvantages from using an ABM.

2.1 Comparison of Modeling Frameworks

In discussing the differences between ABMs and four-step models, there are really two comparisons that need to be made: how the population data is structured, and how travel is organized. Four-step models generally use aggregate population data while ABMs use a synthetic population; and four-step models organize travel into trips with ABMs organizing into activities and tours.

The aggregate population data used in four-step models can be varied in origin and level of detail, but the basic concept is the same. The study area is organized into generally small zones, and certain demographic and socioeconomic data is known or obtained for each zone. This includes data such as number of households, average household income, population, and number of workers, among others. Based on this information, the zone can be segmented along arbitrary variables. For example, since households with more workers and more vehicles tend to make more work trips, it is useful to estimate the distribution of households in each zone along these two variables. Then, average trip production rates are determined for each household category (e.g. for households with each number of workers by vehicles), and the total number of trips produced in a zone is calculated based on these rates.

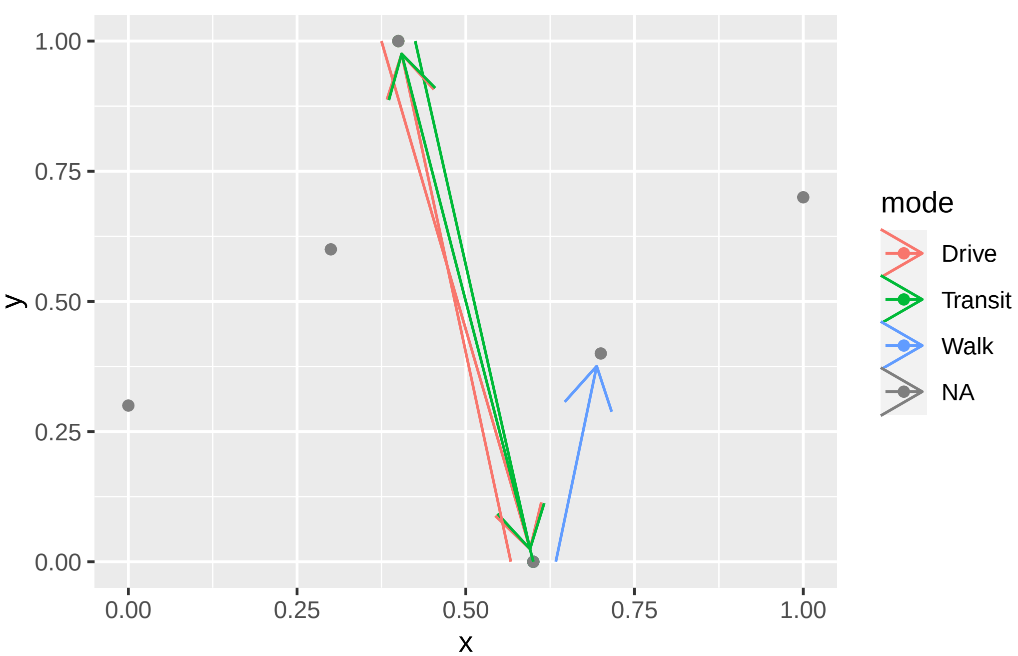
This approach has two major disadvantages. The first is that no other information can be known later in the modeling process other than how the population data was segmented initially. If the population was segmented only along number of workers and number of vehicles, for example, then it would be clear how many trips are produced by a zone's 2-worker, 1-vehicle households, but there is no way to know the distribution of household income for those trips. The second disadvantage is related to the first, which is that as the number of segmentation variables increases, so does the number of trip production rates that are needed. Additionally, since the sample size decreases dramatically with more specific binning, the margin of error on many if not most of these rates would be extremely large. Many models therefore only segment the zones along a small number of variables, but this limits the types of analyses that can be performed.

An alternative approach is to use a synthetic population and regression models to determine trip productions. A synthetic population takes demographic and socioeconomic data at various levels of detail to create a "population" with generally the same distribution as the study area. The goal is to have a population that is functionally similar to the actual population, but without the privacy concerns of using real individual household data. The major advantage with this approach is that the demographic and socioeconomic data is known at the person and household level, rather than the zone level, and this data remains available throughout the modeling process. This allows, for example, an equity analysis to determine the "winners" and "losers" of a proposed development without needing to encode race/gender/etc. into each step of the model.

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.2     v tibble    3.2.1
```

```
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.1
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to b

Attaching package: 'igraph'


The following objects are masked from 'package:lubridate':

    %--%, union


The following objects are masked from 'package:dplyr':

    as_data_frame, groups, union


The following objects are masked from 'package:purrr':

    compose, simplify


The following object is masked from 'package:tidyr':

    crossing


The following object is masked from 'package:tibble':

    as_data_frame


The following objects are masked from 'package:stats':

    decompose, spectrum


The following object is masked from 'package:base':

    union
```

```
Rows: 6 Columns: 3
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl (3): zone, xloc, yloc

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
Rows: 5 Columns: 6
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (3): mode, time, purpose
dbl (3): from, to, number

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.

Warning in format_fortify(model = model, nodes = nodes, weights = "none", :
duplicated edges detected
```



Four-step/trip-based models usually use aggregate data, in part because it is simpler to obtain and work with (?). Trip-based models using a synthetic population—often called trip-based microsimulation models—do exist (cite one or more), but these are relatively rare. There may be several reasons for this, ... ABMs on the other hand need to use a synthetic population since they model the decisions of individual people and households. This has a major advantage of being able to use arbitrary population variables/groupings at each model step, rather than being limited to variables/groupings from upstream steps.

The other main difference between four-step models and ABMs (and the main difference from trip-based microsimulation models) is that ABMs organize travel into "tours", a sequence of trips that begin and end at the home, rather than just trips. In fact, Miller (2023) point out that many current "activity-based" models ought to rather be labelled "tour-based" due to this focus on building tours. This is contrasted with "activity scheduling" models, in which activity participation is modeled explicitly and trips emerge as the means to get from one activity to the next. ***(Miller seems to claim that this is a crucial difference, noting that the "bad rap" ABMs have might be true of tour-based models but need not be true of all ABMs (specifically activity scheduling models). Is this important/worth noting/worth expounding on? Modeling activities versus modeling tours is definitely a different approach, but from my admittedly somewhat ignorant perspective it seems more like a semantic and maybe computational difference rather than much of a conceptual one.)***

2.2 Claimed Advantages

ABMs are often lauded as superior to TBMs due to their usage of synthetic populations and trip-chaining (tours), which is considered to better represent individual choices. According to Rasouli and Timmermans (2014), there are limits to TBMs that are eliminated with ABMs. The TBM does not consider any kind of dependency between trips on the same tour, and it excludes dependency between the members of a household and between mode choices on a tour. Because of the aggregate population data used in the TBM, it can produce results that are unrelated to actual human behavior. These are some of the issues that are claimed to be remedied with the use of synthetic populations and trip-chaining in ABMs.

Unlike TBMs, ABMs take into account the specific members of each household within the model and the tours they will take. This creates dependency between the trips that are taken by each individual. For example, if one member of the household drives their car from home to work in the morning, they can not have a trip going from home to the park until they has a return trip from work to home. This situation is taken into account with ABMs. For similar reasons, the creation of tours in ABMs gives room to possible trips that TBMs may not consider. Many destinations are taken because they are close to another destination on the tour and not necessarily close to the residence. TBMs would make trips based on proximity to the household, while ABMs may make trips based on proximity to the tour.

The dependency that trips have on the tour is similar to the dependency that modes and household members have on each other. If someone drives to work, then a bike can not be used as the mode to return from work. If a child is driven to school by a parent, then the child can not drive a car home without being driven by someone else. These dependencies within the tours and mode choices will not only affect the trip patterns of the household, but they will also affect the model's results in response to policy. Because the people within the household will have travel patterns that depend on the patterns of others in the household, a policy affecting one person in the household can affect everyone in the household no matter how directly the policy connects to them. This type of dependency isn't possible with TBMs.

Changes in accessibility due to policies are often analyzed, and this is where another advantage of ABMs lies. Dong et al. (2006) compare measures of accessibility between ABMs and TBMs, and they explain aspects of accessibility that are overlooked with TBMs. TBMs allow for accessibility analysis, but they only compare specific trips leaving a zone to the possible destinations within a range. For example, "total number of employment opportunities within 30 minutes by transit." Accessibility with ABMs is calculated as a function of all activities pursued throughout the day, which goes much further than TBMs. Along with having a spatial dimension to the accessibility function, ABMs also provide a dimension of individuality. ABMs take into consideration that different groups of people have different levels of accessibility whereas TBMs wouldn't be able to make that consideration. With TBMs, young college students and retired 70-year-olds would be grouped in the same way. The study was performed to compare how these differences might affect the analysis of accessibility using the two models.

In the first part of the study they measured the impact that a peak period toll would have on accessibility using an ABM. Since people have different residential locations and different interests, the toll should have a different impact on different people. As would be suspected, the model showed that the accessibility of people without a car was less affected by the imposition of a toll. In another test, they saw a similar connection. People that are unemployed were not as affected by the toll.

In the second part of the test, they ran models to show the accessibility of a 26-year-old man under certain circumstances: Whether he was employed or not and whether he had a car or not.They also did a simulation taking into account the location of his job. All of these circumstances where able to be analyzed together, while in TBM, each scenario would have to be analyzed separately.

The final part of the study was a side-by-side comparison of the accessibility impacts of a peak period toll given by each model. Because the ABM took into account schedule flexibility, it resulted in lower accessibility impacts than the TBM. This is more congruent with what would actually happen because of people's ability to change their schedules.

The claimed advantages above show some ways in which ABMs give a better picture of reality because of the way they model individual behavior. The ability to more accurately reflect human behavior is largely due to the use of a synthetic population. When individual data from each household is computed, as in ABMs, the results will reflect individual behaviors more accurately than when aggregate household data is gathered, as in TBMs.

Another advantage of a synthetic population has to do with how the data is used after running the model. Once a simulation is run using aggregate data, there is no further analysis that can be done because their is no connection between the resulting trips and the household data that was used for the model. Running a simulation with data from a synthetic population, on the other hand, gives results based on the movements of each individual in the population. Therefore, after the results are obtained, further analysis can be made to see which groups of people took which trips. Even if a piece of information from the data isn't used to determine coefficients during the simulation, it can be referred to after running the model. For example, the race of each person with a certain mode choice can be found to analyze the affect certain decisions may have on minorities. If extra household information like this was taken into consideration while using aggregate data, it would need to be used in the model. As more connections are made to other pieces of information in the model, like race or income, it will become less accurate dues to the smaller sample sizes of each category. The ability to analyze the data after running the model is an advantage of using a synthetic population.

In an effort to see how synthetic populations in ABMs affect equity analysis, Bills and Walker (2017) used the 2000 Bay Area Travel Survey to create a synthetic population. The model consisted of 1454 zones representing nine counties in the San Francisco Bay Area. After running a few different scenarios, the results were compared to see the differences that existed among low income and high income populations. With a 20% reduction in travel cost, they saw that high income workers benefited more than low income workers. They did similar comparisons for scenarios involving reduced travel times for different mode choices and were able to see the effects each scenario had on the high and low income workers.

It is important to note that, while many connect them with ABMs, synthetic populations can be used in running both ABMs and TBMs. The difference is that TMBs can work using either a synthetic population or aggregate data, but ABMs require a synthetic population. The larger computing power needed to work with a synthetic population may be the reason why most TBMs are still being run using aggregate data. The fact that more TBMs don't often use synthetic populations could possibly signify a wider acceptance that ABMs are better. If a company is willing to go through the work to create a synthetic population, then they may jump straight to an ABM because they believe it will be most accurate.

If ABMs are believed to be more accurate, one may wonder why there are so many MPOs that don't make the switch away from TBMs. Miller (2023) gives some possible reasons for the gap between the academic interest of ABMs and the implementation into mainstream operational planning practice. There has been a lot of research done on the subject, and many models have been developed. Some notable ABMs are ADAPTS(cite?), ALBATROSS(cite?), CEMDAP(cite?), FAMOS(cite?), FEATHERS(cite?), MATSim(cite?), and TASHA(cite?). Possible reasons they have not been implemented include the following:

Computational inefficiency and complicated program design: As alluded to in the previous paragraph, ABMs often take more time, more computing power, and more money to run. The synthetic population needed to run an ABM uses much more data. In areas with thousands of TAZs and millions of people, a super computer is needed, and it will cost much more than what is spent to run TBMs. If a city can see similar results using a TBM, they may decide not to invest in an ABM.

Absence of a standard model system: The modeling systems are often designed with different approaches and designed for specific areas making it hard to transfer from one urban area to another.This also makes it difficult for agencies to determine which approach is the best and decide which to implement. In relation to this, Miller also states that the pressures of publishing unique and ground-breaking research in academia can deter researchers from converging towards best theories and methods.

Lack of resources: Most of these models were developed in academic settings which often lack resources, and possibly desire, to put them into practice. This leaves it up to governments and consultants to put the models into practice, but they can be hesitant to promote software development and to invest in new systems.

Even with so many claimed advantages, these disadvantages of ABMs could be the reason for lack of total implementation throughout the world.

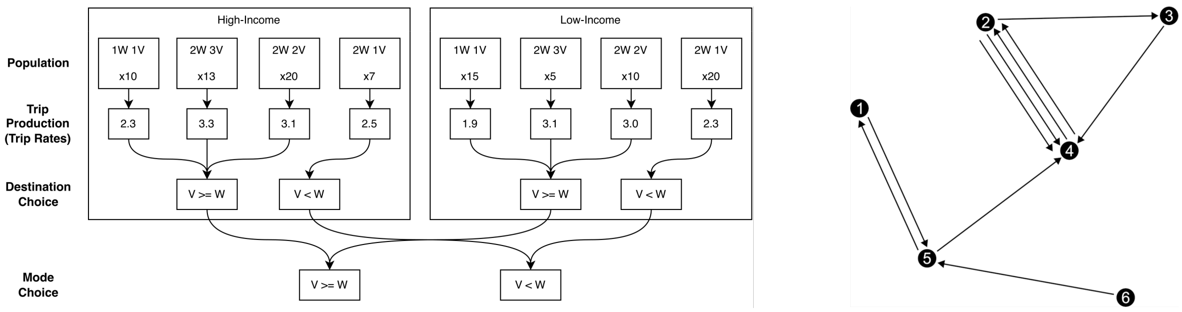
2.3 Research Gap

Though there has been much research on ABMs and their theoretical advantages, to the authors' knowledge there is little in the literature about their practicality in the real world. It is often taken as given that ABMs are unilaterally superior to traditional four-step models due to their better theoretical foundation, but it is not clear if that better foundation always leads to better results. Ferdous et al. (2012) compared the trip- and tour-based model frameworks of the Mid-Ohio Regional Planning Commission and found that the tour-based model performed slightly better at the region level, but about the same at the project level. If this is true more generally, an agency may have no real need to switch to an ABM over their current four-step model since the improvement in results may not outweigh the increase in data, computational, training, and potentially financial requirements.

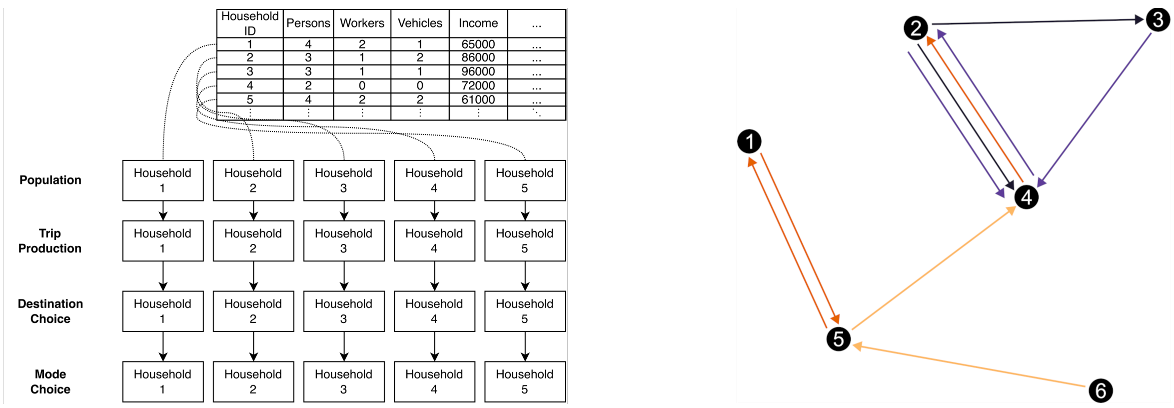
Ferdous et al. (2012), however, mainly compared the *accuracy* of the two frameworks, but did not address the methodological differences between them. What types of data are needed as input to each model type? What computational resources are needed? Are there analyses that can only be done through (or that are made easier by) one of the model types? What would an agency need in order to transition from a four-step model to an ABM? Are certain types of scenarios suited to one model type? These questions remain largely unanswered, and the answers in the current literature are mainly theoretical, with little use to an agency considering the transition.

This research aims to answer these questions by providing a side-by-side comparison of a four-step and activity-based model. Several "proposed development" scenarios are run in each model, and the strengths and weaknesses of each are compared. It is important to note that this paper is not focused on model accuracy, as in any model type this can be adjusted dramatically through calibration efforts. Rather, the focus is on the methodological differences between the approaches, and the types of analyses that can be done with each model type. This paper hopes to be a sort of guide for agencies considering adopting an ABM, providing illustrated comparison that can be used to help inform this decision.

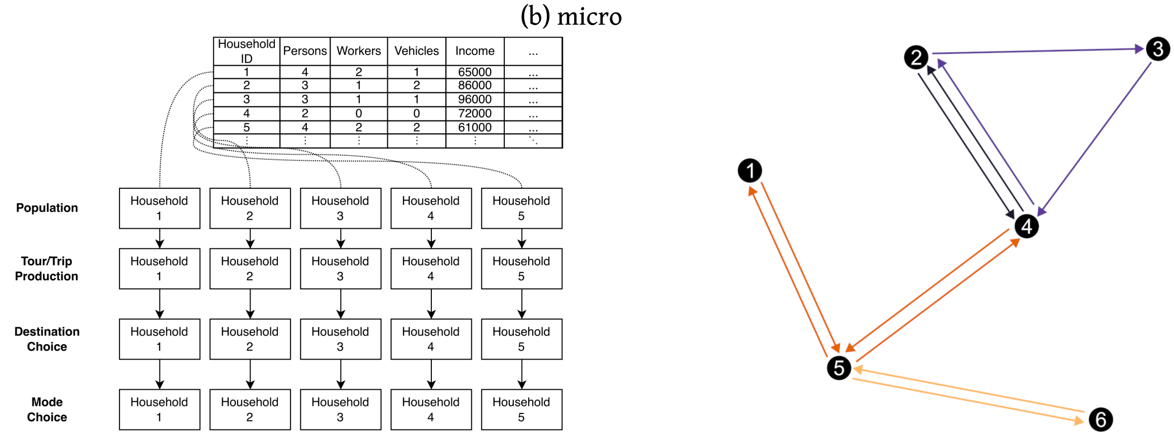
3 flowchart-prelim



(a) tbm



(b) micro



(c) abm

Figure 3.1: Flowcharts

References

Bills, Tierra S., and Joan L. Walker. 2017. "Looking Beyond the Mean for Equity Analysis: Examining Distributional Impacts of Transportation Improvements." *Transport Policy* 54 (February): 61–69. https://doi.org/10.1016/j.tranpol.2016.08.003.

Dong, Xiaojing, Moshe E. Ben-Akiva, John L. Bowman, and Joan L. Walker. 2006. "Moving from Trip-Based to Activity-Based Measures of Accessibility." *Transportation Research Part A: Policy and Practice* 40 (2): 163–80. https://doi.org/10.1016/j.tra.2005.05.002.

Ferdous, Nazneen, Lakshmi Vana, John L. Bowman, Ram M. Pendyala, Gregory Giaimo, Chandra R. Bhat, David Schmitt, Mark Bradley, and Rebekah Anderson. 2012. "Comparison of Four-Step Versus Tour-Based Models for Prediction of Travel Behavior Before and After Transportation System Changes." *Transportation Research Record* 2303 (1): 46–60. https://doi.org/10.3141/2303-06.

Miller, Eric. 2023. "The Current State of Activity-Based Travel Demand Modelling and Some Possible Next Steps." *Transport Reviews* 0 (0): 1–6. https://doi.org/10.1080/01441647.2023.2198458.

Rasouli, Soora, and Harry Timmermans. 2014. "Activity-Based Models of Travel Demand: Promises, Progress and Prospects." *International Journal of Urban Sciences* 18 (1): 31–60. https://doi.org/10.1080/12265934.2013.835118.