A REVIEW OF THEORETICAL AND PRACTICAL ISSUES IN MICROSIMULATING TRANSPORT DEMAND

Eveline Helder, Michiel de bok and Gerard de Jong
Significance
Kurt Verlinden and Cindy Puttemans
MINT nv

1. INTRODUCTION

Disaggregate travel demand models have traditionally been applied by using sample enumeration, that is by calculating the probabilities for each available alternative and applying the computed probabilities to a sample of individuals (possible with expansion factors to make the sample representative of the population). In recent years, especially for activity-based travel demand models, microsimulation has emerged as an alternative way of applying the model to obtain demand forecasts (Walker 2005, Arentze and Timmermans 2000, Goulias and Kitamura 1992). The key difference is that in micro-simulation a discrete choice is allocated to each individual instead of a set of probabilities (that sum to 1 for the individual). This allocation takes place on the basis of random draws from some probability distribution. Having discrete outcomes per individual provides greater flexibility in model application.

An advantage of microsimulation is that, since the decision making in microsimulation is agent based, policy makers will be able to identify with the agents and understand the choices they make.

The strategic passenger model for Flanders (northern half of Belgium) version 4 is currently being developed (see also de Bok et al. 2015, Verlinden et al. 2015), and incorporates such a microsimulation approach. The general set-up of the model is to simulate the entire population of Flanders.

Within the model, tours at the daily level will be simulated for the population of Flanders. These simulations are based on discrete choice models representing the behavior of the strategic passenger model for Flanders estimated mainly on the Flemish Travel Surveys. These models are then used in model application to calculate a probability for each travel alternative the agent can possibly choose, using characteristics of the location of origin and destination and the characteristics of the agent making the trip. From these alternatives and their corresponding probabilities, one alternative is chosen based on a random selection number and stored in memory as a result (this is commonly known as Monte Carlo simulation).

Most transport models assume that the stochastic component of the utility functions follows the Extreme Value Distribution type I (EV I). One way of doing the Monte Carlo simulation then is to take draws from the EV I for each alternative, calculate the utility for each alternative and select the alternative with the highest utility for that individual. This method was used by Horni (2013). Most applications however, take the random component of the utility functions at its mean value (zero) and calculate probabilities first from inserting the model coefficients in the logit probability formula. All alternatives together then form a cumulative distribution function for an individual. Then a draw from the uniform distribution between 0 and 1 is taken and this is mapped onto the cumulative distribution function, which then determines which alternative will be chosen. We will also use this latter, more standard, method of Monte Carlo simulation.

The following problem occurs when implementing a Monte Carlo simulation; the results for an individual agent critically depend on the value that was randomly selected. This means that if the same simulation is replicated, with a different set of random numbers, the results will differ (Castiglione et al., 2003). This is called the simulation error. Several authors have chosen to use fixed random number seeds (usually by individual: microseeding), so that exactly the same results can be obtained when a model run is redone (for an example, see Bowman et al. 2006). However, the basic problem remains that the results of the model depend on the random values that are selected: different random numbers would have given a different outcome, and the choice between sets of random numbers is as arbitrary as can be.

Fortunately, these differences will generally even out when the results are evaluated for a large number of individual agents. However, on smaller scales (especially individual agents and for a choice problem with many alternatives), the decision may vary wildly and for comparison of different scenarios (e.g. policy scenario versus reference scenario) they can even be counter-intuitive. For instance, an agent who, in a reference run, chooses to drive by car to work, might switch to taking the bus in a policy scenario where the prices of public transport increase. This happens because the policy that changes the public transport alternative will also change the location of many other alternatives on the cumulative distribution function, even if these alternatives themselves are not changed. This counter-intuitive result may no longer be noticeable at a zonal or regional level, but nevertheless this is unwanted in a model, as it potentially undermines the trust of the policy makers in the model.

In this paper, we will address these issues from a methodological perspective. We will test the hypothesis that the statistical distribution responsible for the simulation error is the binomial distribution. If this is true, we can predict the size of the simulation error beforehand and therewith calculate the number of replications needed to get a sufficiently accurate outcome before the simulation is started.

The paper is organized as follows. Firstly, we will give an overview over the available literature and the outcomes of our expert interviews. Then, we will characterize the simulation error from a mathematical perspective, using the strategic passenger model for Flanders as an example. We will provide a few equations to predict the number of replications needed to reach a certain accuracy in the model results. Finally, we will describe the implications of randomness in results of microsimulations.

2. A BRIEF REVIEW OF THE LITERATURE OF SIMULATION ERROR IN TRANSPORT MODELS.

The commonly adopted approach towards the fluctuations of results between different replications is to run the model for a number of replications (sets of random draws for each individual) and average the outcomes. Practical examples that average outcomes include: Castiglione et al. (2014), Miller et al. (2003), Vovsha et al. (2008), Freedman et al. (2006), Bowman et al. (2006), Veldhuizen et al. (2000), Cools et al. (2011), Yasmin et al. (2014). Castiglione et al. (2003), give an empirical example how the exact number of times the model needs to be ran critically depends on the level (geographical, population, mode) at which the results will be used in policy studies. As a general rule: the smaller the scale one is interested in, the more replications are needed for the results to reach a substantial accuracy. Usually, the spread in outcomes is evaluated after the model has been run a number of times, and it is determined empirically whether this number was sufficient.

For instance, Castiglione et al. (2003) investigated the number of runs needed to get a reliable result from the San Francisco Model. Their results indicate than 10-20 runs are needed for individual zones, 1 run for a typical neighborhood and 1 run for a typical region.

Cools et al. (2011) took a different approach, they investigate the simulation error for the activity based model FEATHERS (for Flanders). Using a regression analysis, they characterize the relative size of the estimation error. Their results confirm those of Castiglione et al 2003 in that the more detailed the result, the larger the estimation error.

Rasouli and Timmermans (2013) undertook a similar exercise, utilizing the Albatross transport model for the Netherlands. They sort their results according to the size of the traffic flow and find that the coefficient of variation is related to the traffic flow by (traffic flow)$^{-0.3}$.

3. EXPERT INTERVIEWS

Since the published material on how to deal with simulation error in practice is limited, we did asked a number of international experts on the application of activity–based transport models three questions by email. The questions were (somewhat shortened here):

Q1. Do you look at the Monte Carlo simulation outcomes at the individual level?

Q2. Have you come across the issue of counterintuitive results between a reference case and a project case at the individual level?

Q3. What are your experiences with simulation variance in forecasting?

We obtained responses from John Bowman, Peter Vovsha, Mark Bradley and Kay Axhausen/Andreas Horni.

The communis opinio among these experts clearly was that analysis of results at the individual level should be avoided when doing Monte Carlo simulation. This goes for comparing project scenarios against reference scenarios, but also for analyzing a single scenario or for a base year.

In some cases the demand model is applied iteratively in combination with a traffic assignment. This means that several runs are done with the demand model anyway (which can be averaged for the final trip tables), so that there is not really a need to do multiple runs per individual (Vovsha et al. 2008). Other methods for obtaining stable results are microseeding and gradually freezing the results for groups of individuals.

When aggregate study area outcomes are required it might suffice to sample 1 out of 10 individuals and expand the results by 10. For detailed results on the other hand, supersampling (e.g. sample 10 days per individual) could be recommended (response from Bradley).

4. THE STRATEGIC PASSENGER MODEL FOR FLANDERS

Currently, the fourth generation of the strategic passenger transport models for Flanders is being developed. The fourth generation models will be replacing the prevailing model 3.6.1. The fourth generation consists of a freight model (Grebe et al., 2016), and a passenger model.

The main reason for improving on the previous generation model is that the last full population census was in 2001 and there is no new population census planned. This means that new sources of travel data will have to be used. Additionally, both base year and reference forecast year are being updated and the networks and assignment techniques are updated.

The basic set up of the fourth generation of the strategic passenger transport models is therefore substantially different from the previous generation. As a first submodel, the entire Flemish population is being simulated (see de Bok et al., 2015). Starting from the available data, among others the population census of 2001, the evolution of the characteristics of the Flemish population is being simulated up until the base year, and continuing to the reference future year, e.g. 2020.

Taking the simulated population as a starting point, the tours for the population will be simulated. The simulation is based on parameters estimated on the most recent travel surveys (OVG 3-4.5 and OWoWi, see de Bok et al., 2015). The tours for each member of the population are simulated in a few separate sub models, including tour frequency choice and mode destination choice and their discrete results are stored in between. This is done in order to allow for flexible usage of the entire model: one can choose, in applications, to reuse certain intermediate results and continue from there on.

5. SIMULATION ERROR ON THE RESULTS OF STRATEGIC PASSENGER MODEL FOR FLANDERS

5.1 Tour Frequency Submodel

For characterizing the variation of individual results of the strategic passenger model for Flanders, we first focus on the tour frequency submodel. In this submodel, each agent is given a choice between making 0, 1, 2 or 3 tours, by purpose. The choice for the number of tours is incremental; once an agent chooses to make at least one tour, the option to add another tour with the same purpose is offered. A set of random numbers is generated to evaluate the choice made by the agent. Theoretically, one agent can make a substantial amount of tours as 3 tours per purpose can be made per day, but given that the probabilities for making a second or third trip are generally low, this rarely happens in the simulation. This has immediate consequences for the number of tours generated for one zone: the total number possible is probably a lot larger than the number of trips actually generated. For studying the scale of the simulation error, we now focus on the number of tours generated by an origin zone.

For each agent, a set of random numbers is drawn to simulate their choices. In fact, this is not too different from rolling a dice for each agent. If, for instance, the probability of making a certain trip is one third, one can imagine that throwing 5 or 6 will simulate a tour for the agent, 1 to 4 will not. In the real simulation, chances are likely not nice round numbers that can be divided by 1/6, but there is still a similarity with a weighted dice.

The statistics of this dice-rolling experiment can mathematically be described by so-called binomial statistics. For this case in particular, since the number of tours generated is much smaller than the number of tours that can possibly be generated, the statistics are described by a special, more simple, form of binomial statistics: Poisson statistics. Since this is a simpler form, we will first focus on the these statistics, but we will get back to the binomial statistics for the submodel for mode and destination choice in the next paragraph.

Since the different agents in the model have different personal characteristics (gender, age, occupancy, etc.), each of the agents has their own dice, weighed to match their specific characteristics. The number of tours generated from one zone is a sum of all the outcomes of all the throws will all the different dices. We will test below whether the analogy with the mathematical Poisson distribution still holds in this case.

For this paper, we are focusing on the variation of the results, once the experiment is replicated. Generally, this is characterized by the standard deviation on the results. One of the characteristics of the mathematical Poisson distribution is that the standard deviation (σ) on a result is simply the square root of the mean (μ) in the case of the strategic passenger transport models for Flanders:

$$\sigma = \sqrt{\mu}$$

In order to test whether this relation holds, we replicated the outcomes for the number of produced tours per zone of the submodel for tour generation 100 times and calculated the standard deviation on the results. The results are shown in Figure 1.

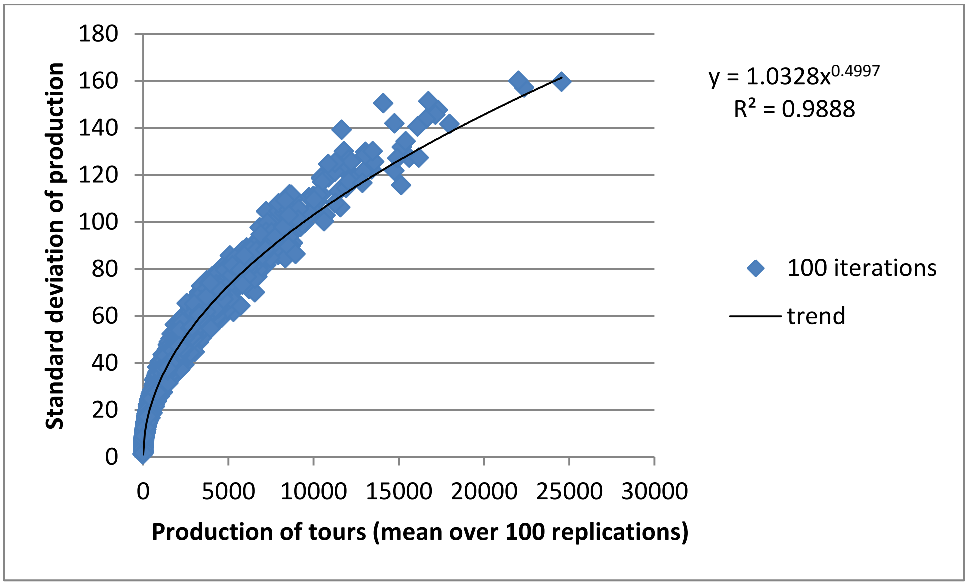

Figure 1: standard deviation over production over 100 replications against the production. The best-fit power law has an index very close to 0.5.

Figure 1 shows the standard deviations of the production for individual zones plotted against the average production itself. A power-law fit to these points gives an index very close to 0.5, this is consistent with the distribution being Poissonian. From the results in Figure 1, we calculated both the square root of the mean and the standard deviation of the productions. We plotted those in Figure 2. If the distribution is really Poissonian, we would expect the square root of the mean for each zone to be similar to the standard deviation. Figure 2 shows this exact relation.

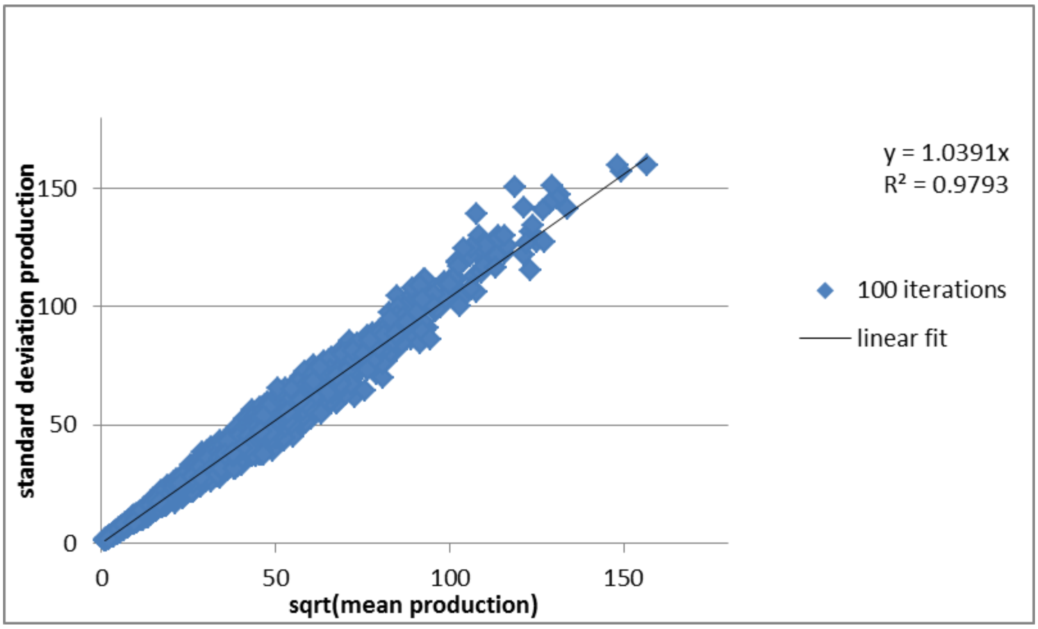

Figure 2: standard deviation against sqrt(mean) of production from all zones for 100 replications.

We can see that the linear fit to the data points has a slope very close to one, and the measure of variability (R) is close to 1. The fact that the slope is very close to one, is an indication that the distribution is indeed very close to a Poisson distribution.

To keep the model as flexible as possible, the results will need to be discretized and stored in memory. This will mean that we do not save averaged outcomes, but only discrete outcomes. This is implemented as follows: per zone and purpose, the average production is determined. From all replications that were used for determining the average, one is chosen that has the closest production to the average of this particular zone and purpose. This will give us a set of discrete outcomes that will be fed into the mode destination model.

5.2 Mode Destination Model

Each of the tours generated in the previous module, will be given a mode and destination combination[1]. This implies that for each tour, there are 6 (modes) times 6756 (zones within Flanders) is 40536 choices. This is the maximum, as some modes are not always available. For instance, the mode walking is available only when the round trip distance is less than 15 km. This is equivalent with rolling a weighted dice with 40536 sides for each person. The standard deviation of the simulation error for this submodel is described by binomial statistics, and has therefore a slightly different form:

$$\sigma = \sqrt{\mu}\sqrt{1-\frac{\mu}{N}}$$

Here, N is the total number of times the dice was rolled. Note that as long as μ is small compared to N, this equation simplifies to the equation from the previous paragraph. Since the number of tours is known and finite, the number of tours choosing one out of the 40536 is finite. The effect will be even stronger: the number of tours that has a sizeable chance for choosing a certain mode-destination combination is rather limited, since people prefer to travel not too long, or some modes might simply be unavailable from the location of departure.

The term $\sqrt{1-\frac{\mu}{N}}$ can be can be calculated exactly, but this is somewhat complicated. Since the number of available alternatives is still large, we expect $\sqrt{1-\frac{\mu}{N}}$ to be relatively close to 1. For evaluating whether the standard

[1] Note that in the final version of the model, the destination and mode will be simulated in two separate modules. The general principles described in this paragraph and paper will remain valid, however.

deviation matches that of a binomial distribution, we will just expect the standard deviation to be slightly less than the square root of the mean and therefore even be relatively close to a Poissonnian distribution. Figure 3 shows the results for all 6 modes, on destination level in a plot similar to Figure 2. Figure 3 shows, again that the square root of the mean is very close to the standard deviation of the results of the 100 replications. It gives a slight overestimation (coefficient in front of x is less than 1 for all modes, especially for walking), but this can be explained by taking into account the $\sqrt{1-\frac{\mu}{N}}$ term.

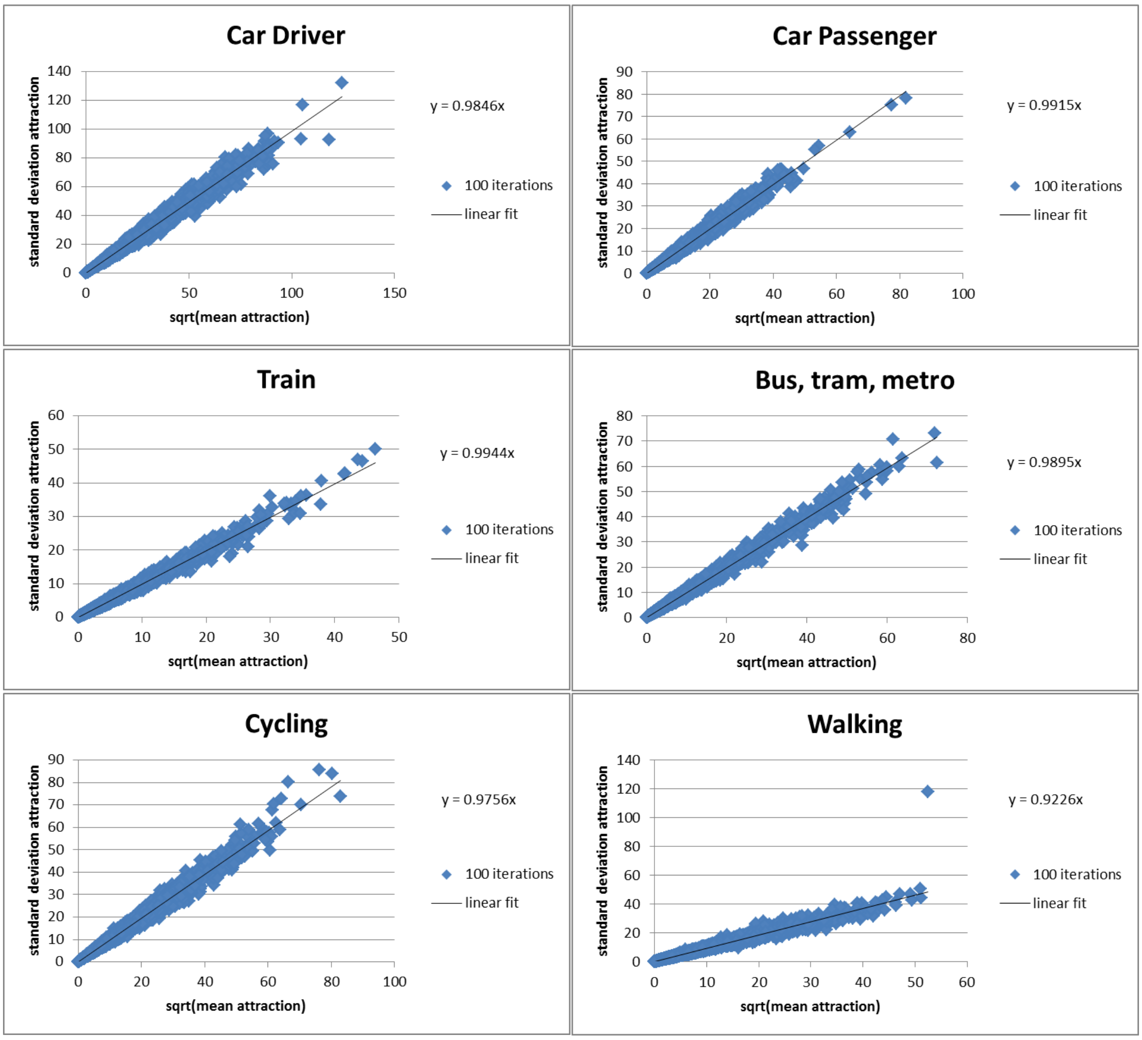

Figure 3: x-axis shows the square root of the number of times a certain destination was chosen, the y-axis shows the standard deviation on the number of times a certain destination was chosen. The model was ran for 100 replications.

Figure 2 and Figure 3 give a very strong indication that we can predict the standard deviation of the results of multiple replications by taking the square root of the mean outcome. This is exactly what one would expect for a binomial distribution where the number of alternatives is much larger than the number of times a certain alternative was chosen.

Again, also for the mode-destination sub model, the strategic passenger model for Flanders requires us to save the results in a discretized manner. For the discretization, we calculate the average destination pattern of tours departing from a certain zone for a certain purpose. From all the replications that went into calculating this average, we choose the replication that provides the closest match to the average pattern. These are the discrete results that will be used.

6. HOW MANY REPLICATIONS DO WE NEED?

There are substantial benefits of using the underlying statistics to calculate the standard deviations.

First of all, being able to predict the standard deviation of a result on a typical outcome puts us in a great position to determine the number of replications that need to be used to get reliable results. The 95% confidence intervals (CI_{95}) of an estimate of a mean from n_{it} replications is the following:

$$\mathrm{CI}_{95} = \frac{1.96\sigma}{\sqrt{n_{it}}} = \frac{1.96\sqrt{\mu}}{\sqrt{n_{it}}}$$

For example, if you expect the number of people taking the bus from zone A to zone B is about 10. However, you want to know this number with a CI_{95} of 1. This means you will have to run the model for $1.96^2 * 10 = 38.42$ replications, rounded off to 39. In general, you would first need to determine which is the smallest scale (for example the smallest number of tours taking a certain mode from zone A to zone B) you will be interested in, as this is going to be the result with the largest uncertainty.

Also, if the model needs to be ran for two scenarios, one can calculate how many replications are needed to determine a statistically significant difference between the two scenarios. When we focus on a certain segment in the results (say, the number of agents taking the bus for a commuting trip between Antwerp and Brussels), scenario 1 gives μ_1 for n_{it1} replications and scenario 2 gives μ_2 for n_{it2} replications. The difference between both scenarios is obviously $\mu_1 - \mu_2$. The CI_{95} of this result is given by:

$$\mathrm{CI}_{95} = 1.96\sqrt{\frac{\mu_1}{n_{it1}} + \frac{\mu_2}{n_{it2}}}$$

7. IMPLICATIONS OF RANDOMNESS

The international experts we consulted in section 3 advised that, in the context of microsimulation, results should not be analysed at an individual level, because of the randomness that will have a too big influence at the most detailed level. Zonal or study area wide forecasts are possible with microsimulation. The question is where the boundary between these two situations will be (e.g. can we predict for a population segment?).

As described before, randomness in results can have some unwanted implications. For instance, some results can become counter-intuitive, therewith potentially undermining the trust of the policy maker in the model.

The rule for determining the standard error on the average outcomes give some handle on when to trust differences between the outcomes on scenarios at face value. This, however, can only be used once we focus on results of the averages of a number of replications.

The strategic passenger model for Flanders usually produces discretized results. As we select the result that is closest to the mean for a certain subselection (often for a certain zone and purpose), a difference in the selected results for this subselection for two scenarios is most likely a real difference. However, once we start evaluating the results of the two scenarios on a scale smaller than the subselection used for determining the best-fit replication the differences are probably not significant.

As an example, we run 100 replications of a certain model for two scenarios. In scenario 1, zone A produces 700 commuting tours and scenario 2 produces 2% more commuting tours: 714. We know that the standard error on this difference is $\mathrm{CI}_{95} = 1.96\sqrt{\frac{700}{100} + \frac{714}{100}} = 7.4$ This difference therefore is significant at more than a 95% level. Also, scenario 1 produces 500 commuting trips for agents with the status "active in work force", and scenario 2 produces only 10% more: 505. Given that this was not a subselection used for selecting the best replication, the aforementioned equation does not apply and we cannot determine whether this difference is significant. However, we do know that the CI_{95} is at least as good as the CI_{95} for one replication. Therefore, we can judge whether the difference is significant by using the following equation: $\mathrm{CI}_{95} = 1.96\sqrt{\frac{500}{1} + \frac{505}{1}} = 62$. Hence, we cannot conclude that the difference is significant.

In contrast, if we focus on a selection larger than the subselection used for selecting the best replication, for instance, the number of tours produced from a certain zone, for all purposes the equation for CI_{95} is valid again. As a general rule, results of microsimulation models can be trusted, as long as you do not focus on results on too small a scale.

In order to minimize the effects of the simulation error on small scales however, the strategic passenger model for Flanders does use the same starting seed for each zone and replication, therewith ensuring that, as long as nothing changes, the tours simulated for the agents will be the same in each replication. This will partially prevent counter intuitive changes in simulated choices, but it will not be prevented for all cases.

CONCLUSIONS AND RECOMMENDATIONS

In this paper, we showed that the results of the strategic passenger model for Flanders can be characterized by a binomial distribution. This results in a simple equation for the 95% confidence interval on the results.

This is particularly useful for an a priori estimation of how many replications are necessary for the desired precision of outcomes. First of all, determine the smallest number of travelers in a certain category you are interested in, then define how accurately you need to know them. This will immediately give you the number of replications needed for producing the result in the desired accuracy.

BIBLIOGRAPHY

Arentze, T., and Timmermans, H. (2000). Albatross: a learning based transportation oriented simulation system. Eindhoven: Eirass.

Bowman, J.L., Bradley, M.A. and Gibb, J. (2006) The Sacramento activity-based travel demand model: estimation and validation results. Paper presented at the European transport Conference, Strasbourg.

De Bok, M., De Jong, G., Baak, J., Helder, E., Puttemans, C., Verlinden, K., Borremans, D., Grispen, R., Liebens, J. and Van Criekinge, M. (2015). A Population Simulator and Disaggregate Transport Demand Models for Flanders. In Transportation Research Procedia, Current practices in transport: appraisal methods, policies and models – 42nd European Transport Conference Selected Proceedings, Vol 8, Pages 168–180.

Castiglione, J., Freedman, J. and Bradley, M. (2003), Systematic Investigation of Variability due to Random Simulation Error in an Activity-Based

Microsimulation Forecasting Model. In Transportation Research Record: Journal of the Transportation Research Board, No. 1831, Transportation Research Board of the National Academies, Washington, D.C., pp. 76–88.

Castiglione, J., Bradley, M. and Gliebe, J. (2014). SHARP 2 Capacity Project C46 Activity-Based travel demand models: A primer, Preliminary draft final report for the Strategic Highway Research Program, Transportation Reserach Board of The National Academies.

Cools, M., Kochan, B., Bellemans, T., Janssens, D., and Wets, G. (2011). Assessment of the effect of micro-simulation error on key travel indices: Evidence from the activity-based model feathers.

Daly, A. (1998). Prototypical sample enumeration as a basis for forecasting with disaggregate models. In Transport planning methods. Proceedings of seminar D held at AET European Transport conference, Loughborough University, , UK, 14-18 September 1998. VOLUME P423.

Freedman, J., Castiglione, J., and Charlton, B. (2006). Analysis of new starts project by using tour-based model of San Francisco, California. Transportation Research Record: Journal of the Transportation Research Board, (1981), 24-33.

Goulias, K. G., and Kitamura, R. Travel Demand Forecasting with Dynamic 43 Microsimulation. In Transportation Research Record, No. 1357, Transportation Research 44 Board of the National Academies, Washington, D.C., 1992, pp. 8–17.

Grebe, S., de Jong, G., Borremans, D., van Houwe, P., and Kienzler and H. P. (2016). Redeveloping the Strategic Flemish Freight Transport Model. In Commercial Transport (pp. 3-21). Springer International Publishing.

Horni, A. (2013). Destination choice modeling of discretionary activities in transport microsimulations (Doctoral dissertation, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 21496, 2013).

Miller, E. and Roorda, M. (2003). Prototype model of household activity-travel scheduling. Transportation Research Record: Journal of the Transportation Research Board, (1831), 114-121.

Rasouli, S., and Timmermans, H. (2013). Probabilistic forecasting of time-dependent origin–destination matrices by a complex activity-based model system: effects of model uncertainty. International Journal of Urban Sciences, 17(3), 350-361.

Veldhuisen, J., Timmermans, H., and Kapoen, L. (2000). Microsimulation model of activity-travel patterns and traffic flows: specification, validation tests, and Monte Carlo error. Transportation Research Record: Journal of the Transportation Research Board, (1706), 126-135.

Verlinden, K., Puttemans, C., de Bok, M., de Jong, G., and Helder, E. (2015) “Micro-simulation with discrete choice models: application in Flanders”, these ETC proceedings.

Vovsha, P., Donnelly, R. and Gupta, S. (2008). Network equilibrium with activity-based microsimulation models: the New York experience. Transportation Research Record: Journal of the Transportation Research Board, (2054), 102-109.

Walker, J. (2005). Making household microsimulation of travel and activities accessible to planners. Transportation Research Record: Journal of the Transportation Research Board, (1931), 38-48.

Yasmin, F., Morency, C. and Roorda, M. J. (2014). Macro-, Meso-, and Microlevel Validation of an Activity-Based Travel Demand Model. In Transportation Research Board 93rd Annual Meeting (No. 14-1257).