Establishment of the Swiss LCV model

Raphaël Ancel Nicole A. Mathys

Federal Office for Spatial Development
July, 26th 2024

Abstract

This report describes the Swiss model for Light Commercial vehicles (LCVs). It describes the behavior of all LCVs registered in Switzerland except (i) those owned by physical persons, which are modeled in the national model for passenger transport (NMPT) and (ii) those registered by companies of the branch "postal and courier activities" (NOGA code 53), which are modelled separately. The Swiss model for LCVs is tour-based and generates a trip matrix. Road assignment takes place outside the LCV model, within the NMPT.

The model structure, the estimation of all parameters as well as the tests carried out to validate the model are described. Guidelines for model applications and a short descriptions of the openly available model scripts are also provided.

1 Introduction

Light commercial vehicles (LCVs) are defined in the European Union and in Switzerland by their gross weight, which must not exceed 3.5 tonnes, and by their targeted usage: the carriage of goods. Their usage goes however beyond the carriage of goods: a survey conducted in 2013 in Switzerland found that 53% of LCVs were used for service trips [Federal Statistical Office, 2015a]. LCVs have gained the attention of public authorities because of their central role in city logistics and their increasing number on the roads: the annual number of LCV registrations in Switzerland has surged by 82% between 1999 and 2019 [Federal Statistical Office, 2022] and the latest forecasts of the Swiss government [Federal Office for Spatial Development, 2021a] projects a 58% increase of LCV vehicle kilometers between 2017 and 2050.

This report documents the establishment of Swiss national LCV-model. This model is designed to (i) provide a reasonable description of the current average daily traffic of LCVs at the road level and (ii) to permit the translation of long-term demographic and economic projections into traffic forecasts suitable for infrastructure planning. The LCV-model integrates with the national models dedicated to the mobility of persons [Federal Office for Spatial Development, 2020] and heavy freight [Federal Office for Spatial Development, 2015]. The applied methodology, inspired from Hunt and Stefan [2007], is vehicle-based and allows for a rather fine description of the touring behavior.

This paper is articulated as follows. Section 2 presents how the LCV model integrates with the other national traffic models. Section 3 provides a more extensive and context-specific definition of a LCV and outlines the main uses of these vehicles. Section 4 covers briefly the literature dedicated to freight and service traffic modeling. Section 5 describes the data available. Section 6 presents the chosen branch-based segmentation. Section 7 describes the model itself. Section 8 describes the model implementation, and covers various issues related to the model application, including an analysis of the variability of the results and a comparison with traffic counts.

2 Integration of the LCV model within the Swiss traffic modeling landscape

The integration of the Swiss LCV model within the Swiss traffic modeling landscape is presented in Fig. 1. The Swiss LCV model describes the behavior of all LCVs except (i) those owned by physical persons (which we refer to as "private LCV") and (ii) those registered by companies of the branche "postal and courier activities (NOGA code 53) which are both modeled separately. We refer hereafter to LCV registered by juridical entities located in Switzerland as "juridical LCV".

The Swiss LCV model and the parcel LCV model (briefly described in Appendix B) both produce a trip matrix. These two matrices are added to form a single matrix with juridical LCVs. This trip matrix for juridical LCV is then sent to NPVM (the national passenger transport model) for traffic assignment. Other inputs of this traffic assignment include the heavy truck matrices (rigid and articulated trucks) of the Swiss model for heavy freight transport (AMG) as well as other trip matrices generated within NPVM (including "private LCV").

Note that there is no integrated feedback loop between the network assignment and the generation of the juridical LCV matrix and heavy freight matrices. In other words, the travel time matrices are considered as constants within the Swiss LCV model, the parcel LCV model and the AMG.

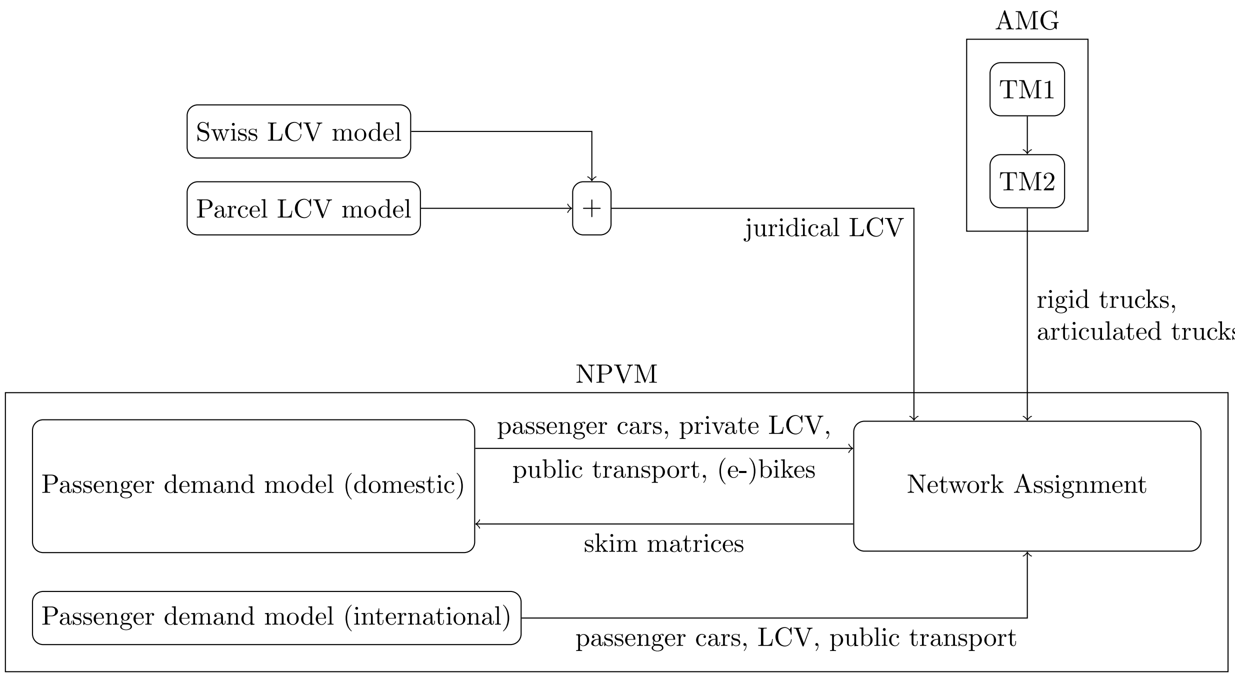

Figure 1: Swiss traffic modelling landscape

3 Scope

In the transport literature, the demand is usually segmented by trip purpose (e.g. specific models for freight transport, for private mobility in general or for more specific purposes, for instance tourism). Here, we focus instead on a vehicle type (LCV) and on an owner type (juridical). This is dictated by the data available. This section provides a more precise definition of LCVs and describes the various trip purposes often associated to the use of such vehicles.

3.1 Definition of a LCV

The definition of light commercial vehicles we adopt corresponds to the homologation class N1 of the Swiss and EU classifications of vehicle types. These vehicles must be "designed to carry goods" and have a gross vehicle weight up to 3.5 tons. The Swiss ordinance on the technical requirements for road vehicles (SR 741.41) provides a more precise definition of what it means for a vehicle to be "designed to carry goods": it "includes the vehicles equipped with additional folding seats in the load compartment for the occasional and non professional transport of persons, provided that the total number of seats, including the driver's seat, does not exceed 9" (translated from French). Minivans like the Renault™ Kangoo™ can be classified either as passengers cars (when they have true rear seats) or as LCVs (when they have no rear seats or only folding ones). Passenger cars are excluded from our analysis.

In terms of the vehicle categories defined in the national vehicle registre [Federal Roads Office, 2022], the adopted definition of LCVs covers the categories 30 (delivery vehicles) and 36 (light articulated vehicles) as well as the vehicles of category 38 (articulated trucks) with a gross weight up to 3.5 tons. Delivery vehicles (category 30) represent 97 % of the total.

3.2 Vehicle purposes

LCVs are mostly used for business purposes. This broad category encompasses *freight*, *service trips with goods*, *service trips without goods* and *passenger transport services* such as coaches and taxis [Federal Office for Spatial Development, 2021b]. While there is no statistics on the distribution of LCVs between these purposes, we expect LCVs to be mostly active in the the sub-segments *service trips with goods* and *freight*, and much less in the sub-segments *service trips without goods* and *passenger transport services*. Within the sub-segment *service with goods*, we expect the LCV to be the most common vehicle type, while a minority of trips might be done with passenger cars. In the *freight* sub-segment, many trips are carried out by heavier trucks, railways, ships, and to some extent, plane. LCVs have however a particular role to play in the *freight* sub-segment, as they are often used for first/last-mile transportation. They only account for a tiny share of ton-kilometers but are nonetheless very significant in terms of vehicle-kilometers traveled and are key for the logistics chain. LCVs may also be of relative importance in the segment of *service trips without goods*, although we expect passenger cars to be more commonly used (together with public transit and soft modes of transportation - for instance when going to a meeting).

4 Literature review

The following literature review is structured according to the type of method used: it describes first the models that are based on the four-step model of travel demand and then the alternative approach followed for the Swiss LCV model, which builds on Hunt and Stefan [2007].

For a more comprehensive literature review specific to freight applications, the reader is referred to de Jong et al. [2021]. For methods addressing business trips in particular, Ellison et al. [2017] is a good reference.

4.1 Adjusted versions of the Four-Step Model

Most freight and business trip models follow the logic of the traditional four-step model of travel demand. The first two steps (generation and distribution) and the last step (network assignment) are usually handled at an aggregate scale (i.e. without distinguishing individual agents), as when modeling the movements of people. The third step, which corresponds to mode choice when modeling people, is where most adjustments need to be done. Some studies simply skip it - e.g. Cambridge Systematics and COMSIS Corporation [1996] - while others expand it by considering additional logistics choices, such as shipment size/frequency (for the transport of goods only), trip-chaining strategy, or departure time. We shortly describe hereafter three approaches from the literature:

- The ADA approach [de Jong and Ben-Akiva, 2007] addresses the third step in a **disagreggate** fashion, hence the name (ADA stands for Aggregate-Disaggregate-Aggregate). It focuses on freight and has been applied to several countries (Norway, Sweden, Denmark and ongoing work in Austria). It encompasses the choice of shipment size/frequency, of the number of trips in the transport chain and of mode and vehicle type for each trip. Yet, with a high level of detail come large data needs. The standard ADA approach requires a commodity flow survey to be available, describing the full trip chain. Such surveys can be expensive to conduct and are not very common. The application of ADA in Austria required such a survey to be conducted specifically for the needs of model development - see Grebe et al. [2020]. We believe that applying such an approach in Switzerland would be prohibitively expansive.

- Wang and Holguín-Veras [2008], Thoen et al. [2020], Sakai et al. [2020] only focused on the third step and proposed detailed solutions for modeling trip chains in a **disaggregate** fashion. These papers assume a known demand (for instance shipments) and build tours sequentially by successively adding destinations. Such approaches ensure that all the demand is satisfied in an efficient manner. It is not appropriate for Switzerland, because we have very little data about the demand for LCV trips.

- The commercial software PTV-VISUM, which is already used in Switzerland for modelling the mobility of persons, provides with its "TB-Freight" module a solution for the modeling of trip-chains. Tour trips are generated in an **aggregate** fashion: the heuristic behind the model produces a trip matrix which is consistent with a pre-determined producer-consumer matrix. A parameter allows to define the number of intermediary trips that are generated for each producing zone, thereby allowing to control the average number of stops per tour [PTV GROUP, 2019]. This method is aggregate in the sense that it only generates trip matrices, without generating individual tours. A drawback is that this method cannot control that the distribution of tour length is reasonable. In cases with isolated zones, TB-Freight tends to produce a large number of internal trips in these zones, thereby implying that the few tours serving these zones have an extremely large number of stops. For an application to Switzerland, the main challenge would be to estimate the producer-consumer matrix.

4.2 Models without underlying demand

To avoid the difficulties associated with the explicit consideration of shipments and the data needs that come with it, Hunt and Stefan [2007] proposed a model of tours which is purely vehicle-based. This type of approach seems to be the most suitable in the Swiss case, given the data available. This model is disagreggate in that it considers individual vehicles, but it is also aggregate in that it considers transport zones instead of individual establishments. Every zone is the base for a number of tours. These tours are associated to various purposes and vehicle types. The total number of tours is estimated based on land-use and economic characteristics of the zone via a regression analysis. Stops are then generated one by one, via an iterative procedure. After every stop, a module called *Next Stop Purpose* determines whether the vehicle goes back to its establishment or continues its tour. If another stop should be made, the module *Next Stop Location* determines the next destination. Yet another logit-model determines the stop duration.

Here, the destination is not dictated by a demand matrix like in the models described in Section 4.1, but it is chosen based on variables such as the population size, the number of employees, land-use, and the generalized travel cost between the current zone and the destination. As all the key inputs required for such a model are also available in Switzerland, we adopted a similar approach.

5 Data and zones

5.1 LCV survey

The main data source is a vehicle-based survey conducted in 2013 in Switzerland [Federal Statistical Office, 2015a]. The survey was conducted with two types of questionnaires. Questionnaires of type 1 only contained a few questions about the mileage and the main use of the vehicle during a reference day. It is useful for calibration purposes and to estimate the total vehicle kilometers travelled by LCVs. Questionnaires of type 2 contained the same questions as those of type 1, but also questions about the individual trips and shipments transported during that day (postal codes of origin and destination, weight, type of good, mileage of the vehicle at the origin and destination). This second part however was only required for vehicles that carried at least 50 kg of goods during the reference day. Overall, according to the survey report [Federal Statistical Office, 2015b]:[1]

- Questionnaires of type 1 were sent to the owners of 40'000 LCVs. The response rate was 76%.
- Questionnaires of type 2 were sent to the owners of 28'000 LCVs. The response rate was 71%, but the part with questions about individual trips and shipments was only completed for 3'874 LCVs.

Besides, to limit the respondents' burden with questionnaires of type 2, a simplified form was proposed for tours with shipments of a single type containing at least four different destinations within a circular area of radius 10 km (this concerns mainly postal and parcel deliveries). This simplified form does not contain the information for every single shipment, but only the overall weight and the mileage at the first and last stop, as well as at the origin.

The 3'874 questionnaires with detailed data correspond to 1.2% of the entire LCV fleet. These surveys contain detailed information about 11'473 individual shipments, and summarizing information (i.e. simplified forms) for 830 tours.

Because the time required to fill the questionnaire of type 2 increases with the number of shipments handled, these questionnaires might be biased. Thus, when it is possible (for instance for the daily distance traveled or the vehicle purpose), we only rely on the answers to the questionnaires of type 1.

5.2 National vehicle register

The second most important data source is the Swiss national vehicle register [Federal Roads Office, 2022]. For this project we used the standard dataset "BEST_R". This dataset contains various pieces of information for each vehicle registered in Switzerland. The attributes that are most relevant for our application are the type of owner (private or business), the vehicle category (e.g. 30, 36 and 38 - see Section 3.1) , the curb weight, the gross weight and the postal

[1] These were the planned numbers. Light variations in the implementation.

code of where the vehicle is registered. The branch of the vehicle owner is not available in the original data but was inferred by coupling this register with the national business register [Ancel and Mathys, 2024].

5.3 Zone definitions

The model estimation and application rely on two different zoning systems:

- For model estimation, we relied on the postal codes (state 2013), which is the most precise geographical information contained in the LCV survey. These correspond to a partition of Switzerland in 3'187 geographical areas.

- For model applications, we use the zones of the national model for passenger transport [Federal Office for Spatial Development, 2020] (about 8'000 zones). This choice simplifies the integration of the results in the national model for passenger traffic (for network assignment in particular).

The two zoning systems have many common limits because the zones of the national passenger transport model were obtained by subdividing the municipalities (state 2010).

To be able to compute skim matrices, the centroids of the geographical areas defined by postal codes must be connected to the nodes of the network via connectors. These were created automatically with VISUM, by ensuring that the average speed and the minimum travel time on connectors were consistent with those defined for the original zones of the Swiss national model for passenger transport.

5.4 Skim matrices

Travel time and distance matrices were obtained using the national model for passenger transport, for the reference state 2017, for both zoning systems.

As trips remaining within a zone cannot be assigned to the road network, we had to rely on some assumptions to define the distances and travel times of internal trips. For the 8'000 zones of the national transport model (which are relatively small), we assumed an internal trip length (resp. travel time) equal to the one to the nearest neighboring zone. For the zones defined by postal codes (which are larger), we assumed an internal trip length (resp. travel time) equal to half the one to the nearest neighboring zone.

The generalized cost was then computed as a weighted sum of the distance and travel time matrices. The weights correspond to the cost figures for LCVs that are recommended by the Swiss norm VSS 74 827 for cost-benefit analyses (0.5553 CHF/km and 0.4890 CHF/h for the year 2016), which are then actualized for the target year using the producer price index for goods transport by road, as reported by the Federal Statistical Office [2023].

6 Segmentation

The chosen segmentation relies on the branch of business owners. This facilitates the establishment of long-term forecasts, in which population and economic forecasts usually serve as inputs. We renounced to integrate additional segmentation criteria such as engine type, gross vehicle weight or vehicle purpose. These could admittedly be useful to test differentiated political measures, but it is not clear yet which criteria are likely to be the most important in future. As a more detailed segmentation comes at the expense of statistical significance, we chose to keep only the branch as segmentation criteria.

The branch segmentation is based on the Swiss General Classification of Economic Activities (NOGA 2008). We used only the first level (sections) of this classification, except for the branch H, in which we isolated the postal and courier activities (2^{nd} level NOGA code (division): 53). The branches corresponding to the first level of the NOGA classification are listed in Table 1.

Table 1: First level (sections) of the NOGA 2008 classification (source: `https://www.kubb-tool.bfs.admin.ch/en`)

A	agriculture, forestry and fishing
B	mining and quarrying
C	manufacturing
D	electricity, gas, steam and air-conditioning supply
E	water supply; sewerage, waste management and remediation activities
F	construction
G	wholesale and retail trade; repair of motor vehicles and motorcycles
H	transportation and storage
I	accommodation and food service activities
J	information and communication
K	financial and insurance activities
L	real estate activities
M	professional, scientific and technical activities
N	administrative and support service activities
O	public administration and defence; compulsory social security
P	education
Q	human health and social work activities
R	arts, entertainment and recreation
S	other service activities
T	activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	activities of extraterritorial organisations and bodies

The distribution of juridical LCVs by owner type and branch estimated in Ancel and Mathys [2024] is summarized in Table 2. To ensure a sufficient number of observations in each segment, we further aggregate the branches with

Table 2: Estimated distribution of LCVs by owner type and branch (state: June 2022), source: Ancel and Mathys [2024]

Owner type	NOGA08	Vehicles	Vehicle share [%]	Surveys (type 2)
Business	A	3’767	1	61
	B	626	0	0
	C	38’999	9	440
	D	4’696	1	40
	E	2’836	1	31
	F	128’807	30	737
	G	37’452	9	484
	H (49-52)	15’857	4	177
	H (53)	7’811	2	45
	I	2’358	1	40
	J	2’328	1	15
	K	2’992	1	21
	L	2’794	1	14
	M	10’776	3	76
	N	30’977	7	172
	O	12’765	3	65
	P	1’161	0	10
	Q	2’877	1	45
	R	1’275	0	12
	S	2’489	1	34
Private	-	109’444	26	1’355*
Total	-	423’086	100	3’874

* This includes some business-owned vehicles, whose branch could not be determined.

less than 100 questionnaires of type 2 (in gray in the table) for all modeling steps which rely on the LCV survey (i.e. all except the first one, *Vehicle Generation*). This is considered acceptable as these branches account individually for only up to 3 % of the fleet and together for 13 %.

The distribution of vehicles across branches is quite similar to the distribution of surveys across branches. This is positive, because it ensures that the most significant segment are also those where we can best characterize the vehicle behavior.

In the LCV survey, the private LCVs cannot be distinguished from juridical LCVs whose branch could not be identified. These vehicles represent a very large share, because the matching of surveys to branches was done at the time of the survey (2013-2014) with a strict matching criterion. The matching procedure considered in Ancel and Mathys [2024] considers alternative matching criteria, which allowed identifying the branch for a larger proportion of vehicles.

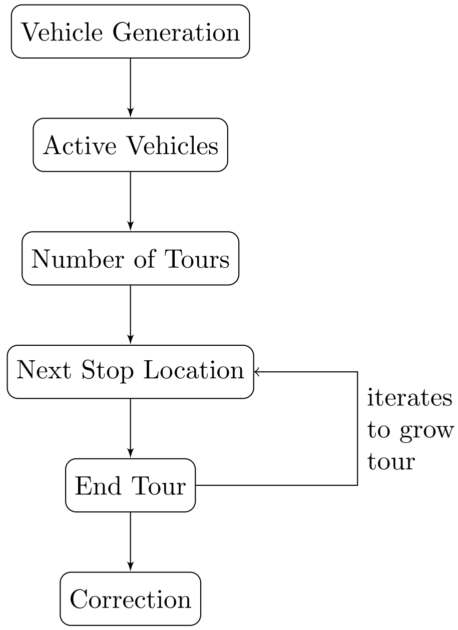

Figure 2: Model structure

7 Model

7.1 Global structure

The adopted model structure, largely inspired from Hunt and Stefan [2007], is illustrated in Fig. 2. It consists of six modules:

- *Vehicle Generation*: provides an estimate of the number of juridical LCVs per zone and by branch (see Section 6).
- *Active Vehicles:* determines which proportion of vehicles is active on a given day.
- *Number of Tours:* assigns a number of tours to each active vehicle.
- *Next Stop Location:* determines which zone is visited next.
- *End Tour:* determines whether the vehicle makes an additional intermediate stop or goes back to its establishment.
- *Correction:* applies a branch-specific correction factor on the trip matrices, to better reproduce the daily distance travelled.

Table 3: Overview of the different modules

Module	Method	Data used for estimation	Explanatory variables
Vehicle Generation	average rates per employee	Vehicle register 2022 at postal code level, matched with business register	jobs
Active Vehicles	average rates	LCV survey 2013, type 1 and 2	none
Number of Tours	average rates	LCV survey 2013, type 2	none
Next Stop Location	multinomial logit	LCV survey 2013, type 2	population, jobs, generalized travel cost, land-use
End Tour	binomial logit	LCV survey 2013, type 2	number of stops, generalized travel cost
Correction	average rates	LCV survey 2013, type 1	none

7.2 Module *Vehicle Generation*

This module converts the number of full-time equivalents per zone and per branch into a number of registered vehicles per zone and branch. This is achieved using the average vehicle rates, derived from empirical data. The LCVs are generated using branch-specific rates expressed in LCV per job in the branch (full-time equivalent, FTE hereafter).

Estimates of the number of vehicle per branch in Switzerland for June 2022 are provided in Ancel and Mathys [2024]. The number of FTE come from from two different sources:

- Branch A (state: 2021). Source: *Labour productivity by by economic sector at current prices* (FSO, 2023);[2]
- All other branches (state: 2$^{\text{nd}}$ quarter 2022). Source: *Full-time job equivalent per sector* (FSO, 2023). Note: this source does not contain any statistics for the branch A (agriculture), where the number of FTE is often more difficult to estimate.

The resulting vehicle generation rates are listed in Table 4.

[2]For methodological reasons, the federal office of statistics provides estimates of the FTE for the branch A one year after the other branches.

Table 4: Vehicle generation rates

NOGA08	LCVs for 1'000 FTE
A	38
B	138
C	63
D	172
E	158
F	380
G	73
H (49-52)	88
H (53)	247
I	12
J	14
K	14
L	60
M	28
N	126
O	75
P	5
Q	5
R	20
S	22

The branch F (construction) is the one with the most LCVs per employee (380 LCVs for 1'000 employees), followed by the branches H(53) (postal and courier activities) with 247 LCVs for 1'000 employees.

7.3 Module *Active Vehicles*

This module determines which proportion of vehicles are used on an average day (Monday-Sunday) or an average weekday (Monday-Friday). This proportion is segment-specific (see Section 6) and was computed using both types of surveys (type 1 and 2). The results are provided in Table 5.

The probabilities are relatively similar across branches. The vehicles which are used the most frequently (p(active)=0.50, i.e. 3.5 days per week in average) are those of the branch H (49-52) (transportation and storage, except postal and courier activities).

7.4 Module *Number of Tours*

The active vehicles were then converted into a number of tours, using segment-specific average rates (see Table 5). To compute these average number of tours, we relied on the observations of LCV surveys of type 2. As they are not originally grouped into tours, the identification of tours relied on various assumptions (see

Table 5: Parameters of the modules *Active Vehicles* and *Number of Tours*

NOGA08	p(active) Mo-Su	p(active) Mo-Fr	Tours/day (when active)
C	0.45	0.59	1.58
F	0.47	0.63	1.43
G	0.46	0.59	1.43
H (49-52)	0.50	0.63	1.3
N	0.45	0.61	1.69
Other	0.47	0.61	1.75

Appendix A.3). Note that these average numbers of tours per vehicle and per day relate to active vehicles only.

7.5 Module *Next Stop Location*

The module *Next Stop Location* consists in a multinomial logit model, where the alternatives are the different zones. In the estimation process, the alternatives are the zones defined by postal codes while in the application, the alternatives are the zones of the national passenger transport model. This precludes the use of zone-specific constants.

To reduce the computational burden in the estimation process, we relied on stratified importance sampling. Instead of considering a choice set made of all possible destination zones[3], we only kept 300, including the one chosen in the empirical data: 100 among the 200 zones which are the closest from the current one (in terms of generalized cost), 100 among the next 600 zones, and 100 among the furthest 2'387 zones (resp. 2'386, see[3]). The differences in terms of probability to be sampled were then accounted for in the estimation process by adding $\log(p_j)$ to the utility function, where p_j must be proportional to the inverse of the probability that alternative j is sampled ($p_j = 2$ for the 200 closest zones, $p_j = 6$ for the next 600 zones and $p_j = 23.87$, resp 23.86 for the remaining 2'837 zones, resp. 2'386 zones) - see McFadden [1978] for a theoretical background to stratified sampling and Li et al. [2005] for a practical application similar to ours.

Let b denote the base of the tour and i the current zone. Let $C_{i,j}$ denote the generalized cost (as defined in Section 5.4) by LCV from zone i to any zone j

[3] There are 3'187 possible destination zones for first trips of a tour and 3'186 for subsequent trips (trips originating from a zone other than the base zone and ending in the base zone are considered as return trips and do not involve any destination choice).

(in CHF). The chosen utility specification is the following:

$$\begin{aligned}
U_j = \sum_{\text{land_use}} & \theta_{\text{land_use}} \times \delta_{\text{land_use}} \\
& + (1 - \delta_{\text{same_ZIP}}) \times [(\theta_C + \theta_{C,\text{first}} \times \delta_{\#\text{stops}=1}) \times C_{i,j}/100 \\
& + \theta_{C>50} \times \max(0, C_{i,j} - 50)/100] \\
& + \theta_{\text{same_ZIP}} \times \delta_{\text{same_ZIP}} \\
& + \log(\text{population in zone } j + \theta_{\text{jobs/pop}} \times \text{jobs in zone } j) \\
& + \log(p_j).
\end{aligned}$$

The θ's represent the coefficients to be estimated while the δ's represent dummy variables. Instead of utilizing zone-specific constants, we have constants relative to the land-use. The following land-use types are considered:[4]

- Low density: at most 100 inhabitants/km^2 and 100 jobs/km^2,
- Residential: more than 100 inhabitants/km^2 and a density of population at least twice as large as the density of jobs,
- Intermediary: at most 3'000 jobs/km^2 and neither a residential area nor a low density area,
- Employment Node: more than 3'000 jobs/km^2 and neither a residential area nor a low density area.

The dummy variables associated to the various land-uses ensure that only the constant $\theta_{\text{land\_use}}$ corresponding to the land-use of the considered destination zone enters in the utility function.

The disutility associated to the generalized cost from the current zone to the candidate destination is captured for trips outside the same ZIP code via a piece-wise linear relation. For trips that are not the first of a tour, every additional monetary unit between 0 and 50 CHF brings a negative utility of $\theta_C/100$, then every additional monetary unit above 50 CHF brings a negative utility of $(\theta_{C,\text{later}} + \theta_{C>50})/100$. A similar piece-wise linear relation applies for first trips of a tour, but θ_C is replaced by $(\theta_C + \theta_{C,\text{first}})$. The factor 1/100 is here to ensure that the estimated coefficients have approximately the same scale.

For trips inside the same ZIP code, the skim matrices are not very reliable and not consistent between the two zoning systems. Therefore, we chose not to use the skim matrices for such trips and estimate instead a constant parameter $\theta_{\text{same\_ZIP}}$. This can be relatively easily transferred to the transport zones (the postal codes are usually closely related to the municipalities and the transport zones were defined by splitting the municipalities of 2010).

The number of potential attractors (inhabitants and employees) in a zone enters the utility function via the so-called "size term". The logarithm accounts for the fact that a zone is actually an aggregation of individual destinations (i.e.

[4] This is adapted from Hunt and Stefan [2007].

houses or establishments) - see McFadden [1978] for the theoretical derivations. The parameter $\theta_{\text{jobs/pop}}$ quantifies the relative attractiveness of jobs compared to inhabitants. Segments where the customers are mostly businesses are expected to have a large $\theta_{\text{jobs/pop}}$. Note that this size term enters the utility function without being multiplied by a coefficient θ_{size} (in other words, θ_{size} is forced to one). This is because we want the model to be transferable to a different zoning system. This would not be possible with a value of θ_{size} different from 1, because this would imply a non-linear dependency between the zone size and the probability that a zone is chosen - see Daly [1982] for more details.

The models for the various segments were estimated with the R package "Apollo" [Hess and Palma, 2019] using a dataset extracted from the LCV survey (see Appendix A.2). During the estimation process, the explanatory variables whose robust t-ratio was found to be smaller in absolute value than 1.8 were discarded, except for the parameters $\theta_{\text{jobs/pop}}$.[5] The resulting models are presented in Table 6.

[5] The parameter $\theta_{\text{jobs/pop}}$ quantifies the relative significance of jobs with respect to inhabitants. Assuming that jobs have no importance at all is certainly not better than using a parameter associated with some sizeable uncertainty.

	C	F	G	H (49-52)	N	Other
$\theta_{\text{land\_use}}$ Low Den	2.25***	1.59***	1.46***	2.07***	0.94**	1.29***
	(0.36)	(0.29)	(0.21)	(0.47)	(0.57)	(0.36)
$\theta_{\text{land\_use}}$ Residential	1.27***	0.55**	0.71***	0.92**	0.53	0.62**
	(0.29)	(0.25)	(0.18)	(0.40)	(0.45)	(0.29)
$\theta_{\text{land\_use}}$ Interm	1.27***	0.46***	0.77***	0.80**	0.61*	0.70***
	(0.23)	(0.18)	(0.16)	(0.36)	(0.38)	(0.23)
$\theta_{\text{land\_use}}$ Emp Node [ref]	0.00	0.00	0.00	0.00	0.00	0.00
$\theta_{\text{same ZIP}}$	0.00	−0.44**	−0.41**	−1.10***	0.00	0.00
		(0.21)	(0.24)	(0.40)		
θ_C [100 CHF]	−9.03***	−9.07***	−9.56***	−8.90***	−9.21***	−10.48***
	(0.50)	(0.42)	(0.52)	(0.94)	(1.02)	(0.58)
$\theta_{C>50}$ [100 CHF]	5.91***	4.29***	5.41***	5.86***	5.11***	7.31***
	(0.68)	(0.65)	(0.63)	(1.05)	(1.33)	(1.02)
$\theta_{C,\text{ first}}$ [100 CHF]	0.00	0.00	1.35***	1.27***	0.00	0.00
			(0.33)	(0.47)		
$\theta_{\text{jobs/pop}}$	3.07*	0.84*	1.86***	3.26*	0.63	1.24*
	(2.08)	(0.51)	(0.63)	(2.51)	(0.57)	(0.79)
No Observations	877	1223	1171	561	337	742
Log Likelihood (Null)	−4198	−7888	−6375	−1566	−2067	−4383
Log Likelihood (Converged)	−2667	−5003	−4131	−1167	−1238	−2485
Rho-squared	0.36	0.37	0.35	0.26	0.40	0.43

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Values in parentheses represent robust standard errors.

Table 6: Estimation results for *Next Stop Location*

The positive land-use coefficients have to be interpreted as a larger attractiveness for the corresponding land-uses **compared to employment nodes** (which was taken as reference by forcing the corresponding $\theta_{\text{land\_use}}$ to be zero), and **all things being equal** (in particular, for a given population size and number of jobs). The relatively high value of $\theta_{\text{land\_use}}$ for low density areas suggests that between 2 candidate destinations with identical distances to the current zone and the base and an identical number of inhabitants and jobs, the zone with the smallest population and job density is more likely to be selected as destination by LCVs. This is true for all demand segments. This is intuitive, because less dense areas are usually better suited for establishments of the primary and secondary sectors, which are more often served by LCVs.

To facilitate the interpretation of the coefficients associated to the generalized cost to the destination, we illustrated the corresponding contribution to the utility function in Fig. 3. We observe that:

- as expected, the utility of each segment is a monotone decreasing function of the generalized cost,
- for all segments, the disutility increases rapidly with the generalized cost until 50 CHF, and then less rapidly,
- for the branches G (wholesale and retail trade) and H (transportation and storage), the utility decreases not as fast for the first trip of a tour as for subsequent trips. This indicates a delivery pattern where several destinations close to each other are grouped within a tour, producing long first and last trips and shorter intermediary trips.

The ρ^2 indicator corresponds to McFadden's pseudo-R squared. It compares the estimated model against the model where all alternatives have equal probabilities to be selected. This is admittedly not ideal as this is discarding also the term $\log(p_j)$, which implies that ρ^2 depends on the sampling strategy chosen. The fact that ρ^2 is not as good for the branch H (transportation and storage) as for other branches might be explained by the larger trip distances that are typically observed for this branch (i.e. the choice of destination is less constrained by the proximity).

7.5.1 Comparison of survey data and simulation results

Fig. 4 compares the trip length and trip duration distributions resulting from a simulation of the year 2013 with those observed in the LCV survey (also from 2013). Note that to facilitate the comparison, we did not use the empirically registered trip distances, but the distances computed with the national model for passenger transport for the trips observed in the data. Note also that we did not consider for this figure the trips remaining within the same ZIP code, because the skims of the two zoning systems are especially inconsistent for such trips.

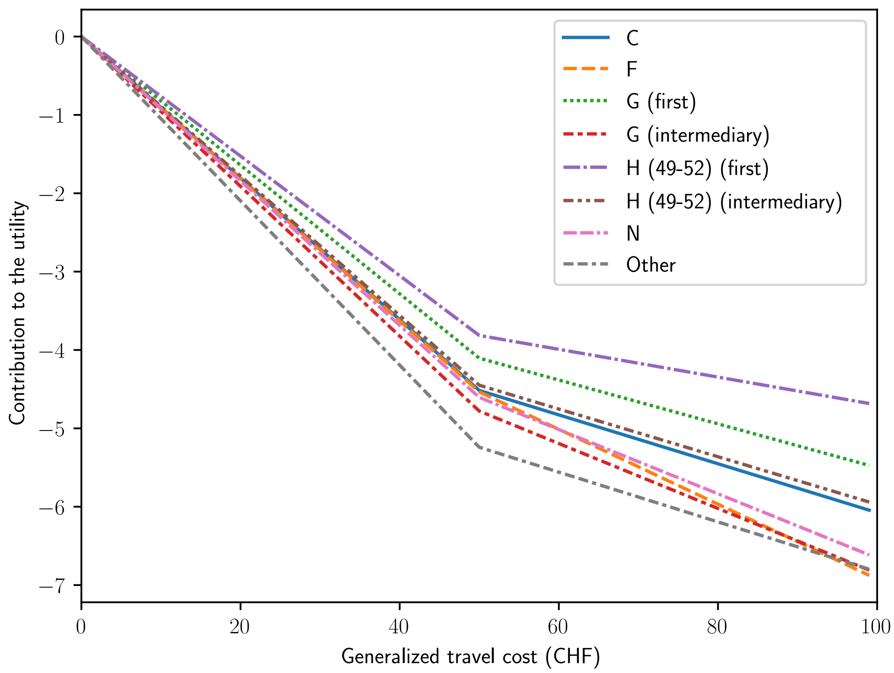

Figure 3: Contribution of the generalized travel cost to the utility.

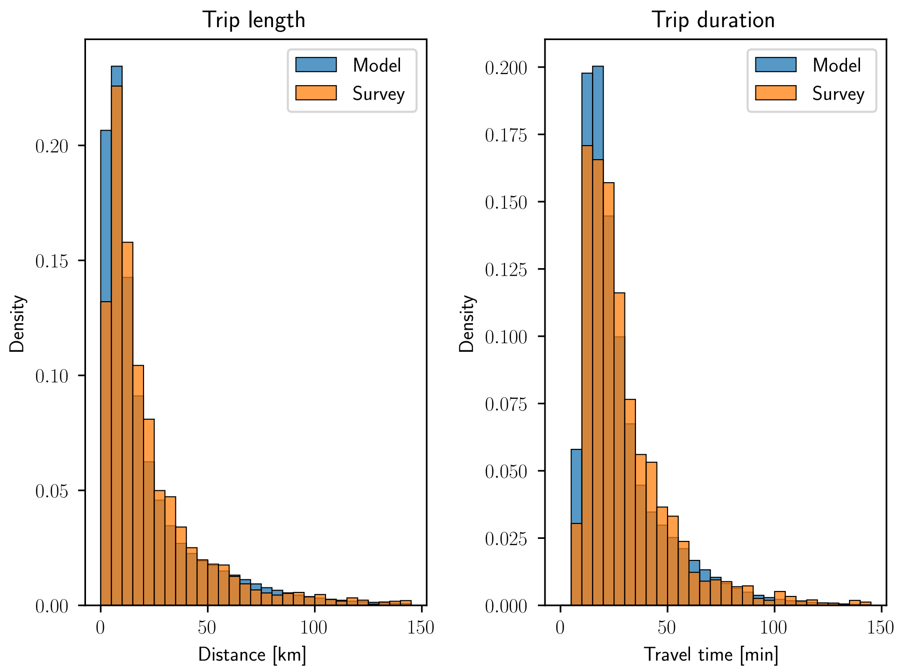

Figure 4: Comparison of trip length and duration distributions between the model results (for the year 2013) and the LCV survey for trips where origins and destinations have different ZIP codes.

Although the trip length and trip duration distributions from the model are relatively close to those observed in the data, there are some noticeable differences: the model tends to overestimate the proportion of short trips (distance up to 10 km or travel times up to 20 min) and to slightly underestimate the proportion of trips in the range 10-40 km, respectively 20-55 min. Several reasons might explain this discrepancy: the different zoning systems, slight differences between the vehicle fleets in the data and in the simulation (see Table 2), and modeling approximations. Overall however, the fit is considered relatively good and the piece-wise linear utility specification with respect to the generalized cost seems to be sufficient.

Although the travel times and distances for trips within the same ZIP code are difficult to compare (these correspond to internal trips for the LCV survey), we can compare the proportion of such trips: in the empirical data, after the data cleaning steps described in Appendix A.2, 13.3 % of trips remain in the same ZIP code. In the simulation, it is 14.2 %. The difference is also considered acceptable.

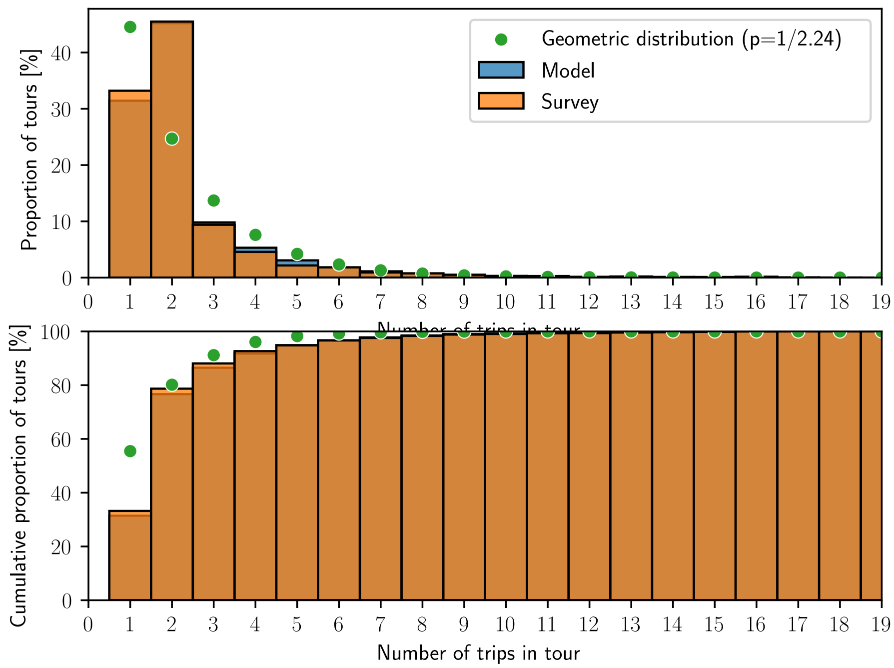

Figure 5: Distributions of the number of trips in a tour in the model results (for the year 2013) and in the LCV survey (type 2).

7.6 Module *End Tour*

The module *End Tour* determines when a tour should end and thus the distribution of the number of trips per tour. This distribution among LCV surveys of type 2 is shown in Fig. 5, together with a geometric distribution having the same average number of trips per tour (2.24). One can see that the empirical distribution has fewer tours with only 1 trip than the geometric distribution, more tours with 2 trips, fewer tours with 3, 4, 5 trips, but more tours with at least 6 trips). The longest observed tour includes 19 trips.

Considering these differences with a geometric distribution, we chose the following specification:

$$\begin{aligned} U_{\text{return-to-establishment}} =& 0 \\ U_{\text{continue}} =& \theta_0 + \theta_{\text{2stops}} \times \delta_{\#\text{stops}=2} + \theta_{\log(\#\text{stops})} \log(\#\text{stops}) \\ &+ \theta_{C,\text{return}} \times \text{Generalized cost from current zone to base}, \end{aligned}$$

where the θ's are coefficients to be estimated and $\delta_{\#\text{stops}=2}$ is a dummy variable equal to 1 if and only if the number of stops made so far (including the base) is equal to 2.[6] The parameter $\theta_{2\text{stops}}$ is here to account for the particularly high probability to make a return trip after exactly two stops (including the base). We expect it to be negative. The number of stops made so far also influences the probability to continue via its logarithm ($\log(\#\text{stops})$). Because tours with a high number of trips are relatively frequent, we expect $\theta_{\log(\#\text{stops})}$ to be positive.

Note that the number of tours with only 1 trip is not affected by the module *End Tour*, but only by the module *Next Stop Location*. Indeed, such tours are produced when the first chosen destination has the same postal code as the base. In this case, we always consider that the tour is over, the module *End Tour* is not used and no return trip is generated.

[6] The number of trips in a tour is equal to the number of stops minus 1.

	C	F	G	H (49-52)	N	Other
θ_0	-2.09^{***}	-0.03	0.60^{***}	0.73^{***}	-0.26	0.23^{*}
	(0.24)	(0.12)	(0.19)	(0.28)	(0.28)	(0.18)
θ_{2stops}	0.00	-1.00^{***}	-0.97^{***}	-1.09^{***}	-1.03^{***}	-1.24^{***}
		(0.15)	(0.17)	(0.30)	(0.28)	(0.25)
$\theta_{\log(\#\text{stops})}$	1.48^{***}	0.00	0.00	0.00	0.00	0.00
	(0.18)					
$\theta_{C,\text{return}}$	0.00	0.00	0.42^{**}	0.51^{**}	1.42^{***}	0.00
			(0.21)	(0.23)	(0.61)	
No Observations	587	897	1032	538	269	475
Log Likelihood (Null)	-352	-724	-671	-173	-208	-337
Log Likelihood (Converged)	-302	-646	-617	-147	-185	-302
Rho-squared	0.14	0.11	0.08	0.15	0.11	0.10

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Values in parentheses represent robust standard errors.

Table 7: Estimation results for *End Tour*

In this case, the ρ^2 indicator compares to a situation where the probability to return to the establishment would always be 0.5. The positive value of ρ^2 indicates that we do better than such a simplistic model. The values of ρ^2 remain however relatively small, i.e. our model only explains a small part of the variability.

The estimated values of θ_{2stops} are all negative, as expected. The only value of $\theta_{\log(\#\text{stops})}$ found to be significantly different from 0 is positive, also as expected. The estimated values of $\theta_{C,\text{return}}$ are positive, i.e. vehicles whose previous stop is far away from their home establishment are more likely to make an additional stop. This is also intuitive, as the incentives to group trips into tours is greater when the destinations to be visited are far away.

7.6.1 Comparison of survey data and simulation results

As can be seen in Fig. 5, the simulated distribution of the number of trips per tour is relatively close to the empirical distribution, which suggests the chosen utility specification is sufficient for our needs. If we consider the distributions more in detail, we see that the proportion of tours with one or two trips is slightly underestimated by the model and conversely, the proportion of tours with at least three trips is slightly overestimated. These differences might be due to modeling limitations and differences in the fleets of modeled vehicles.

7.7 Module *Correction*

Because the most influencing variables when making forecasts is the relative growth of each branch, it is paramount that the model reproduces well the contribution of each branch to the global vehicle-kilometers traveled (VKT). Two important modeling limitations tend to artificially reduce the modeled VKT:

- The recorded distances of the trips retained for model estimation (see Appendix A.2) is on average 13 % larger than the distance according to the model. This bias is due in particular to some trips having a very large recorded distance and a very small modeled one (see Fig. 6). This suggests that intermediary stops were not reported in these cases;

- LCV surveys of type 2 containing tours that are described with the simplified form (see Appendix A.2) could not be used in the model estimation. These LCV tours of type 2 had a daily distance 5.7 % larger than other LCV tours of type 2.

The module *Correction*, placed at the end of the simulation, aims at circumventing these limitations. It associates trips of each segment with an artificial weight (or correction factor), such that the final average daily distance traveled is consistent with the observations in the LCV survey of type 1. This correction factor is given as the ratio of:

- the average daily distance for active vehicles of this segment according to the LCV Surveys of type 1 (consistently with the statistics of the FSO,

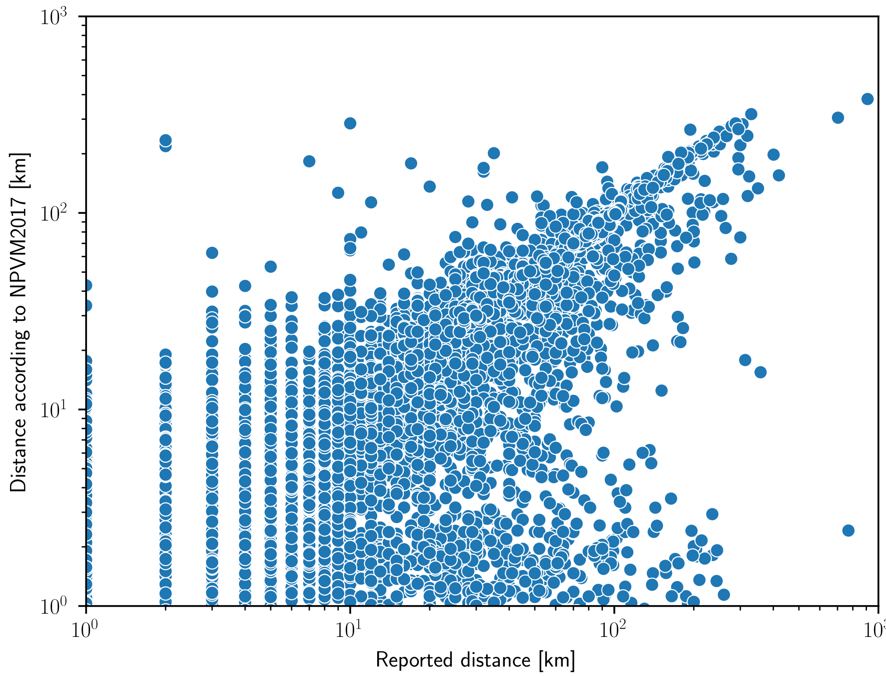

Figure 6: Recorded distance and distance according to our skim matrix for trips of LCV surveys of type 2, after the data cleaning steps described in Appendix A.2.

which relies only on these surveys to estimate the VKT for Switzerland).

- the average daily distance obtained in the simulation, i.e. the ratio of the VKT of that segment (calculated with the skim matrix from the simulation) divided by the number of active vehicles in this segment.

Table 8 shows the average daily distance resulting from the model (before correction) and that from the LCV survey (type 1). As we see, the model consistently underestimates the VKT. The resulting correcting factors range between 1.12 for the branch H (i.e. we artificially increase the number of trips by 12 %) and 1.44 for the branch N (i.e. we artificially increase the number of trips by 44 %). After applying this correction, the model perfectly reproduces the average daily distance per segment.

Table 8: Daily distance traveled per branch in simulation (for the year 2013, before the matrix correction) and in the LCV Surveys of type 1 - active vehicles only.

NOGA08	Simulation [km]	LCV Type 1 [km]	Ratio
C	74.2	92.4	1.25
F	51.1	65.7	1.24
G	86.6	120.9	1.39
H (49-52)	137.1	153.9	1.12
N	51.6	74.6	1.44
Other	54.7	69.8	1.28

8 Model application

8.1 Implementation

The model was implemented in Python, while the code used to estimate the discrete choice models is in R. The Python and R scripts are available online on GitHub (`https://github.com/AREschweiz/LCV_model`). We list hereafter the main scripts:

Repository *Data preparation*:

- `link_IVZ_and_BUR_modular_division.r`: links the vehicle register with the establishment register to compute the number of LCV per branch.
- `construct_tours_from_lcv_data.py`: reads the original dataset with the detailed trip informations from the LCV survey type 2 and generates datasets with shipments, trips and tours.
- `compute_model_parameters_from_surveys.py`: computes the parameters for the modules *Active Vehicles* and *Number of Tours* (Table 5), as well as the average daily distance per branch, required by the module *Correction*.
- `create_estimation_data_for_end_tour.py`: reads the dataset of trips, the skim matrices, the explanatory variables and generates a dataset to be used to estimate discrete choice models for the module *End Tour*.
- `create_estimation_data_for_next_stop_location.py`: reads the dataset of trips, the skim matrices, the explanatory variables and generates a dataset to be used to estimate discrete choice models for the module *Next Stop Location*.
- `estimate_end_tour.r`: estimates the coefficients of the binomial logit model for the module *End Tour*, it needs to be adjusted for each segment.
- `estimate_next_stop_location.r`: estimates the coefficients of the multinomial logit model for the module *Next Stop Location*, it needs to be adjusted for each segment.

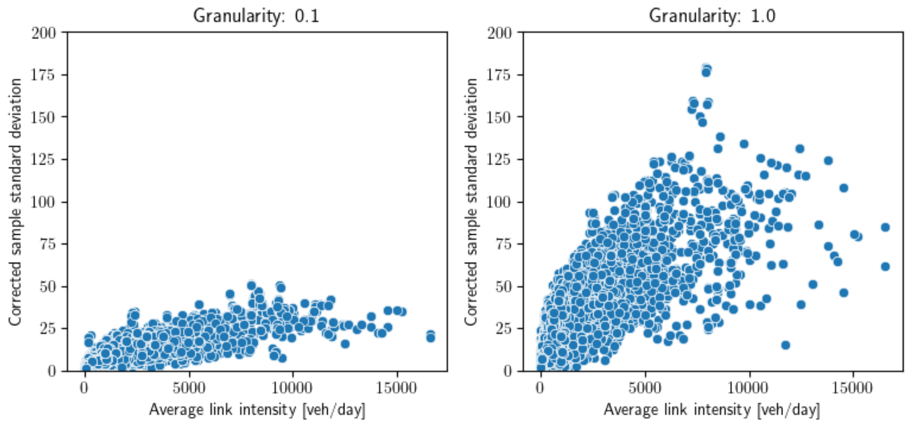

Figure 7: Corrected sample standard deviation as a function of the average link intensity, for five realizations, with $g = 0.1$ (link) and $g = 1$ (right).

Repository *Simulation*:

- `config.yaml`: Configuration file.
- `main.py`: runs the LCV model
- `support.py`: contains auxiliary functions which are called by other scripts.

8.2 Variability of the simulation results

A microsimulation approach inherently comes with simulation variance. To form discrete tours, discrete choices need to be sampled from probabilities calculated with logit models. In this section, we analyze the extent to which this simulation variance causes deviations from expected link intensities (including international traffic and parcel deliveries).

For the analysis we use two different settings for the granularity parameter (g): 1.0 and 0.1. With $g = 1$, every vehicle is simulated once. With $g = 0.1$, small units of 0.1 vehicle are modeled, which is equivalent to simulating every vehicle 10 times and averaging the result. Due to the law of large numbers, the simulations results with $g = 0.1$ are expected to show less variability and to be closer to the expected values than the results with $g = 1$.

Fig. 7 shows for every link in the road network the corrected sample standard deviation as a function of the average link intensity, both being computed for a set of five realizations. As expected, the variability is much smaller with $g = 0.1$ than with $g = 1$. With $g = 0.1$, the standard deviation for a road with 5'000 veh/day is about 20 veh/day. This is considered acceptable.

8.3 Loop detector data for validation or calibration

The official vehicle classification for traffic counts in Switzerland is called Swiss10. Within this classification, the categories 5, 6 and 7 all correspond to LCVs: rigid LCVs, LCVs with trailer, articulated LCVs. The definition of a LCV in the Swiss10 classification is based on measurable characteristics of vehicles, such as their mass or their shape. It differs from the definition used in our model, which is based on the Swiss ordinance 741.41 (see Section 3.1). As a consequence, some vehicles considered in our model as LCV would be considered as passenger cars according to the Swiss10 classification (this is the case for instance for small vans like Renault Kangoo Van, Fiat Doblo Cargo, Ford Transit Courier, Volkswagen Caddy Cargo, Peugeot Partner, Citroen Berlingo Van). Conversely, mini-buses are considered as LCVs within the Swiss10 classification, but not according to the ordinance 741.41.

In addition, the correct identification of a vehicle class requires detectors to be regularly re-calibrated. This re-calibration is generally not done frequently enough to be able to confidently distinguish LCVs from passenger cars. As a result, the different generations of detectors maintained by the Federal Roads Office indicate significantly different shares of LCVs.

In light of these limitations, we believe that the traffic count data available is not appropriate to validate or further calibrate our model. In order to allow such a validation or calibration, a preliminary step would be to distinguish within the vehicle register the LCVs that are also considered as such in the Swiss10 classification from those that are considered as passenger cars. This could possibly be done using model names of the vehicles, or vehicle characteristics (such as the curb weight or gross weight). These different types of LCVs could then be modeled separately and then calibrated/validated with the corresponding Swiss10 categories. This might be done in a future version of the model.

In the current state, in case the final demand matrix would need to be calibrated to better reproduce the traffic counts, we recommend to do this simultaneously for cars and LCVs and aggregate traffic counts for all vehicles having a gross weight up to 3.5 tons.

8.4 Inputs required to apply the model

In order to apply the model, the following inputs for the target year are required:

- Population for each zone of the transport model,
- Number of jobs (FTE) per branch for each zone of the transport model,
- Producer price index for goods transport by road.

The names of the corresponding files must be indicated in the file `config.yaml`. The first two inputs (population and FTE per zone) can be computed for past years using disaggregate data of the STATPOP and STATENT (access is limited and requires a contract).

References

R. Ancel and N. Mathys. Who owns light commercial vehicles? an empirical analysis for switzerland. In *24^{th} Swuiss Transport Resarch Conference*, 2024.

Cambridge Systematics and COMSIS Corporation. Quick response freight manual, 1996. prepared for the United States Federal Highway Administration, Washington, DC, USA.

A. Daly. Estimating choice models containing attraction variables. *Transportation Research Part B: Methodological*, 16(1):5–15, 1982.

G. de Jong and M. Ben-Akiva. A micro-simulation model of shipment size and transport chain choice. *Transportation Research Part B: Methodological*, 41 (9):950–965, 2007.

G. de Jong, M. de Bok, and S. Thoen. Seven fat years or seven lean years for freight transport modelling? Developments since 2013. *Journal of Transport Economics and Policy*, 55(2):124–140, 2021.

R. B. Ellison, C. Teye, and D. A. Hensher. Modelling sydney's light commercial service vehicles. *Transportation Research Part A: Policy and Practice*, 96: 79–89, 2017.

Federal Office for Spatial Development. Aggregierte Methode Güterverkehr (AMG): Methodenbeschrieb, 2015.

Federal Office for Spatial Development. Modelletablierung Nationales Personenverkehrsmodell (NPVM) 2017 - Schlussbericht, 2020.

Federal Office for Spatial Development. Schweizerische Verkehrsperspektiven 2050: Schlussbericht, 2021a.

Federal Office for Spatial Development. Wirtschafsverkehr in urbanen Räumen, 2021b.

Federal Roads Office. National register of vehicles, 2022.

Federal Statistical Office. Erhebung leichte Nutzfahrzeuge 2013, 2015a.

Federal Statistical Office. Erhebung leichte Nutzfahrzeuge 2013: Erhebungsbericht, 2015b. FSO Number: do-d-11.05-LWE13-D21.

Federal Statistical Office. New registrations of road vehicles by vehicle group and vehicle type, 2022. FSO Number: cc-e-11.03.02.02.03.

Federal Statistical Office. Produzentpreisindex dienstleistungen, 2023. URL `https://www.bfs.admin.ch/bfs/en/home/statistics/prices/producer-prices-import-prices/services.assetdetail.24005320.html`.

S. Grebe, G. de Jong, C. Wampera, C. Obermayer, and R. Pompl. VMÖ–a new strategic transport model for austria. *Transportation Research Procedia*, 49: 95–106, 2020.

S. Hess and D. Palma. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, 32:100170, 2019.

J. D. Hunt and K. Stefan. Tour-based microsimulation of urban commercial movements. *Transportation Research Part B: Methodological*, 41(9):981–1013, 2007.

M.-T. Li, L.-F. Chow, F. Zhao, and S.-C. Li. Geographically stratified importance sampling for the calibration of aggregated destination choice models for trip distribution. *Transportation Research Record*, 1935(1):85–92, 2005.

D. McFadden. Modelling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickers, and J. W. Weibull, editors, *Spatial Interaction Theory and Planning Models*. North-Holland Publishing Company, 1978.

PTV GROUP. PTV Visum 2020 Manual, 2019.

T. Sakai, A. Romano Alho, B. Bhavathrathan, G. D. Chiara, R. Gopalakrishnan, P. Jing, T. Hyodo, L. Cheah, and M. Ben-Akiva. Simmobility freight: An agent-based urban freight simulator for evaluating logistics solutions. *Transportation Research Part E: Logistics and Transportation Review*, 141:102017, 2020.

S. Thoen and M. de Bok. Audit and update of the Swiss national LCV model, 2023.

S. Thoen, L. Tavasszy, M. de Bok, G. Correia, and R. van Duin. Descriptive modeling of freight tour formation: A shipment-based approach. *Transportation Research Part E: Logistics and Transportation Review*, 140:101989, 2020.

Q. Wang and J. Holguín-Veras. Investigation of attributes determining trip chaining behavior in hybrid microsimulation urban freight models. *Transportation Research Record: Journal of the Transportation Research Board*, 2066(1):1–8, 2008.

A Appendix - LCV Survey: data preparation

A.1 Description of the dataset

The data of the LCV survey is separated into two main files : a vehicle file (LWE_2013_vehicle.csv) and a transport file (LWE_2013_transport.csv). The vehicle file contains one row per vehicle surveyed. This row contains various information about the vehicle, its main use and its mileage on the reference day. This represents the data that is common to both survey types (1 and 2). The transport file contains one row per "transport". In most cases, a transport is a quantity of goods of a given type (e.g. food) which are transported together (i.e. picked up and dropped off simultaneously). Each row then contains information about the vehicle which carried out these transports (including a unique identifier) and information about the transport itself, such as the weight and type of the transported goods, the postal code of their origin and destination, as well as the vehicle mileage at the origin and destination. "Empty transports" must also be reported. "Empty transports" correspond to vehicle movements without any goods on board. The reporting of such movements allow the reconstruction of the entire trip chain of a vehicle over a day.

In addition to these "regular" and "empty" transports, the transport file also contains lines corresponding to "grouped pick-ups" and "grouped deliveries". In the case of grouped pick-ups, several shipments of the same type of good having the same destination but different origins are combined in a single row. Similarly, grouped deliveries correspond to several shipment of the same type of good having the same origin but several destinations. In such cases, the file does not contain the individual origins and destinations and weights of each shipment. In the case of grouped deliveries, it only contains the total weight of goods, the total number of destinations, the postal code and mileage at the common origin, as well as at the first and last destination.

A.2 Data cleaning

A.2.1 General

Some data cleaning steps where necessary before estimating parameters on LCV surveys of type 2. Three modules rely on this data : *Number of Tours*, *Next Stop Location* and *End Tour*. When estimating their parameters, we always excluded the surveys where the simplified form was used (grouped pick-ups and grouped deliveries), because we do not have enough information to model such tours. These represent 14 % of surveys of type 2. In addition, some additional criteria were used for the modules Next Stop Location and End Tour.

A.2.2 Data for the module *Next Stop Location*

In addition to surveys containing tours described with the simplified form, we excluded:

- **Trips where either the origin or the destination could not be localized:** This can be due to two reasons: either because a given postal code corresponds to some large organizations, and not to a geographical area, or because a given postal code corresponds to a zone outside Switzerland. This corresponds to 5.7 % of trips after excluding surveys with grouped pick-ups and deliveries.

- **Internal trips with long reported distances**: We noticed a large proportion of trips starting and finishing in the zone of the home establishment, some of them having large recorded distances. This suggests that an intermediate stop is missing. We discarded such observations when the recorded trip length was at least twice as large as the distance to the nearest neighboring zone. This corresponds to 6.4 % of trips after excluding surveys with grouped pick-ups and deliveries and trips whose origin or destination could not be localized.

- **All return trips**: The last trips of closed tours do not include any destination choice (the destination is the base of the tour).[7]

The file for the estimation process includes the chosen alternative (i.e. destination zone), various alternative-specific attributes for the chosen alternative as well for other sampled alternatives (such as the number of jobs, inhabitants, travel time from the current zone to the candidate destination), as well as some general data, such as the statistical weight of this questionnaire (reported by the FSO). For more details, see the python script `create_estimation_data_for_next_stop_location.py`.

A.2.3 Data for the module *End Tour*

The data necessary for the estimation of the module End Tour is also derived from the set of trips. If a trip is a return trip (i.e. its destination has the same postal code as the base of this questionnaire), then the decision taken is to end the tour. Otherwise the decision taken is to make another stop. Here as well, we had to discard some observations:

- **Trips originating from the base**: by definition, trips originating from the base zone are first trips of a tour and therefore cannot be return trips.

- **Trips where either the base or the current zone could not be localized**: this corresponds to 6.2 % of trips, after excluding surveys containing grouped pick-ups and deliveries and trips originating from the base.

The file for the estimation process contains the chosen alternative (end tour or continue) as well as various variables related to the choice situation (e.g. distance from the current zone to the base, number of stops made so far). For

[7] Trips originating and ending in the base zone are not considered as return trips and are therefore left in the dataset.

more details, see the python script `create_estimation_data_for_end_tour.py`.

A.3 Reconstructing trip trips and tours

In order to reconstruct the trip trips, we first listed all the stops (i.e. origins and destinations) made by each vehicle. If the same mileage appears several times in the list of origins and destinations of shipments of a vehicle, these are considered as one single stop. We then sorted these stops by increasing mileage, and defined as trips the interval between two consecutive stops.

To reconstruct the tours, we first had to define a base zone. As no base is indicated in the data, we defined the base of a vehicle to be the postal code at which this vehicle started its first trip trip of the day.

For more details, the reader is referred to the python script `construct_tours_from_lcv_data.py`.

B Appendix - Parcel delivery

This appendix provides a brief description of the parcel delivery model, developed by Significance [Thoen and de Bok, 2023]. For more details, refer to the report available in this project's GitHub repository.

B.1 Model description

Unlike the LCV model which starts by generating the vehicles, the parcel delivery model starts from the demand, i.e. the parcels to be delivered. It then assigns them to a courier company, group those to be delivered within the same tour and then constructs a delivery tour (see Fig. 8).

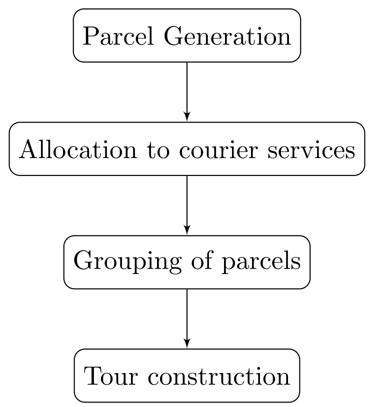

Figure 8: Model structure

The resulting trips are then added to those produced by the LCV model, so that the combined demand can then be assigned to the route network. This last step occurs within the national model for passenger traffic.

The model takes as inputs the population and jobs within each zone of the model, the distance matrix, the market shares of the major courier companies as well as the location of the depots used by each company for parcel delivery. The main parameters are the average number of parcels per person per day, the average number of parcels per job per day, the average number of delivery attempts per parcel and the maximum number of parcels per van.

B.2 Strategy to avoid double-counting

To avoid modeling the same vehicles both in the LCV model and in the parcel delivery model, the entire LCV fleet of the branch 53 is discarded from the LCV model. This means that vehicles of the branch 53 which are used for an activity other than delivering parcels are not modelled. This solution seemed acceptable to us, given that the branch 53 only accounts for about 2 % of LCVs (see Table 2), and that the number of modelled parcel delivery tours (5'844 tours for 2021) is of the same order of magnitude as the number of vehicles of the branch 53 (7'811 in June 2022 - see Table 2). As some vehicles might do two delivery tours per day, the total share of vehicles that are not modelled at all is estimated to be around 1 %.