Description of the Swiss LCV model
Technical report (preliminary version)

Raphaël Ancel Andreas Justen Nicole Mathys

Federal Office for Spatial Development
September, 26th 2023

Abstract

This report describes the Swiss model for Light Commercial vehicles(LCVs). This model covers almost all uses of LCVs, including private usage, freight transport and service trips. Only the postal and courier activities (NOGA code 53) are modelled separately (see Appendix C). The Swiss model for LCVs is tour-based and generates a trip matrix. Road assignment takes place separately, within the national model for passenger transport.

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. They are also becoming increasingly common: 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 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 generic model described in the body of this report describes the behavior of the entire fleet of LCVs registered in Switzerland, except for LCVs of the branch “postal and courier activities”, which are modeled separately (see Appendix C). 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 clarifies the scope of this paper and provides a more extensive and context-specific definition of a LCV. Section 3 provides a brief review of the literature dedicated to freight and service traffic modeling. Section 4 describes the data available. Section 5 presents the chosen branch-based segmentation. Section 6 describes the model itself. Section 7 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 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. This is dictated by the data available. This section provides a more precise definition of the vehicle type we are interested in and of the trip purposes often associated to the use of such vehicles.

2.1 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.

2.2 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] (in German). 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 with goods” and “freight”. Within the sub-segment “service with goods”, we expect the LCV to be the most common vehicle type. In the “freight” sub-segment however, many trips are carried out by heavier trucks, railways, ships, and to some extent, plane. LCVs have a particular role in this 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. LCVs may also be of relative importance in the segment of services without goods, although we expect passenger cars to be more commonly used (together with public transit and soft modes of transportation).

Within the freight segment, the market of parcel delivery is of special interest because of its very dynamic growth and because LCVs often come into use. To be able to better describe their characteristics, the LCVs of the branch “postal and courier activities” are modeled separately - see Appendix C.

3 Literature review

Freight and business trips are rarely modeled together. For a 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. The methods used in these two fields share however many similarities. Here, we treat them together and structure our literature review according to the type of method used: we describe first the freight and business trip models that are based on the four-step model of travel demand and then describe the alternative approach followed for the Swiss LCV model, which builds on Hunt and Stefan [2007].

3.1 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, 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 list hereafter three approaches from the literature:

- The Aggregate-Disaggregate-Aggregate (ADA) freight model system [de Jong and Ben-Akiva, 2007] focuses on freight and has been applied to several countries (Norway, Sweden, Denmark and ongoing work in Austria). It describes in a **disagreggate** fashion the choice of shipment size/frequency, of the number of legs in the transport chain and of mode and vehicle type for each leg. 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]. There is no such survey available in Switzerland.

- Wang and Holguín-Veras [2008], Thoen et al. [2020], Sakai et al. [2020] 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 such that all the demand is satisfied in an efficient manner.

- 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 legs 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 and that guarantees vehicle conservation. 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 generate 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.

3.2 Models without underlying demand

To avoid the difficulties associated with the explicit consideration of shipments, 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 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 3.1, but it is chosen based on variables such as the population size, the number of employees, land-use, and the travel cost between the current zone and the destination.

4 Data and zones

4.1 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 shipments transported during that day (postal codes of origin and destination, weight, type of good, mileage of the vehicle at the origin and destination). Empty legs also had to be registered, thereby allowing for the reconstruction of all vehicle movements. 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 shipments was only completed for 3'874 LCVs.

Besides, to limit the respondent burden with questionnaires of type 2, a simplified form was proposed for tours containing at least four different locations distant from each other from at most 20 km with shipments of a single type. 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.

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

4.2 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 2.1) , the curb weight, the gross weight and the postal 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 (see Appendix B).

4.3 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.

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.

4.4 Skim matrices

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

As trips remaining within a zone cannot be simulated, we had to rely on some assumptions to define the distances and travel times of internal trips. For the 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 scaled for the target year using the producer price index for goods transport by road, as reported by the Federal Statistical Office [2023].

5 Segmentation

The chosen segmentation relies on the owner type (private person or business/institution), as well as 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 owner type and the branch as segmentation criteria.

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

The estimated distribution of LCVs by owner type and branch is summarized in Table 2. To ensure a sufficient number of observations in each segment, we further aggregated the branches with 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 %.

Note that 26 % of LCVs are registered as privately-owned. This is a relatively high share (remember that the letter "C" in LCV stands for "commercial"). Some of these privately-owned vehicles might however be used professionally by self-employed persons.

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.

The proportion of unassigned vehicles is much larger in the LCV survey than in the register. This is because the matching of surveys to branches was done at the time of the survey (2013-2014), with a very strict matching criterion. See Appendix B for a description of the matching procedure we adopted for the national vehicle register, which considers multiple matching criteria.

Table 1: First level of the NOGA 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

Table 2: Estimated distribution of LCVs by owner type and branch (state: June 2022)

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.

6 Model

6.1 Global structure

The adopted model structure, largely inspired from Hunt and Stefan [2007], is illustrated in Fig. 1. It consists of six modules:

- *Vehicle Generation*: This module provides an estimate of the number of LCVs per zone, owner type (private/business) and in the case of a business owner, by branch (see Section 5).
- *Active Vehicles:* This module determines which proportion of vehicles is active on a given day.
- *Number of Tours:* This module assigns a number of tours to each active vehicle.
- *Next Stop Location:* This module determines which zone is visited next.
- *End Tour:* This module determines whether the vehicle makes an additional intermediate stop or goes back to its establishment.
- *Correction:* This module applies a branch-specific correction factor on the trip matrices, to better reproduce the daily distance travelled.

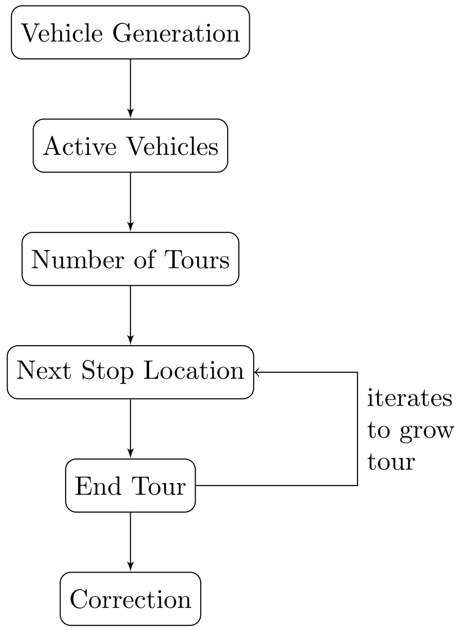

Figure 1: Model structure

Table 3: Overview of the different modules

Module	Method	Data used for estimation	Explanatory variables
Vehicle Generation	average rates per employee or inhabitant	Vehicle register 2022 at postal code level, matched with business register	population, 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

6.2 Vehicle Generation

This module converts the number of inhabitants per zone and the number of full-time equivalents per zone and per branch into a number of registered vehicles per zone, owner type (private or business) and branch (if business). This is achieved using the average vehicle rates, derived from empirical data. The LCVs owned by private persons are generated using a rate in terms of LCV per inhabitant, while the LCVs owned by businesses are generated using branch-specific rates expressed in LCV per job in the branch (full-time equivalent, FTE hereafter).

To estimate the number of vehicle per branch, we used the national vehicle register (state: June 2022), which we linked[2] with the national business register to obtain the branch of each business-owned vehicle. The explanatory variables (FTE or inhabitants) come from the following sources:

- Number of FTE in the branch A (state: 2021). Source: *Labour productivity by by economic sector at current prices* (FSO, 2023);
- Number of FTE in other branches (state: 2nd quarter 2022). Source: *Full-time job equivalent per sector* (FSO, 2023);
- Population (state: 2nd quarter 2022). Source: *Population résidante permanente et non permanente selon la catégorie de nationalité, le sexe et le canton, 2e trimestre 2022* (FSO, 2022).

The resulting vehicle generation rates are listed in Table 4.

[2] As there is no common identifier, we matched the entries based on the name of establishments and their address. See Appendix B for more detail. By combining various matching criteria, 93 % of business-owned LCVs could be assigned to a branch. The remaining 7 % were then distributed between the branches according to the proportions observed among assigned vehicles.

Table 4: Vehicle generation rates

Owner type	NOGA08	LCVs for 1'000 predictor units*
Business	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
Private	-	12

*The predictor corresponds to the branch-specific number of jobs (FTE) for business-owned vehicles and to the number of inhabitants for privately-owned vehicles.

6.3 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 and was computed using both types of surveys (type 1 and 2). 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) (transportion and storage, except postal and courier activities). Those that used the less frequently (p(active)= 0.39, i.e. 2.73 days per week in average) are those registered by private owners or whose branch could not be identified. Note that when applying the model in simulation, we only apply these probabilities to vehicles owned by natural persons (i.e. we do not simulate business-owned vehicles whose branch would be unknown).

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

Owner type	NOGA08	p(active) Mo-Su	p(active) Mo-Fr	Tours/day
Business	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
Private	-	0.39	0.48	1.51

6.4 Number of Tours

The vehicles were then converted into a number of tours, using segment-specific average rates. We rely 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 Appendix A.3). The number of tours per vehicle and per day for the considered users are given in Table 5.

6.5 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 number of zones to be considered in the estimation process and thereby limit the computational time, we relied on stratified importance sampling. Instead of considering a choice set made of 3'187 zones, 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. 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 $p_j = 23.87$ for the remaining 2'837 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_{ij} denote the generalized cost[3] by LCV from zone i to any zone j (in CHF). The chosen

[3]See Section 4.4 for the definition of generalized cost.

utility specification is the following:

$$
\begin{aligned}
U_j = & \sum_{\text{land\_use}} \theta_{\text{land\_use}} \times \delta_{\text{land\_use}} && (1) \\
& + (1 - \delta_{\text{same\_ZIP}}) \times [(\theta_C + \theta_{C,\text{first}} \times \delta_{\#\text{stops}=1}) \times C_{i,j}/100 \\
& + \theta_{C>50} \times \max(0, C_{i,j} - 50)/100] \\
& + \theta_{\text{same\_ZIP}} \times \delta_{\text{same\_ZIP}} && (2) \\
& + \log(\text{population in zone } j + \theta_{\text{jobs/pop}} \times \text{jobs in zone } j) \\
& + \log(p_j).
\end{aligned}
$$

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 a zone enters in the utility function of that zone.

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 legs that are not the first of a tour, every additional monetary unit between 0 and 50 CHF brings an additional utility of $\theta_C/100$ (which is negative), then every additional monetary unit above 50 CHF brings a utility of $(\theta_{C,\text{later}} + \theta_{C>50})/100$ (which is also negative). A similar piece-wise linear relation applies for first legs 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 most problematically, they are 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 $\theta_{\text{same\_ZIP}}$.

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. houses or establishments) - see McFadden [1978] for the theoretical derivations.

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

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 the 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 model specification in Eq. (1) was then 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 $\theta_{\text{jobs/pop}}$. 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. The resulting models are presented in Table 6.

	Private	C	F	G	H	N	Other
$\theta_{\text{land\_use}}$ Low Den	1.62***	2.25***	1.59***	1.46***	2.07***	0.94**	1.29***
	(0.18)	(0.36)	(0.29)	(0.21)	(0.47)	(0.57)	(0.36)
$\theta_{\text{land\_use}}$ Residential	0.59***	1.27***	0.55**	0.71***	0.92**	0.53	0.62**
	(0.16)	(0.29)	(0.25)	(0.18)	(0.40)	(0.45)	(0.29)
$\theta_{\text{land\_use}}$ Interm	0.43***	1.27***	0.46***	0.77***	0.80**	0.61*	0.70***
	(0.14)	(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	0.00
$\theta_{\text{same ZIP}}$	−0.52***	0.00	−0.44**	−0.41**	−1.10***	0.00	0.00
	(0.15)		(0.21)	(0.24)	(0.40)		
θ_C [100 CHF]	−10.38***	−9.03***	−9.07***	−9.56***	−8.90***	−9.21***	−10.48***
	(0.35)	(0.50)	(0.42)	(0.52)	(0.94)	(1.02)	(0.58)
$\theta_{C>50}$ [100 CHF]	6.77***	5.91***	4.29***	5.41***	5.86***	5.11***	7.31***
	(0.49)	(0.68)	(0.65)	(0.63)	(1.05)	(1.33)	(1.02)
$\theta_{C,\text{ first}}$ [100 CHF]	0.00	0.00	0.00	1.35***	1.27***	0.00	0.00
				(0.33)	(0.47)		
$\theta_{\text{jobs/pop}}$	0.95***	3.07*	0.84*	1.86***	3.26*	0.63	1.24*
	(0.35)	(2.08)	(0.51)	(0.63)	(2.51)	(0.57)	(0.79)
No Observations	2039	877	1223	1171	561	337	742
Log Likelihood (Null)	−11517	−4198	−7888	−6375	−1566	−2067	−4383
Log Likelihood (Converged)	−7251	−2667	−5003	−4131	−1167	−1238	−2485
Rho-squared	0.37	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 that are as far away from the current zone and from the base and that have 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.

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. 2. We observe that:

- as expected, the utility of each segment is a monotonically 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 and H, the utility decreases not as fast for the first leg of a tour as for subsequent legs. This indicates a delivery pattern where several destinations close to each other are grouped within a tour, producing long first and last legs and shorter intermediary legs.

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 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).

6.5.1 Analysis

Fig. 3 compares the trip length and trip duration distributions resulting from the simulation with those observed in the data. 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 here the trips remaining within the same ZIP code, because the skims of the two zoning systems are not consistent for such trips.

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, different vehicle

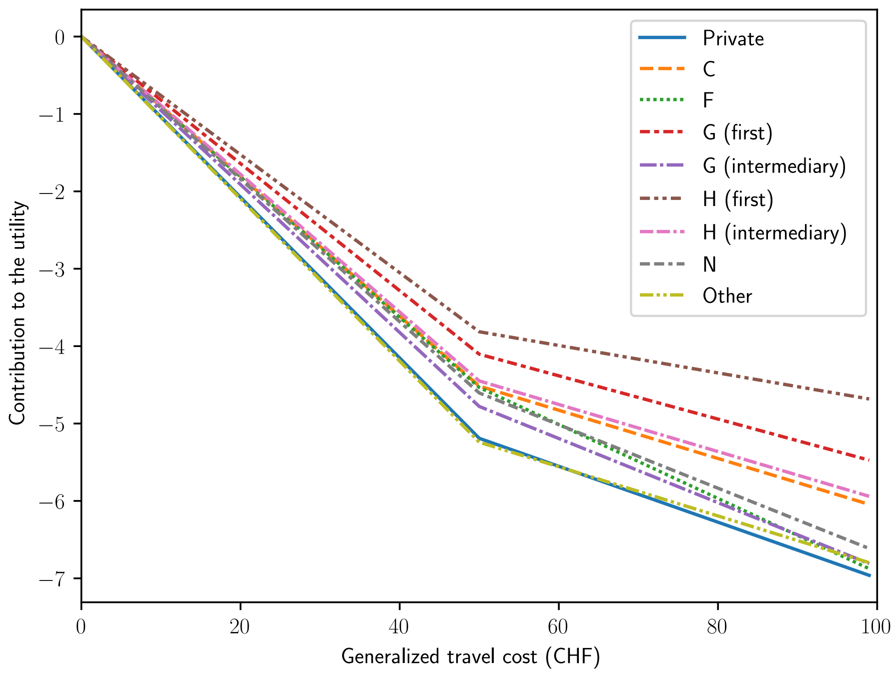

Figure 2: Contribution of the generalized travel cost to the utility.

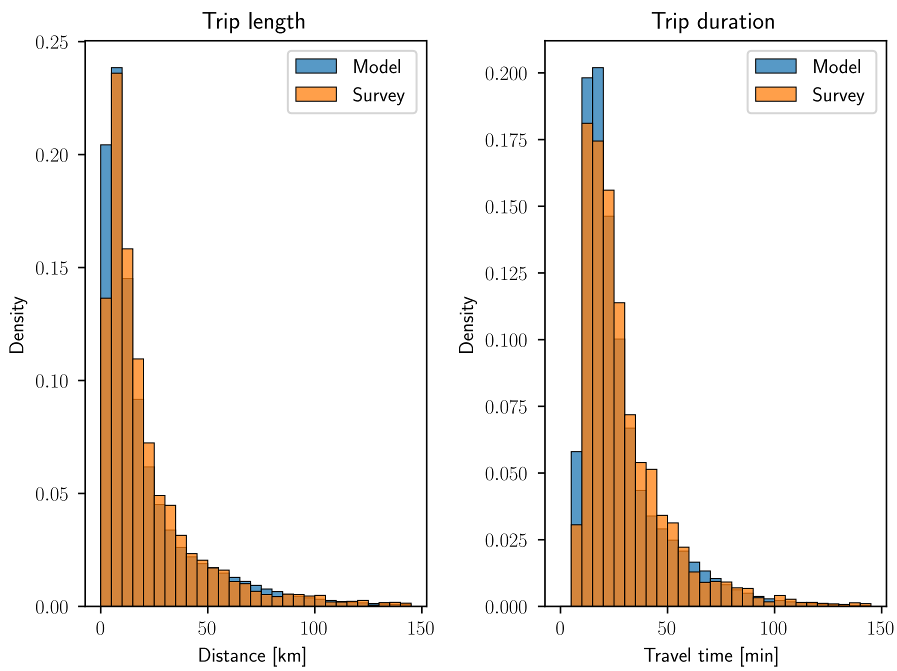

Figure 3: Comparison of trip length and duration distributions between the model results and the LCV survey for trips where origins and destinations have different ZIP codes.

populations in the data and in the simulation, 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, we can compare the proportion of such trips: in the empirical data, after the data cleaning steps described in Appendix A.2, 13.1 % of trips remain in the same ZIP code. In the simulation, it is 14.1 %. The difference is also considered acceptable.

6.6 End Tour

The module *End Tour* is a binary logit model describing the probability to end the tour after the current stop. The retained model specification is the following:

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

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. The number of stops made so far also influences the probability to continue via its logarithm ($\log(\#\text{stops})$).[5]

[5]The choice to include the logarithm rather than simply the number of stops is motivated by the intuition that the absolute difference in probability between stop n and stop $n+1$ should decrease with n.

	Private	C	F	G	H	N	Other
θ_0	0.27^{***}	-2.14^{***}	0.10	0.70^{***}	0.72^{***}	-0.13	0.29^{**}
	(0.12)	(0.23)	(0.11)	(0.18)	(0.28)	(0.29)	(0.17)
$\theta_{2\text{stops}}$	-1.18^{***}	0.00	-1.09^{***}	-1.04^{***}	-1.09^{***}	-1.03^{***}	-1.27^{***}
	(0.14)		(0.14)	(0.17)	(0.30)	(0.29)	(0.24)
$\theta_{\log(\#\text{stops})}$	0.00	1.56^{***}	0.00	0.00	0.00	0.00	0.00
		(0.18)					
$\theta_{C,\text{return}}$	0.36^{***}	0.00	0.00	0.39^{**}	0.53^{**}	1.31^{**}	0.00
	(0.14)			(0.21)	(0.24)	(0.59)	
No Observations	1554	593	910	1042	545	275	485
Log Likelihood (Null)	-1066	-354	-735	-678	-173	-213	-345
Log Likelihood (Converged)	-986	-304	-661	-615	-147	-191	-311
Rho-squared	0.08	0.14	0.10	0.09	0.15	0.10	0.10

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

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 $\theta_{2\text{stops}}$ are all negative. This suggests that the probability to do a third stop when two stops have already be made is smaller than the probability to do a $n+1^{\text{th}}$ stop when n stops have already be made.

Similarly, the estimated values of $\theta_{\log(\#\text{stops})}$ are all positive, i.e. the probability to do a $n+1^{\text{th}}$ stop when n stops have already be made increases with n. This is consistent with the fact that the distribution of the number of stops in a tour tends to be fat-tailed (even if most tours only include 1 or 2 stops, some professions can do dozens of stops within a tour).

The estimated values of θ_{C_return} are positive, i.e. vehicles whose previous stop is far away from their home establishment are more likely to make an additional stop.

6.6.1 Analysis and validation

Fig. 4 shows the distribution of the number of trips in a tour in simulation results and in the LCV surveys of type 2 (after the data cleaning steps described in Section A.2). The empirical and simulated distributions are overall very close, 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 overestimated by the model and conversely, the proportion of tours with at least three trips is slightly underestimated. These differences might be due to modeling limitations and differences in the populations of modeled vehicles.

6.7 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). The depiction of vehicle behavior resulting from the module *Next Stop Location* and *End Tour* suffers from two limitations that affect the VKT:

- the time required to fill the questionnaire of type 2 increases with the number of trips made, surveys of type 2 are likely to be affected by some under-reporting ;
- LCV surveys of type 2 containing tours that are described with the simplified form (see Appendix A.2) were not used in model estimation.

To circumvent these issues, we include a correction step. This module is placed at the end of the simulation. It artificially increases the number of trips by applying a correcting factor to the trip matrix of each segment, so that the

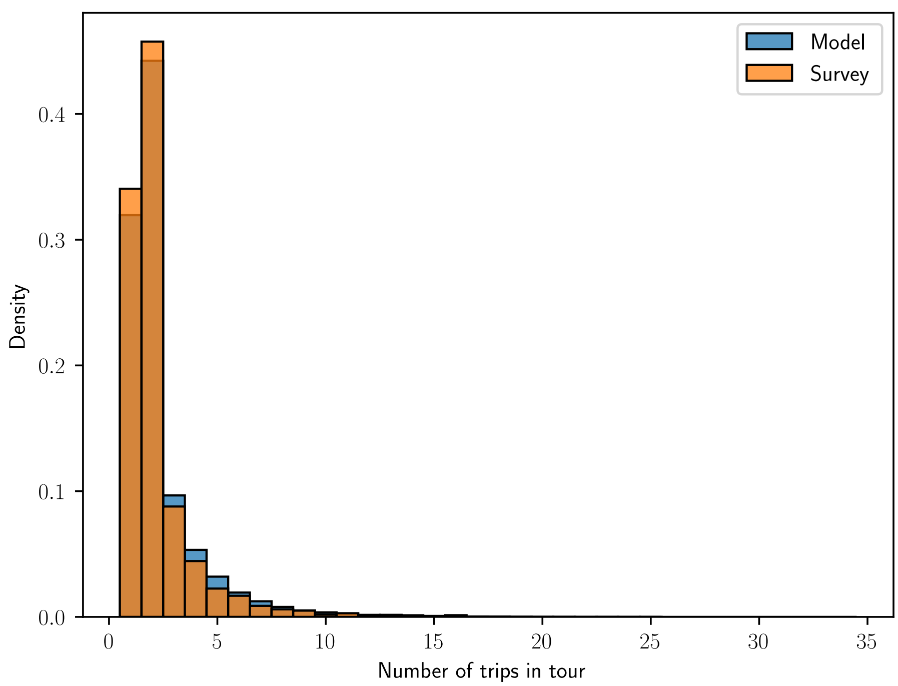

Figure 4: Distributions of the number of trips in a tour in the model results and in the LCV survey (type 2).

final average daily distance traveled is consistent with the observations in the LCV survey of type 1. This correcting 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, 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 1.11 for the branch H (i.e. we artificially increase the number of trips by 11 %) and 1.34 for the branch G (i.e. we artificially increase the number of trips by 34 %). After applying this correction, the model perfectly reproduces the average daily distance per segment.

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

Owner type	NOGA08	Simulation [km]	LCV Type 1 [km]	Ratio
Business	C	75.4	92.4	1.23
	F	53.8	65.7	1.22
	G	89.9	120.9	1.34
	H	138.8	153.9	1.11
	N	56.9	74.6	1.31
	Other	54.5	69.8	1.28
Private		60.8	72.1	1.19

7 Model application

7.1 Implementation

The model was implemented in Python, while the code used to estimate the discrete choice models is in R. The Python and R codes as well as there descriptions are available online on GitHub (`https://github.com/AREschweiz/LCV_model`) [Note: code needs to be updated, description needs to be copied from Significance and adjusted.].

7.2 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

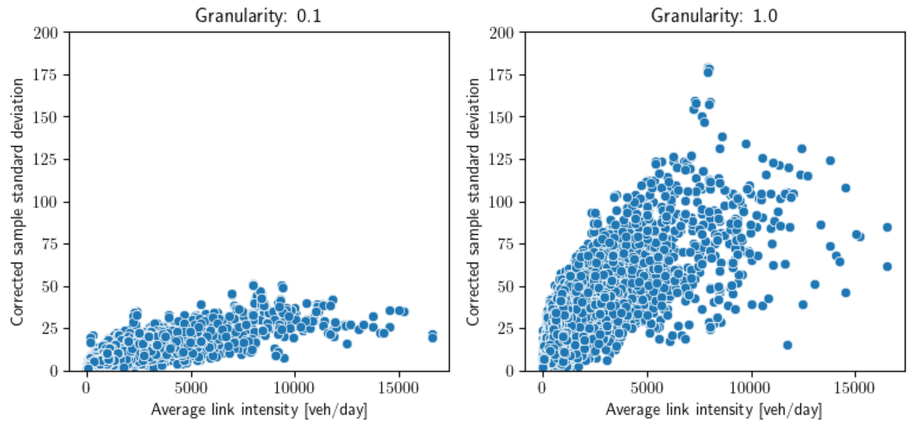

Figure 5: Corrected sample standard deviation as a function of the average link intensity, for five realizations, with $g = 0.1$ (link) and $g = 1$ (right).

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. 5 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.

7.3 Description of available loop detector data

7.4 Calibration

7.5 Comparison with the LCV Survey

References

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 data-sets

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

We carried out some data cleaning steps when estimating models based on LCV surveys of type 2. Three models rely on this data : the number of tours per day, Next Stop Location and End Tour. When estimating these models, 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. In addition, some additional criteria were used for the models Next Stop Location and End Tour.

A.2.2 Data for the module NextStopLocation

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

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

- **Trips with postal code corresponding to large organizations**: some postal codes do not correspond to well-defined geographic areas but to large organizations. The estimates of travel times from or to such postal codes would necessarily be very approximate. We discarded the legs where either the origin or destination had such codes.

- **Postal code corresponding to zones outside Switzerland**: Because of difficulties to obtain comprehensive skim matrices for all the neighboring countries, we focused on trips within Switzerland. The legs where either the origin or destination were abroad.

- **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.

The file generated for the estimation process includes the chosen alternative, 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. For more details, see the python script `create_estimation_data_for_next_stop_location.py`.

A.2.3 Data for the module EndTour

The data necessary for the estimation of the module EndTour is also derived from the set of **Legs**. If the leg is a "return leg" (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:

- **Legs originating from the base**: to avoid stopping tours before they start, we assumed that the first leg of a tour could not be a returned leg. This does not mean that the zone of the base is excluded from the destination choice.

- **Postal code corresponding to large organizations**: here as well, we discarded the observations where either the current zone or the base had postal codes corresponding to large organizations.

- **Postal code corresponding to zones outside Switzerland**: we discarded the observations where either the current zone or the base had postal codes outside of Switzerland.

The file generated 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 more details, see the python script `create_estimation_data_for_end_tour.py`.

A.3 Reconstructing trip legs and tours

In order to reconstruct the trip legs, we first listed all the stops (i.e. origins and destinations) made by each vehicle. If two shipments transported by the same vehicle had the same mileage at their origin, we considered their two origins as only one stop. We then sorted these stops by increasing mileage, and defined as trip legs 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 leg of the day.

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

B Appendix - Linking the national vehicle register with the national business register

In order to identify the number of vehicles registered within each branch, we link the national vehicle register with the national business register. The objective is to add a "branch" attribute to each vehicle. The matching is done based on the establishments' name, their address (a string consisting of a street name and a number) and their postal code. As these fields are not always consistent between the two registers, we test a variety of matching criteria. We start with the strictest (exact same name, address and postal code) and then progressively relax the various constraints (e.g. (a) exact same name and postal code, but no constraint on the address or (b) same address and postal code, but some small differences between the two names are allowed). When several establishments of the national business register fulfill the matching criterion for the same vehicle, we distinguish two cases : if all the establishments fulfilling the matching criterion belong to the same branch, we consider the matching as unambiguous and assigned this common branch to the vehicle. If the establishment belong to different branches, we do not assign any branch to the vehicle.

Once a vehicle is assigned to a branch with a criterion, this assignment is considered definitive and no additional matching criterion is applied. Note also that in some cases, the vehicle register contains two addresses and postal codes : one for the enterprise and one for the particular establishment. In such cases, we first try to match the vehicle owner based on the establishment's coordinates and if it fails, we retry with the enterprise's coordinates.

The criteria we use and the number of vehicles they permitted to assign to a branch are summarized in Table 9. After application of all the matching criteria considered, about 86 % of vehicles could be assigned to a branch. This proportion is further increased by applying some "manual assignment" when the name of the vehicle owner contains some specific words: for instance the word "transport" was assigned to the branch H. Some particularly large vehicle owners were also directly assigned to a branch using this method (this was done for instance for the Swiss Army and for the Swiss post).

Table 9: Number of assigned vehicle per matching criterion

Matching criterion	Assigned vehicles	Share[%]
1) Same name, address and postal code	166'921	52
2) Same name and postal code	46'179	13
3) Same name	14'282	4
4) Same address and postal code and:		
a) Name_BUR$_x$ = Name_IVZ	16'845	5
b) name_BUR = name_IVZ	6'155	2
c) name_BUR$_x$ = name_IVZ	1'471	0
d) L(Name_BUR, Name_IVZ) = 1	2'683	1
e) L(Name_BUR$_x$, Name_IVZ) = 1	2'167	1
f) Two words in common	11'840	4
Same postal code and:		
a) Name_BUR$_x$ = Name_IVZ	6'350	2
b) name_BUR = name_IVZ	2'171	1
c) name_BUR$_x$ = name_IVZ	673	0
d) L(Name_BUR, Name_IVZ) = 1	1'038	0
e) L(Name_BUR$_x$, Name_IVZ) = 1	930	0
f) Two words in common	5'691	2
Manual assignment	15'396	5
Unassigned	25'307	8
Total	320'099	100

Notes:

- This table refers to vehicles owned by business establishments.
- State of vehicle register: December 16$^{\text{th}}$, 2022.
- State of establishment register: April 18$^{\text{th}}$, 2023.
- Name_BUR and Name_IVZ denote the name in the business register (BUR) and vehicle register (IVZ, from the german *Informationssystems Verkehrszulassung*), Name_BUR$_x$ denotes a short version of Name_BUR where we only keep the first x characters (x denote the number of characters in Name_IVZ, which is limited to a maximum of 30) and name_BUR, name_IVZ, and name_BUR$_x$ denote lowercase versions of these names. $L(a, b)$ denotes the Levenshtein distance between the words a and b, i.e. the minimum number of single-character edits (insertions, deletions or substitutions) required to change a into b. Allowing a match when the Levenshtein distance is equal to 1 means that we allow one character edit. "Two words in common" means that Name_BUR and Name_IVZ should have at least two words in common for them to be matched.

C 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.

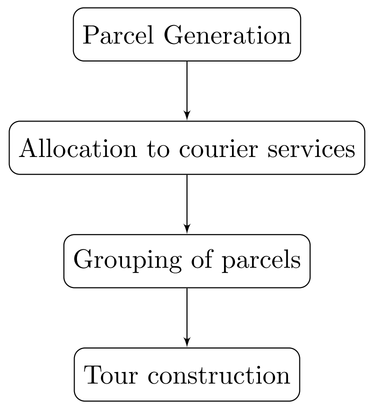

Figure 6: Model structure

C.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. 6).

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.

C.2 Strategy to avoid double-counting

To avoid modeling the same vehicles twice (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 than the branch 53 only accounts for about 2 % of LCVs (see Table 2), and that the number of modelled parcel delivery tours (6'195 tours for 2020) is almost as large 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 %.

D Appendix - International traffic

The modules described in Section 6 are only applied to zones inside Switzerland. Because of limited data availability, the international traffic was modelled in a much simpler way, with conventional trip generation and trip distribution. These steps are briefly presented hereafter.

D.0.1 Trip generation

For lack of better data, trip generation rates were estimated based on the results of the model described in Section 6 for zones within Switzerland. We did a linear regression of the number of trips a zone produces (or, equivalently, attracts) on the population and number of jobs of each zone. The zones used for this purpose are the 7'965 domestic zones of the national passenger transport model. The result is an average rate of 0.11 LCV trip per day per job (full time equivalent) and 0.07 LCV trip per day per inhabitant. The R^2 of the linear regression is 0.92. We then applied these rates on all zones (inside and outside Switzerland) of the national passenger transport model.

D.0.2 Trip distribution

Here as well, the results of the model described in this paper for domestic traffic were used instead of empirical data for estimation purposes. Let us denote X_i the number of trips originating from/arriving into zone i. With the retained functional form, the number of trips from zone i to zone j is given by:

$$\text{Trips from zone } i \text{ to zone } j = \alpha_i \beta_j X_i X_j \exp(\gamma C_{i,j}),$$

where γ is the parameter to be estimated, while α_i and β_i must be computed using the Iterative Proportional Fitting algorithm. The value of γ that minimizes the root-mean-square error for the Swiss zones was found to be -0.086.

This model is applied on the trips generated in the first step (Trip Generation), to obtain an Origin-Destination matrix for all zones (inside and outside Switzerland). The part of the matrix corresponding to trips inside Switzerland is deleted, as this part of the demand is already covered by the more detailed LCV model described in the body of this paper.