An integrated accident injury severity and risk-taking behavior model for the evaluation of Via Sicura

Nicola Ortelli^{*1,2}, Matthieu de Lapparent¹, Silvia F. Varotto³, and Michel Bierlaire²

¹School of Management and Engineering Vaud, HES-SO, Switzerland ²École Polytechnique Fédérale de Lausanne (EPFL), Switzerland ³École nationale des travaux publics de l'État (ENTPE), University Gustave Eiffel, France

SHORT SUMMARY

This study presents a flexible model for risk-taking behavior and accident injury severity. It is specifically designed to evaluate the impact of Via Sicura, a Swiss road safety program, on the severity of accident outcomes. Our proposed model treats the risk-taking behavior of each driver as a latent variable that depends on a number of socioeconomic and contextual factors, and whose manifestation can be measured by means of behavioral indicators. The aggregated risk, a central feature of our framework, represents the combined latent risk-taking behaviors among all drivers within an accident and is successfully identified as explanatory of the severity of injuries sustained by all individuals involved. Our findings reveal that Via Sicura's repressive measures successfully deter risk-taking behavior among drivers, preventing an estimated 63 fatal, 876 major and 2'303 minor injuries over a ten-year period.

Keywords: accident injury severity, driving behavior, latent variable model

1 INTRODUCTION

In Switzerland, considerable efforts in favor of road safety have been carried out for the past 50 years. As a result, between 1970 and 2020 the number of fatalities on Swiss roads has dropped from 1'694 to 227. In 2013, as a further commitment to reach its ultimate target of zero fatalities or major injuries [\(Swiss Council for Accident Prevention, 2002\)](#page-8-0), the Swiss Federal Council has initiated a road safety program called Via Sicura.

Via Sicura consists of a number of preventive measures that aim at improving road safety by ensuring that "only the drivers who have received the necessary level of instruction and possess a full driving capacity drive in safe motor vehicles and on forgiving roads" [\(Federal Roads Office,](#page-8-1) [2005\)](#page-8-1). Together with those, Via Sicura also introduces some repressive measures designed to act as a deterrent against risk-taking behavior. In particular, a new, stricter legislation governing extreme speeding offenders and drunk drivers was introduced in January 2013.

The first official evaluation of the program highlights its positive effect by comparing the yearly totals of major and fatal injuries between 2013 and 2015 with the ones predicted by a counterfactual $\text{model} - i.e.,$ without Via Sicura— estimated on the data from 2000 to 2012 [\(Swiss Federal Council,](#page-8-2) 2017 : according to their analysis, an average of 33 fatal injuries per year were prevented during the three first years of the program. Because it focuses on yearly totals, however, this approach suffers two drawbacks: (i) heterogeneity at the accident, vehicle and individual levels is ignored; and (ii) the identified impact of Via Sicura is actually the combination of two distinct effects, namely on accident occurrence and accident severity, and those are not captured separately.

To overcome these limitations, we propose a model that focuses on assessing the influence of Via Sicura on injury severity at a disaggregate level, namely through the dissuasive effect of its repressive measures on the behavior of drivers and, in turn, through the impact of the resulting change in behavior on accident outcomes. Similar to [Lavieri et al.](#page-8-3) [\(2016\)](#page-8-3), our framework models risk-taking behavior as a latent variable whose value directly impacts the injury severity of all individuals involved in an accident. Our framework also includes an innovative way of aggregating the risk-taking behavior of any number of drivers, which allows for a comprehensive evaluation of Via Sicura. For this purpose, we use a dataset derived from all police reports of accidents that occurred in Switzerland between 1992 and 2022, representing over 3.4 million events and 4.2 million individuals involved.

2 METHODOLOGY

Figure [1](#page-1-0) presents our modeling framework. In a given accident, the risk-taking behavior of each involved driver is modeled as a latent variable that depends on their socioeconomic characteristics and on some context variables. In this context, we define the risk-taking behavior as the propensity to act in a way that deliberately disregards safety, while endangering other persons as well as oneself. Following [Lavieri et al.](#page-8-3) [\(2016\)](#page-8-3), our framework assumes that the risk-taking behavior of a driver not only affects the injury severity of the occupants of their own vehicle, but also that of all other individuals in the accident. Indeed, the behavior of each driver is modeled independently, but when an accident occurs, we assume that it is the conjunction of said behaviors that contributes to the severity of the injuries sustained by all involved individuals. The aggregated risk therefore represents the combination of the risks taken among all drivers. The other variables that are deemed to be explanatory of injury severity relate to the individual, to the vehicle or to the circumstances of the accident.

Figure 1: Ordered logit and latent variable for injury severity and risk-taking behavior.

Structural equations

Let A be a set of reported traffic accidents. We denote as $\mathcal{V}(a)$ the set of vehicles involved in accident $a \in \mathcal{A}$ and as $\mathcal{I}(a, v)$ the occupants of vehicle $v \in \mathcal{V}(a)$. The latent risk-taking behavior of the driver of vehicle $v \in V(a)$, denoted by r_{av1}^* , is defined as a linear combination of variables that are meant to explain said behavior:

$$
r_{av1}^* = \gamma \mathbf{z}_{av1}' + \eta_{av1},\tag{1}
$$

where z_{av1} is a vector that contains the explanatory variables, γ is a vector of parameters to be estimated from the data, and η_{av1} is an error term assumed to be Gumbel-distributed:

$$
\eta_{av1} \stackrel{\text{iid}}{\sim} \text{Gumbel}\left(0, \mu\right). \tag{2}
$$

Risk-taking behavior is measured using behavioral indicators, which rely on some latent continuous variables. Let I_{av1p}^* be one such underlying continuous variable; in addition to the driver's risktaking behavior r_{av1}^* , its value also depends on a vector of variables s_{av1p} that are deemed to be explanatory of the behavioral indicator:

$$
I_{av1p}^* = \boldsymbol{\theta}_p \mathbf{s}_{av1p}' + \lambda_p r_{av1}^* + \nu_{av1p},
$$
\n(3)

where θ_p and λ_p are parameters to be estimated from the data and ν_{av1p} are error terms that follow a logistic distribution with scale parameter $\delta_p > 0$:

$$
\nu_{av1p} \stackrel{\text{iid}}{\sim} \text{Logistic}\left(0, \delta_p\right). \tag{4}
$$

Individual injury severity is influenced by the latent risk-taking behavior of all drivers involved in the accident. The challenge here is to formulate an appropriate synthetic variable that aggregates all these latent behaviors into a single value. In this context, we define the aggregated risk r_a^* as the maximum risk-taking behavior among all involved drivers:

$$
r_a^* = \max_{v \in \mathcal{V}(a)} (r_{av1}^*)
$$
\n⁽⁵⁾

Given that the risk-taking behaviors of all drivers follow the same Gumbel distribution, the aggregated risk r_a^* is also Gumbel-distributed:

$$
r_a^* \stackrel{\text{iid}}{\sim} \text{Gumbel}\left(\mu \sum_{v \in \mathcal{V}(a)} \exp\left(\frac{\gamma \mathbf{z}_{av1}'}{\mu}\right), \mu\right). \tag{6}
$$

Finally, the latent injury severity associated with individual $i \in \mathcal{I}(a, v)$ is denoted by U_{avi} and defined as:

$$
U_{avi} = \beta \mathbf{x}'_{avi} + \alpha r_a^* + \varepsilon_{avi},\tag{7}
$$

where \mathbf{x}_{avi} is a vector containing the observable variables that are deemed to be explanatory of individual levels of injury, β and α are vectors of parameters to be estimated from the data, and ε_{avi} is an error term:

$$
\varepsilon_{avi} \stackrel{iid}{\sim} \text{Logistic}\left(0, \sigma\right). \tag{8}
$$

Measurement equations

Three binary indicators of risk-taking behavior are identified in the available data: (i) one of the presumed causes of the accident being attributable to recklessness; (ii) driving, riding or walking under the influence of alcohol, drugs or impairing medications; and (iii) not wearing a seatbelt or a helmet. In our model, each one is defined as

$$
I_{av1p} = \begin{cases} 0 & \text{if } I_{av1p}^* \le \kappa_{p1}, \\ 1 & \text{otherwise}, \end{cases}
$$
 (9)

where I_{av1p}^* is the underlying continuous variable of indicator I_{av1p} , and κ_{p1} is the associated threshold parameter, to be estimated from the data. Let $\Theta_p = (\theta'_p, \lambda_p, \delta_p, \kappa_{p1})'$ be the vector of unknown parameters; conditional on r_{av1}^* , the probability that I_{av1p} takes value $\ell \in \{0,1\}$ is computed as

$$
\Pr\left(I_{av1p} = \ell \mid \mathbf{s}_{av1p}, r_{av1}^*, \mathbf{\Theta}_p\right) =
$$
\n
$$
F\left(\frac{\kappa_{p\ell+1} - \mathbf{\Theta}_p \mathbf{s}_{av1p}' - \lambda_p r_{av1}^*}{\delta_p}\right) - F\left(\frac{\kappa_{p\ell} - \mathbf{\Theta}_p \mathbf{s}_{av1p}' - \lambda_p r_{av1}^*}{\delta_p}\right),
$$
\n(10)

where F is the cumulative distribution function of the logistic distribution and, by convention, $\kappa_{p0} \equiv -\infty$ and $\kappa_{p2} \equiv +\infty$.

Finally, the severity of the injuries suffered by each individual is reported on the following four-level scale: 0 — none, 1 — minor, 2 — major, and 3 — fatal. In our model, these outcomes are generated by the corresponding latent injury severity U_{avi} , as follows:

$$
y_{avi} = \begin{cases} 0 & \text{if } & U_{avi} \leq \tau_1, \\ 1 & \text{if } & \tau_1 < U_{avi} \leq \tau_2, \\ 2 & \text{if } & \tau_2 < U_{avi} \leq \tau_3, \\ 3 & \text{if } & \tau_3 < U_{avi}, \end{cases}
$$
(11)

where τ_1 , τ_2 and τ_3 are threshold parameters to be estimated from the data. Major injuries are defined as those that require a hospitalization of more than one day; before 2015, however, the definition also included injuries that "preclude all normal home activities for at least 24 hours". This change is taken into account by defining two different parameters for threshold τ_2 : the first is used for accidents that occurred between 1992 and 2014, and the second for those between 2015 and 2022.

Let $\mathbf{B} = (\beta', \alpha, \sigma, \tau_1, \tau_2, \tau_3)'$ be the vector of unknown parameters; for all $a \in \mathcal{A}, v \in \mathcal{V}(a)$ and $i \in \mathcal{I}(a, v)$, the probability — conditional on r_a^* — that the reported level of injury y_{avi} is equal to $k \in \{0, 1, 2, 3\}$ is defined as

$$
\Pr(y_{avi} = k \mid \mathbf{x}_{avi}, r_a^*; \mathbf{B}) =
$$
\n
$$
F\left(\frac{\tau_{k+1} - \beta \mathbf{x}_{avi}' - \alpha r_a^*}{\sigma}\right) - F\left(\frac{\tau_k - \beta \mathbf{x}_{avi}' - \alpha r_a^*}{\sigma}\right),
$$
\n(12)

where $\tau_0 \equiv -\infty$ and $\tau_K \equiv +\infty$.

3 Results and discussion

The considered data are derived from police reports of traffic accidents that occurred in Switzer-land between [1](#page-3-0)992 and 2022.¹ In total, the dataset contains 2.03 million accidents, 3.40 million involved vehicles — including pedestrians — and 4.22 million occupants of said vehicles. Due to missing values in important explanatory variables, approximately one in ten observations cannot be exploited. The two components of the model are estimated sequentially, using different samples: the latent behavior component uses only the observations associated to drivers, whereas the injury severity component is estimated with the observations of all individuals involved in "complete accidents", i.e., accidents for which not a single value is missing. The variables that we consider as explanatory of the risk-taking behavior and of injury severity are described in Table [1](#page-3-1) and Table [2.](#page-3-2)

Latent behavior component

The latent behavior component is estimated first; the results are reported in Table [3.](#page-4-0) The effect of age on the risk-taking behavior is modeled as a piecewise-linear function with, as breakpoints, the values 18, 35 and 65; it is also segmented based on the gender of the driver and the presence of passengers in the vehicle. For ease of interpretation, Figure [2](#page-5-0) illustrates the obtained results. All other parameters in the structural equation of the risk-taking behavior are statistically significant and have the expected signs: learner drivers, child passengers and adverse weather are associated with a reduction in risk-taking behavior, whereas the late night variable is shown to increase it. Drivers also seem to take more risks on highways and rural roads than in urban areas. The year of collection is used to capture the effect of all other efforts made by the Swiss government to promote safe behavior in the past 30 years— awareness-raising campaigns, safety education programs, driver training courses, and so on. The effect is modeled as a second-degree power series. Finally, the variable accounting for the dissuasive effect of Via Sicura is shown to reduce the risk-taking behavior of drivers; its impact is similar in magnitude to the one associated with learner drivers.

Table 1: Description of the explanatory variables of the risk-taking behavior.

¹Verkehrsunfall Jahresdatensatz (DWH-VU), Federal Roads Office.

 $continued$

Parameter	Value	t -test
Risk-taking behavior		
γ -age, 0-18	-0.108	-10.8
γ -age, 18-35	-0.0245	-32.2
γ -age, 35-65	-0.0105	-21.8
γ -age, 65-100	0.00296	2.31
γ -age, with passenger, 0-18	0.0275	27.8
γ -age, with passenger, 18-35	-0.0278	-16.8
γ -age, with passenger, 35-65	-0.00877	-7.58
γ -age, with passenger, 65-100	0.00134	0.448
γ -age, female driver, 0-18	-0.0488	-43.0
γ -age, female driver, 18-35	0.0195	13.0
γ -age, female driver, 35-65	0.00316	3.32
γ -age, female driver, 65-100	0.00619	2.41
γ -age, female driver, with passenger, 0-18	-0.0188	-9.56
γ -age, female driver, with passenger, 18-35	0.0219	6.70
γ -age, female driver, with passenger, 35-65	-0.00174	-0.701
γ -age, female driver, with passenger, 65-100	-0.00342	-0.443
γ -learner driver	-0.239	-16.9
γ -child passenger	-0.254	-13.3
γ -adverse weather	-0.227	-14.3
γ -late night	2.09	88.5
γ -urban	θ	
γ -rural	1.16	76.6
γ -highway	1.21	76.2
γ -year	0.0175	14.1
γ -year, squared	-0.000874	-16.2
γ -via sicura	-0.240	-14.9
μ	1.50	79.7

Table [3](#page-4-0)— continued: Estimation results of the latent behavior component.

Figure 2: Age distribution and effect on risk-taking behavior.

The effect of Via Sicura is also visible in the top half of Figure [3,](#page-6-0) which exhibits the distribution of risk-taking behavior among drivers for each year of data, as obtained via Monte-Carlo simulation. The bottom half of the figure shows the distribution of the *aggregated risk* for each year of data, that is, the maximum risk-taking behavior among the drivers involved in each accident, as defined in [\(6\)](#page-2-0). From 2013 onward, Figure [3](#page-6-0) also shows the counterfactual distribution of risk-taking behavior without the effect of Via Sicura, *i.e.*, as if the measures of the road safety program were never implemented.

Figure 3: Evolution of driver risk-taking behavior and resulting aggregated risks.

Injury severity component

Table [4](#page-7-0) gathers the estimation results of the injury severity component. Age is again modeled as a piecewise-linear function with the values 18, 35 and 65 as breakpoints, and its associated parameters are segmented based on gender; Figure [4](#page-7-1) illustrates its effect. All other parameters are statistically significant and have the expected signs: those relating to the vehicle categories coincide with the level of protection each vehicle provides, wearing a seatbelt is associated with a decrease in injury severity and so are accidents that occurred while parking, whereas single-vehicle accidents and higher speed limits increase the severity of the sustained injuries.

The effect of wearing a helmet is not considered in our model because its effect could not be correctly identified for any of the relevant vehicle categories. Our explanation is that a helmet reduces the chances of major and fatal injuries, but is not as effective in preventing minor injuries. In fact, the available data shows a higher rate of minor injuries for cyclists and motorcyclists that wear a helmet than for those who do not. The correct effect could be captured by replacing the threshold parameters of the ordered logit with functions of relevant exogenous variables as in [Eluru](#page-8-4) [et al.](#page-8-4) [\(2008\)](#page-8-4).

Finally, using a simple iterative procedure, we calibrate the threshold parameters of the injury severity component such that the model replicates the observed shares of the four levels of severity. We then simulate the *counterfactual* shares between 2013 and 2022, that is, without including the effect of Via Sicura in the latent behavior component. The results in Table [5](#page-7-2) indicate a difference of 63 fatal, 876 major and 2'303 minor injuries over the ten-year period, which we attribute to the dissuasive effect of Via Sicura on driver behavior.

Parameter	Value	t -test				
Individual injury severity						
α -aggregated risk	0.136	76				
β -age, 0-18	0.0223	29.9				
β -age, 18-35	0.000874	2.05				
β -age, 35-65	0.00563	20.2				
β -age, 65-100	0.019	26.9				
β -age, female, 0-18	0.0395	94.9				
β -age, female, 18-35	0.0000581	0.0868				
β -age, female, 35-65	$-0.00227 - 5.20$					
β -age, female, 65-100	-0.00696	-6.47				
β -pedestrians	0.635	49.9				
β -slow modes	0.238	18.3				
β -e-bikes & mopeds	-0.0654	-4.23				
β -motorcycles	-0.0692	-5.03				
β -cars $\mathcal C$ vans	-2.12	-161				
β -heavy vehicles	-3.89	-203				
β -seatbelt	-1.34	-215				
β -while parking	-1.34	-82.9				
β -while driving, single-vehicle accident	0.559	138				
β -while driving, speed limit	0.00963	141				
Reported injury level						
T_1	θ					
τ_2 -until 2014	2.38	631				
τ_2 -since 2015	2.65	405				
τ_3	5.37	513				

Table 4: Estimation results of the injury severity component.

Figure 4: Age distribution and effect on injury severity.

Table 5: Observed and counterfactual levels of severity between 2013 and 2022.

Level of injury	Observed shares			Counterfactual shares [90% CI]		
None	702'739	78.4%	699'497	$78.1\,\%$	[698'549, 700'514]	
Minor	156'362	17.5%	158'665	17.7%	[157'835, 159'444]	
Major	34'898	3.9%	35'774	4.0%	[35'454, 36'081]	
Fatal	1'926	0.2%	1'989	0.2%	[1'964, 2'012]	

4 Conclusions

In this study, we propose a flexible framework for risk-taking behavior and accident injury severity modeling, for the purpose of evaluating the effect of the Via Sicura road safety program on accident outcomes.The main novelty of our framework is the introduction of the aggregated risk, defined as the maximum risk-taking behavior among all drivers involved in an accident. This definition implies that our framework can accommodate accidents involving any number of vehicles, which is central to the comprehensive assessment of public policies. Our proposed model is shown to successfully capture the dissuasive effect that the repressive measures of Via Sicura have on the behavior of drivers and, in turn, the impact of this change in behavior on injury severity.

Future work should focus on replacing the ordered logit formulation of the injury severity component by a generalized ordered logit [\(Eluru et al., 2008\)](#page-8-4), which allows for a more realistic modeling of variables that are known to impact specific levels of injury differently, such as the use of a helmet for cyclists and motorcyclists. Compared to the current approach, the generalized ordered logit formulation defines the threshold values of the injury severity scale as functions of relevant exogenous variables; all thresholds may then vary across accidents, vehicles and individuals. Another natural advancement consists in including other behaviors, such as distracted or careless driving, to further improve the realism of our model. Finally, further investigation could also focus on alternative ways of aggregating the risk-taking behavior of drivers.The maximum was chosen for its simplicity, relevance and interpretability, but developing and comparing other candidates could deserve an entire paper in its own rights.

REFERENCES

- Eluru, N., Bhat, C. R., & Hensher, D. A. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis \mathcal{C} Prevention, $40(3)$, 1033-1054.
- Federal Roads Office. (2005). Via sicura. Programme d'action de la Confédération visant à renforcer la sécurité routière.
- Lavieri, P. S., Bhat, C. R., Pendyala, R. M., & Garikapati, V. M. (2016). Introducing latent psychological constructs in injury severity modeling: multivehicle and multioccupant approach. Transportation research record, 2601 (1), 110–118.
- Swiss Council for Accident Prevention. (2002). Elaboration des fondements d'une politique nationale de sécurité routière. Mandat de recherche ASTRA 2000/447 sur demande de l'OFROU.
- Swiss Federal Council. (2017). Évaluation du programme via sicura (No. Q261-0477). Rapport du Conseil Fédéral en réponse au postulat 16.3267 de la Commission des Transports et des Télécommunications du Conseil des États du 14 avril 2016.