Can we infer on behavioral impacts of public policy on accident severity outcomes?

A Swiss case study using historical disaggregate accident reports

Nicola Ortelli
Michel Bierlaire
Matthieu de Lapparent

HEIG-VD, HES-SO / TRANSP-OR, EPFL

September 2021
Can we infer on behavioral impacts of public policy on accident severity outcomes?

Nicola Ortelli, Matthieu de Lapparent
School of Engineering and Management Vaud
University of Applied Sciences and Arts
Avenue des Sports 20 - CP 521
1401 Yverdon-les-Bains, Switzerland
{nicola.ortelli,matthieu.delapparent}@heig-vd.ch

Nicola Ortelli, Michel Bierlaire
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
Station 18
1015 Lausanne
{nicola.ortelli,michel.bierlaire}@epfl.ch

September 2021

Abstract

Reducing the number of fatalities caused by traffic accidents is a major concern of our modern society. The existing literature generally deals with explaining the influence of directly observable variables on the occurrence of crashes or on the severity of the injuries they cause. While these studies provide valuable insights on the vulnerability of passengers in various seat positions, types of vehicles and specific locations, driver behavior is usually not adequately considered. In fact, only few published models include such attitudinal aspects, mainly because of data limitations: crash data generally do not include behavioral characteristics or psychological measurements.

We use the integrated "choice" and latent variable framework in order to account for unobservable driver behavior in injury severity modeling and apply it to a dataset that contains information on road accidents that occurred in Switzerland between 1992 and 2017. Specifically, we build a latent variable that captures the "propensity to taking risks" that drivers may exhibit, based on a number of socioeconomic characteristics and context variables. Among the latter, we include a dummy variable that represents the entry into force of the Via Sicura road safety program. Preliminary results show that our framework has the ability to the appropriately identify the effect of driver behavior on injury severity, as well the influence of Via Sicura thereupon.

Keywords

Road safety; Injury severity; Latent behavior; Ordered logit; hybrid model
1 Introduction

Reducing the number of fatalities caused by traffic accidents is a major concern of our modern, motorized society. Everywhere around the world, efforts are made by governments, auto manufacturers and transportation agencies to enhance road safety. In Switzerland, considerable commitment in favor of safer infrastructure, vehicles and behavior has been carried out for the past 50 years. As a result, the number of fatalities on Swiss roads has dropped from 1’694 in 1970 to 253 in 2015 and the number of severely injured individuals has decreased from 18’314 to 3’830 over the same interval (Swiss Federal Council, 2017).

More recently, in an effort to pursue such commitment, the Swiss Federal Council has initiated a road safety program called Via Sicura. The program aims to further reduce the number of fatalities and severe injuries by ensuring, to the greatest possible extent, that the Swiss road network is used only by well-instructed and capable drivers. Starting in January 2013, a number of legislative measures have come into force on a step-by-step basis. Most of these are preventive, such as the ban on alcohol for new and professional drivers or the compulsory use of lights during the day for all motor vehicles, but a number of repressive measures are also included. Namely, a new, stricter legislation governing extreme speeding offenders and drunk drivers was introduced at the very beginning of the program.

In this paper, we seek to measure the "dissuasive" effect of these repressive measures on the risk-taking behavior of drivers and, in turn, the effect of such change in behavior on the severity of the injuries suffered by individuals involved in road accidents. To this end, we first propose an intuitive ordered logit that models individuals’ injury severity based on various factors related to the individual, to his or her vehicle and to the characteristics of the crash. Notably, we include a dummy variable that represents the entry into force of the Via Sicura repressive measures. In a second phase, we leverage the integrated "choice" and latent variable (ICLV) framework to account for the Via Sicura measures in a more appropriate way: we define a latent variable that captures the propensity to taking risks that drivers may exhibit and base this risk-taking behavior on a number of socioeconomic characteristics and context variables, including the Via Sicura dummy variable. The models are estimated on a dataset that gathers information from police reports of road accidents that occurred in Switzerland between 1992 and 2017.

We organize the remainder of this paper as follows: Section 2 provides a brief overview of the existing literature on injury severity modeling; Section 3 introduces the methodological approach we follow; Section 4 provides a brief description of the available dataset and Section 5 gathers the results obtained from the estimation of our models; finally, Section 6 summarizes the findings of the present study and identifies the future steps of this research.
Can we infer on behavioral impacts of public policy on accident severity outcomes?

2 Literature review

Due to the importance of road safety all around the world, accident occurrence modeling and crash-injury severity modeling have been active fields of research for decades. Since the first accident occurrence model that included a regression component (Weber 1970, 1971), a wide variety of studies have investigated the effect of various factors on the occurrence of accidents or on the injuries they cause. In the context of the latter, the existing literature deals with explaining the influence of directly observable variables on the vulnerability of individuals in different types of vehicles (de Lapparent 2005, 2006; Xin et al. 2017), seat positions (de Lapparent 2008; Bogue et al. 2017), on different road types (Huang et al. 2008; Qiu and Fan 2021; Choudhary et al. 2018) or when involved in different types of accidents (Shankar and Mannering 1996; Kockelman and Kweon 2002).

These studies are valuable in that they provide insights into the complex interactions that vehicle, infrastructure, and human characteristics have on the resulting crash-injury severities; nevertheless, the vast majority entirely omits the yet crucial effects of driver behavior from their models. We see two obvious reasons for these omissions: (i) crash data generally do not include behavioral characteristics or psychological measurements; and (ii) the specification and estimation of models that enable the inclusion of attitudes and other latent constructs can be extremely tedious.

Hence, to the best of our knowledge, and despite an ongoing stream of research concerned with characterizing various driving behaviors (Tasca 2000; Clapp et al. 2011; Scott-Parker and Weston 2017; Hu et al. 2021), only three studies attempt to include such attitudinal aspects into their injury-severity models. We briefly discuss these, as they are of direct relevance to the current study.

Nevarez et al. (2009) appear to be the first to account for driving behavior in their binary model of injury severity. They do so by introducing an "aggressive driving" dummy variable that is based on whether the driver was "speeding, tailgating, failed to yield right of way, changed lanes improperly, or disregarded other traffic control". This simple approach is certainly an improvement with regards to the abovementioned studies; however, by treating the aggressive behavior as exogenous, the model is not capable of providing insights into measures that seek to decrease injury severity by reducing aggressiveness in driving behavior.

In comparison, Paletti et al. (2010) include a similar dummy variable in their model, but treat it as endogenous — their "latent aggressive driving act propensity" is defined as a function of observed environmental, vehicle, crash and driver factors, in addition to an error term — and
interact it with a number of explanatory variables in the injury-severity model. The framework proposed by Paleti et al. (2010) is therefore analogous to a latent segmentation scheme, the segmentation being based on the "aggressive driving propensity" binary variable.

Finally, the more recent work of Lavieri et al. (2016) provides an additional example of injury-severity models that accounts for driver behavior. In this case, two distinct behaviors — risky and distracted driving — are modeled as latent variables and their manifestation is measured by means of binary indicators. Not only is the use of indicators helpful in model identification and in increasing the efficiency of the estimated parameters (Walker, 2001), but it also mitigates the risks of endogeneity-related issues: in fact, the two indicators considered in conjunction with the risky behavior — seat belt use and alcohol consumption — are known in the literature for being particularly problematic when treated as exogenous.

3 Modeling approach

Our modeling approach is based on a rather straightforward conceptual framework. We assume that the level of injury suffered by any person involved in a road accident may be explained by a number of variables, each related to one of three hierarchical levels: (i) individual; (ii) vehicle; and (iii) accident. The first level includes variables such as gender, age or seat belt use. The second level gathers variables that are common to all occupants of a vehicle: the type of said vehicle, the year of its entry into circulation, the number of children aboard, or the driver’s blood alcohol level. Finally, the third level includes variables that are related to the accident itself, and may therefore be common to the occupants of several vehicles. These include meteorological, traffic and road conditions, the day-of-week and time of the accident, etc.

In addition to these directly observable factors, and in accordance with the existing literature, we also include a latent variable that accounts for risk-taking behavior. Latent variables are typically used to capture the effect of unobservable psychological constructs such as attitudes and perceptions; in our case, we assume the risk-taking behavior to be explained by a number of socioeconomic characteristics and context variables, among which we include the entry into force of the Via Sicura repressive measures. We use seat belt use, substance consumption and driving without a license as indicators, i.e., "indirect measurements" of the latent behavior.

Despite the hierarchical organization of the variables, we model injury severity at the individual level. An appropriate way of taking into consideration the inherent correlation among observations — both at the vehicle and accident levels — is currently being investigated.
3.1 Ordered logit model

Let \( n = 1, \ldots, N \) be an index representing individuals involved in road accidents; their observed injury levels \( y_n \) are reported on a discrete scale \( j = 1, \ldots, J \). For example, as in the data considered for this study, index \( j \) may take values of "no injury" \((j = 1)\), "minor injury" \((j = 2)\), "major injury" \((j = 3)\) and "fatal injury" \((j = 4)\).

We treat \( y_n \) as the expressions of an underlying process characterized as

\[
y^*_n = \sum_{k=1}^{K} \beta_k x_{nk} + \varepsilon_n = u_n + \varepsilon_n, \tag{1}
\]

where \( y^*_n \) is the actual severity of injuries suffered by individual \( n \), \( x_{nk} \) are \( K \) attributes deemed as "explanatory" of that injury severity, \( \beta_k \) are their associated coefficients, and \( \varepsilon_n \) is an error term for which we assume \( \varepsilon_n \sim \text{Logistic}(0, 1) \).

The dependent, continuous variable \( y^*_n \) is mapped to the discrete injury levels by means of \( J + 1 \) strictly increasing thresholds, denoted as \( \{\tau_0, \tau_1, \ldots, \tau_J\} \), in the usual ordered-response fashion: given that \( \varepsilon_n \) follows a logistic distribution, the probability of \( y^*_n \) being reported as an injury of level \( j \) may be computed as

\[
P_n(j) = P(\tau_{j-1} < y^*_n < \tau_j) = P(\tau_{j-1} < u_n + \varepsilon_n < \tau_j) = F(\tau_j - u_n) - F(\tau_{j-1} - u_n), \tag{2}
\]

where \( F(\cdot) \) denotes the cumulative distribution function of a logistic distribution with location and scale parameters 0 and 1. The first and last thresholds are fixed to \( \tau_0 = -\infty \) and \( \tau_J = \infty \), whereas \( \{\tau_1, \ldots, \tau_{J-1}\} \) are estimated. In the case where \( J = 4 \), as in the data considered for this study, we make sure that the estimated thresholds verify \( -\infty < \tau_1 < \tau_2 < \tau_3 < \infty \) by defining two strictly positive parameters, \( \delta_1, \delta_2 > 0 \), and expressing \( \tau_2 \) and \( \tau_3 \) as

\[
\begin{cases}
\tau_2 = \tau_1 + \delta_1, \\
\tau_3 = \tau_1 + \delta_1 + \delta_2.
\end{cases} \tag{3}
\]

Together with all \( \beta_k \), parameters \( \tau_1, \delta_1 \) and \( \delta_2 \) are estimated through maximum likelihood.

3.2 Latent variable component

We now describe the inclusion of a latent variable and associated indicators in the ordinal logit model of Section 3.1. Two sets of equations are needed for this purpose: (i) a structural equation
Can we infer on behavioral impacts of public policy on accident severity outcomes?

September 2021

that includes the factors influencing the latent variable; and (ii) several measurement equations, each describing the relationship between the latent variable and one of its indicators.

The structural equation characterizing a single latent variable $z_n^*$ is written as

$$
z_n^* = \gamma_0 + \sum_{\ell=1}^{L} \gamma_{\ell} x_{n\ell} + \omega_n,  \tag{4}
$$

where $\gamma_0$ is an intercept, $\gamma_{\ell}$ are $L$ coefficients, each associated with one explanatory factor $x_{n\ell}$, and $\omega_n \sim N(0, \sigma_{\omega})$ is an error term. It is worth noting that $x_{nk}$ from (1) and $x_{n\ell}$ may share a number of common variables, when such variables are assumed to have an impact both on the injury severity $y_n^*$ and the latent variable $z_n^*$. Also, because the latent construct is typically considered as an additional explanatory variable in the former model, we can rewrite (1) as

$$
y_n^* = \sum_{k=1}^{K} \beta_k x_{nk} + \beta_z z_n^* + \epsilon_n,  \tag{5}
$$

where $\beta_z$ is the coefficient capturing the effect of the latent variable on the injury level $y_n^*$.

As regards the measurement equations, we define a number of continuous indicators $I_{in}^*$ that, in our case, solely depend on the latent variable $z_n^*$. Each indicator is expressed as

$$
I_{in}^* = \alpha_{i,0} + \alpha_{i,1}z_n^* + \nu_{in}, \tag{6}
$$

where $\alpha_{i,0}$ is an intercept, $\alpha_{i,1}$ is the coefficient measuring the effect of the latent variable on the $i$-th indicator and $\nu_{in}$ is an error term for which we assume $\nu_{in} \sim N(0, \sigma_{\nu})$. The manifestations of the latent variable, denoted $I_{in}$, are typically encoded as discrete. The indicators $I_{in}^*$ therefore need to be "discretized", which can be achieved in a similar fashion as explained in Section 3.1; the only difference resides in the probabilistic distribution the error terms are assumed to follow in each case, but that is seamlessly taken care of. For instance, as in this study, the considered indicators are binary, which is equivalent to dealing with a two-level discrete scale. Given that $\nu_{in}$ follows a normal distribution, the outcome probabilities of $I_{in}$ are computed as

$$
\begin{cases}
P(I_{in} = 0) = P(I_{in}^* < \mu_i) = P(\alpha_{i,0} + \alpha_{i,1}z_n^* + \nu_{in} < \mu_i) = \Phi \left( \frac{\mu_i - \alpha_{i,0} - \alpha_{i,1}z_n^*}{\sigma_i} \right), \\
P(I_{in} = 1) = 1 - P(I_{in} = 0) = 1 - \Phi \left( \frac{\mu_i - \alpha_{i,0} - \alpha_{i,1}z_n^*}{\sigma_i} \right),
\end{cases}  \tag{7}
$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a normal distribution $N(0, 1)$ and $\mu_i$ is a threshold associated with the $i$-th indicator; its value is to be estimated.
3.3 Estimation and identification

The ordered logit model is straightforwardly estimated through maximum likelihood estimation. One should nevertheless note that (1) does not include an intercept, for identification purposes: in fact, all threshold parameters are estimated instead. The decision to estimate all thresholds rather than fixing one and estimating an intercept instead is arbitrary.

In comparison to the ordered logit, the estimation of the hybrid model is more complex. Despite the higher computational cost, we choose to estimate both components of the model simultaneously rather than sequentially. Indeed, simultaneous estimation is known to be superior because it produces consistent and efficient parameter estimates (Walker, 2001). We also prefer numerical integration over simulation due to the size of our dataset, but this is only possible because our model includes a single latent variable. Finally, a number of parameters need to be normalized for identification purposes; these are discussed in Section 5.2.

4 Data description

The data under consideration are derived from police reports of road accidents that occurred in Switzerland between 1992 and 2017. In total, after discarding irrelevant and incomplete observations, the dataset used for model estimation contains information about 434'038 crashes, 662'903 vehicles — pedestrians are treated as such — and about the 831'372 occupants of these vehicles. Almost 95% of all accidents involve either one or two vehicles, whereas only 4.9% of the vehicles have more than two occupants.

As mentioned above, the dataset contains a variety of features that are organized into three hierarchical levels. At the highest level, the variables describe the context of the accident, which is common to all individuals involved in the crash. The intermediate level gathers the attributes of the vehicles and their drivers; these are therefore shared by all their occupants. Finally, the lowest level groups characteristics relative to single individuals. The severity of the injuries suffered by individuals involved in the accidents are reported on a four-level scale: (i) no injury; (ii) minor injury; (iii) major injury; and (iv) fatal injury. Table 1 presents an overview of the descriptive characteristics of the dataset.

---

1 Verkehrsunfall Jahresdatensatz (DWH-VU), Federal Roads Office FEDRO.
5 Modeling results

5.1 Ordered logit model

We begin by presenting the intuitive — i.e., oversimplified — ordered logit model. Individual injury severity is assumed to be a function of the following explanatory variables:\(^2\)

\(^2\)The exact specification is provided in the Appendix.
• **Via Sicura** — we define a binary variable that takes value 1 for all crashes that occurred on the 1st of January 2013 or later. The date corresponds to the entry into force of the Via Sicura repressive measures. We denote by $B_{\text{VIA\_SICURA}}$ the parameter that captures the effect of said variable;

• **accident type** — we include three dummy variables that equal to 1 for crashes documented as frontal collisions, as "while parking" or as involving pedestrians, respectively. The parameters associated with these are $B_{\text{FRONTAL\_COLLISION}}$, $B_{\text{WHILE\_PARKING}}$ and $B_{\text{PEDESTRIAN\_INVOLVED}}$;

• **speed limit** — this serves as a proxy for the actual vehicle speed. For it to be more realistic, we interact it with a variable that describes the traffic volume at the time of the crash — normal, high and unknown. The parameters are $B_{\text{MAX\_SPEED\_TRAFFIC\_NORM}}$, $B_{\text{MAX\_SPEED\_TRAFFIC\_HIGH}}$, and $B_{\text{B\_MAX\_SPEED\_TRAFFIC\_UNK}}$;

• **vehicle type** — captured by means of three parameters, namely $B_{\text{TWO\_WHEELER}}$, $B_{\text{SOFT\_MODE}}$ and $B_{\text{PEDESTRIAN}}$. The two-wheeler category includes all two-wheel motorized vehicles, whereas soft modes include all non-motorized vehicles. Cars are used as the reference, together with all other private and motorized vehicles;

• **vehicle’s year of entry into service** — newly-built motorized vehicles are ever safer. We simplistically assume the effect to be linear and capture it by means of $B_{\text{YEAR\_ENTRY}}$;

• **seat belt use** — encoded as a binary variable. Its effect is captured by parameter $B_{\text{belt}}$;

• **gender** — captured by $B_{\text{FEMALE}}$;

• **age** — $B_{\text{AGE\_FEMALE}}$ and $B_{\text{AGE\_MALE}}$ capture the effect of age differently, depending on the individual’s gender. In addition, $B_{\text{AGE\_SQ\_MALE}}$ captures the effect of the square of age for men. Its equivalent for women is not statistically significant.

Table 2 provides the estimation results of the ordered logit model. All parameter estimates appear to be significant — which was to be expected, given the number of observations — and all display the expected signs. As a reminder, parameters $\tau_1$, $\delta_1$ and $\delta_2$ define the thresholds used to map the continuous injury severity to the reported discrete levels.

### 5.2 Hybrid model

We now discuss our hybrid model. The main difference with the previous one is the inclusion — according to the method introduced in Section 3.2 — of a latent variable that represents the risk-taking behavior that some drivers may exhibit. The behavior is modeled as a continuous variable and is included in the model among the factors that explain the injury severity. It is important to note that as opposed to the previous model, the risk-taking behavior should be modeled at the vehicle level. In fact, it is reasonable to assume that the severity of injuries of all
Table 2: Ordered logit model. Parameter estimates.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Rob. t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_VIA_SICURA</td>
<td>−0.151</td>
<td>−23.8</td>
</tr>
<tr>
<td>B_FRONTAL_COLLISION</td>
<td>0.455</td>
<td>34.9</td>
</tr>
<tr>
<td>B_WHILE_PARKING</td>
<td>−1.84</td>
<td>−77.2</td>
</tr>
<tr>
<td>B_PEDESTRIAN_INVOLVED</td>
<td>−0.973</td>
<td>−46.9</td>
</tr>
<tr>
<td>B_MAX_SPEED_TRAFFIC_NORM</td>
<td>0.536</td>
<td>53.2</td>
</tr>
<tr>
<td>B_MAX_SPEED_TRAFFIC_HIGH</td>
<td>0.324</td>
<td>30.6</td>
</tr>
<tr>
<td>B_MAX_SPEED_TRAFFIC_UNK</td>
<td>1.45</td>
<td>82.6</td>
</tr>
<tr>
<td>B_TWO_WHEELER</td>
<td>1.71</td>
<td>110</td>
</tr>
<tr>
<td>B_SOFT_MODE</td>
<td>1.80</td>
<td>62.0</td>
</tr>
<tr>
<td>B_PEDESTRIAN</td>
<td>2.86</td>
<td>78.4</td>
</tr>
<tr>
<td>BYEAR_ENTRY</td>
<td>−0.102</td>
<td>−8.22</td>
</tr>
<tr>
<td>B_BELT</td>
<td>−1.54</td>
<td>−121</td>
</tr>
<tr>
<td>B_FEMALE</td>
<td>0.411</td>
<td>29.2</td>
</tr>
<tr>
<td>B_AGE_FEMALE</td>
<td>0.00432</td>
<td>17.9</td>
</tr>
<tr>
<td>B_AGE_MALE</td>
<td>0.000768</td>
<td>3.67</td>
</tr>
<tr>
<td>B_AGE_SQ_MALE</td>
<td>0.0000632</td>
<td>6.60</td>
</tr>
<tr>
<td>(\tau_1)</td>
<td>0.600</td>
<td>20.3</td>
</tr>
<tr>
<td>(\delta_1)</td>
<td>2.41</td>
<td>376</td>
</tr>
<tr>
<td>(\delta_2)</td>
<td>2.87</td>
<td>157</td>
</tr>
</tbody>
</table>

Sample size: 831'372  
Final log likelihood: −431'404  
Number of est. parameters: 18

Occupants of a vehicle depend on the driver’s behavior, rather than theirs. That would be but an issue had the latent variable component to be estimated on its own; however, as discussed in Section 3.3, the two components need to be estimated simultaneously. For this reason, we define a number of variables that solely relate to the drivers and propagate them to all occupants of their respective vehicles, so as to model both the injury level and the risky driving behavior at the individual level. A potential flaw of this approach is that the behavior of drivers accompanied by several passengers is overrepresented, which may lead to biased estimates in the measurement equation of the latent variable; we still go along with it for its ease of implementation and because the vast majority of vehicles in the dataset have a single occupant. Finally, we also assume that the risk-taking latent variable is always equal to zero for pedestrians.

We begin by describing the structural equation of the latent variable. We assume the risk-taking behavior of drivers to be explained by the following factors:\(^3\)

\(^3\)Once again, the exact specification is provided in the Appendix.
Can we infer on behavioral impacts of public policy on accident severity outcomes?  

- **Via Sicura** — taken into account as in the previous model, except it now affects the latent behavior rather than injury severity. We denote by $\text{coef}_\text{via_sicura}$ its associated parameter;

- **time of the accident** — we define a binary variable that equals to 1 for accident that occurred between 10PM and 5AM and to 0 otherwise. The influence of this variable on the risk-taking behavior is measured by coefficient $\text{coef}_\text{late_night}$;

- **visibility** — similarly, we define a dummy variable that takes value 1 if the visibility was reduced at the time of the accident and 0 otherwise. Its associated parameter is $\text{coef}_\text{bad_visibility}$;

- **road conditions** — we assume the effects of wet, snowy and icy roads to be similar in magnitude on the driver’s propensity to taking risks. These are therefore captured by a single coefficient, denoted $\text{coef}_\text{bad_road}$;

- **passengers’ age** — we define a binary variable that is equal to 1 if at least one of the occupants of the vehicle is 12 years old or younger. The associated parameter is $\text{coef}_\text{child_aboard}$;

- **driver’s gender** — captured by $\text{coef}_\text{female_driver}$.

As regards the measurement equations, the three indicators are defined as follows:

- **driver’s substances consumption** — we define a variable that takes value 1 for drivers who were positively tests for alcohol, medicines or drugs, and 0 otherwise;

- **driver’s protection** — similarly, we define a binary variable that equals 1 for car drivers who do not use the seat belt or for two-wheeler drivers who do not wear a helmet;

- **driver’s license** — we define a dummy variable that is equal to 1 for drivers who drive vehicles that require a license without ever possessing one.

Table 3 provides the estimation results of the hybrid model. While the magnitudes of the parameter estimates cannot be directly compared with the ones in the previous table, we may still note that they have the expected signs. In particular, $B_{RISKY}$, the coefficient associated with the risk-taking behavior, is positive: the higher the risks taken by the driver, the more severe the injuries of all occupants of the vehicle. The signs of the parameters associated with the explanatory variables of the latent behavior are also realistic; it is reasonable that drivers take more risks at night, whereas limited visibility and poor road conditions have the opposite effect. Additionally, the results suggest that female drivers and drivers accompanied by children display less risky behavior. The coefficient capturing the effect of the Via Sicura repressive measures, $\text{coef}_\text{via_sicura}$, is also negative and its magnitude is comparable to the other parameters. For identification purposes $\alpha_{\text{substances},0}$ and $\alpha_{\text{substances},1}$ are normalized; $\sigma_{\text{substances}}$, $\sigma_{\text{no_license}}$ and $\sigma_{\text{no_protec}}$ cannot be estimated because the indicators are binary.
Can we infer on behavioral impacts of public policy on accident severity outcomes?  

September 2021

Table 3: Hybrid model. Parameter estimates.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Rob. t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>coef_intercept</td>
<td>54.4</td>
<td>68.5</td>
</tr>
<tr>
<td>coef_via_sicura</td>
<td>−71.0</td>
<td>−67.5</td>
</tr>
<tr>
<td>coef_late_night</td>
<td>96.0</td>
<td>68.2</td>
</tr>
<tr>
<td>coef_bad_visibility</td>
<td>−14.0</td>
<td>−40.9</td>
</tr>
<tr>
<td>coef_bad_road</td>
<td>−9.01</td>
<td>−58.2</td>
</tr>
<tr>
<td>coef_child_aboard</td>
<td>−30.5</td>
<td>−58.5</td>
</tr>
<tr>
<td>coef_female_driver</td>
<td>−28.4</td>
<td>−68.9</td>
</tr>
<tr>
<td>$\sigma_\omega$</td>
<td>67.7</td>
<td>−68.8</td>
</tr>
<tr>
<td>$\alpha_{\text{substances},0}$</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>$\alpha_{\text{no_license},0}$</td>
<td>−2.20</td>
<td>−568</td>
</tr>
<tr>
<td>$\alpha_{\text{no_protec},0}$</td>
<td>−1.41</td>
<td>−1100</td>
</tr>
<tr>
<td>$\alpha_{\text{substances},1}$</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>$\alpha_{\text{no_license},1}$</td>
<td>0.00296</td>
<td>36.9</td>
</tr>
<tr>
<td>$\alpha_{\text{no_protec},1}$</td>
<td>0.00326</td>
<td>58.6</td>
</tr>
<tr>
<td>$\sigma_{\text{substances}}$</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>$\sigma_{\text{no_license}}$</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>$\sigma_{\text{no_protec}}$</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>$\mu_{\text{substances}}$</td>
<td>95.9</td>
<td>68.1</td>
</tr>
<tr>
<td>$\mu_{\text{no_license}}$</td>
<td>0.590</td>
<td>153</td>
</tr>
<tr>
<td>$\mu_{\text{no_protec}}$</td>
<td>0.143</td>
<td>111</td>
</tr>
<tr>
<td>B_RISKY</td>
<td>0.00293</td>
<td>42.5</td>
</tr>
<tr>
<td>B_FRONTAL_COLLISION</td>
<td>0.466</td>
<td>35.5</td>
</tr>
<tr>
<td>B_WHILE_PARKING</td>
<td>−1.81</td>
<td>−75.7</td>
</tr>
<tr>
<td>B_PEDESTRIAN_INVOLVED</td>
<td>−0.948</td>
<td>−45.4</td>
</tr>
<tr>
<td>B_MAX_SPEED_TRAFFIC_NORM</td>
<td>0.600</td>
<td>59.5</td>
</tr>
<tr>
<td>B_MAX_SPEED_TRAFFIC_HIGH</td>
<td>0.416</td>
<td>38.8</td>
</tr>
<tr>
<td>B_MAX_SPEED_TRAFFIC_UNK</td>
<td>1.07</td>
<td>56.1</td>
</tr>
<tr>
<td>B_TWO_WHEELER</td>
<td>1.82</td>
<td>117</td>
</tr>
<tr>
<td>B_SOFT_MODE</td>
<td>1.96</td>
<td>67.0</td>
</tr>
<tr>
<td>B_PEDESTRIAN</td>
<td>3.03</td>
<td>82.6</td>
</tr>
<tr>
<td>B_YEAR_ENTRY</td>
<td>−0.0692</td>
<td>−5.58</td>
</tr>
<tr>
<td>B_BELT</td>
<td>−1.45</td>
<td>−115</td>
</tr>
<tr>
<td>B_FEMALE</td>
<td>0.503</td>
<td>35.3</td>
</tr>
<tr>
<td>B_AGE_FEMALE</td>
<td>0.00455</td>
<td>18.6</td>
</tr>
<tr>
<td>B_AGE_MALE</td>
<td>0.000899</td>
<td>4.26</td>
</tr>
<tr>
<td>B_AGE_SQ_MALE</td>
<td>0.000108</td>
<td>11.0</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>0.956</td>
<td>32.0</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>2.42</td>
<td>377</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>2.88</td>
<td>157</td>
</tr>
</tbody>
</table>

Sample size: 831372
Final log likelihood: −940907
Number of est. parameters: 34
6 Conclusion

In this study, we show how the integrated choice and latent variable framework allows for the inclusion of behavioral constructs in a standard injury-severity model. The resulting hybrid model enables the dissuasive effect of repressive measures to be captured, in a way that is more appropriate than including it among the explanatory variables of injury severity. The estimation results show that our model is coherent, realistic and in line with the findings of the existing literature.

This study constitutes a first step toward a comprehensive accident-occurrence and injury-severity model that could serve as a tool for the assessment of the impact of public policies in a global way. Intended future work specific to the developed hybrid model includes the definition of additional driving behaviors, such as aggressive, distracted or defensive driving. Also, further investigation should focus on better taking into consideration the inherent correlation among observations at the vehicle and accident levels.
7 References


Can we infer on behavioral impacts of public policy on accident severity outcomes?  


Appendix

Ordered logit model: specification of the continuous injury severity

\[ u = \text{B\_VIA\_SICURA} \times (\text{year} \geq 2013) + \]
\[ \quad \text{B\_FRONTAL\_COLLISION} \times (\text{acc\_type} = \text{'frontal\_collision'}) + \]
\[ \quad \text{B\_WHILE\_PARKING} \times (\text{acc\_type} = \text{'while\_parking'}) + \]
\[ \quad \text{B\_PEDESTRIAN\_INVOLVED} \times (\text{acc\_type} = \text{'pedestrian\_involved'}) + \]
\[ \quad \text{B\_MAX\_SPEED\_TRAFFIC\_NORM} \times (\text{max\_speed/100})^2 \times (\text{traffic} = \text{'normal'}) + \]
\[ \quad \text{B\_MAX\_SPEED\_TRAFFIC\_HIGH} \times (\text{max\_speed/100})^2 \times (\text{traffic} = \text{'high'}) + \]
\[ \quad \text{B\_MAX\_SPEED\_TRAFFIC\_UNK} \times (\text{max\_speed/100})^2 \times (\text{traffic} = \text{'unknown'}) + \]
\[ \quad \text{B\_TWO\_WHEELER} \times (\text{veh\_type} = \text{'two\_wheeler'}) + \]
\[ \quad \text{B\_SOFT\_MODE} \times (\text{veh\_type} = \text{'soft\_mode'}) + \]
\[ \quad \text{B\_PEDESTRIAN} \times (\text{veh\_type} = \text{'pedestrian'}) + \]
\[ \quad \text{B\_YEAR\_ENTRY} \times (\text{year\_entry/1000}) + \]
\[ \quad \text{B\_BELT} \times (\text{protection} = \text{'belt'}) + \]
\[ \quad \text{B\_FEMALE} \times (\text{gender} = \text{'female'}) + \]
\[ \quad \text{B\_AGE\_FEMALE} \times (\text{gender} = \text{'female'}) \times \text{age} + \]
\[ \quad \text{B\_AGE\_MALE} \times (\text{gender} = \text{'male'}) \times \text{age} + \]
\[ \quad \text{B\_AGE\_SQ\_MALE} \times (\text{gender} = \text{'male'}) \times (\text{age}-40)^2 + \]

Hybrid model: specification of the risk-taking behavior

\[ r = (\text{veh\_type} != \text{'pedestrian'}) \times [\text{coef\_intercept} + \]
\[ \quad \text{coef\_via\_sicura} \times (\text{year} \geq 2013) + \]
\[ \quad \text{coef\_late\_night} \times ((\text{time} \geq 22) \text{ OR } (\text{time} < 5)) + \]
\[ \quad \text{coef\_bad\_visibility} \times (\text{bad\_visib} = 1) + \]
\[ \quad \text{coef\_bad\_road} \times (\text{road\_cond} = \text{'wet'}) \text{ OR } (\text{road\_cond} = \text{'snowy'}) \text{ OR } (\text{road\_cond} = \text{'icy'}) + \]
\[ \quad \text{coef\_child\_aboard} \times (\text{child\_aboard} = 1) + \]
\[ \quad \text{coef\_female\_driver} \times (\text{female\_driver} = 1)] \]