RSS 2022
8th Road Safety & Simulation International Conference
Road Safety and Digitalization
08-10 June 2022 • Athens, Greece

The effect of speed limits on accident frequency on the German Autobahn - A causal machine learning approach

Maike Metz-Peeters*

Ruhr University Bochum, RWI Essen, Maike.Metz-Peeters@rub.de

Abstract

The presented study analyzes the effect a 120 km/h speed limit has on accident rates on German motorways. To do so, various geo-spatial data sources are combined to a data set containing rich information on roadway characteristics for 500-meter segment of large parts of the German motorway network. A causal forests, a novel non-parametric regression technique is used under the assumption that crash frequency and speed limit probability are positively correlated, which leads to the estimation of a lower bound of the absolute size of the effect. To respect the spatial nature of the data, spatial lags are included and standard errors are clustered on the location level. To overcome spatial over-fitting, target oriented cross validation and forward-feature selection are used. As target oriented cross validation requires the respective speed limit to be present in various locations, this only seems fully possible for speed limits of 120 km/h. Therefore, the focus lies on estimating the effect of this speed limit as compared to the absence of any limit. The causal forest further allows learning about effect heterogeneity on segments of different congestion levels.

Keywords: crash frequency, causal machine learning, causal forest, speed limits, German Autobahn, spatial machine learning

*Corresponding author. Tel.: +49-234 32-27563;
E-mail address: Maike.Metz-Peeters@rub.de

1. Introduction

The German Autobahn is famous for being the only public road system in an advanced economy where there does not exist any binding speed limit in large parts of the network. Driving speeds above and beyond 150 km/h are not uncommon on these roads [1] and there has been a vital public debate about the sensibility of a general speed limit of 130 or even 120 km/h for many years [2, 3]. Supporters emphasize environmental and alleged safety benefits, as well as a reduction in traffic jams and economic gains [4] while opponents challenge these arguments, claim irrelevant emission and no crash reductions, high economic costs, and a loss of time and driving freedom should such a speed limit be introduced [5, 2].

Despite this considerable public interest and the significance of the German motorway network in the European transport infrastructure, current studies of the effects speed limits have on accident frequency on these roads, are missing. At the same time, studies looking into the effects of variations in mandatory speed limit in other countries are not directly transferable to this environment with completely absent limits. The reasons for this lack in research are manifold. Large scale experimental studies aiming at investigating this question are all older than 35 years, a time span in which there have been significant changes not only with respect to vehicle safety and roadway construction [6]. Observational studies, on the other hand, come with a number of methodological challenges, two of which stand out. First, there is no readily available data set suitable for analyzing this question. Second, the problem of unobserved heterogeneity is inherent to this task, as has long been acknowledged in accident research [7, 8, 9].

The present paper takes on the first challenge by processing and combining various geospatial data sources in order to obtain a numerical data set containing rich segment-wise information on roadway geometrics, traffic properties, environmental characteristics and socioeconomic factors of 500 m motorway segments from 12 out of 16 German states. Data is aggregated over the years $2017 - 2019$ to reduce volatility in crash counts and covers more than 50% of the German motorway network. The data processing is implemented using ArcGIS and Python and is fully automated, reproducible, and extendable. As the individual data sources are almost all open data, an open version of this newly constructed data set and the corresponding code will be available from the author.

Identification of the causal effect of speed limits on accident frequency with standard regression techniques would require controlling all factors jointly influencing the probability of speed limit installation and crash occurrence (conditional independence assumption (CIA)), as well as close approximation of the underlying data generating process (DGP). As speed limits are often installed as a result of higher (expected or observed) crash incidence, this is clearly infeasible and unobserved heterogeneity is inevitable [10]. This particular relationship, however, is exactly what allows for meaningful conclusions about the size of the true causal effect. By introducing speed limits on segments with higher crash incidence, unobserved factors that increase (decrease) crash frequency also increase (decrease) speed limit probability. Under this assumption, the estimated effect for speed limits on accident frequency is upward biased [11]. To have full control of this *upper-bound assumption*, a causal forest, an efficient non-parametric method from the emerging field of causal machine learning is employed [12]. It works entirely non-parametrically, and thus does not require the researcher to make strong assumptions about the underlying functional forms. In addition, causal forests help uncover treatment effect heterogeneity according to pre-specified hypotheses, without requiring explicit modeling of interaction terms.

The remainder of this paper is structured as follows: In chapter 2 a short overview is given over existing research investigating dealing with the effects of speed limits on accident frequency. Chapter 3 presents the used data sources and describes the construction of the new data set. Section 4 introduces the causal forest and explain how it is useful for our setting and under the upper bound assumption. Results are presented in chapter 5. Last, chapter 6 gives an outlook and concludes.

2. Literature review

To our best knowledge, there do not exist international studies analyzing the effects of speed limits as compared to the absence of any mandatory speed limit on crash frequency. The uniqueness of this German environment thus requires to look into national research, largely commissioned by governmental bodies. On the federal level, the German Federal Highway Research Institute (BASt, Bundesanstalt für Straßenwesen) is the research institute of the German federal government in the field of road engineering. As such, the BASt conducts and commissions road safety research in Germany to provide the Federal Ministry of Transport and Digital Infrastructure (BMVI) with scientific aids for transport management and policy decisions [13].

Since 1952, large parts of Germany's motorways have been without any mandatory speed limits almost continuously over time [14, 15]. Only a general *recommended* maximum speed of 130 km/h was introduced on a trial basis in 1974 and has been in place permanently since 1978. In this context, the BASt conducted a large experiment from 1974 to 1976 to evaluate the effect of a binding speed limit of 130 km/h compared to only a recommended maximum speed [15]. For two years, a (mandatory) speed limit was installed on half of the almost 3000*km* of experimental road[1], whereby test- and control segments were swapped after one year. Through this design, they were able to eliminate time- as well as segment-specific effects. They found mandatory speed limits of 130 km/h to reduce crashes causing injuries by roughly 10% and the number of fatalities or severely injured by about 20%. Based on these results, it was assumed in [16] that a speed limit would lead to a reduction in fatalities of around 20% and a reduction in severely injured by 10.6% and of lightly injured by 8.7% was derived. Furthermore, the authors deducted theoretically that a speed limit of 100 km/h would reduce crashes by 37%. [2]

In addition to this thoroughly designed large scale experiment, a few smaller studies exist. From 1981 to 1986, Durth et al. [17] introduced speed limits of 100 and 120 km/h on selected test roads in the federal state of Hesse. In a pre-post analysis of treated roads, they find substantial reductions of crash rates and even more so accident cost rates, which reflects stronger declines of more severe crashes.[3] Similarly, Schnüll et al. [18] evaluated the experimental introduction of 100 km/h and 120 km/h speed limits on 54 km of selected roads in Lower Saxony for two years. Compared to the year before the introduction, they found substantial decreases in crash numbers, crash rates and crash cost rates, while control sections showed slight increases. In an observational study, Scholz et al. [19] evaluate the introduction of a speed limit of 130 km/h on 124 km of motorways in the state of Brandenburg at the end of 2002. By comparing the reduction in crash cost rates on the treated roads before and after the introduction to the reduction on 340 km of control roads[4], they attribute a 26.5% decrease in crash cost rates to the speed limit introduction. In a theoretical exercise, Bauernschuster and Traxler [14] apply findings about the relationship between the reduction of driven speed and the number of crashes from other countries to German motorway data. For currently unconstrained sections, they derive a reduction of fatalities by $15-47\%$, of severely injured by $11-38\%$, and of lightly injured by $5-27\%$, would a speed limit of 130 km/h be introduced, whereby they consider the larger values more plausible.

As we have seen, the only thoroughly controlled large scale experiment conducted on the question of speed limit effects on crash frequencies is now more than 45 years old. Over this time span, roadway construction and design have improved, traffic counts have increased, and cars have become not only safer but also faster, which may have increased not only average driven speeds but also speed variance [14]. Furthermore, driver collective composition and drivers' attitudes may have changed. Additional studies of the introduction of various speed limits have confirmed the general direction of these early findings. They are conducted on small motorway collectives, that in most cases were selected due to higher crash incidence and do often not account for different developments in traffic counts between treatment and control segments. Reliable recent empirical evidence is missing. The various data sources and the construction of a new dataset to address this lack in research, are presented next.

3. Data

To analyze the determinants of crash frequency and the effect speed limits have in this context, a variety of geospatial data sources was combined, most of which are available openly. The goal is to control for as many factors influencing crash frequency - and thereby speed limit probability - as possible to obtain an estimate for the upper bound that is as close to a true causal effect estimate as possible. All data processing was implemented and fully automated using advanced geoprocessing tools from ArcGIS and Python to ensure reproducibility.[5] Due to data availability limitations, motorways of 12 out of 16 German states are included and aggregated over a period of three years ($2017-2019$) to reduce volatility in this rare-event data.[6]

Since most data sources are available aggregated over a calender year, one data set per year was constructed and aggregated over the three years for the analysis. The shape of the motorway network as well as information on speed limits, number of lanes, and the presence of bridges and tunnels was taken from Open Street Maps (OSM) [20]. Therefore, the state of the OSM network at the middle of each year was used. The information about motorway entrance and exit ramps was also derived from OSM data. Following the findings from Balck et al. [10], the motorway network was partitioned into segments of 500 m length in two dimensions.[7] Compared to using segments that are homogeneous in their attributes, this has the advantage that one can look into the effect changes within a segment have on accident frequencies, as well. Furthermore, previous research found this length to be optimal for the analysis of injury accidents as it reduces high volatility of too short segments but still preserves a meaningful

[1] If not stated otherwise, road lengths are always given as divided by direction.
[2] It is also noted in [16], that a temporary speed limit of 100 km/h in winter 1973/74 that was introduced during the oil crisis, is estimated to have led to a reduction of 40% in crashes resulting in fatalities or severe injuries, and of 35% in any injuries, when accounting for the reduced traffic flows. However, as the authors note, this data originates from a time of crisis, where not only compliance may have differed from normal times.
[3] The authors find reductions to be larger than on control segments, but find crash rates to increase on parallel roads, which may reflect certain groups of drivers to evade to unconstrained roads.
[4] That were constrained or unconstrained throughout the entire study period (2000-2006).
[5] The respective Python scripts are available from the author on request. They can be used to reproduce and extend the data set.
[6] Missing states are: North Rhine-Westphalia, Thuringia, Mecklenburg-West Pomerania, and Berlin
[7] Variations in actual road length arise due to differing elevation profiles.

Table 1: Average number of accidents by severity and speed limit per 500 m segment.

	100	120	130	None
Light	1.0069	0.7472	0.5129	0.6345
Severe	0.1905	0.1760	0.1852	0.2070
Fatal	0.0119	0.0180	0.0134	0.0205
Total	1.2093	0.9412	0.7115	0.8620
No of segments	1008	2449	1123	15873
Segment share	4.93%	11.97%	5.49%	77.61%

mapping between accident occurrence and road characteristics near this location that would be lost for longer segments [10].

Additional data is merged to these 500 m segments, based on their location. Accidents with injuries or fatalities are taken from the Unfallatlas [21]. Vehicle counts are taken from the yearly data of permanent counting stations [22] and interpolated for sections without a permanent counting station. A network sector ID was added [23], which was also used in the identification of ramps. Road condition measures were processed [24] in the form of averages over each 100 m segment of the network, separated by lane. A road substance and a serviceability index are used. In both indices, measures of longitudinal and cross sectional road roughness are included. The substance index includes measures of general road surface appearance and surface condition. The serviceability index includes measures of road friction and informs on driving safety and comfort [25]. Lower values indicate better road conditions. A digital elevation model on a 30 m grid is taken from [26] and changes in elevation are derived along the road segments. Furthermore, annual weather data [27] and information on the respective region from regional shapefiles [28] were employed. These were also utilized to compute sector-wise measures for various socio-demographic characteristics [29]. For this, an area-weighted average was derived over a 25 km buffer around the network sector for each socio-demographic characteristic. A list of all derived variables can be found in table 2.

To aggregate the data over the three years, crash counts are summed and traffic count and weather variables are averaged. Road condition measures are collected once in 2017 and 2018, sociodemographic variables in 2017 and the elevation profile in 2000. Roadway properties taken from OSM are required to stay constant over time or else omitted from the analysis.[8]

Table 1 gives some first insights into the constructed data set. Accidents resulting in light injuries only are more frequent the lower the speed limit is. This is likely caused by the more dangerous environment that has led to the installation of the speed limit and underlines the need to control for those factors. Severe and fatal accidents appear to be somewhat positively correlated with the absence of speed limits. Furthermore, graphic 1 shows the posted speed limits on the German motorway as of 2019, and figure 2 the number of accidents between 2017 and 2019 per 500 m segment on the 12 included states. It should be noted that while speed limits of 100 and 120 km/h can be found all over the network, speed limits of 130 km/h are mainly found in three regions on contiguous road segments.

4. Random Forests

The causal forest was first introduced in [30] and generalized and slightly adjusted in [12]. It builds on the classic random forest algorithm [31], which is a popular machine learning algorithm to estimate the conditional mean function $\mu(x) = \mathbb{E}[Y_i|X_i = x]$ of an outcome $Y_i \in \mathbb{R}$, given some covariate vector $X_i \in \mathbb{R}^p$, in a training sample of n i.i.d. instances, labeled $i = 1,...,n$. In this standard regression context, a random forest prediction at test point x is the average over the predictions of an ensemble of B individual regression trees. Regression trees aim at recursively partitioning the data in order to find subgroups that are homogeneous along the dimensions predicting the outcome so that using the mean of each resulting partition (or node) as a predictor for the whole subgroup is justified.

For each tree, a bootstrap sample is drawn from the available training data. Starting with the whole bootstrap sample as parent node, the tree is grown by recursively splitting each node into two child nodes, at the point on a single covariate that minimizes the selected loss function, when the mean of each node is assigned as its predictor. The algorithm is thus greedy by choosing the maximum improvement of fit at each individual step typically evaluated via the mean squared error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu}(X_i))^2 \quad (1)$$

To reduce correlation between the individual trees and thereby the variance of the ensemble, only $mtry < p$ available explanatory variables are randomly drawn at each step to be considered in the choice of split. This splitting routine

[8] This led to a large number of missing observations and will likely be handled differently in the future. Currently this leads to missing data for 17937 of the 38390 500 m segments in the 12 included states.

Table 2: Variable Description. For variables marked with *, lags are included as well. Variables marked with $^+$ are considered strongly spatial variables.

Variable name	Description
total	Total number of crashes between 2017 and 2019 on segment.
fatal	Number of fatal crashes.
severely_injured	Number of crashes resulting in severly injured.
lightly_injured	Number of crashes resulting in lightly injured.
maxspeed_120	Indicator of speed limit of 120 being the predominant speed limit.
tunnel*	Share of segment that is a tunnel.
bridge_total*	Share of segment that is any bridge. (bridge+small_bridge)
main_Entry*	Share of road with a main entrance ramp.
sec_Entry*	Share of road with a secondary entrance ramp (e.g. from service area).
main_Exit*	Share of road with a main exit ramp.
sec_Exit*	Share of road with a secondary exit ramp (e.g. to service area).
node_area*	Share of road at motorway node, including $500m$ before and $300m$ behind node.
straight	100$\cdot$(linear distance between start and end point$/$segment length).
right_turn*	Measured by cutting segments into $100m$ pieces and summing up angle changes.
left_turn*	Left turn measured as right turn but looking at other direction.
n_lanes	Average number of lanes on the segment.
AADT	Average annual daily traffic, all motor vehicles and weekdays$/$1000.
TVSVms	Average annual daily traffic heavy traffic.
HT_share	Share of heavy traffic in total traffic (TVSVms/AADT).
monday	Monday factor.
bSo	Sunday factor.
SV5Kfzms	Design traffic volume (volume on 50th most congested hour of year), all vehicles.
bSV50ms	Heavy traffic share in design traffic volume.
holiday	Holiday factor *cross section over both directions (cs)*.
Mt	Average hourly traffic volume by day (6am-10pm). *(cs)*
Mn	Average hourly traffic volume by night (10pm- 6am). *(cs)*
night_ratio	(Average hourly traffic by night)/(average hourly traffic by day) *(cs)*
asphalt*	Indicator of asphalt as predominant road surface (else concrete).
surface_het*	Indicator of road surface heterogeneity withing the segment.
sub_mean*	Mean of substance index.
sub_var	Variance of substance index.
sub_max	Maximum (worst) substance index.
perf_mean*	Mean of performance/serviceability index.
perf_var	Variance of performance/serviceability index.
perf_max	Maximum (worst) performance index.
down_change*	Take elevation at 21 points along segment and sum up downward changes between points.
up_change*	Sum up upward changes between these points.
max_slope*	From elevation map, derive slope and take the maximum at any of the points along the segment.
mean_slope$^{*\,+}$	Average slope over all segment points.
elevation$^+$	Average elevation on all segment points.
air_temperature_mean$^+$	Average annual air temperature at the location of the segment.
frost_days$^+$	Average annual number of days with minimum air temperature $< 0C$.
ice_days$^+$	Average annual number of days with maximum air temperature $< 0C$.
snowcover_days$^+$	Average annual number of days with snowcover $> 1cm$ in the morning.
precipGE10mm_days$^+$	Average annual number of days with $>= 10mm$ precipitation.
precipGE20mm_days$^+$	Average annual number of days with $>= 20mm$ precipitation.
precipGE30mm_days$^+$	Average annual number of days with $>= 30mm$ precipitation.
precipitation$^+$	Average annual sum of precipitation in *mm*.
summer_days$^+$	Average annual number of days with maximum air temperature $>= 25C$.
sunshine_duration$^+$	Average annual sunshine duration in hours.
wind$^+$	Average wind speed $10m$ above ground over the years 1981-2000
emp_quo$^+$	Regional employment rate (average over region with radius of 25km around segment).
pop_18_25$^+$	Regional share of population older or equal to 18 and younger than 25 years.
pop_ol_65$^+$	Regional share of population aged 65 and older.
fem_share$^+$	Regional share of women in total population.
hh_inc$^+$	Regional average monthly disposable household income in euro.
pop_dens	Regional population density.
pcar_dens$^+$	Regional density of passenger cars.
gdp_p_cap$^+$	Regional GDP per capita.
rurality$^+$	Regional rurality index.
location	Identifier for consecutive road segments of same motorway number within one state.

is stopped when the fit cannot be further improved or some minimum node size min_n is undercut. For the tree prediction $\hat{\mu}_b$ of some test point x, this point is send down the tree following the previously constructed splitting rules. To reduce the variance of these noisy estimates, all B trees are averaged to obtain the random forest predictor: $\hat{\mu}_B = \sum_{b=1}^{B} \hat{\mu}_b$ [32].

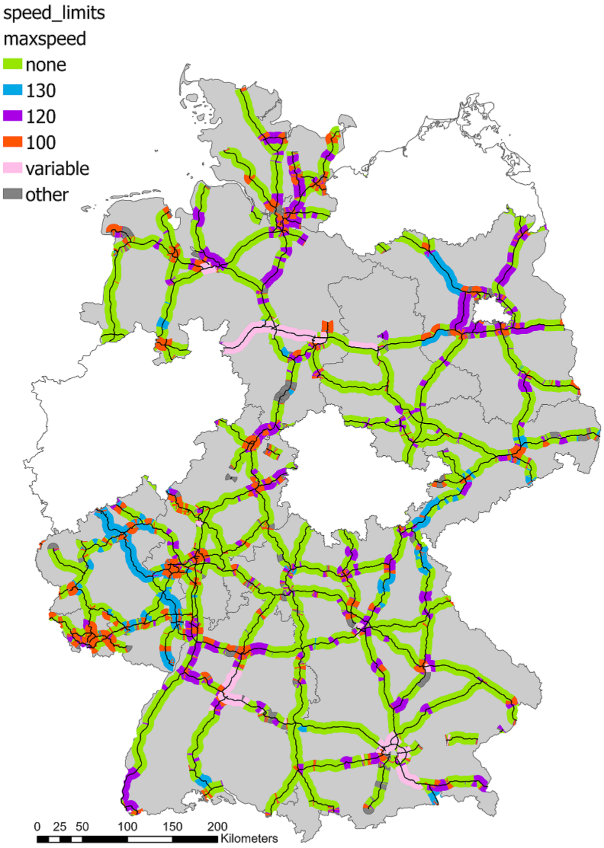

Figure 1: Speed limits in the 12 considered states in 2019.

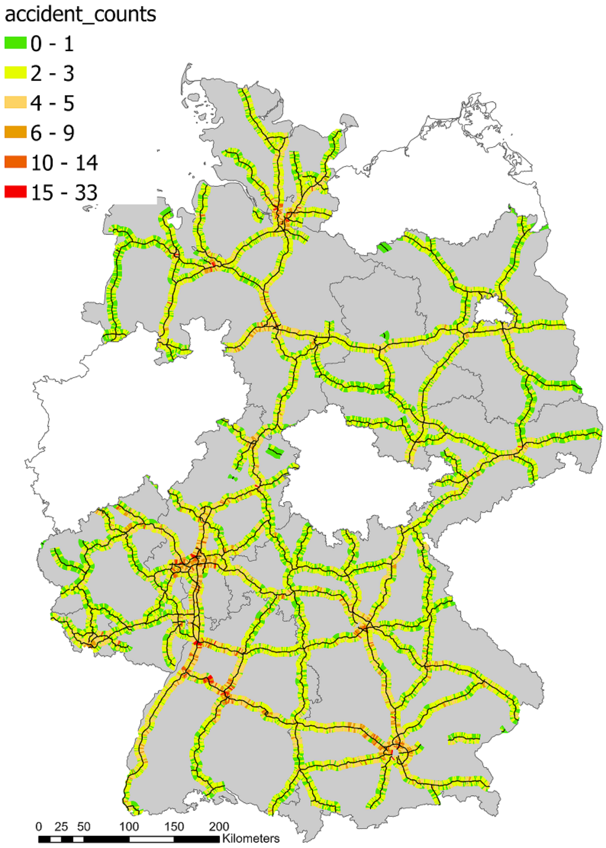

Figure 2: Accident counts per 500m segment from 2017-2019.

For the generalized random forest framework developed in [12], it is important to see that the random forests can also be represented as a weighted average of the training sample: $\hat{\mu}_B = \sum_i^n \alpha_i(x) Y_i$. With $\alpha_i(x) = \frac{1}{B}\sum_{b=1}^{B} \alpha_{bi}(x)$ and $\alpha_{bi}(x) = \frac{\mathbb{1}(\{X_i \in L_b(x)\})}{|L_b(x)|}$ where $\mathbb{1}(\cdot)$ denotes the index function and $L_b(x)$ is the set of training samples that fall into the same terminal node (or leaf) as the test point. The weights $\alpha_i(x)$ sum to 1 and are larger for observations i that fall into the same leaf as x more frequently. In this sense, random forests may also be interpreted as adaptive nearest neighbor estimators where the weights $\alpha_i(x)$ define the adaptive neighborhood of x [12]. While the performance of nearest neighbor methods typically suffers from the *curse of dimensionality*, adaptive forest-based neighborhoods are wider along dimensions irrelevant for outcome prediction and narrower along dimensions with strong signals [33]. The derived weights are thus specific to the problem at hand.

4.1. Potential outcomes framework

Causal forests define treatment effects under the potential outcomes framework [34] [30]. With a binary treatment indicator $W_i \in \{1,0\}$, it is assumed that for each observation i, there exist two potential outcomes that could have occurred under the respective treatment regimes: $Y_i^{(0)}$ if no treatment was received, and $Y_i^{(1)}$ if it was. The conditional average treatment effect (CATE) function is then defined by:

$$\tau(x) = \mathbb{E}[Y_i^{(1)} - Y_i^{(0)} | X_i = x] \tag{2}$$

Naturally, only the outcome corresponding to the realized treatment $W_i \in \{0,1\}$ can be observed: $Y_i = Y_i^{(W_i)}$.

This view inherently entails a number of assumptions, when the aim is to estimate equation 2 via local methods such as causal forests. **Unconfoundedness** (or CIA) assumes that potential outcomes are independent ($\perp\!\!\!\perp$) of treatment assignment conditional on the covariates [35]:

$$\{Y_i^{(1)}, Y_i^{(0)}\} \perp\!\!\!\perp W_i | X_i. \tag{3}$$

Observations that are sufficiently close to each other with respect to confounding covariates can thus be considered as draws from a randomized experiment [30]. To ensure that local estimators, that aim at controlling for confounding factors, are feasible, **common support** is required:

$$\varepsilon < P[W_i = 1|X_i = x] < 1 - \varepsilon \tag{4}$$

for all x and some $\varepsilon > 0$. As mentioned above, these two assumptions cannot simply be assumed to hold in the analysis of speed limit effects but are replaced by the upper bound assumption defined above.

4.2. Causal Forests

Causal forests essentially estimate an empirical representation of equation 2 in the forest-based adaptive neighborhood of x, defined by weights $\alpha_i(x)$ [12]. As in the standard random forest algorithm described above, weights are obtained from an ensemble of trees, which are grown via recursive partitioning. As direct counterparts of function 1 are not available for the case of heterogeneous treatment effect estimation, [30] note that minimizing the MSE in the parent node amounts to maximizing the heterogeneity between nodes. Therefore, a splitting criterion is optimized that maximizes the heterogeneity of treatment effect estimates in the two resulting child nodes.[9] Therefore, causal forests essentially identify neighborhoods in which treatment effects are relatively constant and differ from other nodes, so that it estimates heterogeneous treatment effects.

While an early version of the causal forest simply subtracted average outcomes in the treatment from average outcomes in the control group [30], the causal forest's performance was later improved by employing a more robust estimator [12] :

$$\hat{\tau}(x) = \frac{\sum_{i=1}^{n} \alpha_i(x)(Y_i - \hat{m}^{(-i)}(X_i))(W_i - \hat{e}^{(-i)}(X_i))}{\sum_{i=1}^{n} \alpha_i(x)(W_i - \hat{e}^{(-i)}(X_i))^2}. \tag{5}$$

Y_i and W_i are orthogonalized by regressing out their main effects $m(X_i)$ and $e(X_i)$, respectively. More concretely, $e(X_i) = \mathbb{E}[W_i|X_i]$ denotes the treatment probability of segment i and $m(X_i) = \mathbb{E}[Y_i|X_i]$ its expected crash rate, both assumed to be a function of the covariates. Therefore, two separate regression forests are trained to predict outcome and treatment assignment and to estimates $\hat{m}^{(-i)}(X_i)$ and $\hat{e}^{(-i)}(X_i)$ "out-of-bag", i.e. using only the trees grown without observation i. This local centering step makes the estimator more robust to selection bias from observed characteristics, and is thus important in the presented application.

To ultimately make the causal forest pointwise consistent and asymptotically Gaussian, two major changes are made to the original tree building process in [31]. First, trees are grown on subsamples of subsample size s, rather than bootstrap samples of the training data. And second, trees have to satisfy an concept called *honesty* [30].[10]

5. Results

5.1. Propensity Score Estimation

To employ the criterion defined in 5, an estimator for the propensity score is required. The standard approach implemented in [36] is to fit a separate random forest for this purpose and to use out-of-bag estimates for each observation. The spatial nature of our data does however require a somewhat different strategy for the task. The problem arises from the fact that not only many of the exploratory variables, especially socio-demographic and weather characteristics, are strongly auto-correlated across space, but also speed limits are clustered together on contiguous or parallel road segments over several kilometers, as we can see in figure 1. This is not necessarily caused only by underlying (observed or unobserved) explanatory factors but may be the result of pragmatism, traffic planning and political decisions to name a few. Therefore, the forest can exploit the variance in such strongly spatial variables to identify treated segments, especially when combined with further spatial dummies. This holds for the socio-demographic and weather characteristics, but may also arise - in a somewhat less severe sense - with the various traffic variables, that are constant between two motorway nodes.

As the spatial clustering of 130 km/h limits is much more severe than for the other two speed limits and since segments restricted by 100 km/h limits can be assumed to be affected by unobserved heterogeneity much more strongly, the presented study focuses on the effect of 120 km/h limits. Many of the strongly spatially auto-correlated variables are also strongly correlated with each other and can be assumed to have a rather mild individual effects on

[9] For computational efficiency, a gradient-based approximation of the loss criterion is used in the generalized random forests algorithm used here. Technical details thereof are described in Section 2.3 of [12].

[10] In building an individual tree, the response Y_i of each training unit i can be used either to build the tree (select splitting variables and place splits) or to estimate the target, but never for both [30]. This is implemented by splitting the subsample for each tree into two parts, one of which is used for split selection, and the other for treatment effect estimation.For additionally required regularity assumptions, it is referred to [12].

crash frequency and speed limit probability. These variables are marked with a $^+$ symbol in table 2. As including all of these variables would be redundant, a principal component analysis (PCA) algorithm is run that extracts 4 principal components (p_comp_0-p_comp_1) to use in place of all the individual variables in the remainder of this paper. These variables still contain close to 80% of the variation of these variables and should thus cover their main features.

To respect the spatial nature of the data, separate locations are identified in the motorway network by considering contiguous segments belonging to the same motorway (number) and lying in the same federal state as one 'location'. To illustrate the problem of spatial over-fitting when ignoring the spatial nature of the data, we first set aside 20% of the data as hold-out set that only contains complete locations. For the rest of the data, $mtry$ and min_n are chosen via standard 4-fold cross-validation (CV). The aim is to get an estimate for speed limit probability for segments having no or a 120 km/h speed limit. Instead of using accuracy, the share of correct predictions, as model fit score, the F1-score is more appropriate for the unbalanced classification task at hand. It balances the precision and recall scores that intuitively evaluate the ability of the classifier to avoid false positives and false negatives, respectively. We can see the results of evaluating the selected model on our hold-out set in the first row of table 3. Strikingly, 3 out of 4 principal components are among the 6 most important variables to predict treatment propensity. In the presence of various traffic and geometric features, this is highly implausible and can be regarded an indication of spatial over-fitting.

Table 3: Cross validation results evaluated on a hold-out set with only unseen location.

	Accuracy	Precision	Recall	F1-score	All or 6 most important variables
Standard CV	0.85	0.27	0.07	0.11	p_comp_1, p_comp_2, pop_dens, holiday,
LLO CV	0.83	0.24	0.11	0.15	Mt, p_comp_0
LLO CV with FFS	0.81	0.36	0.49	0.41	Mt, TVSVms

Two remedies for this behavior come from the geographic literature, namely [37]. First, they propose leave-location-out CV (LLO-CV), which means putting only complete locations into CV folds, so that the prediction error is always evaluated only on locations that the algorithm has not seen any instances of during training. However, the model training itself, aims only at optimizing the fit of the training data itself. Therefore, aside from giving more realistic error estimates, spatial CV may affect model regularization by choosing different tuning parameters, but it does not affect the choice of splitting variables itself. Our results of this approach for 4-fold CV can be seen in the second row of table 3. While the F1-score slightly improves due to target-oriented parameter tuning, the individual forests still put a strong emphasis on spatial variables, which is reflected in obtaining the same set of most important variables.

To overcome this issue, Meyer et al. [37] propose forward-feature-selection (FFS), which works as follows: First, random forests considering only 2 variables are grown on the training data for all possible 2-variable combinations. The model resulting in the best prediction score on the test data with only unseen locations, is selected. Then the number of variables is increased by iteratively adding the one variable that leads to the largest performance increase on unseen locations. If this can achieve no further improvement, variable selection is terminated. We implement this approach again using 4-fold LLO CV to select variables and then tuning parameters and evaluate model performance on our hold-out set. As can be seen in the last row of table 3, only 2 variables (both containing information on traffic flows) are selected by this approach, leading to a considerable improvement of the F1-score.

To obtain the final propensity score estimates, all data is split into 5 folds. For each fold, a random forest containing only the selected variables and tuning parameters is trained on the 4 other folds. This model is then used to obtain predictions for the left-out fold. This is done for each fold to get estimates for all observations. The procedure is repeated 10 times with different (LLO) splits and predictions for each observation are averaged, to obtain more robust propensity score estimates. The results are presented in figure 3, where $\hat{e}^{(-g(i))}$ indicates that each estimate is obtained without considering the location g observation i belongs to. Selected quantiles are represented in the figure showing a high share of segments with a very low speed limit probability and only few segments with very high propensity scores. We can see that most segments are associated with relatively low probabilities of receiving a speed limit, while only for few segments, the environment seems to be dangerous enough to inevitably require a speed limit, reflected by a high propensity score. Other reasons for the installation may be noise protection or the closeness to more dangerous sections.

We argue that while some degree of spatial over-fitting likely arises when predicting $\hat{m}(X_i)$, it does not lead to the same kind of problem. The regression forest for predicting this property may also exploit spatial features to explain spatial clustering of a certain level of crash rates, not otherwise explained by observed characteristics. But in contrast to the speed limit probability problem, these clusters are likely caused by some unobserved spatial characteristics, that truly are relevant for crash rate prediction and are not a result of human intention. Therefore,

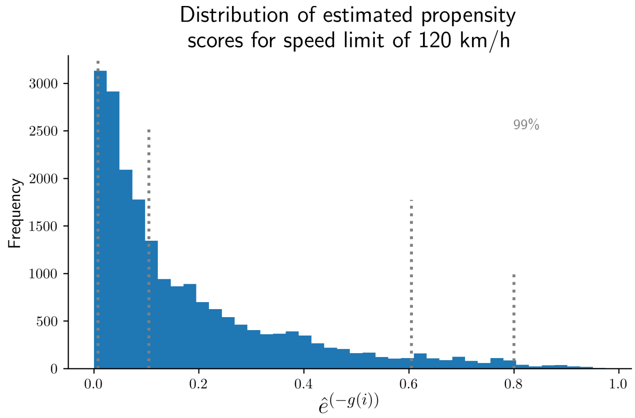

Figure 3: Estimated propensity scores and respective quantiles when using spatial CV and FFS.

the forest may actually use spatially auto-correlated as a sort of flexible spatial dummy generator, which in this case may even improve crash rate prediction. If the aim was to predict accident frequencies for unknown locations, this behavior would not be desirable and a similar approach as before would be required to predict the outcome, as otherwise spatial overfitting would lead to poor generalization performance.

5.2. Causal Forest Results

Non-parametric local methods such as random forests tend to perform poorly in fitting strong smooth signals [38]. In the application at hand, traffic density clearly has such a strong effect not only on expected crash frequency, but also on conditional treatment effects. The effect on crash counts has been found to be close to linear - yet slightly regressive - by previous studies [10]. Therefore, it seems beneficial to use crash rates, e.g. defined as number of accidents per 10.000 vehicles per day, rather than crash counts as outcome in our analysis. While such accidents rates have been considered censored in some applications [39], here the observed zeros are viewed as *real* zeros, and not a representations of some unobservable latent outcome. As discussed in Chapter 3.4.2 of [11], estimation via empirical representations of 2 is thus appropriate to directly identify causal effects. As causal forests do not extrapolate, but work with weighted averages, they avoid predicting outside of the range of the outcome in a non-parametric way. In addition, it is described in more detail in section 6 of [12], that only the *expected* outcome and treatment assignment need to be Lipschitz continuous in x for the causal forest to be applicable. For $\mathbb{E}[Y_i|X_i = x]$, this holds true, as long as the probability of observing a specific number of crashes is smaller than 1 for all segments. Even the safest roads have a small but positive probability of observing a crash due to major driving mistakes, at any day with positive traffic counts. Therefore, this assumption is well justified and the causal forest seems to be an appropriate modeling choice in this regard.

To avoid violations of the upper bound assumption by comparing potentially very dangerous unrestricted segments with potentially less dangerous segments with lower treatment probability, all 72 segments with an estimated propensity score >= 0.95 are dropped. Then, a causal forest is fit for each severity level of crash rates. Results are presented in table 4. To give the results a meaningful interpretation, in addition to conditional average treatment effects (CATEs), individual semi-elasticities are reported that were derived by:

$$SE_{W_i}^{\hat{Y}_i} = \frac{\hat{Y}_i^{(1)} - \hat{Y}_i^{(0)}}{\hat{Y}_i^{(0)}}. \tag{6}$$

We can see that the estimated CATE is statistically significant at the 5% level only for fatal and severe crash rates. Speed limits of 120 km/h are found to reduce severe crash rates by 21% and fatal crash rates by 24%. Note that under the upper bound assumption, these reductions can be thought of as minima for the absolute value of an exact treatment effect estimate. For the total and light crash rates, less pronounced negative values are not found to be significantly different from 0.

As mentioned above, causal forests are designed to gain insights into treatment effect heterogeneity. A plausible assumption would be that a roads congestion changes the effect of speed limits. For example, it may be the case that roads with a high AADT would benefit more from setting a speed limit, as higher driven speed is most dangerous on crowded roads. Therefore, CATEs are derived separately for segments with above and below median AADT. This hypothesis can be confirmed by our results. For more congested roads, strongly significant negative effects are found for severe and fatal crash rates, jointly and individually, indicating a reduction of 21% and 24%, respectively.

Table 4: Causal forest regression results with standard error clustered at location level.

Outcome:	total rate	fatal rate	severe rate	light rate	fatal & severe
ATE	-0.03	-0.003**	-0.017**	-0.009	-0.019**
Std.err.	(0.025)	(0.001)	(0.008)	(0.018)	(0.008)
CI (95%)	-0.079;0.02	-0.005;0	-0.032;-0.001	-0.045;0.027	-0.034;-0.004
$\bar{SE} \cdot 100$	**-10.72%**	**-24.2%**	**-21.13%**	**-6.3%**	**-21.2%**
$\bar{Y}$	0.363	0.009	0.091	0.264	0.099
Share of treated			13.4%		
$\text{CATE}_{(AADT>q_{50,AADT})}$	-0.062**	-0.004***	-0.024***	-0.031	-0.028***
Std.err	(0.025)	(0.001)	(0.006)	(0.021)	(0.006)
CI (95%)	-0.111;-0.013	-0.006;-0.002	-0.035;-0.013	-0.071;0.01	-0.039;-0.016
$\bar{SE} \cdot 100$	-11.4%	-22.6%	-22.6%	-6.9%	-22.3%
$\bar{Y}\vert\text{AADT}> q_{50,AADT}$	0.356	0.009	0.077	0.271	0.086
Share of treated			17%		
$\text{CATE}_{(AADT<=q_{50,AADT})}$	0.002	-0.001	-0.009	0.013	-0.011
Std.err	(0.04)	(0.002)	(0.014)	(0.027)	(0.013)
CI (95%)	-0.077;0.081	-0.005;0.002	-0.036;0.018	-0.041;0.067	-0.037;0.015
$\bar{SE} \cdot 100$	-10.1%	-25.7%	-20%	5.7%	-20%
$\bar{Y}\vert\text{AADT}<= q_{50,AADT}$	0.369	0.009	0.105	0.256	0.113
Share of treated			9.8%		
95% $\text{CI}_{\hat{\tau}_{AADT_h}-\hat{\tau}_{AADT_l}}$	-0.157;0.029	-0.006;0.002	-0.044;0.014	-0.111;0.023	-0.045;0.013
			N=20485,	*: p=0.1, **: p=0.05, ***: p=0.01	

It seems that this is also what drives the overall crash rates to be significant at the 5% level, as the light crash rate does not show a significant effect. For less congested roads, no statistically significant effects are identified. Even an estimated reduction of 26% of fatality rates on these roads seems to be overburdened with uncertainty and would require further investigation.

6. Conclusion

Upper bound for the effects of a speed limit of 120 km/h were estimated on different levels of crash severity rates. It was found that such a speed limit only influences fatal and severe crash rates, while no statistically significant effect was found for light crash rates. Fatal crash rates are found to be reduced by $22\% - 26\%$ and severe crash rates by $20\% - 23\%$. For both severity levels, the effect was statistically significant only for more congested roads. As the absolute effect size, does however not differ systematically, this might be a result of overall lower numbers of observed crashes where fewer vehicles are found. As more data becomes available, a closer look at this relationship may shed more light on these different effects.

In addition this paper showcased the pitfalls and potential solutions for using machine learning in the context of spatial data, an issue that - to our best knowledge - has not previously been discussed in the causal machine learning literature. It could be shown how methods from the geographic toolset can be used to avoid spatial over-fitting and get more reliable estimates of the spatial targets.

The findings of this study are overall in line with previous research, that has repeatedly shown a much larger impact of different speed limits on severe and fatal crashes than on light crashes, or even crashes leading to property damage only. The analysis indicates strong safety benefits from the introduction of a 120 km/h speed limit and thus underlines the importance of this discussion. However, empirical studies always rely on relatively strict assumptions, as the upper bound assumption made here. New, thoroughly conducted, large-scale experiments would therefore be of great interest to gain further insights into the effects of 120 as well as 130 km/h speed limits.

References

1. Tim Holthaus, Claus Goebels, and Bert Leerkamp. Evaluation of driven speed on German motorways without speed limits. Technical report, Bergische Universität Wuppertal, 2020.

2. Katrin Bennhold. Impose a Speed Limit on the Autobahn? Not So Fast, Many Germans Say. *The New York Times*, February 2019, 2019. URL https://nyti.ms/2DR79Os.

3. Ben Knight. German government shows cracks over autobahn speed limit. *Deutsche Welle*, December 2019, 2019. URL https://p.dw.com/p/3VM0H.

4. VCD. Tempolimit auf Autobahnen für Verkehrssicherheit und Klimaschutz. Technical report, Verkehrsclub Deutschland e.V., 2019.

5. VDA. Fakten gegen ein generelles Tempolimit. Technical report, Verband der Automobilindustrie e.V., 2021. Version 4.0.

6. Martin Bärwolff, Ralf Berger, Christian Lippold, Martin Schmotz, Juliane Martin, and Bettina Schröter. Maßnahmen zur vermeidung von abkommensunfällen auf bab. Technical report, Bundesanstalt für Straßenwesen, 2019.

7. Panagiotis Ch. Anastasopoulos and Fred L. Mannering. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis & Prevention*, 41(1):153–159, 2009.

8. Dominique Lord and Fred Mannering. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5):291–305, 2010.

9. Fred L. Mannering, Venky Shankar, and Chandra R. Bhat. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*, 11:1–16, 2016.

10. Henning Balck, Hagen Schüller, Michael Balmberger, and Christian Rossol. Verfahren zur zusammenführung von informationen unterschiedlicher netzanalysesysteme. Technical Report Verkehrstechnik - Heft V 298, Bundesanstalt für Straßenwesen, 2017.

11. Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

12. Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized Random Forests. *Annals of Statistics*, 47(2): 1148–1178, 2019.

13. Bundesanstalt für Straßenwesen. The bast - tasks, objectives and history. https://www.bast.de/BASt_2017/EN/BASt/BASt_node.html, 2021. Accessed: 2021-05-26.

14. Stefan Bauernschuster and Christian Traxler. Tempolimit 130 auf Autobahnen: Eine evidenzbasierte Diskussion der Auswirkungen. *Perspektiven der Wirtschaftspolitik*, 22(2):86–102, 2021.

15. BASt. Auswirkung einer richtgeschwindigkeit im vergleich zu einer höchstgeschwindigkeit von 130 km/h auf autobahnen. Technical report, Bundesanstalt für Straßenwesen, Projektgruppe "Autobahngeschwindigkeiten", Köln, 1977.

16. BASt. Abschätzung der auswirkungen einer senkung der höchstgeschwindigkeit auf das unfallgeschehen im straßenverkehr". Technical report, Bundesanstalt für Straßenwesen, Bergisch Gladbach, 1984.

17. Walter Durth, J.S. Bald, K. Habermehl, H. Hanke, R. Lutz, and R. Roos. Großversuch zur geschwindigkeitsbeschränkung auf autobahnen in hessen - untersuchung zum fahr- und unfallverhalten. Technical report, TU Darmstadt, Darmstadt, 1986. im Auftrag des Hessischen Ministers für Wirtschaft und Technik.

18. Robert Schnüll, Norbert Handke, Stephan Hoffmann, and Fabian Schütte. Auswirkungen von verkehrsbeschränkungen an hochbelasteten bundesautobahnen. *Straßenverkehrstechnik*, 39(4):162–168, 1995.

19. Thomas Scholz, Antje Schmallowsky, and Tilmann Wauer. Auswirkungen eines allgemeinen Tempolimits auf Autobahnen im Land Brandenburg. Technical report, 2007.

20. OpenStreetMap contributors. Retrieved via https://overpass-turbo.eu/. https://www.openstreetmap.org, 2021. [License: CC BY-SA 2.0 - https://www.openstreetmap.org/copyright].

21. Statistische Ämter des Bundes und der Länder. Unfallatlas. https://unfallatlas.statistikportal.de/, 2020. [License: dl-de/by-2-0 - http://www.govdata.de/dl-de/by-2-0].

22. Bundesanstalt für Straßenwesen. Automatische Zählstellen 2017-2019. https://www.bast.de/BASt_2017/DE/Verkehrstechnik/Fachthemen/v2-verkehrszaehlung/zaehl_node.html;jsessionid=9BE1F5EC97952EEFAB21C06D3F0BFD7B.live11314, 2021.

23. Bundesanstalt für Straßenwesen. Bundesfernstraßennetz. https://www.bast.de/BASt_2017/DE/Verkehrstechnik/Fachthemen/Daten/Daten-BISStra.html?nn=1817946, 2020.

24. Bundesanstalt für Straßenwesen. Zustandserfassung und -bewertung (ZEB), BAB 2017 & 2018, 2021.

25. Bernhard Steinauer, Andreas Ueckermann, and Günther Maerschalk. Analyse voliegender messtechnischer Zustandsdaten und Erweiterung der Bewertungsparameter für Innerortsstraßen. Technical Report Straßenbau - Heft S 46, Bundesanstalt für Straßenwesen, 2006.

26. NASA. Shuttle Radar Topography Mission (SRTM) 1 arc second. https://www2.jpl.nasa.gov/srtm/, 2015.

27. Deutscher Wetterdienst. Climate Data Center, annual grids. https://opendata.dwd.de/climate_environment/CDC/grids_germany/annual/, 2020.

28. Bundesamt für Kartographie und Geodäsie. Verwaltungsgebiete 1:250 000 mit Einwohnerzahlen (Ebenen), Stand 31.12. (VG250-EW 31.12.). https://gdz.bkg.bund.de/index.php/default/open-data/verwaltungsgebiete-1-250-000-mit-einwohnerzahlen-ebenen-stand-31-12-vg250-ew-ebenen-31-12.html, 2021. [License: dl-de/by-2-0 - http://www.govdata.de/dl-de/by-2-0].

29. BBSR Bonn. INKAR - Indikatoren und Karten zur Raum- und Stadtentwicklung. https://www.inkar.de/, 2021. [License: dl-de/by-2-0 - http://www.govdata.de/dl-de/by-2-0].

30. Stefan Wager and Susan Athey. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

31. Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

32. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data mining, inference, and prediction*. Springer Series in Statistics. Springer, New York, NY, 2 edition, 2009. URL https://hastie.su.domains/ElemStatLearn/printings/ESLII_print12_toc.pdf.

33. Yi Lin and Yongho Jeon. Random Forests and Adaptive Nearest Neighbors. *Journal of the American Statistical Association*, 101(474):578–590, 2006.

34. Donald Rubin. Estimating causal effects of treatment in randomized and non-randomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

35. Donald Rubin and Paul Rosenbaum. The central role of the propensity score in obxervational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

36. Julie Tibshirani, Susan Ahtey, Erik Sverdrup, and Stefan Wager. *grf - Generalized Random Forests*. https://cran.r-project.org/package=grf, r package version 2.0.2 edition, 2021.

37. Hanna Meyer, Christoph Reudenbach, Tomislav Hengl, Marwan Katurji, and Thomas Nauss. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, 101:1–9, 2018.

38. Rina Friedberg, Julie Tibshirani, Susan Athey, and Stefan Wager. Local linear forests. *Journal of Computational and Graphical Statistics*, 30(2):503–517, 2021.

39. Panagiotis Ch. Anastasopoulos, Fred L. Mannering, Venky N. Shankar, and John E. Haddock. A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Analysis & Prevention*, 45:628–633, 2012.