RUHR
ECONOMIC PAPERS

Maike Metz-Peeters

The Effects of Mandatory Speed Limits on Crash Frequency - A Causal Machine Learning Approach

RUB rwi #982

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Ludger Linnemann
Technische Universität Dortmund, Department of Business and Economics
Economics – Applied Economics
Phone: +49 (0) 231/7 55-3102, e-mail: : Ludger.Linnemann@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Manuel Frondel, Prof. Dr. Torsten Schmidt, Prof. Dr. Ansgar Wübker
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #982

Responsible Editor: Thomas Bauer

ISSN 1864-4872 (online) – ISBN 978-3-96973-147-5

Ruhr Economic Papers #982

Maike Metz-Peeters

The Effects of Mandatory Speed Limits on Crash Frequency - A Causal Machine Learning Approach

RUB

RWI

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973147
ISSN 1864-4872 (online)
ISBN 978-3-96973-147-5

Maike Metz-Peeters[1]

The Effects of Mandatory Speed Limits on Crash Frequency - A Causal Machine Learning Approach

Abstract

This study analyzes the effects of binding local speed limits on crash frequency on German motorways. Various geo-spatial data sources are merged to a new data set providing rich information on characteristics for 500-meter segments of large parts of the German motorway network. The empirical analysis uses a causal forest, which allows to estimate the effects of speed limits on crash frequency under fairly weak assumptions about the underlying data generating process and provides insights into treatment effect heterogeneity. The paper is the first to discuss the issue of spatial over-fitting and potential solutions in the context of causal machine learning. Substantial negative effects of three levels of speed limits on accident rates are found, being largest for severe, and especially fatal crash rates, while effects on light crash rates are rather moderate. The heterogeneity analysis suggests that the effects are larger for less congested roads, as well as for roads with entrance and exit ramps, while heterogeneity regarding shares of heavy traffic is inconclusive.

JEL-Codes: R41, R42, R48

Keywords: Crash frequency; speed limits; German Autobahn; causal machine learning; causal forest; spatial machine learning

January 2023

1 Maike Metz-Peeters, RUB and RWI. - *I would like to thank Thomas K. Bauer for his valuable comments. Furthermore, I would like to thank Marco Irzik, Martin Pöppel-Decker and Andreas Schepers of the Federal Highway Research Institute (BASt, Bundesanstalt für Straßenwesen) for a helpful discussion and useful directions. In addition, I would like to thank the Federal Ministry for Digital and Transport (BMVI, Bundesministerium für Digitales und Verkehr) and the BASt for the provision of the road condition data (ZEB). I acknowledge financial support from the Mercator Research Center Ruhr.* *– All correspondence to: Maike Metz-Peeters, Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany, e-mail: Maike.Metz-Peeters@ruhr-uni-bochum.de*

1 Introduction

The German Autobahn is famous for being the only public road system in an advanced economy without binding speed limits on large parts of the network. With only a general *recommended* maximum speed of 130 km/h in place, driving speeds above and beyond this threshold are not uncommon (Holthaus *et al.*, 2020). The sensibility of a general mandatory speed limit of 130 or even 120 km/h has been the subject of a vital and emotional public debate for many years (Bennhold, 2019; Knight, 2019). Supporters emphasize environmental and alleged safety benefits, as well as a reduction in traffic jams and economic gains (VCD, 2019). Opponents challenge these arguments, claiming irrelevant emission and crash reductions, high economic costs, and a loss of time and personal freedom in case of the introduction of a general speed limit (VDA, 2021; Bennhold, 2019). As the question of the effects of speed limits on crash frequency is central in this discussion, this study aims at providing recent empirical evidence on the effects of different levels of mandatory local speed limits on crash frequency, as compared to the complete absence of binding limits.

Despite the heated public debate and the significance of the German motorway network in the European transport infrastructure, international studies on the effects of mandatory speed limits against the absence of binding speed limits on crash frequency do not exist. On the national level, the only large scale experiment analyzing the effects of a speed limit of 130 km/h compared to the recommended speed limit on crash occurrence has been conducted more than 45 years ago, a time span in which there have been significant changes with respect to vehicle safety and roadway construction (Bärwolff *et al.*, 2019). The study by BASt (1977) started in 1974 and was conducted on roughly $3,000$ km of motorway.[1] A speed limit of 130 km/h was installed on half these segments. After one year, test- and control segments were exchanged to remove time- and segment-specific effects. The authors found a reduction of crashes involving injuries of about 10% and of fatalities and severely injured by roughly 20% through the mandatory limit. Based on these results, BASt (1984) theoretically deducted that a speed limit of 100 km/h would reduce crashes by 37%. As is noted in BASt (1984), this result may be considered a lower

[1] In the present paper, all road lengths are derived by considering both directions separately.

bound for the absolute size of the effect, as the model assumes that only crash severity is affected, not crash frequency.

Similar environments allowing for the empirical analysis of mandatory speed limits compared to the absence of binding limits do not exist in other countries. However, a number of international studies investigate the effects of variations in binding speed limits on crash frequencies. Ashenfelter and Greenstone (2004) analyze an increase in speed limits on rural interstate roads from 55 to 65 mph in some US federal states in 1987. Using a difference-in-differences approach, they find an increase in fatality rates of about 35%. Analyzing these and additional speed limit changes in the year 1996, van Benthem (2015) finds an increase in speed limits by 10 mph to generally increase crashes by $9 - 15\%$, and fatal crashes by $34 - 60\%$. The author finds increases in speed limits to lead to a shift towards more severe crashes. Aarts and van Schagen (2006) provide a review on the relationship between driven speeds and crash occurrence. They conclude that an increase in driven speeds generally increases crash rates for individual vehicles as well as for road segments. Furthermore, they find a larger increase for smaller or urban roads than for larger or rural roads and that larger speed variances lead to higher crash rates. In this strand of literature, relational models, defining the number of crashes after a change in driven speeds as a function of the previous number of crashes and the previous as well as the new average driven speed, play an important role (Elvik *et al.*, 2019). The exact form of these models depends on the crash severity level considered.

In a German policy paper, Bauernschuster and Traxler (2021) aim at obtaining insights into the relationship between mandatory speed limits and crash numbers, based on these models. In absence of reliable measurements of driven speeds for the German motorway, they assume that a speed limit of 130 km/h would result in a 5-10 km/h reduction in the average driven speed of light traffic, as compared to no speed limit. They derive reductions of $15 - 47\%$ in fatalities, of $11 - 38\%$ in severely injured, and of $5 - 27\%$ in lightly injured through a speed limit of 130 km/h, whereby they consider the larger values in these wide ranges more plausible. Bauernschuster and Traxler (2021) stress, however, that their estimates are impaired by the lack of data and the high uncertainty regarding the used functional forms. The latter is also caused by the fact that evidence allowing for

the validation and parameterization of these models for the German case does not exist. This lack is in turn also due to the absence of a data set that would allow for analyzing the relationship between speed limits and crashes on the German Autobahn.

To close this gap, this paper processes and combines various geo-spatial data sources in order to obtain a data set containing rich segment-wise information on roadway geometry, traffic properties, environmental characteristics and regional socioeconomic information of 500 m motorway segments from 12 out of 16 German states. The data is aggregated over the years $2017 - 2019$ to reduce volatility in crash counts. The final data set covers about 50% of the over $26,000$ km of German motorway network and thus vastly exceeds all previous studies on the topic in spatial extent.[2]

An additional challenge for the identification of causal effects of speed limits on crash frequency is the problem of unobserved heterogeneity, that is inherent to such questions (Lord and Mannering, 2010; Mannering *et al.*, 2016). Identification of causal effects requires to correctly control all factors jointly influencing the probability of speed limit installation and crash occurrence. As speed limits are often installed as a result of higher (expected or observed) crash incidence, this is clearly unfeasible and unobserved heterogeneity is inevitable. Instead, it is assumed in this paper, that unobserved factors that increase (decrease) crash frequency also increase (decrease) speed limit probability, so that the estimated effects of speed limits on accident frequencies are upward biased (Angrist and Pischke, 2009). Under this *upper-bound assumption*, an estimated negative effect reflects an upper bound of the true causal effect and a lower bound in absolute values. To avoid violations of this assumption, it is only crucial to control for variables that appear dangerous (safe) and thus influence the traffic planners' decision to install precautionary speed limits, but that actually lead to more cautious (risky) driving and thus ultimately to lower (higher) crash probability. As such characteristics are observable to traffic planners, accounting for them with observational data is likely much easier than controlling for all potential confounders.

To have full control of the upper bound assumption, a causal forest is employed (Athey *et al.*, 2019). The approach is entirely non-parametric, and, thus, does not require strong

[2]An open version of this newly constructed data set and the corresponding code are available under the CC BY-SA 4.0 license on https://github.com/maikemp/motorwayData.

assumptions about the data generating process (DGP). In addition, causal forests help to explore treatment effect heterogeneity and to formally test pre-specified hypotheses regarding effect heterogeneity without the necessity to explicitly model interaction terms. To further investigate the established finding that an increase in the variance in driven speeds increases crash frequency, it is tested whether treatment effects differ for segments with low and high traffic counts, low and high shares of heavy traffic in total traffic, as well as for the presence or absence of entrance or exit ramps.

For inherently spatial data, spatial auto-correlation in outcome and control variables may lead to strong relations between nearby observations, not only caused by similar covariate values. When applying machine learning methods to such data, this can essentially cause a "leak" in the algorithm, giving rise to spatial over-fitting (Meyer *et al.*, 2019). Research defining conditions where this causes biases in the estimation, as well as potential solutions, is still scarce (Kopczewska, 2021; Meyer *et al.*, 2018). Spatial cross-validation (CV) as well as a forward-features selection procedures have been proposed to alleviate this problem (Meyer *et al.*, 2018). The present study explores in which steps of the application spatial over-fitting may arise and investigates the potential of the proposed solutions. It is shown that cluster robust forests implement spatial CV and thus provide a straightforward tool in the analysis of spatial data. Cluster robust causal forests at the location-year level have been used in an application on spatial data by McCullough *et al.* (2022), in the field of agricultural economics. Furthermore, in a non-causal application of the generalized random forest algorithm, the causal forest is a variant of, Brokamp (2022) employs cluster robust estimation on the location level in a fine particulate matter concentrations prediction model. However, in both applications, the aim of cluster robust estimation is to obtain reliable prediction errors for the trained model's predictions on new location-year combinations or locations, respectively. The issue of spatial over-fitting and its consequences, however, has not been addressed before in the context of causal machine learning. The results of this study suggest that spatial over-fitting is especially of concern when the degree of spatial auto-correlation in the outcome variable is high, and when the outcome on one segment directly affects the outcome on a neighboring segment. In the presented analysis, spatial lags furthermore account for spatial spillovers.

It is found that speed limits have strong negative effects on crash rates, especially for severe and fatal crashes. Less frequented roads are shown to be associated with higher effectiveness of speed limits, likely caused by higher driven speeds and speed variances. While the effects for road segments with different levels of heavy traffic shares are not conclusive, speed limit effects are found to be larger in the presence of entrance and exit ramps, whereby the differences are, however, not statistically significant.

The remainder of this paper is structured as follows: Section 2 presents the data sources and describes the construction of the new data set. Section 3 introduces the causal forest, and discusses the challenges arising when applying machine learning methods to spatially auto-correlated data as well as potential solutions. Results are presented in section 4, and section 5 concludes.

2 Data

To analyze the causal effects of speed limits on crash frequency, a variety of geo-spatial data sources were combined. To estimate an upper bound of the causal effect of a speed limit on crash frequency that is as close to the true causal effect as possible, it is important to obtain extensive information on factors potentially influencing crash frequency - and thereby the probability that a speed limit is in place. The focus, however, lies on factors that might lead to violations of the upper bound assumption, as described above. Due to data availability, motorways of 12 out of 16 German states are included, which account for roughly $19,000$ km of the total motorway network.[3] All data is aggregated over a period of three years ($2017 - 2019$) to reduce volatility in this rare-event data (crashes), while minimizing the number of segments that have to be excluded due to fundamental changes in the motorway, such as a change in the posted speed limit.

The shape of the motorway network as well as information on speed limits, number of lanes, the presence of bridges and tunnels, as well as information on the absence of a shoulder, posted overtaking bans and conditional speed limits was taken from Open Street Map (OSM) (OpenStreetMap contributors, 2021). The state of the OSM network was extracted at the middle of each year. Information about entrance and exit ramps, as

[3]Missing states are: North Rhine-Westphalia, Thuringia, Mecklenburg-West Pomerania, and Berlin.

well as on road curvature was then derived from the OSM network. Following Balck *et al.* (2017), the network was partitioned into segments of 500 m length on a two dimensional map. Balck *et al.* (2017) find this length to be optimal for the analysis of injury accidents as it reduces volatility compared to shorter segments, and preserves a meaningful mapping between accident occurrence and road characteristics that would be lost for longer segments. To these 500 m segments, I merged the number of crashes involving injuries or fatalities, provided by Statistische Ämter des Bundes und der Länder (2020). Vehicle counts are taken from the yearly data of permanent counting stations provided by BASt (2021a) and extrapolated for sectors without a permanent station.[4,5] A network sector ID from BASt (2020) was used to facilitate this extrapolation and to help in the identification of entry and exit ramps.

Road condition measures were provided by BASt (2021b) in the form of a substance and a serviceability index, as well as the road surface material for each 100 m segment of the network, separated by lane. Both indices include measures of longitudinal and cross sectional road roughness. The substance index further measures the general road surface appearance and surface condition. The serviceability index includes measures of road friction and informs on driving safety and comfort (Steinauer *et al.*, 2006). A digital elevation model on a 30 m grid is taken from NASA (2015) and changes in elevation are derived along the road segments. Annual weather data is available from DWD (2020) and sector-wise measures for various socio-demographic characteristics were derived from BBSR (2021), employing also regional shapefiles provided by BKG (2021). 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 with a description and basic descriptive statistics can be found in Table A1 in the Appendix. As shown there in detail, a number of likely less relevant, but highly correlated variables, such as weather and socio-demographic variables, are reduced to four principal components that capture about 80% of their variation.

[4]A sector is typically the motorway segments in both directions, that lie between two network nodes.
[5]For sectors without a permanent counting station (CS), derivations from BASt (2022) are used that transfer the results of a manual counting at all sectors conducted in 2015 to the year 2019. To impute counts for the other two years, the relation observed at a CS between the updated values of 2019 and the counts of the respective year is transferred to sectors of the same motorway in the same direction, without a CS. For sectors lying between two sectors with a CS, an inverse distance weighting is applied to these relations. For roads without a CS on an adjacent sector, the average relation of similar other roads is used.

To aggregate the data over the three years, crash counts are summed and traffic count and weather variables are averaged. Only the wind variable represents an average of the years $1981 - 2000$ and is assumed to show a similar pattern as would be observed in recent measurements. Road condition measures were collected once from 2017 to 2018, and socio-demographic variables in 2017. The elevation profile was measured in the year 2000 and is assumed not to have changed substantially since then. Roadway properties taken from OSM are required to have changed at most marginally over time or else lead to the omission of an observation.[6,7]

The final data set contains $13,000.5$ km of motorway, and thus roughly half of the whole German motorway network and almost 70% of the network in the included states. Table 1 shows the average number of crashes of each severity level by the posted speed limit. A simple t-test is derived, that, for each crash severity level, tests the difference between the average number of crashes on segments with the respective speed limit and without a speed limit. The symbol $>$ ($<$) denotes a positive (negative) difference. Table 1 suggests lower average numbers of fatal crashes on segments with mandatory speed limits. The same holds true for severe crashes, but with no significant difference for the speed limit of 100 km/h. For the light crashes, the picture is more mixed.

	Speed limit	100		120		130		None	Total	Count
Crash severity	Light	0.903	(>***)	0.758	(>***)	0.572	(<***)	0.657	0.678	17,633
	Severe	0.187	(<)	0.183	(<***)	0.184	(<**)	0.209	0.203	5,270
	Fatal	0.012	(<***)	0.016	(<**)	0.015	(<**)	0.021	0.019	507
	Total	1.103	(>***)	0.957	(>***)	0.771	(<***)	0.887	0.900	23,410
	Speed limit share	5.33 %		12.91 %		6.07 %		75.69 %		
	N	1,385		3,357		1,578		19,681	26,001	

Table 1: Average number of accidents by severity and speed limit per 500 m segment in the final analysis data set. $<$ and $>$ denote the sign of a t-test for a simple mean comparison of the respective average crash numbers between the respective speed limit and the absence of a speed limit. The stars denote the significance levels: *: p=0.1, **: p=0.05, ***: p=0.01.

[6]A change in one variable that affects x% of the segments is denoted by x. These change values of all variables are summed up. Based on this, an aggregate value of 20 is allowed for the whole segment. When this value is exceeded, the segment is omitted from the analysis.

[7]Data processing was implemented and fully automated using advanced geo-processing tools from ArcGIS and Python to ensure reproducability and extendability.

3 Methodology

The causal forest was first introduced by Wager and Athey (2018) and generalized and advanced further by Athey *et al.* (2019). It builds on the classical random forest algorithm by Breiman (2001), 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.

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 fit improvement at each individual step, typically evaluated via the mean squared error (MSE):

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

To reduce the correlation between the individual trees and thereby the variance of the ensemble, only $mtry \leq p$ of the available explanatory variables are randomly drawn at each step to be considered in the choice of split (Hastie *et al.*, 2009). This splitting routine is stopped when the fit cannot be improved further or some minimum node size is undercut. For the tree prediction $\hat{\mu}_b(x)$ 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: $\hat{\mu}_B(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{\mu}_b(x)$ (Hastie *et al.*, 2009).

For the generalized random forest framework developed in Athey *et al.* (2019), it is important to see that the random forests can also be represented as a weighted average of the training sample: $\hat{\mu}_B(x) = \sum_i^n \alpha_i(x)Y_i$. 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 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 (Athey *et al.*, 2019). While the performance of classical 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 (Lin and Jeon, 2006). The derived weights are thus specific to the problem at hand.

3.1 Causal Forests

Causal forests define treatment effects under the potential outcomes framework by Rubin (1974) (Wager and Athey, 2018). With a binary treatment indicator $W_i \in \{1, 0\}$, it is assumed that for each observation i, two potential outcomes that could have occurred under the respective treatment regimes exist: $Y_i^{(0)}$ if no treatment was received, and $Y_i^{(1)}$ in case of treatment. The conditional average treatment effect (CATE) 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 framework inherently entails a number of assumptions, when the aim is to estimate equation (2) via local methods, such as causal forests. **Unconfoundedness** assumes that potential outcomes are independent of treatment assignment conditional on the covariates (Rubin and Rosenbaum, 1983):

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

Observations that are sufficiently close to each other with respect to confounding covariates can thus be considered as draws from a randomized experiment (Wager and Athey, 2018).

To ensure the feasibility of local estimators aiming at controlling for confounding factors, **common support** is furthermore 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. This assumption states that as speed limits are typically set as a result of an accumulation of crashes in a segment, the confounders that cannot be controlled for are of such a nature that their correlation with crash frequency and speed limit probability has the same sign. Then, an upper bound for the true causal effect is estimated, which, in case of a negative bound, amounts to a lower bound in absolute values.

Causal forests essentially estimate treatment effects in the forest-based adaptive neighborhood of x, defined by the weights $\alpha_i(x)$ (Athey *et al.*, 2019). As in the standard random forest algorithm, weights are obtained from an ensemble of trees, which are grown via recursive partitioning. As direct counterparts of equation (1) are not available for the case of heterogeneous treatment effect estimation, Wager and Athey (2018) note that minimizing the MSE in the parent node amounts to maximizing the heterogeneity between child nodes. Therefore, a splitting criterion is optimized that maximizes the heterogeneity of treatment effect estimates in the two resulting nodes.[8] Hence, causal forest trees essentially identify neighborhoods in which treatment effects are relatively constant and differ from other nodes, so that the algorithm focuses on treatment effect heterogeneity.

To make the causal forest more robust to potential confounding from observed characteristics, Athey *et al.* (2019) employ a local centering step before training the causal forest. Y_i and W_i are residualized by regressing out their main effects $m(X_i)$ and $e(X_i)$, respectively. For this, two separate regression forests are trained to predict the expected outcome, marginalizing over treatment assignment, $\hat{m}^{(-i)}(X_i)$, and the expected treatment assignment (propensity score) $\hat{e}^{(-i)}(X_i)$, respectively. As denoted by the $(-i)$ notation,

[8]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 Athey *et al.* (2019).

both properties ares estimated "out-of-bag" (oob), i.e. using only the trees grown without observation i. This allows the forest to concentrate on treatment effect heterogeneity without spending a lot of splits to account for observed confounding. After obtaining the forest weights through the described procedure, the treatment effect at a point x is also estimated using a more robust formulation:

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

To ultimately make the causal forest pointwise consistent and asymptotically Gaussian, two major changes are made to the original tree building process in Breiman (2001). First, trees are grown on subsamples of size s, rather than bootstrap samples of the training data. Second, trees have to satisfy a concept called *honesty* (Wager and Athey, 2018). It means that 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 (Wager and Athey, 2018). 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.[9]

To make forests cluster robust, these subsamples cannot contain instances from the same cluster, but are sampled along cluster boundaries (Athey and Wager, 2019). Furthermore, a prediction for observation i is considered out-of-bag only if the whole cluster that observation i belongs to was not used in its estimation. Conditional average treatment effects are estimated with a doubly-robust augmented inverse-propensity weighting (AIPW) correction.

As a general test for effect heterogeneity, the rank-weighted average treatment effect (RATE), developed by Yadlowsky *et al.* (2021) is used. For this test, a scoring function $S(X_i)$ is required that ranks all observations according to a treatment prioritization rule, such as the expected treatment benefit. Based on this, the targeting operator characteristic $TOC(q)$ is derived. At each q, this function gives the CATE of the $q \cdot 100\%$ of observations with the highest treatment prioritization minus the average treatment effect (ATE). The RATE is then a weighted integral over this curve. For the Qini metric employed here,

[9]For additionally required regularity assumptions, see Athey *et al.* (2019).

each point is weighted by its corresponding quantile q. This metric thus gives less weight to large deviations from the ATE for few, highly prioritized observations and has larger statistical power than alternatives if the effect heterogeneity is believed to be prevalent in the entire population (Yadlowsky *et al.*, 2021). For valid statistical inference, prioritization scores cannot be formed on the same observations as the evaluated CATEs. Thus, for this test, the sample is split into two parts of equal size. One part is only used to train a causal forest to use as prioritization rule. The second part is then sorted according to this rule and used to estimate a second causal forest to get the CATEs returned by the TOC.

The performance of non-parametric local methods such as random forests tends to be impaired when fitting strong smooth signals (Friedberg *et al.*, 2021). In the application at hand, traffic volume clearly has such a strong effect on the outcome variable. For German motorways, the effect on crash counts has been found to be close to linear - yet slightly regressive - by previous studies (Balck *et al.*, 2017). Therefore, it seems beneficial to use crash rates, here defined as number of accidents per $30,000$ vehicles per day, rather than crash counts as outcome in the analysis. While such accidents rates have been considered censored in some applications (Anastasopoulos *et al.*, 2012), here the observed zeros are viewed as *real* zeros, and not a representations of some unobservable latent outcome. As discussed in Angrist and Pischke (2009), estimation via empirical representations of equation (2) is thus appropriate to directly identify causal effects. As causal forests do not extrapolate, but essentially represent weighted averages of the training data, they naturally avoid predicting outside of the range of the outcome. In addition, as described in more detail in Athey *et al.* (2019), only the *expected* outcome and treatment assignment need to be Lipschitz continuous in x for the causal forest to be consistent. As long as the probability of observing a crash over a fixed time interval is Lipschitz continuous in x, this assumption is well justified and the causal forest is an appropriate modeling choice.

Furthermore, spatial lags are included for selected variables, as indicated in Table A1. This is especially important in high-speed environments like German motorways as the reason for a crash or even the true point of collision itself may lie well before the point where the accident was documented.

3.2 Spatial prediction

The data used in the analysis is inherently spatial. The implications from applying machine learning methods in such settings are widely unexplored in the causal machine learning literature. However, geographic literature suggest, that spatial auto-correlation in explanatory and outcome variables may lead to spatial over-fitting and thus bias the results (Meyer *et al.*, 2018). A frequently suggested solution is to use target oriented CV where predictions errors are evaluated only on locations that have not been included in training the model. Furthermore, Meyer *et al.* (2018) argue that certain spatially auto-correlated variables can be harmful in spatial predictions, which may be solved by a forward-features selection (FFS) procedure. This procedure starts fitting all possible two-variable models and then chooses the one resulting in the best fit on unseen locations. It then recursively adds the variable that most increases out-of-location fit. When no further improvements are possible, the variable selection is terminated.

As cluster robust causal forests consider an observation oob, only if it does not belong to the same cluster, they implement such leave-location-out (LLO) CV in the estimation of the propensity score and main effect functions, when separate locations are used as clusters. To learn about the risk of spatial over-fitting and to explore the potential of cluster robust causal forests to overcome this issue, a practical experiment is conducted for the prediction of the two nuisance components on the data at hand. 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'.

The experiment is conducted by setting aside a validation set of 20% of the data, containing only complete locations. Regression forests with the following CV and variable selection procedures are trained on the remaining data, whereby *mtry* and the minimum node size are chosen via CV:[10]

(a) Random CV with all variables,

(b) Spatial CV with all variables, and

(c) Spatial CV with FFS.

[10]The grf package by Tibshirani *et al.* (2022) is used for all causal forests in this paper. The RATE test is also implemented in this package.

All forests in this experiment are trained with 4,000 trees, only in the FFS selection procedure, small forests of 150 trees are used for computational feasibility. All tress fitted in the presented paper are honest. To reduce variability of the results, the experiment is repeated 10 times and results are averaged. Instead of the main effect function, the conditional expected outcome $\mathbb{E}[Y|X, W]$ is considered as the prediction target - as only for this quantity, a direct empirical counterpart is observable.

Results for the propensity score are displayed in Table 2. They indicate that for the propensity score estimates, random CV not only leads to overly optimistic oob errors, but also produces slightly larger validation errors, at least for speed limits of 120 and 130 km/h. To explore the reasons for this behavior, Figure 1 to 3 display the oob estimates obtained from forests with 10,000 trees, estimated on the whole sample for setup (a) and (b).[11] Actually restricted segments are displayed with a striped outline. It appears that the algorithm with random CV performs very well at pointing to the actually restricted segments, especially for the two larger speed limits. As indicated by the larger validation errors in Table 2, these models, however, do not seem to capture the relevant features of the underlying DGP as they don't generalize to unseen locations.

		Oob error			Validation error			5 most important variables		
	Setup	a	b	c	a	b	c	a	b	c
Speed limit	100	0.0356	0.0526	0.0510	0.0518	0.0526	0.0525	sunday, HT_share, total_turn, total_turn_lag, holiday	sunday, HT_share, holiday, total_turn, total_turn_lag	HT_share, sunday, total_turn, total_turn_lag, node_area
	120	0.0477	0.1212	0.1189	0.1257	0.1213	0.1243	pop_dens, pc_1, AADT_day, MSV50, HT_share	pop_dens, pc_2, pc_1, HT_share, AADT_day	pc_1, pop_dens, pc_2, AADT_day, pc_0
	130	0.0188	0.0696	0.0687	0.0733	0.0682	0.0691	pc_1, MSV50, AADT_day, AADT_night, AADT	pc_1, pc_0, pop_dens, night_share, holiday	AADT_night, night_share, MSV50, pc_2, AADT_day

Table 2: Cross validation results for propensity score estimates: MSE evaluated on a hold-out set with only unseen location.

As indicated by the striped outlines in Figures 1 and 2, respectively, the 120 and 130 km/h limits are strongly clustered together on contiguous or parallel road segments over several kilometers. 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. As nearby segments are also similar with respect

[11]The Figures for setup (c) look very similar to the graphics for setup (b). They are provided in Figure A1 of the Appendix.

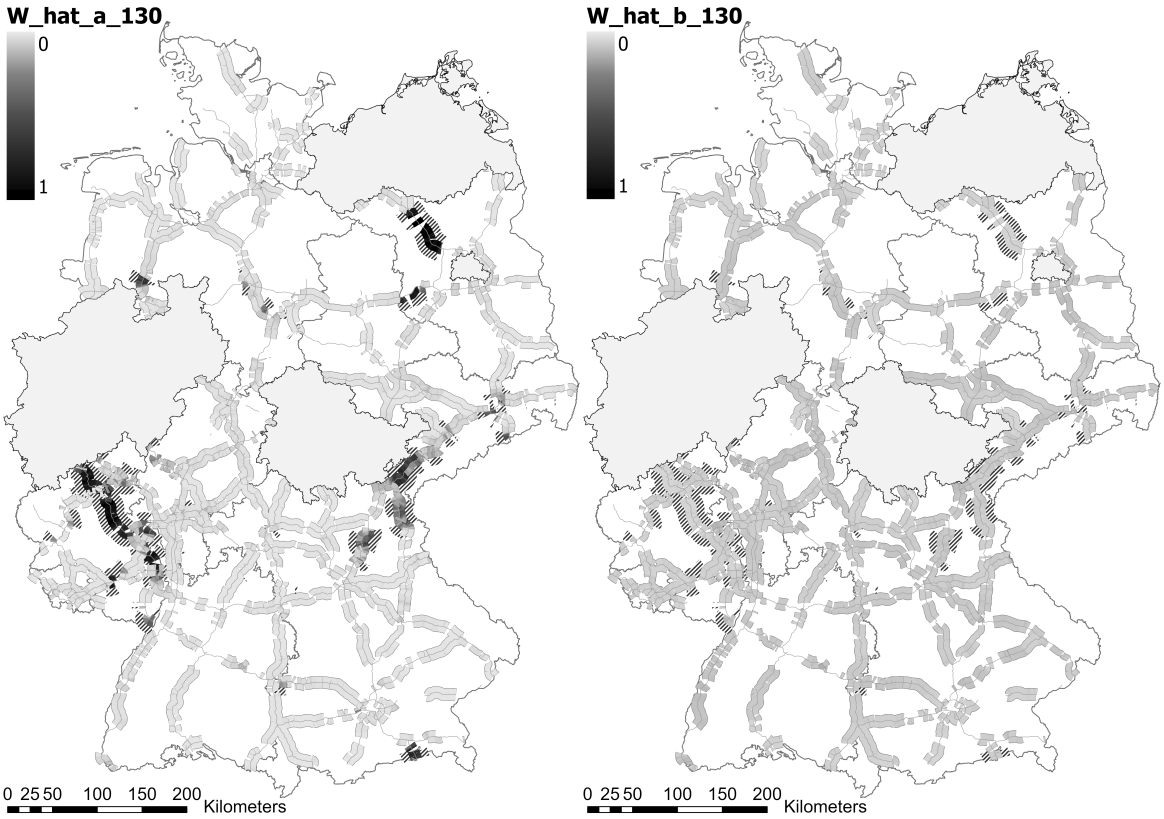

Figure 1: Propensity scores for the speed limit of 130 km/h, estimated according to setup a and b. Actually restricted segments displayed with a striped outline.

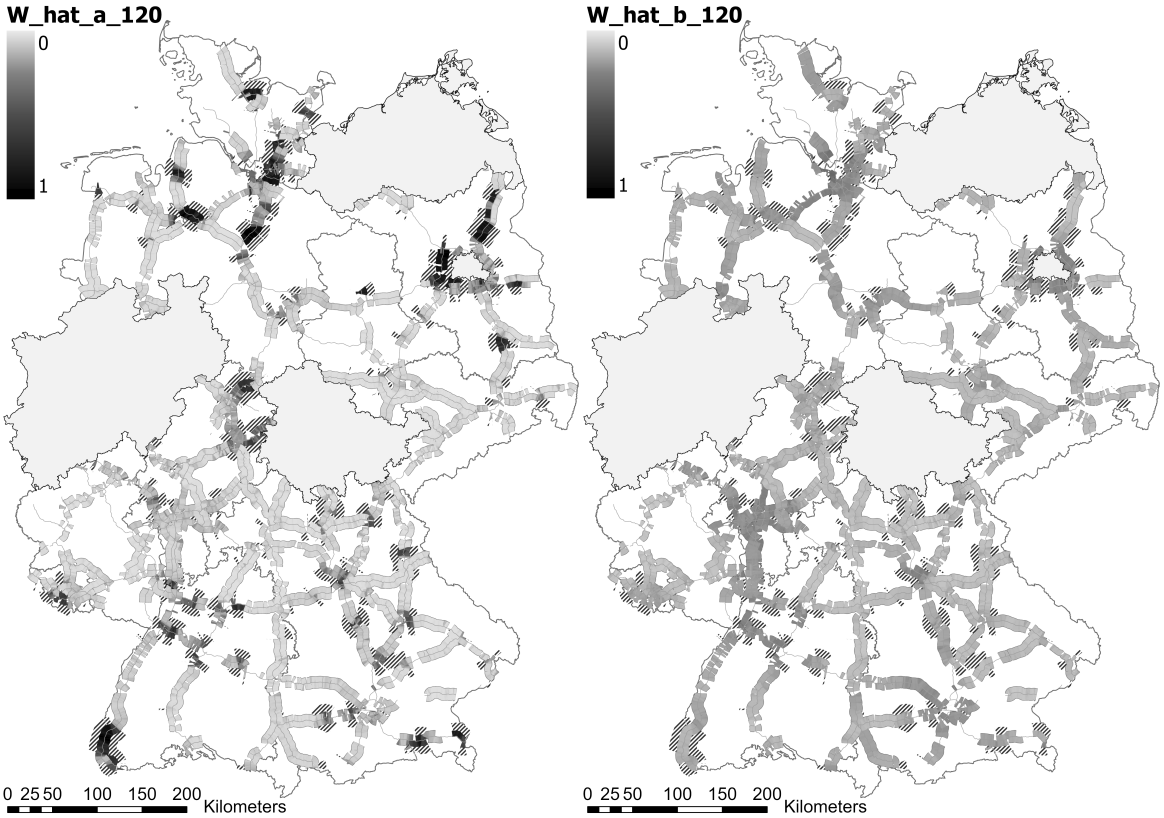

Figure 2: Propensity scores for the speed limit of 120 km/h, estimated according to setup a and b. Actually restricted segments displayed with a striped outline.

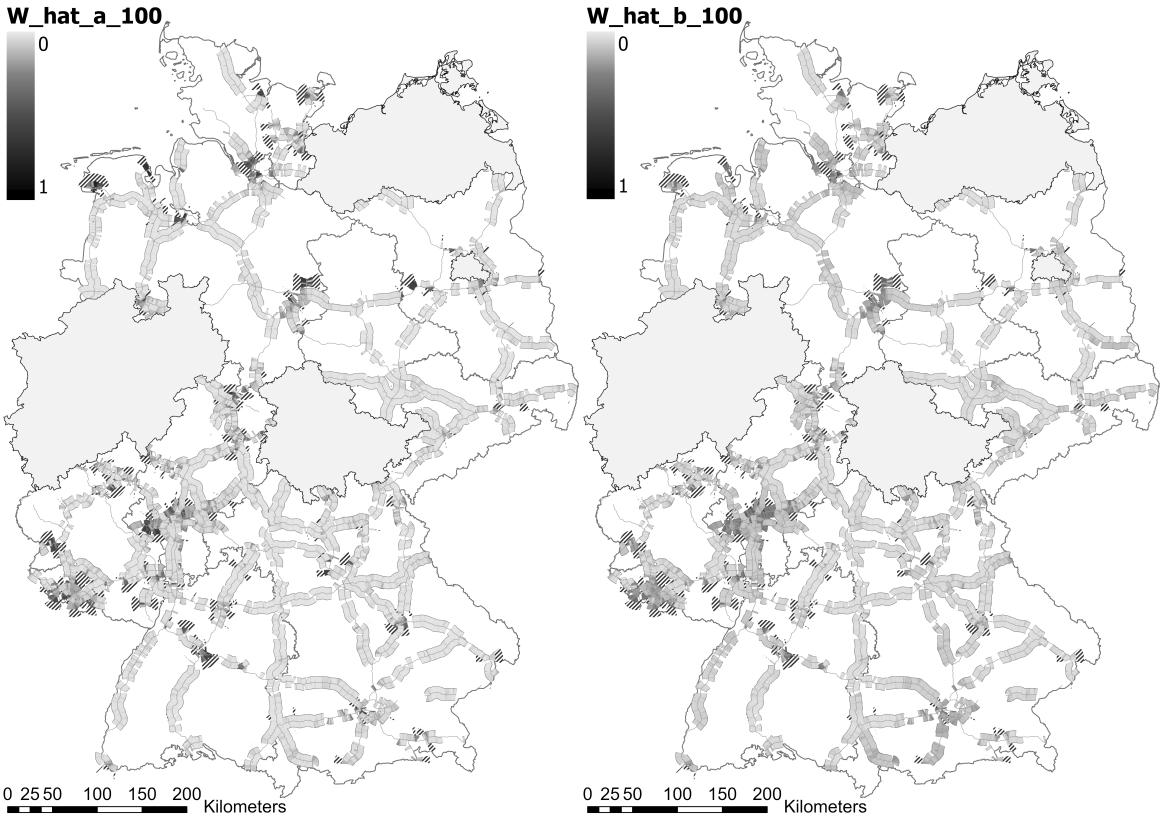

Figure 3: Propensity scores for the speed limit of 100 km/h, estimated according to setup a and b. Actually restricted segments displayed with a striped outline.

to other spatially auto-correlated variables, the forests may easily exploit their variations to identify treated segments. For the propensity score estimate, the target of interest is to get a good estimator of the probability of road segments to be speed constrained, when controlling for all variables that may also affect crash occurrence. Associations that are only driven by spatial proximity, but have no potential to be related to crashes, should not be accounted for.[12]

Based on this, there are at least two possible reasons for the finding, that propensity score estimation of the 100 km/h limit is not improved by LLO CV. First, the degree of spatial clustering is much lower, as can be seen by the striped outlines in Figure 3, as this speed limit seems to be installed primarily on relatively short sections, thus limiting the potential for spatial over-fitting. Second, as it is the most restrictive limit, it may be driven by a much stronger signal that the algorithm can pick up on. This is also supported by

[12]That these results are actually driven by spatial over-fitting, and not some genius property of the algorithm, can be further (i.e. beyond out-of-location fit) supported by two points. First, the propensity score is larger for longer restricted segments, which is likely explained by the algorithm having more segments with similar spatial variables and the same treatment assignment available. Second, the largest propensity scores are estimated for the speed limit of 130 km/h, even though this is the least restrictive limit. This is likely caused by its high degree of spatial clustering.

the observation that in all three setups, selected variables mostly reflect traffic properties as well as road curvature rather than likely more indirectly related variables such as socio-demographic and weather characteristics, represented by the respective principal components. For all three speed limits, the FFS does not lead to further improvements in the validation errors. This eases concerns that strongly spatially auto-correlated variables cause major issues in the estimation.

		Oob error			Validation error			5 most important variables		
	Setup	a	b	c	a	b	c	a	b	c
Crash rate	FR	0.0047	0.0047	0.0047	0.0045	0.0045	0.0045	MSV50, HT_share, night_share, AADT_day, AADT	MSV50, HT_share, night_share, AADT, pc_0	sub_max, up_change, asphalt_lag, maxspeed, surf_het
	SR	0.0535	0.0540	0.0539	0.0500	0.0500	0.0501	pc_1, AADT_day, AADT, MSV50, pc_3	pc_1, AADT_day, AADT, MSV50, pc_3	pc_1, AADT_day, main_Entry, AADT, MSV50
	LR	0.1787	0.1826	0.1824	0.1696	0.1697	0.1699	node_area, ramps, main_Entry, pc_3, pc_1	node_area, pc_1, ramps, pc_3, HT_share	node_area, pc_3, pc_1, main_Entry, ramps

Table 3: Cross validation results for expected outcome estimates: MSE evaluated on a hold-out set with only unseen location. FR: fatal rate, SR: severe rate, LR: Light rate.

Table 3 shows the results for the estimated expected outcome function. It appears that neither LLO CV, nor FFS do anything to improve validation errors. If anything, they even slightly worsen validation errors for the light crash rate. If spatial over-fitting in propensity score estimation is driven by the fact that close segments often have the same speed limit due to their proximity, rather than the underlying DGP, this finding is plausible. Crash clustering on nearby roads are either spurious or driven by observed or unobserved road characteristics. While a speed limit on road segment i increases the probability of observing the same limit on the next segment, the same does not hold true for the expected crash rate, when abstracting from rare cases of follow-up crashes. To some degree, this result eases concerns of omitting crucial variables affecting crash rates, as such unobserved variables could cause a clustering of crashes that would be unaccounted for by the data, which may in turn lead to spatial over-fitting.

The experiment showed that the propensity score estimation with non-cluster robust regression forests can lead to spatial over-fitting, leading to potentially serious biases in the estimation. As cluster robust forests with location clusters essentially implement LLO CV, this is a straightforward way to help the algorithm to concentrate on generalizable patterns in the data rather than relations arising through spatial auto-correlation. As

was discussed, crash rates are driven by some probability process determined by roadway characteristics rather than crash rates on adjacent segments, so that spatial over-fitting is not a major concern here. To estimate the main effect function as marginalizing over treatment assignment, cluster robust estimation of this property is still sensible to keep the algorithm from indirectly controlling for speed limit presence via spatially auto-correlated variables. As accounting for the clustering of treatment assignment is relevant in the standard error estimation (Abadie *et al.*, 2023), causal forests will also be estimated in a cluster robust way.

4 Results

For each combination of mandatory speed limit and crash rate, a cluster robust causal forest with 15,000 trees is fit to the data.[13] The terminal node size and *mtry* are chosen by cluster robust CV. Results are presented in Table 4. To provide a meaningful interpretation, semi-elasticities are reported in addition to average treatment effects, that were derived by dividing the respective doubly-robust treatment effects by the mean expected outcome under no treatment.

Table 4 shows that all speed limits have strong negative effects on total crash rates, that are significant at the 1% level. Fatal crash rates are estimated to be reduced by 40% through imposing a speed limit of 100 km/h, by 37% through 120 km/h limits and by 32% through 130 km/h limits, whereby the latter is not statistically significant and standard errors for all effects on fatal crash rates are relatively large. This is an expected result from the rare event nature of the observed outcome. Severe crash rates are estimated to be reduced by 34% through a limit of 100 km/h, by 31% through a limit of 120 km/h, and by 24% through a limit of 130 km/h, all of which are statistically significant at least at the 5% level. The reductions in light crash rates are statistically significant only for speed limits of 120 km/h and 130 km/h, with estimated reductions of 12% and 26%, respectively.

[13]The nuisance parameters are fit with 10,000 trees. Note also that multiarm forests implemented in the grf package would have the potential to estimate the effects of multiple treatments on multiple outcomes in a single forest. The adaptive neighborhoods would be formed to simultaneously estimate the various treatment effects. As the derived weights would thus not be tailored to the individual comparison, individual forests are used.

Speed limit	100				120				130			
Crash rate	total	fatal	severe	light	total	fatal	severe	light	total	fatal	severe	light
ATE	-0.051***	-0.003***	-0.030***	-0.019	-0.063***	-0.003**	-0.028***	-0.031***	-0.097***	-0.003	-0.021**	-0.070***
se	0.016	0.001	0.005	0.015	0.014	0.001	0.005	0.012	0.028	0.003	0.010	0.017
$\bar{SE}$	**-14**%	**-40**%	**-34**%	**-7**%	**-17**%	**-37**%	**-31**%	**-12**%	**-27**%	**-32**%	**-24**%	**-26**%
$\bar{Y}$	0.373	0.009	0.092	0.272	0.366	0.009	0.090	0.267	0.366	0.009	0.093	0.265
$\hat{ACS}$	1,894	99	908	827	2,030	92	904	939	3,267	59	694	2,368
$\hat{APS}$	-	109	1,266	1,420	-	101	1,260	1,614	-	65	967	4,068
RATE	-0.021***	-0.000	-0.004**	-0.016***	-0.018***	0.000	-0.004**	-0.014***	-0.025***	0.001*	-0.010***	-0.018***
se	0.006	0.001	0.002	0.005	0.004	0.000	0.002	0.003	0.009	0.001	0.004	0.006
	N=21,066	$\bar{W}$=0.066			N=23,038	$\bar{W}$=0.146			N=21,259	$\bar{W}$=0.074		

*: p=0.1, **: p=0.05, ***: p=0.01

Table 4: Causal forest results estimated with cluster robust forests, on the location level. ATE gives the average treatment effect for the whole analysis sample for the respective speed limit and crash rate. $\bar{SE}$ denotes the semi-elasticitiy of the effect, $\bar{Y}$ is the average outcome and $\bar{W}$ the average treatment assignment for the analysis sample. $\hat{ACS}$ gives the estimated annual crash reduction and $\hat{APS}$ is the estimated annual number of people saved from the respective crash, as described in the text. RATE gives the RATE statistic as described in section 3.1 and *se* denotes the respective estimated standard errors.

Due to different levels of unobserved heterogeneity, the effects between different speed limits cannot directly be compared. The finding that the largest reduction in light crashes rates is estimated for the least restrictive speed limits is thus likely caused by the fact that the road segments with this limit are inherently less dangerous than those restricted with a lower speed limit and that the upper bounds estimated for the higher speed limits are likely closer to the true causal effects than for the lower speed limits.

The RATE test employing the Qini metric is conducted as described above. As displayed in Table 4, the test provides strong evidence of generalizable effect heterogeneity, except for fatal crash rates. To further investigate differences in speed limit effectiveness, heterogeneity is next explored with respect to the pre-specified hypotheses. Namely, it will be explored whether CATEs differ between segments with low and high average annual daily traffic (AADT), low and high shares of heavy traffic, and in the presence and absence of entrance or exit ramps. For the first two, the cut-off is chosen such that aggregated AADT in the two groups is roughly equal. This is intended to avoid forming subgroups with very volatile effects due to low crash occurrence. For the same reason, fatality rates are excluded from this analysis. Segments with ramps present on at least 1% of the segment form one subgroup for the heterogeneity analysis of this variable and contain 28% of aggregated traffic.

Results are shown in Figures 4–6, and with more details and formal difference tests in Tables A2–A4 of the Appendix. All three speed limits exhibit larger absolute effects on all severity levels for segments with low traffic volume. This contrasts the believe that speed limits are obsolete in low traffic situations and is likely explained by the higher driven speeds and speed variances in these situations. However, formal t-tests for the difference in these effects are not statistically significant in most cases. Only for the effect of a speed limit of 100 km/h on severe crash rates, the statistical significance of the difference between CATEs for segments with low and high AADT would even survive a Bonferroni correction for testing as many as $(3 \cdot 3 \cdot 3)$ hypotheses.

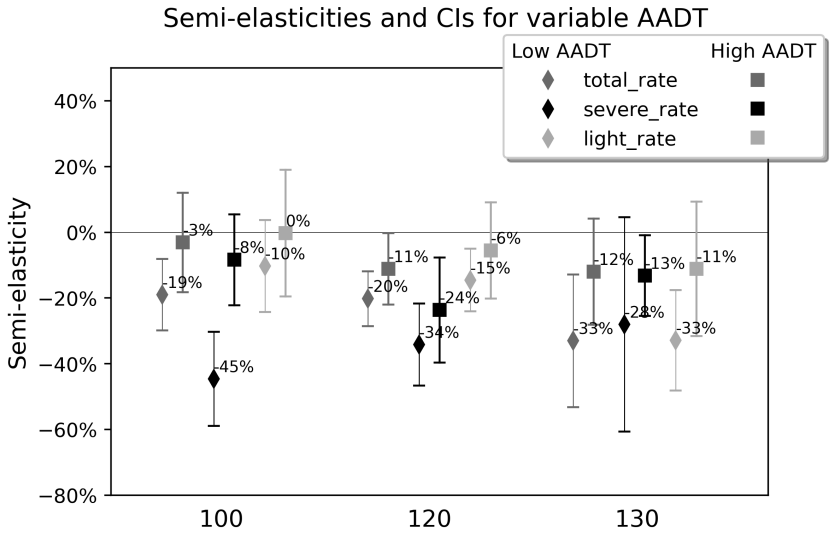

Figure 4: CATEs for segments with low and high traffic volume.

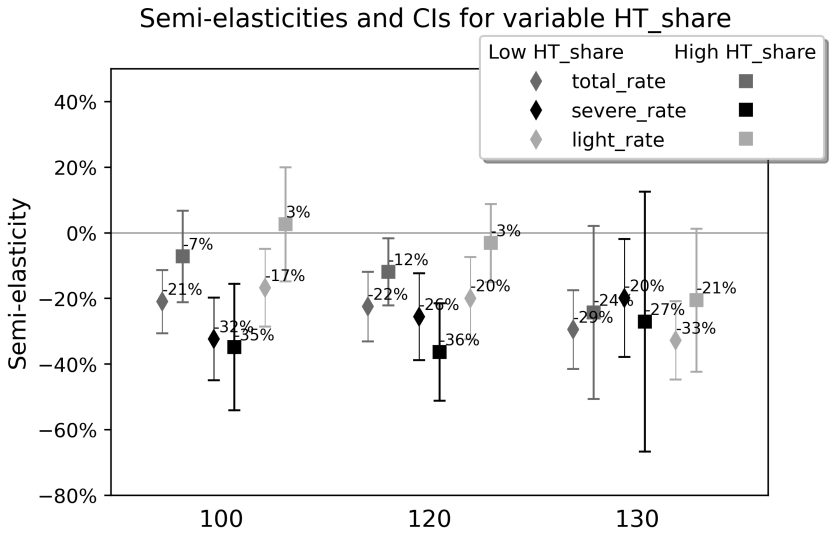

Figure 5: CATEs for segments with low and high shares of heavy traffic.

For heavy traffic shares, results may suggest that speed limits are more successful at reducing light crashes on segments with low heavy traffic shares, and more effective at reducing severe crash rates on segments with high shares. Both directions could be explained as, on the one hand, more heavy traffic may reduce average driven speeds and increase driver caution, which may reduce the speed limit effects. On the other hand, it may increase variance in driven speeds and thereby the probability for traffic conflicts, which in turn may increase speed limit effects. Results could indicate that the first explanation drives the effects on light crashes as here, driver caution could play a bigger role, while for severe crashes, heavy traffic leads to more dangerous situations, in which speed limits may prevent more severe crashes. The differences are however neither large, nor would the occasional statistical significance survive any multiple testing correction, so that these relationships remain purely descriptive.

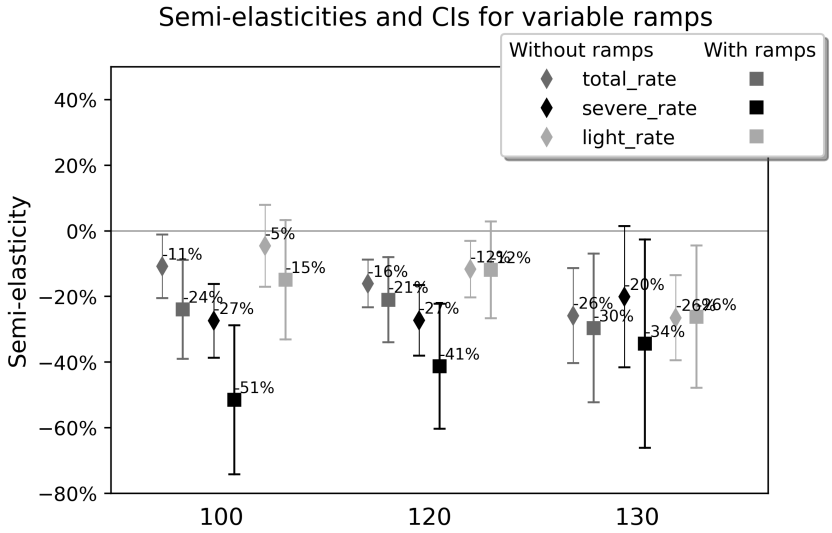

Figure 6: CATEs for segments without and with entry or exit ramps.

Regarding the presence of access and exit lanes, results show larger effects of speed limits in the presence of such ramps. Standard errors for this group are, however, relatively large, likely caused by its smaller size. Differences between the two groups are not statistically significant.

Similar to Athey and Wager (2019), the same analysis is also conducted without cluster robustness to explore the implications. While the overall direction of the results is similar, there are substantial differences, especially for the 130 km/h speed limit, which showed the largest degree of spatial over-fitting in the propensity score estimation without cluster

robustness in section 3.2. For the 100 km/h speed limit, the crash reducing effects are somewhat larger, for the 120 km/h very similar, and for 130 km/h limit, the effects on fatal and severe crashes are substantially closer to 0. This again shows that the cluster robust estimation can have significant impacts on the results when there is strong spatial clustering in the treatment assignment. For the reasons discussed in section 3.2, this non-cluster robust estimation is considered to be biased, especially for the 130 km/h limit. The heterogeneity analysis for the non-cluster robust estimation shows very similar patterns to the cluster robust setting.[14]

In the presented analysis, violations of the overlap assumption are probable, especially for lower speed limits: if specific segments are unambiguously considered dangerous (safe), there may be a lack of segments to compare them to in the data. The algorithm would then be forced to use control (treatment) sections that are as similar as possible, but inherently less dangerous (more dangerous). Ignoring these violations, as was done up to now, should thus be covered by the upper bound assumption, as it would force comparing more dangerous restricted sections with less dangerous unrestricted segments. To explore these consideration, the analysis is repeated using a sample that only uses data of segments with estimated propensity scores between 0.05 and 0.95, and further excludes segments that have been set to the prevalent speed limit but contain a change in speed limits, as well as segments with conditional speed limits or overtaking bans. In fact, the effects estimated with these restricted samples are somewhat larger for fatality rates and very similar for the other crash rates. The patterns detected in the heterogeneity analysis again remain strongly evident.[15] This analysis lends credibility to the upper bound assumption and the robustness of the results. As this approach is however wasteful in its use of data and induces additional uncertainty into the analysis, more credibility is attributed to the analysis of the unrestricted data set.

Next, an estimate is derived for the annual reduction in crashes that would have been expected through the installation of the respective speed limit on all unrestricted road segments. For this, the semi-elasticity of the average treatment effect of the untreated is derived and multiplied with the observed number of crashes on the unrestricted roads

[14]These results are available in Table A5 and Figures A2–A4 of the Appendix.
[15]These results are available in Table A6 and Figures A5–A7 of the Appendix.

in the analysis sample. To also get a doubly robust estimate of the treatment effect on currently unrestricted segments for the unobserved road segments, an approach proposed by Dahabreh *et al.* (2020) is used.[16] For this, a full extent data set was constructed that contains the employed covariates for 25,978.5 km of the motorway, and thus close to the whole network. To construct this data set, the state of the network in 2019 was used in case of changes in the variables that are required to be constant over time. Also, less stringent imputation techniques are used to replace all missing values. As the number of crashes is largely missing for these segments, the aggregate crash counts for these segments are derived by taking the crash counts reported in the official statistics for the years 2017-2019, for road segments without speed limits and construction sites (Destatis, 2018, 2019, 2020). From this, the number of observed crashes on unrestricted samples in the analysis data set is subtracted and this value is multiplied with the derived semi-elasticity for the unobserved, unrestricted samples. The two resulting crash reductions are summed to derive the estimate $\hat{ACS}$ of annual crashes saved for all currently unrestricted segments in the network. To obtain an estimate $\hat{APS}$ of the annual people saved from crashes, these values are multiplied with the average number of harmed people per crash, for each severity level, as reported in the official statistics.[17] The respective values are reported in Table 4. For a speed limit of 130 km/h, a decline of 65 annual fatalities, and 967 annual severely injured is derived for the considered years. For a speed limit of 120 km/h, 101 saved fatalities, and 1,260 saved severely injured are estimated. The reductions for the maximum speed of 100 km/h are very similar.

5 Conclusion

The present analysis studied the effects of local speed limits of 100, 120 and 130 km/h, each compared to the absence of any mandatory speed limit, on total injury crash rates, as well as on fatal, severe, and light crash rates, respectively. A novel data set was constructed

[16]This is done following the documentation of the grf package (Tibshirani *et al.*, 2022).

[17]The number of average people harmed per crash is an exact translation of the $\hat{ACS}$ estimate only for fatal crashes. As more severe crash types may also involve individuals with less severe injuries, and as the more severe crash types mostly have larger reduction effects, $\hat{APS}$ is slightly underestimated for the less severe crash types. This underestimation is larger for the light crashes with much lower treatment effects than the corresponding severe crashes, but is likely negligible for severe crashes due to the small numbers of fatal crashes.

that builds almost entirely on open data sources. It covers the years 2017–2019 for roughly 50% of the German motorway network and presents a rich and extendable source to answer a large number of questions related to road safety on the motorway. In its spatial extent, the data set vastly exceeds all existing studies on the issue. The analysis was conducted under the assumption that any unobserved roadway properties are related to speed limit probability only through crash occurrence, leading to the estimation of an upper bound of the effect. Thereby, characteristics that increase (decrease) the speed limit probability, even though they would have caused drivers to be more cautious (more risky), are assumed to be observable to traffic planners and are as such covered by the data. To avoid distortions due to false functional form approximation, and to easily explore treatment effect heterogeneity, a causal forests approach was used.

It was found that all three speed limits have large negative effects on fatal crash rates, with a reduction of 40% for a speed limit of 100 km/h, 37% for 120 km/h, and 32% for 130 km/h, whereby the last effect was not statistically significant. Especially for this least restrictive limit, this effect was estimated with a considerable degree of uncertainty due to the rare event nature of fatal crashes, a problem that would also be present under a fully controlled experiment. Severe crash rates are estimated to be reduced by 34% due to the most restrictive, and by 31% and 24% for 120 km/h and 130 km/h, respectively. Light crash rates were found to be less affected by the installation of speed limits, except for the 130 km/h limit, where a reduction of about 26% was found.

The effects of lower speed limits can be assumed to be most affected by unobserved heterogeneity, so that the bounds estimated for less restrictive limits are likely closer to the true effects. Also, the high degree of spatial clustering of 130 km/h limits may still harm the estimation of propensity scores for this limit. As the algorithm only has a limited number of different locations with this restriction, little variation in the spatially auto-correlated variables is present in this data. It may thus be difficult for the algorithm to pick up on the relevant features of the data generating process. As a result, most credibility is attributed to the effect estimates of the 120 km/h speed limit, that also constitutes the largest treatment group.

Evidence further suggests that speed limits would be more effective on less congested roads, as well as in the presence of access and exit lanes, while the heterogeneity regarding heavy traffic shares was inconclusive. These results provide additional support for the hypothesis that higher speed variance leads to higher crash rates. The first finding also casts doubt on the sensibility to install flexible speed limit that remove restrictions in low traffic situations. However, as most of these differences were not statistically significant, they should be treated with caution.

The issue of spatial over-fitting was illustrated and spatial CV and FFS were discussed as potential solutions. This issue has not been previously discussed in the causal machine learning literature. It was shown that spatial over-fitting may lead to severe biases in the estimated nuisance functions. However, LLO CV, as implemented in cluster robust forests, helps the algorithm concentrate on generalizable features of the data instead of over-fitting to spatially auto-correlated variables. Furthermore, it was found that the degree of spatial auto-correlation in the outcome variables matters for the potential for spatial over-fitting. Results suggested that concerns on this issue vanish when the outcome at one location itself does not influence the probability distribution at a neighboring location, but when spatial auto-correlation in the outcome is purely driven by spatial auto-correlation in the covariates.

The results are well in line with previous findings and in fact fall into the upper region of the effect range estimated by Bauernschuster and Traxler (2021). This strengthens trust that the analysis succeeded in controlling for a large range of relevant confounders, so that the estimated bounds can be assumed to be close to the true causal effects. Whether these results are fully transferable to an environment with general mandatory speed limits in the entire network is difficult to predict. For example, speed limits may lose their ability to signal more dangerous road sections, which could increase crash frequency on currently restricted segments. On the other hand, drivers may get used to the lower driving speed and might increase their compliance, which could in turn lead to an even stronger reduction in crash frequency (Bauernschuster and Rekers, 2022).

References

Aarts, L. and van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis and Prevention*, **38**, 215–224.

Abadie, A., Athey, S., Imbens, G. W. and Wooldridge, J. (2023). When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, **138**.

Anastasopoulos, P. C., Mannering, F. L., Shankar, V. N. and Haddock, J. E. (2012). A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Analysis and Prevention*, **45**, 628–633.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ashenfelter, O. and Greenstone, M. (2004). Using mandated speed limits to measure the value of a statistical life. *Journal of Political Economy*, **112** (1), 226–267.

Athey, S., Tibshirani, J. and Wager, S. (2019). Generalized random forests. *Annals of Statistics*, **47** (2), 1148–1178.

— and Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational studies*, **5**, 36–51.

Balck, H., Schüller, H., Balmberger, M. and Rossol, C. (2017). *Verfahren zur Zusammenführung von Informationen unterschiedlicher Netzanalysesysteme*. Tech. Rep. Verkehrstechnik - Heft V 298, Bundesanstalt für Straßenwesen.

Bärwolff, M., Berger, R., Lippold, C., Schmotz, M., Martin, J. and Schröter, B. (2019). *Maßnahmen zur Vermeidung von Abkommensunfällen auf BAB*. Tech. rep., Bundesanstalt für Straßenwesen.

BASt (1977). *Auswirkung einer Richtgeschwindigkeit im Vergleich zu einer Höchstgeschwindigkeit von 130 km/h auf Autobahnen*. Tech. rep., Bundesanstalt für Straßenwesen, Projektgruppe "Autobahngeschwindigkeiten", Köln.

— (1984). *Abschätzung der Auswirkungen einer Senkung der Höchstgeschwindigkeit auf das Unfallgeschehen im Straßenverkehr*. Tech. rep., Bundesanstalt für Straßenwesen, Bergisch Gladbach.

BASt (2020). Bundesfernstraßennetz. https://www.bast.de/BASt_2017/DE/Verkehrstechnik/Fachthemen/Daten/Daten-BISStra.html?nn=1817946.

BASt (2021a). Automatische Zählstellen 2017-2019. https://www.bast.de/BASt_2017/DE/Verkehrstechnik/Fachthemen/v2-verkehrszaehlung/zaehl_node.html;jsessionid=9BE1F5EC97952EEFAB21C06D3F0BFD7B.live11314.

BASt (2021b). Zustandserfassung und -bewertung (ZEB), BAB 2017 & 2018.

BASt (2022). Fortschreibung/Hochrechnung der Ergebnisse der SVZ 2015 und der termporären Messung 2016 bis 2019 auf das Jahr 2019. https://www.bast.de/DE/Statistik/Verkehrsdaten/Manuelle-Zaehlung.html.

Bauernschuster, S. and Rekers, R. (2022). Speed limit enforcement and road safety. *Journal of Public Economics*, **210**, 1–48.

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

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

Bennhold, K. (2019). Impose a speed limit on the autobahn? Not so fast, many Germans say. *The New York Times*, **February 2019**.

BKG (2021). 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, [License: dl-de/by-2-0 - http://www.govdata.de/dl-de/by-2-0].

Breiman, L. (2001). Random forests. *Machine Learning*, **45** (1), 5–32.

Brokamp, C. (2022). A high resolution spatiotemporal fine particulate matter exposure assessment model for the contiguous United States. *Environmental Advances*, **7**, 100155.

Dahabreh, I. J., Robertson, S. E., Steingrimsson, J. A., Stuart, E. A. and Hernán, M. A. (2020). Extending inferences from a randomized trial to a new target population. *Statistics in Medicine*, **29** (14), 1999–2014.

Destatis (2018). Verkehr - Verkehrsunfälle 2017. https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00040387/2080700177004_Erg16082018.pdf.

— (2019). Verkehr - Verkehrsunfälle 2018. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/verkehrsunfaelle-jahr-2080700187004.pdf?___blob=publicationFile.

— (2020). Verkehr - Verkehrsunfälle 2019. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/verkehrsunfaelle-jahr-2080700197004.pdf;jsessionid=FFB4E3351A232191CB07015FE586822C.live742?___blob=publicationFile.

DWD (2020). Climate Data Center, annual grids. https://opendata.dwd.de/climate_environment/CDC/grids_germany/annual/.

Elvik, R., Vadeby, A., Hels, T. and van Schagen, I. (2019). Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. *Accident Analysis and Prevention*, **123**, 114–122.

Friedberg, R., Tibshirani, J., Athey, S. and Wager, S. (2021). Local linear forests. *Journal of Computational and Graphical Statistics*, **30** (2), 503–517.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, New York, NY: Springer, 2nd edn.

Holthaus, T., Goebels, C. and Leerkamp, B. (2020). *Evaluation of driven speed on German motorways without speed limits*. Tech. rep., Bergische Universität Wuppertal.

Knight, B. (2019). German government shows cracks over autobahn speed limit. *Deutsche Welle*, **December 2019**.

Kopczewska, K. (2021). Spatial machine learning: new opportunities for regional science. *The Annals of Regional Science*.

Lin, Y. and Jeon, Y. (2006). Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association*, **101** (474), 578–590.

Lord, D. and Mannering, F. (2010). 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.

Mannering, F. L., Shankar, V. and Bhat, C. R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*, **11**, 1–16.

McCullough, E. B., Quinn, J. D. and Simons, A. M. (2022). Profitability of climate-smart soil fertility investment varies widely across Sub-Saharan Africa. *Nature Food*, **3**, 275–285.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. and Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, **101**, 1–9.

—, —, Wöllauer, S. and Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications - moving from data reproduction to spatial prediction. *Ecological Modelling*, **411**, 108815.

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

OpenStreetMap contributors (2021). OpenStreetMap data retrieved via https://overpass-turbo.eu/ and https://download.geofabrik.de/europe/germany.html. [License: CC BY-SA 2.0 - https://www.openstreetmap.org/copyright].

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

— and Rosenbaum, P. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70** (1), 41–55.

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

Steinauer, B., Ueckermann, A. and Maerschalk, G. (2006). *Analyse vorliegender messtechnischer Zustandsdaten und Erweiterung der Bewertungsparameter für Innerortsstraßen*. Tech. Rep. Straßenbau - Heft S 46, Bundesanstalt für Straßenwesen.

Tibshirani, J., Ahtey, S., Sverdrup, E. and Wager, S. (2022). *grf - Generalized Random Forests*. https://cran.r-project.org/package=grf, r package version 2.2.0 edn.

van Benthem, A. (2015). What is the optimal speed limit on freeways? *Journal of Public Economics*, **124**, 44–62.

VCD (2019). *Tempolimit auf Autobahnen für Verkehrssicherheit und Klimaschutz*. Tech. rep., Verkehrsclub Deutschland e.V.

VDA (2021). *Fakten gegen ein generelles Tempolimit*. Tech. rep., Verband der Automobilindustrie e.V., version 4.0.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, **113** (523), 1228–1242.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E. and Wager, S. (2021). Evaluating treatment prioritization rules via rank-weighted average treatment effects. https://arxiv.org/abs/2111.07966.

A Appendix

Variable name	Description	Mean	Std.	AC
Crash data				
total	Total number of crashes between 2,017 and 2,019 on segment.	0.9	1.32	0.31
fatal	Number of fatal crashes.	0.02	0.14	-0.0
severely_injured	Number of crashes resulting in severely injured.	0.2	0.48	0.09
lightly_injured	Number of crashes resulting in lightly injured.	0.68	1.13	0.31
Open Street Map data				
maxspeed_100	Indicator of speed limit of 100 being the predominant speed limit.	0.05	0.22	0.82
maxspeed_120	Indicator of speed limit of 120 being the predominant speed limit.	0.13	0.34	0.89
maxspeed_130	Indicator of speed limit of 130 being the predominant speed limit.	0.06	0.24	0.93
tunnel*	Share of segment that is a tunnel.	0.0	0.03	0.6
bridge*	Share of segment that is a bridge.	0.04	0.11	0.29
no_shoulder	Share of segment that has no shoulder.	0.01	0.09	0.63
main_Entry*	Share of road with a main entrance ramp.	0.03	0.1	0.03
sec_Entry*	Share of road with a secondary entrance ramp (e.g. from service area).	0.01	0.05	0.06
main_Exit*	Share of road with a main exit ramp.	0.03	0.11	0.02
sec_Exit*	Share of road with a secondary exit ramp (e.g. to service area).	0.01	0.06	0.06
ramps*	main_Entry + sec_Entry + main_Exit + sec_Exit	0.09	0.17	0.27
node_area*	Share of road at motorway node (from 500 m before to 300 m behind node).	0.29	0.42	0.68
straight	100·(linear distance between start and end point/segment length in meter).	1.0	0.0	0.4
right_turn*	Sum of rigth-angle changes in ° between 100 m pieces of segment.	4.76	7.76	0.5
left_turn*	Measured as right turn but looking at other direction.	4.75	7.73	0.49
total_turn*	righ_turn + left_turn	9.51	9.18	0.64
n_lanes	Average number of lanes on the segment.	2.28	0.47	0.93
(ms_cond)	Share of road with any conditional speed limits.	0.04	0.18	0.89
(ms_change)	Indicator of segment having a change in speed limit.	0.04	0.19	0.07
(overtaking_ht)	Share of segment with posted overtaking ban for heavy traffic or trailers.	0.04	0.18	0.89
Counting station data				
AADT	Average annual daily traffic count, all motor vehicles and weekdays.	24,542	11,417	0.97
AADT_HT	Average annual daily traffic of heavy traffic.	4,020	2,282	0.98
HT_share	Share of heavy traffic in total traffic.	0.16	0.06	0.98
MSV50	Traffic volume on 50th most congested hour of year, all vehicles.	2,572	1,037	0.97
HT_share_MSV	Heavy traffic share in design traffic volume.	7.63	4.96	0.98
AADT_day	Average hourly traffic volume by day (6am-10pm) *(cs)*.	2,723	1,252	0.97
AADT_night	Average hourly traffic volume by night (10pm- 6am) *(cs)*.	682	367	0.98
night_share	AADT_night/(AADT_night + AADT_day) *(cs)*	0.19	0.04	0.99
sunday	Sunday factor: AADT Sundays/AADT Tuesdays to Thursdays.	0.88	0.22	0.98
holiday	Holiday factor: AADT in holidays/AADT no holiday (both Mo.-Sat.) *(cs)*.	1.03	0.12	0.96
Road surface conditions data				
asphalt*	Share of road that is asphalt.	0.74	0.43	0.95
sub_mean*	Mean of substance index.	2.36	0.88	0.8
sub_var	Variance of substance index.	0.53	0.66	0.6
sub_max	Maximum (worst) substance index.	3.4	1.28	0.71
perf_mean*	Mean of performance/serviceability index.	2.15	0.49	0.8
perf_var	Variance of performance/serviceability index.	0.23	0.3	0.64
perf_max	Maximum (worst) performance index.	2.87	0.83	0.69
Elevation data				
down_change*	Sum of downward changes in m between elevation at 21 points along segment.	6.43	5.88	0.47
up_change*	Sum up upward changes between these points.	6.5	5.97	0.48
max_slope*	Derive slope from elevation map and take maximum at points along segment.	4.46	3.11	0.65
mean_slope* $^+$	Average slope over all segment points.	2.35	1.83	0.67
elevation$^+$	Average elevation on all segment points.	246	187	1.0
Weather data				
air_temp$^+$	Average annual air temperature at the location of the segment in $°C$.	103	7	0.99
frost_days$^+$	Average annual number of days with minimum air temperature $< 0°C$.	72	17	0.99
ice_days$^+$	Average annual number of days with maximum air temperature $< 0°C$.	11	6	0.98
snowcov_days$^+$	Average annual number of days with snowcover $> 1cm$ in the morning.	19	14	0.99
precip10mm$^+$	Average annual number of days with $>= 10mm$ precipitation.	18	6	0.99
precip20mm$^+$	Average annual number of days with $>= 20mm$ precipitation.	4.46	2.56	0.99
precip30mm$^+$	Average annual number of days with $>= 30mm$ precipitation.	1.38	0.99	0.99
precipitation$^+$	Average annual sum of precipitation in mm.	684	159	0.99
summer_days$^+$	Average annual number of days with maximum air temperature $>= 25°C$.	60	13	0.99
sunshine_dur$^+$	Average annual sunshine duration in hours.	1,835	93	1.0
wind$^+$	Average wind speed $10m$ above ground over the years 1,981-2,000.	32	5	0.91
Regional and socio-demographic data				
pop_dens	Regional population density.	293	225	0.99
emp_quo$^+$	Regional employment rate.	84	1	1.0
pop_18_25$^+$	Regional share of population older or equal to 18 and younger than 25 years.	7.11	1.32	1.0
pop_ol_65$^+$	Regional share of population aged 65 and older.	22	2	1.0
fem_share$^+$	Regional share of women in total population.	50	0	0.99
hh_inc$^+$	Regional average monthly disposable household income in Euro.	1,891	155	1.0
pcar_dens$^+$	Regional density of passenger cars.	601	32	1.0
gdp_p_cap$^+$	Regional GDP per capita.	33	5	0.99
rurality$^+$	Regional rurality index.	39	20	0.99
location	ID for consecutive road segments of same motorway number within a state.			

Table A1: Variable description and descriptives. Auto-correlation (AC) derived after sorting data according to generated ID, numbering adjacent segments successively. *: spatial lags are included. $^+$: reduced to four principal components. In parentheses: only used for selection of a restricted sample. *cs*: cross section over both directions.

A.1 Spatial prediction

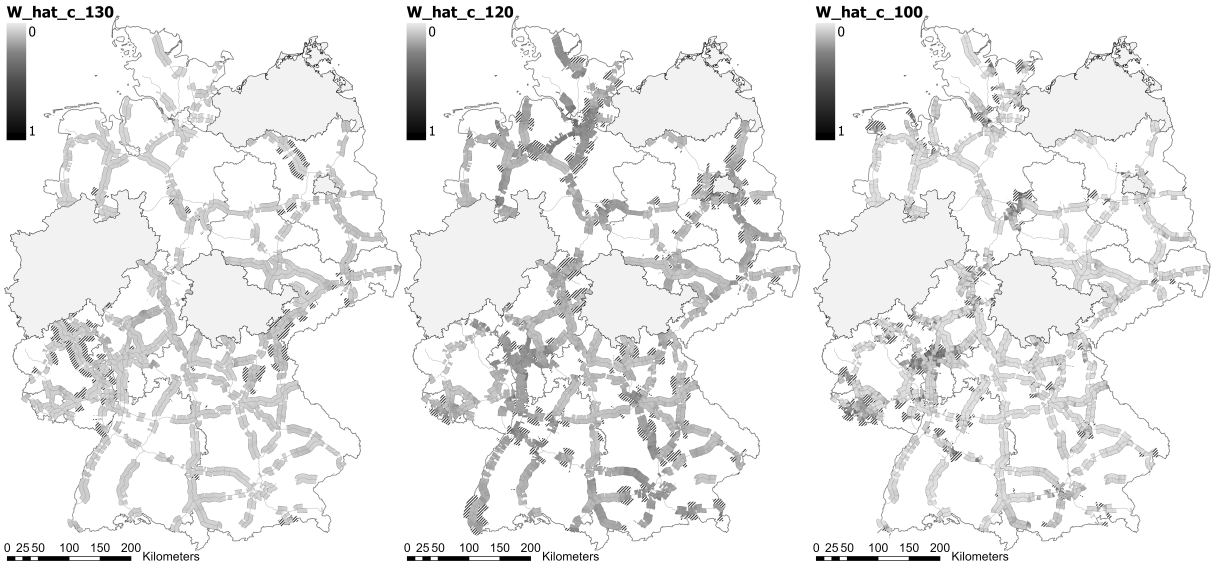

Figure A1: Propensity scores for speed limit of 130, 120, and 100 estimated according to setup c. Actually restricted segments displayed with a striped outline.

A.2 Further details on CATEs

	Crash severity	total		severe		light	
	Subgroup	Low	High	Low	High	Low	High
Variable							
AADT	CATE	-0.069***	-0.011	-0.040***	-0.007	-0.027	-0.001
	se	0.020	0.028	0.007	0.006	0.019	0.026
	$\bar{SE}$	**-19**%	**- 3**%	**-45**%	**- 8**%	**-10**%	**0**%
	$\bar{Y}$	0.366	0.389	0.097	0.081	0.259	0.300
	$\bar{W}$	0.055	0.091	0.055	0.091	0.055	0.091
	Diff.	-0.058* (0.034)		-0.032*** (0.009)		-0.027 (0.032)	
HT_share	CATE	-0.076***	-0.026	-0.029***	-0.031***	-0.044***	0.007
	se	0.018	0.026	0.006	0.009	0.016	0.023
	$\bar{SE}$	**-21**%	**- 7**%	**-32**%	**-35**%	**-17**%	**3**%
	$\bar{Y}$	0.396	0.349	0.085	0.100	0.304	0.238
	$\bar{W}$	0.093	0.037	0.093	0.037	0.093	0.037
	Diff.	-0.050 (0.031)		0.002 (0.01)		-0.051* (0.028)	
ramps	CATE	-0.039**	-0.087***	-0.025***	-0.046***	-0.012	-0.039
	se	0.018	0.028	0.005	0.010	0.017	0.025
	$\bar{SE}$	**-11**%	**-24**%	**-27**%	**-51**%	**- 5**%	**-15**%
	$\bar{Y}$	0.345	0.453	0.087	0.108	0.250	0.335
	$\bar{W}$	0.049	0.114	0.049	0.114	0.049	0.114
	Diff.	0.048 (0.033)		0.022* (0.012)		0.027 (0.03)	

*: p=0.1, **: p=0.05, ***: p=0.01

Table A2: CATEs for speed limit = 100 km/h.

	Crash severity	total		severe		light	
	Subgroup	Low	High	Low	High	Low	High
Variable							
AADT	CATE	-0.074***	-0.041**	-0.031***	-0.021***	-0.039***	-0.015
	se	0.016	0.020	0.006	0.007	0.013	0.020
	$\bar{SE}$	**-20**%	**-11**%	**-34**%	**-24**%	**-15**%	**- 6**%
	$\bar{Y}$	0.360	0.380	0.095	0.079	0.255	0.294
	$\bar{W}$	0.118	0.206	0.118	0.206	0.118	0.206
	Diff.	-0.033 (0.026)		-0.010 (0.009)		-0.024 (0.024)	
HT_share	CATE	-0.082***	-0.044**	-0.023***	-0.033***	-0.053***	-0.008
	se	0.020	0.019	0.006	0.007	0.017	0.016
	$\bar{SE}$	**-22**%	**-12**%	**-26**%	**-36**%	**-20**%	**- 3**%
	$\bar{Y}$	0.385	0.346	0.083	0.097	0.295	0.239
	$\bar{W}$	0.175	0.115	0.175	0.115	0.175	0.115
	Diff.	-0.039 (0.028)		0.010 (0.009)		-0.045* (0.024)	
ramps	CATE	-0.059***	-0.077***	-0.025***	-0.037***	-0.031***	-0.032
	se	0.014	0.024	0.005	0.009	0.012	0.020
	$\bar{SE}$	**-16**%	**-21**%	**-27**%	**-41**%	**-12**%	**-12**%
	$\bar{Y}$	0.339	0.446	0.085	0.106	0.246	0.330
	$\bar{W}$	0.135	0.178	0.135	0.178	0.135	0.178
	Diff.	0.018 (0.028)		0.013 (0.01)		0.001 (0.023)	

*: p=0.1, **: p=0.05, ***: p=0.01

Table A3: CATEs for speed limit = 120 km/h.

	Crash severity	total		severe		light	
	Subgroup	Low	High	Low	High	Low	High
Variable							
AADT	CATE	-0.119***	-0.043	-0.025*	-0.012**	-0.087***	-0.029
	se	0.037	0.030	0.015	0.006	0.021	0.028
	$\bar{SE}$	**-33**%	**-12**%	**-28**%	**-13**%	**-33**%	**-11**%
	$\bar{Y}$	0.359	0.383	0.097	0.082	0.253	0.294
	$\bar{W}$	0.073	0.077	0.073	0.077	0.073	0.077
	Diff.	-0.076 (0.048)		-0.013 (0.016)		-0.058* (0.034)	
HT_share	CATE	-0.106***	-0.088*	-0.018**	-0.024	-0.087***	-0.054*
	se	0.022	0.049	0.008	0.018	0.016	0.029
	$\bar{SE}$	**-29**%	**-24**%	**-20**%	**-27**%	**-33**%	**-21**%
	$\bar{Y}$	0.394	0.341	0.087	0.098	0.299	0.233
	$\bar{W}$	0.044	0.101	0.044	0.101	0.044	0.101
	Diff.	-0.019 (0.053)		0.006 (0.02)		-0.032 (0.034)	
ramps	CATE	-0.093***	-0.107**	-0.018*	-0.030**	-0.070***	-0.069**
	se	0.027	0.042	0.010	0.014	0.017	0.029
	$\bar{SE}$	**-26**%	**-30**%	**-20**%	**-34**%	**-26**%	**-26**%
	$\bar{Y}$	0.340	0.447	0.087	0.110	0.245	0.326
	$\bar{W}$	0.069	0.091	0.069	0.091	0.069	0.091
	Diff.	0.014 (0.049)		0.013 (0.017)		-0.001 (0.034)	

*: p=0.1, **: p=0.05, ***: p=0.01

Table A4: CATEs for speed limit = 130 km/h.

A.3 Results without cluster robustness

Speed limit	100				120				130			
Crash rate	total	fatal	severe	light	total	fatal	severe	light	total	fatal	severe	light
ATE	-0.073***	-0.005***	-0.033***	-0.035***	-0.083***	-0.003**	-0.035***	-0.044***	-0.077***	-0.000	-0.011	-0.064***
se	0.009	0.001	0.003	0.008	0.007	0.001	0.003	0.006	0.011	0.003	0.008	0.008
$\bar{SE}$	**-20**%	**-62**%	**-36**%	**-13**%	**-22**%	**-29**%	**-38**%	**-16**%	**-21**%	**-3**%	**-12**%	**-24**%
$\bar{Y}$	0.373	0.009	0.092	0.272	0.366	0.009	0.090	0.267	0.366	0.009	0.093	0.265
$\hat{ACS}$	2,330	148	937	1,152	2,473	12	997	1,367	1,977	-156	-234	2,372
$\hat{APS}$	-	163	1,307	1,978	-	13	1,390	2,348	-	-172	-326	4,073
RATE	-0.018***	-0.000	-0.003***	-0.013***	-0.013***	0.000	-0.001	-0.008**	-0.015***	0.000	-0.002*	-0.011***
se	0.004	0.000	0.001	0.003	0.003	0.000	0.001	0.004	0.004	0.000	0.001	0.003
	N=21,066	$\bar{W}$=0.066			N=23,038	$\bar{W}$=0.146			N=21,259	$\bar{W}$=0.074		

*: p=0.1, **: p=0.05, ***: p=0.01

Table A5: Causal forest results estimated without cluster robustness.

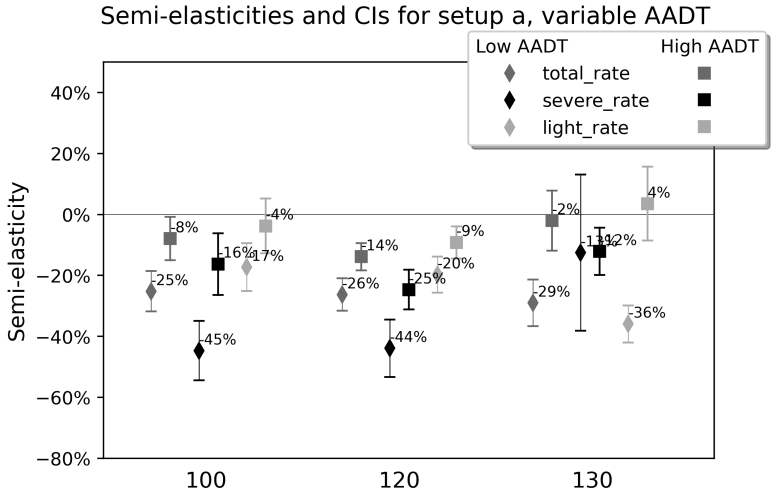

Figure A2: CATEs for segments with low and high traffic volume, without cluster robustness.

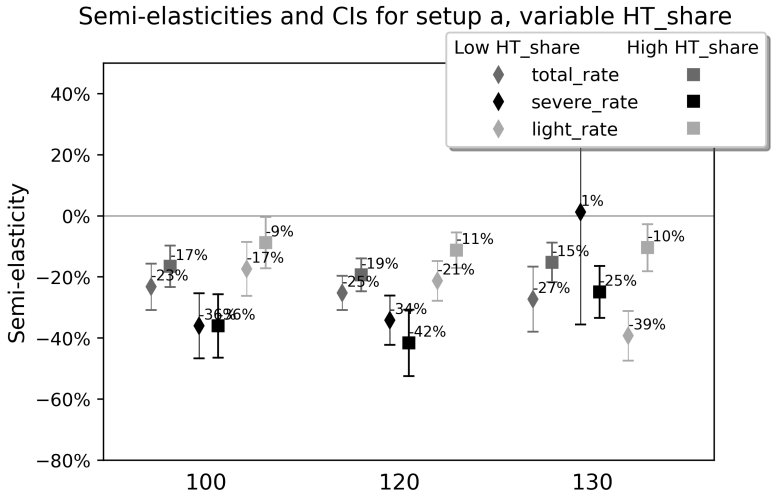

Figure A3: CATEs for segments with low and high shares of heavy traffic, without cluster robustness.

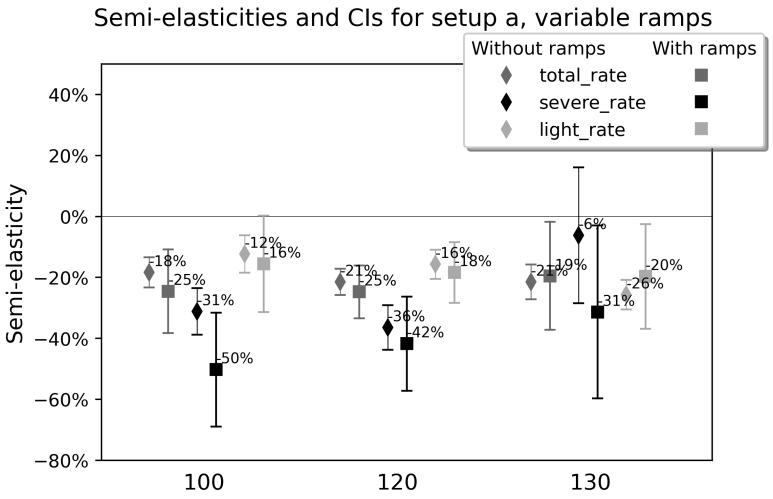

Figure A4: CATEs for segments without and with entry or exit ramps, without cluster robustness.

A.4 Restricted sample results

Speed limit	100				120				130			
Crash rate	total	fatal	severe	light	total	fatal	severe	light	total	fatal	severe	light
ATE	-0.065***	-0.005***	-0.035***	-0.023	-0.068***	-0.003**	-0.026***	-0.038***	-0.101***	-0.003	-0.026***	-0.070***
se	0.019	0.001	0.007	0.016	0.015	0.001	0.005	0.012	0.024	0.003	0.010	0.015
$\bar{SE}$	**-17**%	**-57**%	**-39**%	**- 8**%	**-19**%	**-39**%	**-29**%	**-14**%	**-28**%	**-40**%	**-29**%	**-27**%
$\bar{Y}$	0.405	0.008	0.096	0.301	0.363	0.009	0.091	0.263	0.361	0.009	0.093	0.259
$\hat{ACS}$	2,391	148	991	1,093	2,180	108	854	1,137	3,598	104	900	2,539
$\hat{APS}$	-	164	1,381	1,877	-	120	1,191	1,953	-	115	1,254	4,361
RATE	-0.033***	-0.001*	-0.006*	-0.027***	-0.022***	0.001	-0.007***	-0.011***	-0.022	-0.001	-0.002	-0.020***
se	0.008	0.000	0.003	0.007	0.005	0.001	0.002	0.004	0.014	0.001	0.005	0.007
	N=9,363	$\bar{W}$=0.092			N=19,310	$\bar{W}$=0.13			N=17,815	$\bar{W}$=0.067		

*: p=0.1, **: p=0.05, ***: p=0.01

Table A6: Causal forest results with cluster robust estimation on the location level, using the restricted sample only.

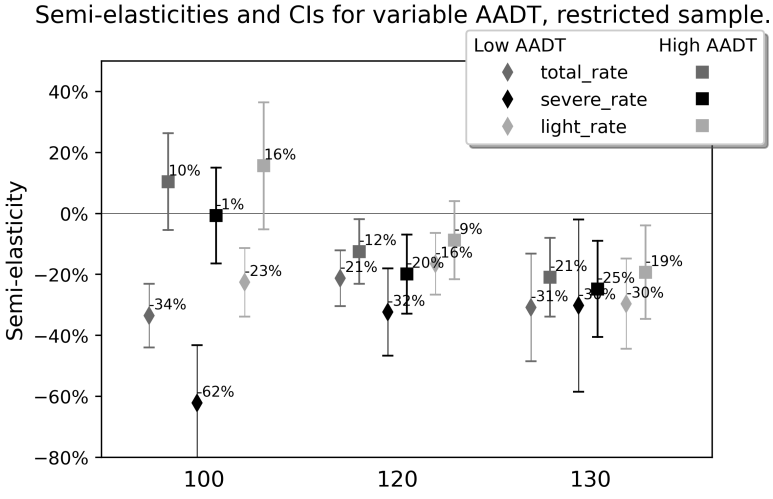

Figure A5: CATEs for segments with low and high traffic volume for the restricted sample only.

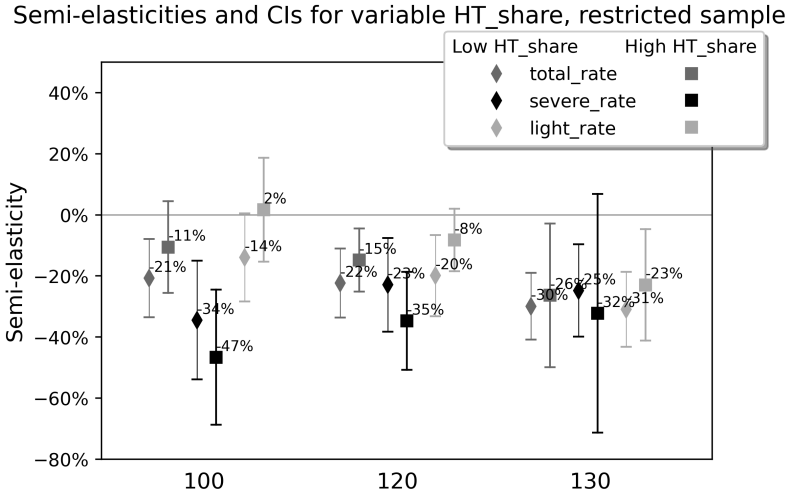

Figure A6: CATEs for segments with low and high shares of heavy traffic for the restricted sample only.

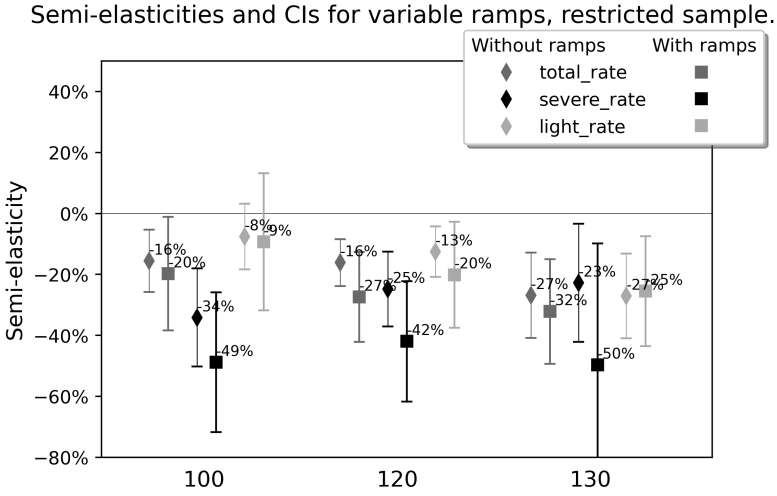

Figure A7: CATEs for segments without and with entry or exit ramps for the restricted sample only.