POLITECNICO DI TORINO Repository ISTITUZIONALE

Updating and transferring Random Effect models: The case of operating speed percentile estimation

Original

Updating and transferring Random Effect models: The case of operating speed percentile estimation / Tremblay, Jean-Michel; Cirillo, Cinzia; Bassani, Marco. - In: TRANSPORTATION RESEARCH. PART A, POLICY AND PRACTICE. -ISSN 0965-8564. - ELETTRONICO. - 148:(2021), pp. 286-304. [10.1016/j.tra.2021.01.008]

Availability: This version is available at: 11583/2896333 since: 2021-04-21T15:40:10Z

Publisher: Elsevier Ltd

Published DOI:10.1016/j.tra.2021.01.008

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright Elsevier postprint/Author's Accepted Manuscript

© 2021. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/.The final authenticated version is available online at: http://dx.doi.org/10.1016/j.tra.2021.01.008

(Article begins on next page)

Transportation Research Part A Updating and transferring Random Effect models: the case of operating speed percentile estimation --Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Keywords:	Operating speed, speed quantiles (percentile), random effects regression, jack-knife resampling technique, out-of-sample prediction
Corresponding Author:	Cinzia Cirillo University of Maryland College Park, MD United States
First Author:	Jean-Michel Tremblay, PhD
Order of Authors:	Jean-Michel Tremblay, PhD
	Cinzia Cirillo
	Marco Bassani, PhD
Abstract:	Random Effect (RE) models are used for analyzing data that are non-independent or when data are characterized by a hierarchical structure. In traffic and highway engineering, RE models have been successfully employed to estimate free-flow speed distributions from data containing observations that are naturally nested according to different levels (i.e. direction, sections, roads). Empirical studies conducted on both urban arterials and rural two-lane highways have shown that RE models, by properly accounting for the survey design, are superior to traditional Fixed Effect (FE) models. In this paper, the transferability of RE models to road sections that were not in the original sample used for model estimation was studied, under the assumption that for these additional sections very few observations are available or can be collected. This problem poses two challenges. First, random effects for the new road sections should be estimated in order to make out-of-sample predictions. Second, the original model formulation makes use of speed quantiles as predictors of the linear model which are not readily available for the new sections. The method proposed estimates an auxiliary model, in which the RE of the original model are correlated to the RE to be defined for the new section, with the former being used to predict the latter. The RE pairs are modeled jointly, taking advantage of their potential mutual correlation. The model coefficients obtained are also validated using a jackknife technique. Results show that the method converges quite fast and that a handful of observations for the new road section are sufficient for good RE estimates.
Suggested Reviewers:	
Opposed Reviewers:	

8

Updating and transferring Random Effect models: the case of operating speed percentile estimation Jean-Michel Tremblay¹, Cinzia Cirillo², and Marco Bassani³ ¹Edgybees Ltd., jeanmi.tremblay@gmail.com ²University of Maryland, Department of Civil and Environmental Engineering, ccirillo@umd.edu ³Politecnico di Torino, Department of Environment, Land and Infrastructure Engineering, marco.bassani@polito.it

April 20, 2020

1

• Abstract

Random Effect (RE) models are used for analyzing data that are non-independent or when data are characterized by a hierarchical structure. In traffic and highway engineering, RE models have been successfully employed to estimate free-flow speed distributions from data containing observations that are naturally nested according to different levels (i.e. direction, sections, roads). Empirical studies conducted on both urban arterials and rural two-lane highways have shown that RE models, by properly accounting for the survey design, are superior to traditional Fixed Effect (FE) models.

¹⁷ In this paper, the transferability of RE models to road sections that were not in the original ¹⁸ sample used for model estimation was studied, under the assumption that for these additional ¹⁹ sections very few observations are available or can be collected. This problem poses two ²⁰ challenges. First, random effects for the new road sections should be estimated in order to ²¹ make out-of-sample predictions. Second, the original model formulation makes use of speed ²² quantiles as predictors of the linear model which are not readily available for the new sections.

The method proposed estimates an auxiliary model, in which the RE of the original model are correlated to the RE to be defined for the new section, with the former being used to predict the latter. The RE pairs are modeled jointly, taking advantage of their potential mutual correlation. The model coefficients obtained are also validated using a jackknife technique. Results show that the method converges quite fast and that a handful of observations for the new road section are sufficient for good RE estimates.

Keywords: Operating speed, speed quantiles (percentile), random effects regression, jack knife resampling technique, out-of-sample prediction.

31 **Introduction**

Among the parameters characterizing vehicular flows, speed is used in multiple applications 32 including traffic analysis, speed management, road design, and road safety. Speeds are col-33 lected through spot speed observations or by recurring to permanent acquisition units in the 34 field (Garber and Hoel 2020; Catani et al. 2017). This data may be used to calibrate models 35 to explain how speed is related to some significant variables depicting the road scenario. As 36 a result, analysts and road designers can select the most appropriate road features to mod-37 ulate risk perception and compel drivers to adopt consistent speed decisions and behaviors. 38 Starting from the '80s, a conspicuous quantity of papers proposed models to predict the 39 85th speed percentile (i.e., V85) of operating speeds (OS), and OS differential (e.g., $\Delta V85$, 40 85MSR) between road elements (Dimaiuta et al. 2011). 41

V85 is conventionally considered representative of OS distribution since it separates the speed 42 of prudent drivers from that of more aggressive ones. V85 models may include geometric 43 characteristics of roads (e.g., lane width, radius or curvature), environmental conditions (e.g., 44 lighting, weather, land use), and driving regulations affecting driver behavior (Himes et al. 45 2013). The variation in operating speed between road elements has repeatedly been used to 46 support design decisions (Lamm et al. 1988). Park and Saccomanno (2006) evidenced that 47 speed variations must be evaluated for individual drivers (i.e., disaggregated data) rather than 48 from aggregated data for the observed group, in order to prevent the so called "ecological 49 fallacy" problem. However, this approach is challenging due to the need to monitor individual 50 vehicles along entire road segments (McFadden and Elefteriadou 2000). 51

However, when attention is focused on a section, the use of V85 becomes controversial since different OS distributions may exhibit the same 85th percentile. To address this issue, Shankar and Mannering (1998) proposed the use of simultaneous equations to model the average and standard deviation of speed in each lane of multilane highways. Later on, Figueroa-Medina and Tarko 2005 introduced a model to predict any speed percentile combining the mean and the standard deviation in a linear regression equation.

Most of the available literature has proposed models of the Fixed Effect (FE) type, in which 58 each speed observation along a road section is assumed to be dependent on the predictors 59 included in that model only (Dimaiuta et al. 2011). This is only acceptable when speed 60 clusters used to calibrate the model are not distinguished per direction, are from segments 61 that do not belong to the same road, and are sufficiently distant from each other. However, 62 when speed observations are clustered and spatially close to each other, each of them may 63 share unobserved effects. Thus, it is not possible to assume independence of errors for 64 individual observations without considering Random Effects (RE) for these groups. 65

Tarris et al. (1996) carried out a panel analysis of free-flow speed data collected from individual drivers with speed values recorded at sensor locations. In the proposed model, RE were associated with groups and speed location. Islam and El-Basyouny (2015) used RE to account for differences in hourly free-flow speed data related to site and community in a pilot study aimed at reducing OS. More recently, Cheng et al. (2018) evidenced that RE
are fundamental in predicting the speed and speed deviation along lanes in multilane highways. They used RE to account for the variation between adjoining lanes, between adjacent
segments, and among segments.

RE models reach a higher coefficient of determination than those obtained assuming FE
coefficients for groups and sensor locations (Bassani, Dalmazzo, et al. 2014). RE models for
OS was proposed by the authors (Bassani, Dalmazzo, et al. 2014; Bassani, Cirillo, et al. 2016;
Bassani, Catani, et al. 2016) in multiple observations for the same direction (d) of a section

(s), on several sections of a road (r), and on several roads of the network. In the model:

$$V_{rsd,i} = \beta_0 + \beta_k X_{rsd,k} + \beta_j Z_p X_{rsd,j} + \alpha_r + \alpha_{s|r} + \alpha_{d|rs} + \epsilon_{rsd,i}, \tag{1}$$

⁷⁹ where β_0 is the general model intercept, β_k and β_j are calibration parameters for the k and ⁸⁰ *j* variables affecting the estimated mean $X_{rsd,k}$, and the estimated standard deviation $X_{rsd,j}$ ⁸¹ respectively, and Z_p is the standardized normal variable. In eq. 1, α_r , $\alpha_{s|r}$ and $\alpha_{d|rs}$ are the ⁸² three nested RE accounting for the variability introduced by the random selection of roads ⁸³ in the network, the section within a road, and the directions in the section.

The objective of this study is to transfer RE models calibrated on a given sample of lanes, 84 sections and roads to road sections that were not included in the estimation sample and 85 for which few speed observations were collected or available to the analyst. The problem 86 has relevant practical implications, as the method proposed will facilitate the use of existing 87 models on different sections without the need to collect a significant number of new obser-88 vations, which is usually a lengthy and costly process. In order for this transferability to be 89 effective and to have realistic out-of-sample predictions, RE need to be predicted for the new 90 road section. Also, the model specification is based on speed quantiles, which are not readily 91 available for the new road sections. 92

REs in most situations are assumed to have zero mean and therefore the best a priori predictor 93 for REs in a new road section is zero as well. However, it is likely that better predictors can 94 be produced by considering a simpler, auxiliary RE model whose purpose is to overcome the 95 unavailability of quantiles in the validation sample. Specifically, following McCulloch et al. 96 (2008), it is possible to build an auxiliary model which, although inferior to the actual model, 97 takes advantage of the potential correlation existing across RE pairs in order to provide good 98 RE predictors. The method is based on the assumption that REs of the original model are 99 correlated with the REs to be defined for the new section and that the first one can be used 100 to predict the latter. The methodology is first developed in the case of one RE, and then 101 extended to the case of a model including two REs. The best predictor is derived and the 102 convergence is tested on empirical data. Finally, validation is performed on the quantiles of 103 all the road sections considered using a jackknife technique (Efron and Tibshirani 1993). 104

The remaining of this paper is articulated as follows. Model formulation for one RE and two REs is presented in Section 2. Numerical results derived from real data collected in the North-West of Italy are reported in Section 3. The case where predictors are multiplied by the normal quantiles is solved in Section 4. In Section 5, a Jackknife re-sampling technique is
used to analyze the causes of poor estimation results. Conclusions and suggestions for future
research are given in Section 6.

111 2 Model formulation

In this Section, we derive the statistical method to transfer an estimated random effect 112 model to a new road section for which very few observations are available to the analyst. 113 The problem is that for this new road section, only the model predictors are available, while 114 the random effect(s) are unknown. Under the hypothesis that the random effect of the new 115 section are correlated to those of the section for which the model has been estimated, we 116 derive the conditional mean of the Best Linear Unbiased Predictor (BLUP) of the error in 117 the new section. The method is developed first for a one random effect model (Section 2.1) 118 and then generalized to a two random effect model (Section 2.2). 119

¹²⁰ 2.1 One random effect model

Following the formulation in eq.1, we first develop the proposed methodology for a simple case that contains one random effect. Let's assume that speed data is available for several sections s, and that the response variable $V_{s,i}$ is affected by a set of predictors $X_{s,k}$. The model contains one random effect α_s and an error term $\epsilon_{s,i}$:

$$V_{s,i} = \beta_0 + \beta_k X_{s,k} + \alpha_s + \epsilon_{s,i},\tag{2}$$

¹²⁵ where α_s and $\epsilon_{s,i}$ are independent and follow a normal distribution:

$$\alpha_s \sim N(0, \sigma_s^2),$$

$$\epsilon_{s,i} \sim N(0, \sigma^2).$$

while β is defined as the vector of fixed coefficients to be estimated and that includes both β_0 and β_k .

Once the model is calibrated, estimates for the model parameters (denoted as $\hat{\beta}$, $\hat{\sigma}_s^2$, and $\hat{\sigma}^2$) are available to the analyst, and predictions for the random variables α_s and $V_{s,i}$ can be obtained (McCulloch et al. 2008).

Assuming the analyst is interested in predicting random effects for a new section s' with unknown random effect $\alpha_{s'}$, the best *a priori* estimate of $\alpha_{s'}$ is zero since $\alpha_{s'} \sim N(0, \sigma_s^2)$. In some contexts this may be satisfying, but other methods can be explored under the assumption that it is possible to sample a few observations for the new section s' in order to improve the knowledge about $\alpha_{s'}$ and build a better *a posteriori* prediction. This is precisely the ultimate objective of our method.

¹³⁷ Suppose that *n* observations $V_{s',1}, V_{s',2}, \ldots, V_{s',n}$ are collected in the new section. $\alpha_{s'}$ cannot ¹³⁸ be observed directly, because we always observe the *sum* of the errors. Hence according to ¹³⁹ eq. 2 the sum of the residuals is:

$$r_{s',i} = \alpha_{s'} + \epsilon_{s',i} = V_{s',i} - \beta X_{s',k}.$$

The difference $V_{s',i} - \beta X_{s',k}$ cannot be measured because the value of the parameters in β is unknown, so the estimated value $\hat{\beta}$ from model calibration is used to approximate the total residuals r'_s, i .

Let's define $u = (\alpha_{s'})$ and let's assume that it follows a normal distribution with mean zero and variance $D = (\sigma_s^2)$. In this case, u refers to a single random effect, but as we will see in the next Section, the methodology applies to any number of effects to be predicted. The residuals $r = (r_{s',1}, r_{s',2}, ..., r_{s',n})$ have zero mean and variance W:

$$W = \begin{bmatrix} \sigma_s^2 + \sigma^2 & \sigma_s^2 & \dots & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 + \sigma^2 & \dots & \sigma_s^2 \\ \dots & \dots & \dots & \dots \\ \sigma_s^2 & \sigma_s^2 & \dots & \sigma_s^2 + \sigma^2 \end{bmatrix}.$$

 147 W can be written as follows:

$$W = \sigma^2 I + \sigma_s^2$$

where I is the identity matrix.

The covariance between the random effect $\alpha_{s'}$ and one given total residual is:

$$Cov(\alpha_{s'}, r_{s',i}) = \sigma_s^2.$$

and more generally, the covariance C between u and r is:

$$C = \begin{bmatrix} \sigma_s^2 & \sigma_s^2 & \dots & \sigma_s^2 \end{bmatrix} = \sigma_s^2 \mathbf{1}_{1 \times n}$$

151 and therefore:

$$\begin{bmatrix} u \\ r \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} D & C \\ C^T & W \end{bmatrix} \right).$$

At this stage, it is not necessary to formulate hypotheses on the joint distribution of (u, r); it is sufficient to know the first and second moments of (u, r) to derive the *best linear unbiased predictor* (BLUP) of u (McCulloch et al. 2008). If the joint distribution were normal, the BLUP would be the overall best predictor of u. This not the case for our real application.

Once r is observed, the BLUP of u is the conditional mean. The expectation of u|r is given by:

$$\hat{\alpha}_{s'} = E(u|r) = CW^{-1}r.$$

The estimated values of the coefficients are used to produce numerical values for the BLUP. The predicted value for an observation i in section s' will be:

$$\hat{V}_{s',i} = \hat{\beta} X_{s',k} + \hat{\alpha}_{s'}.$$

¹⁶⁰ β can be estimated in a relatively easy way, so the main challenge for this problem is to ¹⁶¹ predict $\hat{\alpha}_{s'}$. The objective here is to investigate what is the smallest sample in section s' that ¹⁶² we can use to predict $\alpha_{s'}$ satisfactorily. In general, it can be observed that the prediction ¹⁶³ converges faster when σ^2 is low relative to σ_s^2 ; this is because one single observation of r is ¹⁶⁴ expected to be less noisy and more correlated with the unknown realized value $\alpha_{s'}$.

¹⁶⁵ 2.2 Two random effects models

¹⁶⁶ The case with two nested random effects is formulated in eq. 3:

$$V_{sd,i} = \beta_0 + \beta_k X_{sd,k} + \alpha_s + \alpha_{d|s} + \epsilon_{sd,i},\tag{3}$$

where the subscript d stands for direction, so speed data can be distinguished into two different directions for the same section. The random effect $\alpha_{d|s}$ is nested within the levels of α_s .

We make the assumption that this new random effect is normally distributed:

$$\alpha_{d|s} \sim N(0, \sigma_d^2)$$

As a result, in this model a sample from the new section s' also includes some levels of the direction effect. For illustration purposes, one section, two directions and n observations per direction are assumed. The RE to be predicted is:

$$u = (\alpha_{s'}, \alpha_{d1}, \alpha_{d2}),$$

¹⁷³ and the total residuals are given by:

$$r_{s'd,i} = V_{s'd,i} - \beta X_{s'd,k}.$$

where β is the vector of coefficient to be estimated.

Similar to the one RE model, the RE cannot be directly observed. In this case, the covariance
 structure for the total residuals is:

$$var(r_{s'd,i}) = \sigma_s^2 + \sigma_d^2 + \sigma^2,$$

$$cov(r_{s'd,i}, r_{s'd,j}) = \sigma_s^2 + \sigma_d^2,$$

$$cov(r_{s'd,i}, r_{s'd',i}) = \sigma_s^2$$

¹⁷⁷ Two residuals in the same section but for different directions only share the α_s term so their ¹⁷⁸ covariance is σ_s^2 . Residuals in the same direction also share $\alpha_{d|s}$ so their covariance is the ¹⁷⁹ sum $\sigma_s^2 + \sigma_d^2$. Finally, the total variance of r is the sum of its three components $\sigma_s^2 + \sigma_d^2 + \sigma^2$. ¹⁸⁰ These results are a consequence of the independence assumption made on the three REs.

As in the previous section, we want to derive the joint moments of (u, r). The covariance matrices D, C and W of (u, r) are respectively:

$$D = \begin{bmatrix} \sigma_s^2 & 0 & 0 \\ 0 & \sigma_d^2 & 0 \\ 0 & 0 & \sigma_d^2 \end{bmatrix}, W = \sigma_s^2 + \begin{bmatrix} \sigma_d^2 + \sigma^2 & \sigma_d^2 & 0 & 0 & 0 \\ \sigma_d^2 & \sigma_d^2 + \sigma^2 & \sigma_d^2 & 0 & 0 & 0 \\ \sigma_d^2 & \sigma_d^2 & \sigma_d^2 + \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_d^2 + \sigma^2 & \sigma_d^2 & \sigma_d^2 \\ 0 & 0 & 0 & \sigma_d^2 & \sigma_d^2 + \sigma^2 & \sigma_d^2 \\ 0 & 0 & 0 & \sigma_d^2 & \sigma_d^2 + \sigma^2 & \sigma_d^2 \end{bmatrix}$$

with **0** being a matrix of zeros and I the identity matrix:

$$W = \sigma_s^2 + \begin{bmatrix} \sigma_d^2 + \sigma^2 I & 0_{n \times n} \\ 0_{n \times n} & \sigma_d^2 + \sigma^2 I \end{bmatrix}$$

C is finally given by:

$$C = \begin{bmatrix} \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\ \sigma_d^2 & \sigma_d^2 & \sigma_d^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_d^2 & \sigma_d^2 & \sigma_d^2 \end{bmatrix} = \begin{bmatrix} \sigma_s^2 \mathbf{1}_{1 \times n} & \sigma_s^2 \mathbf{1}_{1 \times n} \\ \sigma_d^2 \mathbf{1}_{1 \times n} & 0_{1 \times n} \\ 0_{1 \times n} & \sigma_d^2 \mathbf{1}_{1 \times n} \end{bmatrix}$$

¹⁸⁵ The predicted random effects are given by the expected mean:

$$E(u|r) = CW^{-1}r.$$

The predicted value for an observation i in section s' and direction d will be:

$$\hat{V}_{s'd,i} = \hat{\beta}_{s'di} X_{s'd,k} + \hat{\alpha}_{s'} + \hat{\alpha}_{d|s}$$

¹⁸⁷ 3 Numerical examples on OS data

Speed data used in this study were collected in road sections of two-lane rural highways 188 in the North-West part of Italy. Individual speeds of isolated vehicles were collected under 189 free-flow conditions in sections where vehicles travel at constant speed (i.e. in the center of 190 tangents and curves). Speeds were included in the database only when a minimum headway 191 of six seconds was observed. The data for model estimation were extracted from a larger 192 database already used by the authors in (Bassani, Cirillo, et al. 2016). The density and the 193 presence of elements along the road section was evaluated along one km across the sample 194 sections. Table 1 lists the values assumed by the variables that were found to be significant 195 in the calibration of the model reported in eq. 1. In the table, the variables are divided 196 into those affecting the average (X_k) and the dispersion (X_i) of predictors. The latter are 197 also divided into numerical and Boolean variables. Finally, the last five columns summarize 198 the minimum (V_{min}) , the maximum (V_{max}) , the 50th (V50) and the 85th (V85) percentile of 199 speeds included in the database, while n_{obs} indicates the number of data available for each 200 section. The notes at the end of table 1 describe the acronyms used to identify the variables. 201

$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		X_{μ_i}						X	ii (numer	ical)							X_{ii} (1	oolea	(u						
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	1 "	R 1-2	PedD n/km	m II	SLW m	SRW	LG KG	TRDLS n/km	TRDRS	DDRS n/km	IDLS n/km	IDRS n/km	PSL km/h	$\left \frac{\Delta PSL}{km/h} \right $	- TBL	S LRR	STS	SRS -	SBLS -	SBRS -	V_{min} km/h	V_{50} km/h	V_{85} km/h	V_{max} km/h	Z¤
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	0.0	0	0	3.60	0.40	0.40	1.50	0	0	2	-	0	20	0				0	0	0	39.0	66.0	76.0	97.0	429
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	0.0	0	0	3.75	0.40	0.40	0.00	0	1	2	0	0	50	0	1	1	0	0	1	1	32.0	74.0	85.0	157.0	972
	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	0.0	0	0	3.70	0.00	0.00	1.00	0	0	1	0	0	6	0	0	1	0	0	1	0	38.0	74.0	89.0	130.0	699
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	3.2	9e-3	0	3.75	1.50	1.50	5.14	0	0	0	0	0	20	0	1	1	0	0	1	1	24.0	82.0	94.0	114.0	312
			0.0	0	0	3.75	1.50	1.50	2.09	0	0	0	0	0	20	0		1	0	0	1	1	57.0	86.0	98.0	124.0	101
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$ [7.7c-3] \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	3.1	.0e-4	0	3.75	1.50	1.50	4.69	0	0	0	1	0	20	0	0	1	0	0	1	0	52.0	80.0	91.0	114.0	101
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	6.6	57e-3	0	3.25	0.00	0.00	8.50	0	1	ŝ	0	2	50	0	0	0	0	0	1	1	46.0	62.0	70.1	77.0	87
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	0	00	0	3.00	0.50	0.50	1.50	0	0	4	0	က	70	0	0	0	0	0	0	0	42.6	71.0	90.3	129.4	107
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	0	00	0	3.00	0.50	0.50	0.00	0	0	1	0		70	0	0	0	0	0	0	0	46.7	81.5	94.3	128.6	120
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Ö.	00	0	3.00	0.50	0.50	0.00	0	0	1	0	0	70	0	0	0	0	0	0	0	50.0	80.2	90.4	114.0	108
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	0	00.	0	3.00	0.50	0.50	2.00	0	0	2	0	0	20	0	0	0	0	0	0	0	49.6	83.0	101.7	129.4	125
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	0	00.	0	3.00	0.50	0.50	0.00	0	0	0	0	1	20	0	0	0	0	0	0	0	49.0	82.7	95.9	127.8	127
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	0	00.	0	3.00	0.50	0.50	0.00	0	0	1	0	0	20	0	0	0	0	0	0	0	46.3	83.2	98.9	128.3	138
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	0	00.	0	3.00	0.50	0.50	0.00	0	0	9	2	1	70	0	0	0	0	0	0	0	49.4	78.5	95.3	149.6	128
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	0	00	0	3.50	1.20	1.20	0.50	0	0	1	0	0	6	0	0	0	0	0	0	0	42.0	67.2	88.5	115.2	41
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	4	.44e-4	0	3.50	1.20	1.20	0.50	0	0	0	1	1	6	0	0	0	0	0	0	0	70.0	97.0	107.0	118.0	26
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	0	00.	0	3.50	1.20	1.20	0.50	1	0	0	4	4	00	0	0	0	0	0	0	0	58.0	99.5	112.9	127.0	28
66e-4 0 3:50 1:20 1:20 1:00 0 0 1 1 1 3:20 5:0.0 5:7.4 7:2.0 3:7.	66e-4 0 3.50 1.20 1.20 1.00 0 1 1 1 3.20 5.0 5.7.4 7.2.0 3.2.0 10e-3 0 3.56 1.20 1.20 1.20 0.00 0 1 0 0 1 1 0 59.0 77.0 92.6 110.0 37.5 37.5 37.0 35.5 98.0 37.5 37.6 35.5 98.0 35.5 38.0 35.6 35.6 37.6 37.6 35.5 38.0 35.5 38.0 38.5 35.6 37.0 36.6 37.6 37.6 35.5 38.0 38.5 38.6 38.6 38.5 38.6 38.5 38.6		0	00.	0	3.50	1.20	1.20	0.00	0	0	J.	1	1	6	0	0	1	0	0	0	0	55.0	77.0	100.3	130.0	30
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Ö	.66e-4	0	3.50	1.20	1.20	1.00	0	0	1	1	4	50	0	0	0	0	0	1	1	32.0	50.0	57.4	72.0	32
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Ö	00	0	3.50	1.20	1.20	0.00	0	0	×	0	n	20	0		0	0		0	0	59.0	77.0	92.6	110.0	27
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	÷	10e-3	0	3.50	1.20	1.20	1.00	0	0	1	0	0	20	0	0	0	0	0	1	1	44.0	69.0	78.0	96.0	37
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	<u> </u>	20e-4	0	3.50	1.20	1.20	0.00	0	0	9	0	2	20	0		0	0		0	0	47.0	75.0	85.5	98.0	38
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	сi	21e-3	4	3.50	1.20	1.20	0.00	0	0	4	n	က	50	0	0				0	0	44.0	51.0	57.5	70.0	36
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	6.	65e-4	0	3.50	1.20	1.20	1.00	0	0	5	0	0	60	0	0	0	0	0	0	0	56.0	87.0	97.0	130.0	43
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	o.	00	0	3.50	1.20	1.20	1.00	1	0	1	0	0	00	0	0		0	0	0	0	65.0	91.0	107.9	119.5	38
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	o.	00		3.50	1.20	1.20	0.00	0	0	7	c,	ю	20	0	0	0	-		0	0	54.7	69.3	85.1	108.1	42
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Ö.	00	0	3.50	1.20	1.20	0.00	0	0	9	0	1	20	0		0	0		0	0	44.3	73.1	83.6	120.2	140
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	3	.33e-3	2	3.20	0.80	0.80	0.00	0	0	4	4	2	20	0	0	1	0	0	0	0	27.6	66.7	75.8	93.3	116
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	0	00.	0	3.80	1.26	1.26	0.00	0	0	က	-	0	6	0	0	0	0	-	0	0	42.6	67.4	75.3	106.6	67
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	00 $\begin{vmatrix} 1 \\ 3.50 \\ 0.7$	00 $\begin{vmatrix} 1 \\ 3.50 \\ 0.7$	0	00	en en	3.50	0.40	0.40	0.00	0	0	ъ	1	4	50	0	0	1	-		0	0	34.7	55.0	68.5	105.4	154
L = curvature, PedD = pedestrian density, LW = lane width, SLW = shoulder left width, SRW = shoulder right width, LG = longitudinal grade, TRDLS = total ramp side, TRDRS = total ramp density right side, DDRS = driveway density right side, IDLS = intersection density left side, IDRS = intersection density right side, LBLS = total ramp density right side, DDRS = driveway density right side, LBLS = intersection density left side, IDRS = intersection density right side, LBLS = total ramp density right side, DDRS = driveway density right side, LBLS = intersection density left side, IDRS = intersection density right side, LBLS = total ramp density right side, LBLS = driveway density right side, LBLS = intersection density left side, IDRS = intersection density right side, LBLS = total ramp density right side, DBRS = driveway density right side, LBLS = intersection density left side, IDRS = intersection density right side, LBLS = total ramp density right side, LBLS = driveway density right side, LBLS = intersection density left side, LBLS = intersection density right side, LBLS = total ramp density right side, LBLS = driveway density right side, LBLS = intersection density left side, LBLS = intersection density right side, LBLS	z = curvature, PedD = pedestrian density, LW = lane width, SLW = shoulder left width, SRW = shoulder right width, LG = longitudinal grade, TRDLS = total ramp side, TRDRS = total ramp density right side, DDRS = driveway density right side, IDLS = intersection density left side, IDRS = intersection density right side, LBLS = side, LBRS = lay-by right side, SLS = sidewalks right side, SBLS = safety barriers left side, SBRS = safety barriers right side, PSL = posted	z = curvature, PedD = pedestrian density, LW = lane width, SLW = shoulder left width, SRW = shoulder right width, LG = longitudinal grade, TRDLS = total ramp side, TRDRS = total ramp density right side, DDRS = driveway density right side, IDLS = intersection density left side, IDRS = intersection density right side, LBLS = side, LBRS = lay-by right side, SLS = sidewalks right side, SBLS = safety barriers left side, SBRS = safety barriers right side, PSL = posted , Vmin = minimum speed value, V50 = 50 th percentile of speed, V85 = 85 th percentile of speed, Vmax = maximum speed value, N = number of speed data. Binomial values	0	00.	-	3.50	0.70	0.70	0.00	0	0	×	1	1	50	0	1	1	0	0	0	0	36.9	69.1	82.5	134.8	161
side, TRDRS = total ramp density right side, DDRS = driveway density right side, IDLS = intersection density left side, IDRS = intersection density right side, LBLS =	side, TRDRS = total ramp density right side, DDRS = driveway density right side, IDLS = intersection density left side, IDRS = intersection density right side, LBLS = side, LBRS = lay-by right side, SLS = sidewalks left side, SLS = sidewalks right side, SBLS = safety barriers left side, SBRS = safety barriers right side, PSL = posted	side, TRDRS = total ramp density right side, DDRS = driveway density right side, IDLS = intersection density left side, IDRS = intersection density right side, LBLS = side, LBRS = lay-by right side, SLS = sidewalks right side, SBLS = safety barriers left side, SBRS = safety barriers right side, PSL = posted side, LBRS = lay-by right side, V50 = 50^{th} percentile of speed, V85 = 85^{th} percentile of speed, Vmax = maximum speed value, N = number of speed data. Binomial values	ЦЦ,	$t = curv_{\tilde{c}}$	uture, F	edD =	= ped	estrian	densit	y, LW =	lane widt	h, SLW	= sho	ulder l	eft wid	lth, SRV	V = sh	oulder	right v	vidth,	LG = I	ongitudi	nal grae	de, TR	DLS =	total r	amp
	side, LBRS = lay-by right side, SLS = sidewalks left side, SLS = sidewalks right side, SBLS = safety barriers left side, SBRS = safety barriers right side, PSL = posted	side, LBRS = lay-by right side, SLS = sidewalks left side, SLS = sidewalks right side, SBLS = safety barriers left side, SBRS = safety barriers right side, PSL = posted $Vmin = minimum$ speed value, V50 = 50 th percentile of speed, V85 = 85 th percentile of speed, Vmax = maximum speed value, N = number of speed data. Binomial values	£	side, TF	DRS =	= total	ramp	densit	y right	side, DD	RS = dri	veway d	lensity	right si	ide, ID	LS = in	tersecti	ion den	sity lef	it side,	IDRS =	= interse	ection d	ensity 1	right sid	e, LBL	S S

Table 1: Summary of characteristics of the selected road sections

202 3.1 Results: one RE model

The first case study models speed data with only one RE, as formulated in 2.1; specifically, it accounts for the RE related to sections but ignores the RE related to directions. A model is calibrated using thirty sections out of the thirty one available in the database and listed in Table 1. Once the model is fitted on the thirty sections, the RE effect is predicted for the section that was left out from model estimation. This is repeated for all the thirty one sections in the dataset. This procedure provides a series of predicted REs; these results are used to assess the convergence of the method proposed.

Figures 1 and 2 illustrate the results obtained for each section. Each subplot represents the predicted RE in the section used for validation. The x-axis corresponds to the number of observations that were used from the validation section in order to predict the REs. For example, an x-value of 10 for section 3 means that a model with all sections but the third one was estimated, and that ten observations in the third section were used in order to predict the (realized) effect of section 3.

Each subplot contains a solid line, a dashed line, and a dotted line. The solid line shows the predicted effects obtained with the **full** model, for which the quantile information in the validation sample is assumed to be known. This is ultimately the effect that we aim to predict. The dotted line shows the predicted effects obtained with a **simple** auxiliary model that only contains the predictors $X_{s,k}$ that affect the mean speed. The dashed line shows the predicted effects for the **conditional** model obtained using information from the auxiliary model. This is the prediction that ultimately is going to be used.

From figures 1 and 2, the following three remarks can be made: (i) a relative convergence in 223 the predicted effects as the number of observations grows is observed; (ii) there is a substantial 224 difference between the dotted line and the black solid line; meaning that the closer the two 225 lines are, the more likely that the one can be predicted from the other; (iii) the dashed line 226 does not approximate the solid line and it is mostly super-posed to the dotted line. Therefore, 227 it can be concluded that the predicted effect of the full model using the auxiliary model are 228 no better approximation than just the predicted effects of the auxiliary model. This results 229 might be seen as disappointing, but later it will be proved that correctly accounting for the 230 sample design solves the problem observed. 231







Figure 2: Sections 23-36

number of observations

²³² 3.2 Results: two RE model

The second case study models two REs as formulated in 2.2; this time both section and direction effects are taken into account. The scope here is to assess the convergence of $\hat{\alpha}_{s'}$, $\hat{\alpha}_{d1}$ and $\hat{\alpha}_{d2}$ in the presence of two REs; this model formulation is fully consistent with the survey design.

Figures 3 and 4 should be read in the same way as figures 1 and 2. The thick solid line corresponds to the predicted section effect, and the two fine solid lines correspond to the predicted direction effects. The dashed lines correspond to the predicted effects using the auxiliary model and those are used to compute the REs in the validation section.

For both sections and directions, the predicted effects using the auxiliary model are very close to the ones using the full model. However, there are still noticeable differences when very few observations are used for the prediction. For example, predictions in sections 8 and 27 are not very precise for only two or three observations.

It is worth noting that striking differences exist between predictions with one and two REs. Effects with only the section component are not close to the *true* effects of the full model; incorporating the direction effects, thus accounting for the design of the sample, drastically improves the predictions.

²⁴⁹ 4 Computation of the residuals

We now turn our attention to the case where the model includes variables $X_{sd,j}$ multiplied by the normal quantile Z_p in addition to the regular predictors $X_{sd,k}$, (see Bassani, Dalmazzo, et al. 2014; Bassani, Cirillo, et al. 2016; Bassani, Catani, et al. 2016). This is done to calculate any percentile speed as a linear combination of variables affecting both the central tendency and the dispersion of the collected speed data. Those quantiles are not available for the validation sample, which is too small to allow for the calculation of variance indicators. Note that $Z_{sd,j}$ is a scalar so the term $Z_p X_{sd,j}$ is simply a scalar multiplied by a vector.

$$V_{sd,i} = \beta_0 + \beta_k X_{sd,k} + \beta_j Z_p X_{sd,j} + \alpha_s + \alpha_{d|s} + \epsilon_{sd,i}$$

In this case, it is not possible to isolate the sum of REs and we must rely on alternative methods. The strategy that is proposed here makes use of the estimated total residuals from a restricted model that only contains observable variables $(X_{sd,k})$ to predict the REs of an unrestricted model (with both $X_{sd,k}$ and $X_{sd,j}$).



Figure 3: Sections 1-22

Figure 4: Sections 23-36



$$V_{sd,i} = \beta_k X_{sd,k} + \alpha_s^* + \alpha_{d|s}^* + \epsilon_{sdi}^*,$$

$$\alpha_s^* \sim N(0, \sigma_s^{*2}),$$

$$\alpha_{d|s}^* \sim N(0, \sigma_d^{*2}),$$

$$\epsilon_{sd,i}^* \sim N(0, \sigma^{*2}).$$

$$r_{sd,i}^* = V_{sd,i} - \beta_k X_{sd,k}$$

$$V_{sd,i} = \beta_k X_{sd,k} + \beta_j X_{sd,j} + \alpha_s + \alpha_{d|s} + \epsilon_{sdi},$$

$$\alpha_s \sim N(0, \sigma_s^2),$$

$$\alpha_{d|s} \sim N(0, \sigma_d^2),$$

$$\epsilon_{sd,i} \sim N(0, \sigma^2).$$

$$r_{sd,i} = V_{sd,i} - \beta_k X_{sd,k} - \beta_j X_{sd,j}$$
(5)

The restricted (eq. 4) and unrestricted models (eq. 5) are calibrated and parameters are estimated, together with the empirical correlation (ρ_{α_s} and $\rho_{\alpha_{d|s}}$,) across the effects of both models. Figures 5a and 5b plot the predicted section and direction effects of the restricted model and the unrestricted model respectively. As anticipated, the more correlated these effects are, the easier will be to predict one using the other.







$$Var(r_{sd,i}^{*}) = \sigma_{s}^{*2} + \sigma_{d}^{*2} + \sigma^{*2}.$$

²⁶⁸ The covariance of two residuals is given by:

$$Cov(r_{sdi*}, r_{sdj*}) = \sigma_{s*}^2 + \sigma_{d*}^2$$
 in the same direction, and $Cov(r_{sd,i*}, r_{sd'i*}) = \sigma_{s*}^2$ not in the same direction.

²⁶⁹ The covariance between residuals and REs are given by:

$$Cov(r_{sdi}^*, \alpha_s) = Cov(\alpha_s^* + \alpha_{d|s}^* + \epsilon_{sdi}^*, \alpha_s) = Cov(\alpha_s^*, \alpha_s) = \rho_{\alpha_s}\sigma_s\sigma_s^* = \sigma_{ss}^*,$$
$$Cov(r_{sdi}^*, \alpha_{d|s}) = Cov(\alpha_s^* + \alpha_{d|s}^* + \epsilon_{sdi}^*, \alpha_{d|s}) = Cov(\alpha_{d|s}^*, \alpha_{d|s}) = \rho_{\alpha_{d|s}}\sigma_d\sigma_d^* = \sigma_{dd}^*.$$

Therefore, the joint covariance of r_* and u is described by the following variance components:

$$D = \begin{bmatrix} \sigma_s^2 & 0 & 0 \\ 0 & \sigma_d^2 & 0 \\ 0 & 0 & \sigma_d^2 \end{bmatrix},$$

$$W = \sigma_{s*}^{2} + \begin{bmatrix} \sigma_{d}^{*2} + \sigma^{*2} & \sigma_{d}^{*2} & \sigma_{d}^{*2} & 0 & 0 & 0 \\ \sigma_{d}^{*2} & \sigma_{d}^{*2} + \sigma^{*2} & \sigma_{d}^{*2} & 0 & 0 & 0 \\ \sigma_{d}^{*2} & \sigma_{d}^{*2} & \sigma_{d}^{*2} + \sigma^{*2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{d}^{*2} + \sigma^{*2} & \sigma_{d}^{*2} & \sigma_{d}^{*2} \\ 0 & 0 & 0 & \sigma_{d}^{*2} & \sigma_{d}^{*2} + \sigma^{*2} & \sigma_{d}^{*2} \\ 0 & 0 & 0 & \sigma_{d}^{*2} & \sigma_{d}^{*2} + \sigma^{*2} \end{bmatrix},$$
$$C = \begin{bmatrix} \sigma_{ss}^{*} & \sigma_{ss}^{*} & \sigma_{ss}^{*} & \sigma_{ss}^{*} & \sigma_{ss}^{*} & \sigma_{ss}^{*} & \sigma_{ss}^{*} \\ \sigma_{dd}^{*} & \sigma_{dd}^{*} & \sigma_{dd}^{*} & 0 & 0 \\ 0 & 0 & 0 & \sigma_{dd}^{*} & \sigma_{dd}^{*} \end{bmatrix}.$$

Note that D is the same as before, but W is different because the random effects in r are from the simpler model, and C is also affected.

Tables 2 through 9 report predicted speed deciles (pred.) and compare them with the observed ones (obs.).

		Secti	ion 1			Sect	ion 3			Sect	ion 5			Sect	ion 7	
quantile	dir	· 1	dir	2	dir	1	dir	2	dir	1	dir	2	dir	1	dir	2
	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.
10	50	56	48.6	54	65.2	58	70.4	63	59.3	60	53.5	57	65.4	66	66.1	72.6
20	56.7	59	56.2	58	68.7	63	73.1	68	64.7	64	61	63	70.7	69	71.8	76
30	61.5	62	61.6	60	71.2	66	75	71	68.6	68	66.3	67	74.5	73	76	79.8
40	65.6	65.8	66.3	62	73.3	69	76.7	74	72	71	70.9	70	77.8	77	79.5	82
50	69.5	67	70.7	65	75.3	72	78.2	77	75.1	73	75.2	75	80.9	80	82.8	85
60	73.3	69	75	67	77.3	74	79.8	79	78.3	77	79.5	77	84	82	86.1	87
70	77.4	71	79.7	70	79.4	77	81.5	82	81.6	81	84.1	81.9	87.3	86.8	89.6	89
80	82.3	73	85.2	74	81.9	81	83.4	85	85.5	85	89.5	87	91.1	90	93.8	93
90	88.9	77.3	92.8	79	85.3	88	86.1	89	91	90	97	94	96.5	95	99.5	97.4

Table 2: Predicted quantiles sections 1,3,5 and 7

		Sect	ion 8			Sect	ion 9			Secti	on 10			Secti	on 11	
quantile	di	r 1	di	r 2	dir	1	dir	: 2	dir	1	dir	2	dii	: 1	dir	2
	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.
10	68.2	69.4	67.9	74.2	63.2	60	65.2	70	47.4	50	54.5	56	56.4	59.3	56.4	59.3
20	73.9	75.4	73.5	77	68.4	67	70.4	72.8	51.8	51.2	57.2	60	62.5	63.2	62.5	61.7
30	78	81	77.5	77.6	72.1	68	74.1	76	54.9	54.6	59.1	63	66.9	65.1	66.8	67.1
40	81.5	84	81	81.6	75.2	73	77.3	79.6	57.6	55	60.8	64	70.6	68.1	70.6	71
50	84.8	86	84.2	85	78.2	75	80.3	82	60.1	56	62.4	65	74.2	70.6	74	72.5
60	88.1	88.2	87.4	87.2	81.1	79	83.3	86.4	62.6	57	63.9	67	77.7	71.9	77.5	74
70	91.6	90	90.9	90	84.3	86	86.5	88	65.3	59	65.6	68	81.5	75.9	81.2	76.4
80	95.7	94.2	94.9	94.8	87.9	87	90.2	91.2	68.4	60.8	67.5	71.2	85.9	81.6	85.6	81.6
90	101.4	103.5	100.5	100.6	93.1	90	95.4	94.1	72.7	64.4	70.2	73	92	95.6	91.6	93

Table 3: Predicted quantiles sections 8,9,10,11

Table 4: Predicted quantiles sections 12,13,14,15

		Sect	ion 12			Secti	on 13			Secti	on 14			Sectio	on 15	
quantile	dir	1	di	r 2	dir	1	dir	2	di	r 1	di	r 2	di	r 1	dir	2
	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.
10	60.9	59.7	61.1	59	62.9	62.7	62.2	65.8	64.5	67.1	63.4	67.8	64.6	64.3	63.7	65.7
20	66.9	64.2	67.4	67.5	68.6	67.4	68.3	68.5	71.2	70.6	70.6	72.1	70.7	69.5	70.2	73.8
30	71.3	71.3	71.8	72	72.7	70.7	72.7	72	76	76.3	75.7	75.3	75	74.3	74.8	75.6
40	75	73.9	75.7	78.3	76.3	75.4	76.5	74.3	80.1	79.6	80.2	79.1	78.8	78	78.7	80.2
50	78.5	76.9	79.3	82.9	79.6	79.8	80	81.9	83.9	83.1	84.3	83	82.2	81.9	82.4	83.8
60	81.9	85.7	82.8	86	82.9	84.2	83.5	84.4	87.7	85.9	88.5	90.6	85.7	84.3	86.1	85.2
70	85.7	88.8	86.7	90	86.4	87.8	87.3	84.8	91.8	91	92.9	94.1	89.4	88.6	90.1	88.7
80	90	92.1	91.2	91	90.5	89.1	91.7	86.9	96.6	96.3	98.1	97.9	93.8	92.6	94.7	93.7
90	96.1	94.8	97.4	106.8	96.3	93.2	97.8	93.4	103.3	107.1	105.3	105.9	99.8	103.1	101.1	97.2

Table 5: Predicted quantiles sections 16,17,18,21

		Secti	on 16			Secti	on 17			Secti	on 18			Sect	ion 21	
quantile	di	r 1	di	r 2	dir	1	dir	2	dii	:1	dir	2	dir	1	dii	r 2
	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.
10	65	71.7	64.4	71.3	62.4	66.2	64.2	67.8	51.4	51.9	51.2	52.6	85.8	73	86.2	81.7
20	71.8	75.2	71.1	74.6	69.1	69.2	70.3	70.9	57.8	54.2	57.8	56.6	89.3	88	90.3	89.2
30	76.6	80.2	76	76.3	74	72.9	74.8	74.4	62.4	58.4	62.6	60.9	91.9	92.5	93.2	95.3
40	80.8	82.5	80.2	79.6	78.1	74.5	78.6	77.4	66.3	65.3	66.7	62.2	94	97	95.6	98.4
50	84.7	85	84	82.2	82	76.9	82.2	80	70	67.2	70.5	68.5	96.1	98	98	104
60	88.6	88.3	87.9	84.4	85.9	85.4	85.7	83.3	73.7	69.7	74.3	71.8	98.1	99	100.3	105
70	92.7	90.5	92.1	91.5	90.1	92.3	89.6	88.7	77.6	74.9	78.4	75.4	100.3	99.5	102.8	105
80	97.6	93.5	96.9	94.3	95	96.3	94	92.4	82.2	80.3	83.1	84.9	102.8	100	105.7	111.8
90	104.3	108.6	103.7	100.2	101.7	99.3	100.2	94.1	88.6	91.2	89.7	91.7	106.3	105	109.7	114.9

		Sectio	on 22			Secti	on 23			Secti	on 24			Secti	on 25	
quantile	di	r 1	dir	2	dir	1	dir	2	dir	1	dir	2	dir	1	dir	2
	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.
10	60.8	64.8	56.3	64	45.5	39.1	43.7	46	63.6	67	66.6	59.7	53.7	54	55.5	50.4
20	68	69	64	70	48.3	42.4	47.6	47	69.1	68.6	71.2	66	58.7	56.2	60.5	55.2
30	73.2	71.4	69.6	71	50.3	46.1	50.5	48	73.1	69	74.6	75.3	62.2	59.8	64.1	65.4
40	77.6	78	74.4	72	52	46.8	52.9	50	76.4	72.4	77.5	77.4	65.3	62.6	67.2	71.4
50	81.8	82	78.9	72	53.6	47.5	55.2	50.5	79.6	74	80.1	79	68.1	66	70.1	73
60	86	94	83.3	75	55.2	48.2	57.4	51.8	82.7	79.4	82.8	81.8	71	68.6	73	78.4
70	90.4	97.8	88.1	76	57	48.9	59.8	54.1	86.1	81	85.7	88.1	74	71	76	81.2
80	95.6	101	93.7	79	59	50.2	62.7	57.4	90	86	89	90.2	77.6	75	79.7	84.2
90	102.8	108.4	101.5	82	61.7	51	66.6	61.5	95.5	99	93.7	96.1	82.5	75	84.7	89

Table 6: Predicted quantiles sections 22,23,24,25

Table 7: Predicted quantiles sections 26,27,28,29

		Section	on 26			Secti	on 27			Sect	ion 28			Secti	on 29	
quantile	dir	· 1	dir	2	dir	· 1	dir	2	dir	1	dii	2	di	r 1	di	r 2
	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.
10	57.8	59.4	60.8	57.2	34.8	44.6	35.4	45.8	66.5	67	68.3	76.6	68.2	72.3	67.2	78.5
20	63.5	65.6	65.4	60.6	40.4	46.2	41	47.6	73.3	75	74.7	78	75.2	77.1	74.8	80.1
30	67.5	67	68.8	65.5	44.4	47.8	45	48.4	78.1	77	79.4	83.8	80.2	84	80.3	82.6
40	71	67.6	71.6	68.8	47.9	49.4	48.4	49.2	82.3	78	83.4	86.4	84.5	89.2	85	85.1
50	74.2	73	74.3	76	51.1	51	51.6	51	86.2	87	87.1	87	88.5	93.1	89.4	85.1
60	77.5	75.8	76.9	76.8	54.3	53.6	54.8	53.6	90	87	90.8	87.6	92.6	93.1	93.8	93.7
70	81	77.1	79.8	77.3	57.7	54.2	58.2	55.6	94.2	88.5	94.8	91	96.9	93.1	98.5	101.2
80	85	80.8	83.1	80.4	61.7	55	62.2	56.8	99	89	99.4	96.8	101.9	94.9	103.9	103.9
90	90.7	86.7	87.7	86.6	67.3	60	67.8	60.2	105.8	103	105.9	100.8	108.9	109.7	111.5	108.8

Table 8: Predicted quantiles sections 30,31,32,33

		Secti	on 30			Secti	on 31			Section	on 32			Secti	on 33	
quantile	dir	1	dir	2	dir	· 1	dir	2	dir	· 1	diı	: 2	dir	· 1	dir	2
	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.
10	62.6	56.3	65.3	57.4	56.8	61.2	60.6	67.2	46.2	52.1	47.4	55.2	54	52.7	56.5	58.2
20	64.2	56.3	66.5	62.1	62.9	66.5	65.6	68.9	52.9	57.5	53.7	59.2	58.8	58.1	60.6	60.9
30	65.3	57.5	67.4	62.1	67.2	68	69.3	70	57.7	59.6	58.3	62.2	62.2	60	63.6	64
40	66.3	61.4	68.1	63.8	70.9	70.3	72.3	73	61.9	59.6	62.2	63.9	65.2	62.7	66.1	67.4
50	67.3	65.9	68.8	71.8	74.4	71.9	75.2	74.4	65.7	64.2	65.8	67.4	67.9	66.2	68.5	67.4
60	68.2	69.3	69.5	74.1	77.9	73.2	78.1	77.1	69.6	66.7	69.4	69.3	70.7	67.6	70.9	67.4
70	69.2	69.3	70.2	74.1	81.6	74.6	81.2	79.6	73.7	69.5	73.3	71.4	73.6	71.1	73.4	71.1
80	70.3	87.2	71.1	79.2	85.9	76.9	84.8	83.4	78.5	79.4	77.9	73.5	77.1	73.8	76.4	71.9
90	71.9	93.2	72.3	79.2	92	84.1	89.8	87.5	85.2	83.4	84.2	77.3	81.9	81.3	80.5	75.3

		Secti	on 34			Secti	on 36	
quantile	dir	1	dir	2	dir	1	dir	2
	pred.	obs.	pred.	obs.	pred.	obs.	pred.	obs.
10	45.6	42.5	43.9	43.1	45.3	44.6	46.5	56.7
20	49.3	46.5	48.7	50.6	53.1	50.9	53.8	59.1
30	52	49.1	52.2	52.7	58.8	59.2	59.1	62.2
40	54.3	51.2	55.2	56.2	63.6	64.2	63.6	65.2
50	56.5	53.2	57.9	58.8	68.1	68.2	67.8	69.5
60	58.6	54.7	60.7	60.2	72.6	70.9	72	72.4
70	60.9	58	63.7	64.9	77.5	74.9	76.5	76.8
80	63.6	62.2	67.2	70.3	83.1	79.3	81.8	78.8
90	67.3	68.4	72	76	91	94.6	89.1	82.5

Table 9: Predicted quantiles sections 34,36

When analyzing Tables 2 to 9, a number of patterns can be observed. First, the speed 275 quantiles of most sections are predicted accurately, which attests the validity of our statistical 276 methodology. For example, most of the sections have predictions that are within an interval 277 of 2-3 kilometer per hour (km/h) and for almost all the quantiles. Most central deciles (40^{th}) 278 to 60^{th} quantiles) have very good predictions and therefore it is possible to say that additional 279 observations collected to predict the REs can be successfully used to predict the mean speed 280 of the section with some accuracy. The worst value obtained for the predicted median (50^{th}) 281 quantile) is the one related to section 29, direction 1, with an observed median of 93.1 km/h 282 and a prediction of 88.5 km/h. This compares with the 10^{th} quantile that predicts 68.2 km/h 283 for an observed value of 72.3 km/h. 284

Second, some sections are likely to have an error in part due to the prediction error of the 285 random effects. One way to assess this is to observe the extreme quantiles of those sections 286 for which both the 10^{th} and the 90^{th} quantiles are underestimated. Example of this type is 287 section 22, direction 1. Looking back at the plots of predicted REs in figure 3, it is possible 288 to observe that the direction effects are not very precise with only five observations and this 289 is likely a case where the prediction has created a small error. For section 23 (figure 4) the 290 section effect and one of the direction effect are overestimated with five observations, which 291 would explain this component of the prediction error. 292

Third, the most obvious prediction error is the overestimation of low deciles and the under-293 estimation of high deciles, or vice-versa. This can be observed for example in section 10, 294 direction 1 for which the 10^{th} quantile is underestimated by 2.6 km/h while the 90^{th} quantile 295 is overestimated by 8.3 km/h. There might still be an error in the predicted RE for this 296 section but it cannot be fixed because the RE is a constant added to all predictions in the 297 same section and direction. The most likely cause for this kind of error lies in the estimated 298 coefficients. This possible source of error will be investigated in the next section using a 299 re-sampling technique that will determine if data from a particular section has a relevant 300 effects on the values of the coefficients estimated. 301

302 5 Jackknife coefficients

A jackknife re-sampling technique (Efron and Tibshirani 1993) is adopted to explore the poor predictions obtained for some combinations of sections and directions; we are particularly interested in predictions of high and low speed quantiles. We have already discussed in the previous section how to compute out of sample predictors for REs. However, the prediction of speed quantiles in a new section also requires the model's coefficients to be estimated for the validation sample.

The use of a jackknife procedure is based on the following observations: (1) the modeling approach suggested in this paper involves the fitting of a model on a set of road sections and the application of it to another section for which data is not available; (2) some sections are poorly predicted by the suggested model but (3) the majority of sections are predicted with ³¹³ remarkable accuracy.

We are specifically interested in comparing the jackknife coefficients obtained by excluding 314 a generic section s and the quality of the forecasts for that section. We hypothesize that 315 significant discrepancies among the values predicted and observed indicate the existence of 316 one or more of the following problems: (i) inaccuracies in the data, (ii) errors in data collection 317 (perhaps these sections have been sampled by operators who did not get proper training, 318 and/or did not assess accurately if the vehicle was traveling under free flow conditions), (iii) 319 different driving conditions might alter motorists' behavior, (iv) the omission of important 320 covariates that were overlooked and that affect some sections. These scenarios would generally 321 cause some sections to be poorly modeled by the approach suggested in this paper and the 322 assessment of such conditions is expected to greatly simplify the investigation task as the 323 model's user validates his data. 324

In a model that is appropriate for its data, it is expected that the exclusion of any observation 325 does not change the estimates of the model coefficients in a relevant way. The re-sampling 326 scheme suggested discards a substantial number of observations. This way to proceed makes 327 it impossible to directly use the standard re-sampling literature to quantify the effect of a 328 specific coefficient, and ultimately to identify the specification issue. Furthermore, the use 329 of such a criteria would involve a different threshold for each section due to the variability 330 of their size. Simulations could be used to derive a more systematic identification method 331 for sections requiring attention, however this is outside the scope of our analysis and we rely 332 instead on a qualitative evaluation of the effect of each section on coefficients' estimates. 333

Tables 10, and 11 in Appendix A present the jackknife coefficients used for speed data 334 validation and should be read together. Each line corresponds to estimates obtained by 335 removing one section, except the first one that contains the estimations on the whole sample. 336 Each column provides the estimated coefficients for a specific predictor. The cells in bold 337 indicate the coefficients that are very different from the full sample equivalent (outside two 338 standard deviations). Ideally, the estimated coefficients in a given column would be stable and 339 comparable to the coefficient estimated using the entire sample. Large differences between 340 a coefficient estimated by excluding one section and those obtained with the whole sample 341 raise concerns about the specific section or the predictor. 342

When reading the results, it can be observed that the jackknife coefficients for Z*SRW are 343 very similar across the thirty one jackknife samples and are very similar to the coefficients 344 estimated on the overall sample, whereas the coefficients for Z*1/R appear to be off only 345 for sections 7 and 10. Sections 1, 3, 7, and 36 generate some of the extreme jackknife 346 coefficients. We can probably conclude that these sections (1,3,7, and 36) behave according 347 to a different model, or that the measurement of predictors and speed data was less reliable 348 on these sections. This is also confirmed in the comparison between predition and observation 349 reported in the Tables 2 to 9. 350

351 6 Conclusions

In this paper we have explored methods to predict speed distributions using mixed linear models. Difficulties associated with this problem are twofold. First we rely on model with random effects to make prediction on new road sections. Second, we are using normal quantiles of the dependent variables as predictors.

We have discussed that it is necessary to observe at least some speed data in the section 356 for which decile predictions were computed. To overcome the problem created by the use of 357 normal quantiles in the calculation of residuals, we propose the use of an auxiliary model. It 358 has been shown how the relationship between this auxiliary model and the full actual model 359 can be used in prediction in order to derive the best linear unbiased predictor (BLUP) in 360 this context. It was also observed that this method was not performing well when used to 361 make random effect predictions for a model that ignored the sampling design of the data, 362 but turned out to be very precise when the full sampling design was accounted in the model. 363

The approach was further tested to validate the models' coefficients associated with the variables of the model. The estimates obtained were roughly stable except for some variables that generated more extreme coefficients for some sections. The sections with extreme coefficients were consistently the same for all the affected variables.

A jackknife technique was used to understand what road sections caused poor predictions of the random effects. It was found that most road sections were predicted satisfactorily. Large errors in the prediction of random effect were mostly caused by errors in the coefficients of the model. Some sections appeared to have a disproportionate effect on the model, suggesting that they should be modeled in a different way.

Overall, this paper has contributed to the transferability of RE models, has identified the problems arising in the estimation of Operating Speeds, has developed a theory to calculate the best predictors and to validate the results obtained. Future efforts should be directed towards the use of the method proposed in practice and possibly to different model types that include REs.

- 378
- 379

380 ACKNOWLEDGEMENTS

The research work included in this paper was made possible thanks to the joint project between the Politecnico di Torino (Italy) and the University of Maryland (US). The authors acknowledge funding received from the Compagnia di San Paolo (Italy) and from the Politecnico di Torino (Rectoral Decree n. 208 of the 24th of May, 2013). The Città Metropolitana di Torino, and the Provinces of Vercelli and Alessandria are greatly acknowledged for the access to the GIS databases used to collect roadway inventory information.

³⁸⁷ Appendices

Table 10: Jackknife coefficients - Part 1

Г

section out	Intercept	PedD	$1/\mathrm{R}$	$\mathbf{Z^{*}Lane}$	$\rm Z^{*1}/R$	STS*Z	Z*SBLS	Z*IDRS	Z*SRS	Z*SBRS	Z*SRW	MIS*Z
none	79.33	-7.59	-1946.46	20.54	-274.07	-3.46	-1.56	-0.83	-2.88	-0.64	-13.38	13.85
1	79.8	-7.77	-2042.39	25.07	-227.05	-1.95	-2.82	-0.91	-0.79	-1.36	-13.62	13.09
3	79.51	-7.67	-1983.47	22.96	-170.71	-2.49	-1.19	-0.97	-2.03	-0.31	-12.56	14.29
5	79.48	-7.65	-1976.76	19.71	-279.35	-3.38	-2.24	-0.95	-3.1	0.12	-13.41	13.74
7	79.21	-7.55	-2277.82	21.89	-289.06	-3.45	-0.93	-0.83	-2.72	-0.47	-13.36	14.13
8	79.09	-7.5	-1896.77	20.86	-283.21	-3.44	-1.44	-0.83	-2.86	-0.57	-13.36	13.9
9	79.3	-7.58	-1943.43	20.46	-274.26	-3.48	-1.64	-0.84	-2.88	-0.59	-13.4	13.87
10	79.2	-7.54	-1208.28	21.33	-248.33	-3.31	-1.48	-0.81	-2.82	-0.54	-13.34	13.78
11	79.53	-7.67	-1987.44	19.97	-267.14	-3.45	-1.5	-0.8	-2.88	-0.71	-13.37	13.85
12	79.28	-7.57	-1936.44	18.9	-271.29	-3.47	-1.53	-0.82	-2.94	-0.66	-13.37	13.89
13	79.33	-7.59	-1946	21.35	-283.83	-3.46	-1.58	-0.84	-2.88	-0.58	-13.39	13.85
14	79.12	-7.51	-1901.82	20.27	-274.76	-3.45	-1.57	-0.83	-2.91	-0.61	-13.36	13.84
15	79.22	-7.55	-1922.98	20.74	-273.66	-3.45	-1.59	-0.86	-2.88	-0.61	-13.38	13.85
16	79.12	-7.51	-1901.59	21.76	-266.84	-3.43	-1.58	-0.87	-2.83	-0.64	-13.38	13.82
17	79.23	-7.55	-1924.69	20.26	-270.36	-3.46	-1.6	-0.86	-2.88	-0.62	-13.38	13.85
18	79.68	-7.72	-2017.04	21.32	-276.61	-3.41	-1.49	-0.82	-2.84	-0.56	-13.4	13.84
20	78.42	-7.24	-1757.97	19.7	-278.04	-3.52	-1.65	-0.83	-2.93	-0.72	-13.37	13.87
21	78.71	-7.36	-1819.11	21.88	-244.19	-3.32	-1.37	-0.9	-2.78	-0.51	-13.39	13.82
22	79.33	-7.59	-1944.68	20.45	-274.83	-3.46	-1.5	-0.81	-2.94	-0.7	-13.37	13.89
23	80.41	-8	-2168.41	20.52	-276.31	-3.46	-1.55	-0.83	-2.88	-0.64	-13.38	13.86
24	79.33	-7.59	-1945.62	20.49	-273.89	-3.55	-1.54	-0.82	-2.86	-0.66	-13.4	13.84
25	79.73	-7.74	-2028.57	20.52	-273.66	-3.47	-1.57	-0.83	-2.89	-0.66	-13.39	13.85
26	79.58	-7.69	-1997.16	20.52	-273.93	-3.59	-1.46	-0.79	-2.74	-0.71	-13.38	13.88
27	79.48	-8.74	-1976.04	20.47	-269.15	-3.5	-1.59	-0.86	-2.83	-0.59	-13.35	13.85
28	79.09	-7.5	-1895.65	19.77	-272.6	-3.5	-1.67	-0.85	-2.91	-0.7	-13.36	13.85
29	78.93	-7.44	-1863.92	20.07	-267.62	-3.49	-1.65	-0.85	-2.93	-0.73	-13.34	13.86
30	79.37	-7.56	-1954.05	19.1	-214.59	-4.27	-2.05	-0.99	-3.53	-1.15	-13.43	13.63
31	79.52	-7.66	-1983.91	20.34	-256.36	-3.1	-1.63	-0.87	-2.74	-0.71	-13.26	13.93
32	79.47	-7.43	-1973.31	21.44	-295.67	-3.71	-1.6	-0.7	-3.11	-0.74	-13.36	13.82
33	79.78	-7.76	-2037.72	21.32	-276.67	-3.33	-1.5	-0.81	-3.05	-0.6	-13.39	13.83
34	79.34	-7.66	-1947.25	19.84	-275.81	-3.59	-1.67	-0.79	-3.01	-0.76	-13.35	13.87
36	79.39	-7.52	-1958.78	17.03	-257.65	-4.26	-2.33	-0.91	-3.62	-1.11	-13.31	13.82

section out	ST4*Z	Z*DDRS	Z*TRDLS	Z^*LW	Z*IDLS	Z*LG	$Z^*\Delta PSL$	Z^*PedD	Z*TRDRS	Z^*LBRS	Z*LBLS
none	0.06	0.43	-0.29	-3.52	-0.43	0.07	0.07	1.62	0.41	1.31	1.16
1	0.04	0.26	-0.78	-4.16	-0.3	0.09	0.19	0.11	0.03	2.57	2.52
3	0.16	0.6	-0.82	-6.73	-0.52	0.06	0.03	2.13	-0.3	1.03	1.25
5	0.07	0.45	-0.27	-3.37	-0.4	0.11	0.06	1.71	0.34	1.53	0.87
7	0.07	0.44	-0.86	-4.13	-0.46	-0.08	0.05	1.66	0.04	1.67	1.09
8	0.06	0.43	-0.4	-3.67	-0.43	0.06	0.07	1.62	0.31	1.36	1.17
9	0.06	0.42	-0.31	-3.48	-0.42	0.05	0.07	1.62	0.42	1.33	1.13
10	0.06	0.44	-0.25	-3.77	-0.49	0.13	0.07	1.58	0.5	1.19	1.17
11	0.06	0.43	-0.27	-3.37	-0.49	0.08	0.07	1.63	0.4	1.32	1.12
12	0.06	0.47	-0.24	-3.17	-0.42	0.07	0.08	1.61	0.4	1.35	1.26
13	0.06	0.42	-0.3	-3.75	-0.39	0.07	0.07	1.59	0.4	1.3	1.18
14	0.06	0.44	-0.26	-3.48	-0.42	0.09	0.07	1.6	0.39	1.3	1.22
15	0.06	0.43	-0.29	-3.58	-0.41	0.07	0.07	1.62	0.41	1.29	1.18
16	0.06	0.39	-0.32	-3.78	-0.46	0.07	0.07	1.65	0.41	1.26	1.06
17	0.06	0.42	-0.28	-3.44	-0.42	0.07	0.07	1.63	0.41	1.29	1.17
18	0.06	0.44	-0.3	-3.83	-0.42	0.07	0.08	1.59	0.37	1.39	1.27
20	0.06	0.4	-0.28	-3.18	-0.43	0.07	0.07	1.63	0.44	1.24	1.03
21	0.06	0.46	-0.46	-4	-0.51	0.07	0.07	1.67	0.17	1.38	1.3
22	0.06	0.45	-0.27	-3.53	-0.45	0.06	0.07	1.6	0.41	1.41	1.08
23	0.06	0.42	-0.3	-3.5	-0.43	0.07	0.07	1.61	0.39	1.31	1.14
24	0.06	0.41	-0.3	-3.5	-0.45	0.07	0.07	1.67	0.41	1.34	1.11
25	0.06	0.43	-0.28	-3.52	-0.43	0.06	0.07	1.62	0.42	1.33	1.16
26	0.06	0.44	-0.29	-3.53	-0.48	0.06	0.07	1.59	0.39	1.38	1.08
27	0.06	0.36	-0.33	-3.5	-0.44	0.07	0.06	2.12	0.46	1.33	1.03
28	0.06	0.4	-0.28	-3.19	-0.44	0.08	0.07	1.65	0.43	1.23	1.03
29	0.06	0.43	-0.12	-3.43	-0.45	0.07	0.08	1.65	0.57	1.28	1.14
30	0.05	0.34	-0.37	-2.7	-0.53	0.06	0.1	1.99	0.43	1.44	1.18
31	0.06	0.46	-0.27	-3.39	-0.46	0.07	0.08	1.38	0.38	1.21	1.12
32	0.06	0.38	-0.33	-3.75	-0.41	0.08	0.08	1.76	0.38	1.6	1.13
33	0.06	0.43	-0.31	-3.81	-0.43	0.07	0.08	1.63	0.37	1.41	1.24
34	0.06	0.41	-0.3	-3.19	-0.44	0.07	0.08	1.25	0.39	1.28	1.1
36	0.04	0.43	-0.35	-1.76	-0.38	0.1	0.1	1.81	0.28	0.82	1.29

Table 11: Jackknife coefficients - Part2

References

Bassani, M., L. Catani, C. Cirillo, and G. Mutani (2016). "Night-time and daytime operating speed distribution in urban arterials." In: *Transportation Research Part F: Traffic Psychology and Behaviour* 42.1, pp. 56–69.

Bassani, M., C. Cirillo, S. Molinari, and J.M. Tremblay (2016). "Random effect models to
 predict operating speed distribution on rural two-lane highways." In: *Journal of Trans- portation Engineering* 142.6, p. 04016019.

- Bassani, M., D. Dalmazzo, G. Marinelli, and C. Cirillo (2014). "The effects of road geometrics and traffic regulations on driver-preferred speeds in northern Italy. An exploratory analysis." In: Transportation Research Part F: Traffic Psychology and Behaviour 25, pp. 10– 26.
- Catani, L., J.M. Tremblay, M. Bassani, and C. Cirillo (2017). "Methodology to backcalcu late individual speed data originally aggregated by road detectors." In: *Transportation Research Record: Journal of the Transportation Research Board* 2659.1, pp. 1–14.
- Cheng, W., G.S. Gill, T. Sakrani, D. Ralls, and X. Jia (2018). "Modeling the endogeneity
 of lane-mean speeds and lane-speed deviations using a bayesian structural equations approach with spatial correlation." In: *Transportation Research Part A: Policy and Practice*116, pp. 220, 221
- 405 116, pp. 220–231.
- ⁴⁰⁶ Dimaiuta, M., E. Donnell, S.C. Himes, and R. J. Porter (2011). "Modeling operating speed."
 ⁴⁰⁷ In: *Transportation Research E-circular* E-C151.
- ⁴⁰⁸ Efron, B. and R.J. Tibshirani (1993). An Introduction to the Bootstrap. Chapman et Hall.
- Figueroa-Medina, A. and A.P. Tarko (2005). "Speed factors on two-lane rural highways in
 free-flow conditions." In: *Transportation Research Record: Journal of the Transportation Research Board* 1912, pp. 39–46.
- Garber, N. and L. Hoel (2020). *Traffic and highway engineering*. 5th Edition. Cengage Learning.
- Himes, S., E. Donnell, and R. Porter (2013). "Posted speed limit: To include or not to include
 in operating speed models." In: *Transportation Research Part A: Policy and Practice* 52,
 pp. 23–33.
- Islam, M.T. and K. El-Basyouny (2015). "Multilevel models to analyze before and after speed
 data." In: Analytic Methods in Accident Research 8, pp. 33–44.
- Lamm, R., E.M. Choueiri, J.C. Hayward, and Paluri. A. (1988). "Possible design procedure to promote design consistency in highway geometric design on two-lane rural road." In: *Transportation Research Record: Journal of the Transportation Research Board* 1195, pp. 111–122.
- McCulloch, C.E., S.R. Searle, and J.M. Neuhaus (2008). Generalized, Linear, and Mixed
 Models. Second. Hoboken, New-Jersey: John Wiley & Sons.
- ⁴²⁵ McFadden, J. and L. Elefteriadou (2000). "Evaluating horizontal alignment design consis-⁴²⁶ tency of two lane rural highways: development of new procedure." In: *Transportation*
- 427 Research Record: Journal of the Transportation Research Board 1737, pp. 9–17.

- ⁴²⁸ Park, Y.J. and F.F. Saccomanno (2006). "Evaluating speed consistency between successive
- elements of a two-lane rural highway." In: *Transportation Research Part A: Policy and Practice* 40.5, pp. 375–385.
- Shankar, V. and F. Mannering (1998). "Modeling the endogeneity of lane-mean speeds and
 lane-speed deviations: a structural equations approach." In: *Transportation Research Part*
- ⁴³³ A: Policy and Practice 32.5, pp. 311–322.
- 434 Tarris, J., C. Poe, J. Mason, and K. Goulias (1996). "Predicting operating speeds on low-speed
- ⁴³⁵ urban streets: Regression and panel analysis approaches." In: *Transportation Research Pagendi*: *Journal of the Transportation Research Pagend* 1522, pp. 46–54
- Record: Journal of the Transportation Research Board 1523, pp. 46–54.