Predicting Operating Speed, Linear Regression versus Structural Equation Models

Author 1 (**Corresponding author**)

- Behzad Bamdad Mehrabani, Ph.D. candidate
- Faculty of Architecture, Architectural Engineering and Urban Planning, Université catholique de Louvain, Tournai, Belgium, E-mail: behzad.bamdad@uclouvain.be
- ORCID: 0000-0001-8585-7879

Author 2

- Babak Mirbaha, Associate Professor
- Civil Engineering Department, Imam Khomeini International University, Qazvin, Iran, E-mail: mirbaha@eng.ikiu.ac.ir
- ORCID: 0000-0003-1004-7673

Author 3

- Luca Sgambi, Assistant Professor
- Faculty of Architecture, Architectural Engineering and Urban Planning, Université catholique de Louvain, Tournai, Belgium, E-mail: luca.sgambi@uclouvain.be
- ORCID:0000-0002-4132-2087

Author 4

- Ali Abdi Kordani, Associate Professor
- Civil Engineering Department, Imam Khomeini International University, Qazvin, Iran, E-mail: aliabdi@eng.ikiu.ac.ir
- ORCID: 0000-0003-3175-2566

Full contact details of corresponding author.

- E-mail : behzad.bamdad@uclouvain.be
- Tel. +32 497 29 22 14
- Address: Rue du Glategnies 6, 7500 Tournai

Number of words in the main text: 5073

Number of tables: 7

Number of figures: 7

Abstract

The present research aims to predict operating speed by considering geometric and roadside factors. Although most previous studies have employed Linear Regression Modelling (LRM) to predict operating speed, this study recommends Structural Equation Modelling (SEM) to the prediction of operating speed in rural multi-lane highways. In addition to geometric variables, LRM takes roadside variables into account. When employing SEM, two latent variables, namely “roadside effects” and “geometric effects” were defined. The first latent variable is the combination of land-use type, land-use density, and access per segment, while the latter is extracted from segment length, highway grade, curvature, the presence of guardrail and flat roadside slope, and the posted speed limits. The residual analysis and R-squared values for LRM and SEM suggest that SEM demonstrates superior modelling performance compared to LRM. These results show the significant role of latent variables in predicting speed, which cannot be achieved through ordinary LRM.

Keywords

Traffic engineering, Roads & highways, Transport planning

List of notations

Y_i	dependent variable
$\beta_0,\beta_1,\ldots,\beta_n$	model coefficients
X_i	independent variables
ε	error term

1. Introduction

Operating speed is a significant traffic parameter that explains the effects of the geometric design of roads and roadside factors on driver behaviour. Operating speed is the speed at which drivers are observed operating their vehicles during free-flow conditions. The 85th percentile of the distribution of observed speeds is the measure of the operating speed. Operating speed has various applications in road traffic and safety studies. Predicting the safety conditions of a particular highway depends mainly on operating or average speed (Gargoum and El-Basyouny 2016).

Numerous studies have attempted to predict the operating speed of rural highways. The majority of such studies have aimed to predict operating speed in rural two-lane highways (Yagar and Van Aerde 1983; Abbas et al. 2011; Esposito et al. 2011; Bella 2013; Garcia et al. 2013; Bassani et al. 2015; Boroujerdian et al. 2016; Hashim et al. 2016; Shallam and Ahmed 2016; Eboli et al. 2017) rather than multi-lane highways (Gong and Stamatiadis 2008; Himes and Donnell 2010; Semeida 2013; Gargoum and El-Basyouny 2016). Previous studies have demonstrated that operating speed in rural roads is determined by various factors, including geometric, traffic, and roadside parameters (Park and Saccomanno 2006; Bella 2013; Garcia et al. 2013; Boroujerdian et al. 2016). Most such studies have addressed geometric factors (Park and Saccomanno 2006; Gong and Stamatiadis 2008; Boroujerdian et al. 2016; Eboli et al. 2017; Sil et al., 2020). Moreover, several studies have concentrated on curves while neglecting tangents. (Park and Saccomanno 2006; Gong and Stamatiadis 2008; Abbas et al. 2011; Hassan and Sarhan 2011; Echaveguren et al. 2015; Shallam and Ahmed 2016).

On the other hand, the majority of previous studies have only considered observed variables and overlooked the effects of latent variables on operating speed and endogeneity issues. According to Schumacker and Lomax (2004), Latent variables can only be observed or measured indirectly. Hence, they are regarded as constructs inferred from observed variables. Structural Equation Modelling (SEM) is a method for estimating latent variables (specified as a linear combination of observed variables) impacts. The model can handle both endogenous and exogenous variables and help investigate the direct and indirect impacts of exogenous variables on endogenous variables in causal systems.

Therefore, the present research explores two distinct approaches to modelling in order to predict operating speed in multi-lane highways. The first approach attempts to provide a prediction by employing Linear Regression Modelling (LRM), which only considers observed variables. The second approach uses SEM to investigate the role of latent variables on operating speed modelling. Both approaches to modelling have taken geometric and roadside factors into account. The results obtained through the two methods are compared in order to determine whether considering latent variables and endogeneity relationships in modelling operating speed improves prediction results and fit indices.

2. Literature Review

Research on operating speed can be divided into two groups based on the type of effective variables: studies based on geometric and traffic factors and studies based on roadside factors. It should be noted that two-lane and multi-lane highways are both considered in the literature review due to the existence of common variables that exert an influence on both road types.

2.1. Studies Based on Geometric and Traffic Factors

2.1.1. Research on Curve

Although numerous studies had been conducted in order to predict operating speed in rural two-lane and multi-lane highways, most such studies present models for horizontal curves (Gong and Stamatiadis 2008; Abbas et al. 2011; Hassan and Sarhan 2011; Garcia et al. 2013; Shallam and Ahmed 2016). K. Shallam and M. Ali Ahmed (2016) only considered radius, deflection angle (i.e. the angle at which the preceding and succeeding tangents of the curve meet) and sight distance to predict operating speed. Others such as P. Garcia et al. (2013) include variables such as longitudinal grade, side of the curve (left curve or right curve), and grade direction (upgrade or downgrade) in addition to the radius. Another study by Hassan and Sarhan (2011) suggests that different methodologies such as SEM should be explored due to the limitations of regression (Hassan and Sarhan 2011).

2.1.2. Research on Tangent

Another group of studies has only considered geometric factors in order to predict the operating speed of tangents. For instance, Boroujerdian et al. (2016) investigated the relationship between the mean and standard deviation of speed and geometric features. They presented separate models for downgrade and upgrade tangents and determined that the initial speed of vehicles at the starting point of tangents and grades was the most effective variable in both downgrade and upgrade speed prediction models. Esposito et al. (2011) presented three classes of models for tangents and curves, as well as a single model applicable to both. Geometric variables such as segment length, curvature change rate, grade, and radius were considered in this study.

Linear and regression models are the most common methodologies employed by previous studies (Yagar and Aerde, 1983; Fitzpatrick et al., 1997; Fitzpatrick et al., 2001; Gong and Stamatiadis, 2008; Esposito et al., 2011; Dell'Acqua and Russo, 2011; Abbas et al., 2012; P. Garcia and T. Nodari, 2013; K. Shallam and M. Ali Ahmed, 2014; H. Hashim et al., 2016; Boroujerdian et al., 2016; Eboli et al., 2017). However, few studies have employed other methods such as SEM (Shankar and Mannering, 1998; A. Gargoum and El-Basyouny 2016; Sadia et al. 2017; Cheng et al., 2018), Artificial Neural Networks (ANN) (Semeida 2013; Ke et al., 2019), and panel data analysis (Figueroa Medina and Tarko 2005; Cruzado and Donnel 2011).

Some studies have employed SEM to investigate endogeneity issues between lane mean speed and lane speed deviation. However, such studies have not considered geometric and roadway characteristics simultaneously (Shankar band Mannering 1998; Himes and Donnell, 2010; Cheng et al., 2018). For instance, Himes and Donnell (2010) used simultaneous equations to model operating speed in multi-lane highways. This study has taken both geometric design and traffic flow parameters into account. The research investigates the endogenous relationship between the lane mean speed and deviations from the lane mean speed. The results indicate that there is an endogenous relationship between the mean speed and deviations from the mean speed.

Using SEM, A. Gargoum and EL-Basyouny (2016) explored the relation between speed and safety, concluding that average speed, volume, segment length, medians, and horizontal curves have statistically

significant effects on the frequency of collisions. Moreover, shoulder, bus stops, posted speed limits, segment length, volume, and average vehicle length significantly influenced driving speeds. It should be noted that their research only takes observed variables into account in order to model operating speed and the frequency of collisions. However, Sadia et al. (2017) conclude that latent variables have a significant effect on the selected speed in four-lane highways.

2.2. Studies Based on Roadside Factors

Few studies have evaluated the effects of roadside factors on operating speed (Yagar and Van Aerde, 1983; Fitzpatrick et al., 2001; Figueroa Medina and Tarko, 2005; Dell'Acqua and Russo, 2011). For example, Yagar and Van Aerde (1983) showed that land use and posted speed limits have the most significant impact on the speed in rural two-lane highways. However, the land-use type was not considered in their study. Using a regression model, Fitzpatrick et al. (2001) examined different types of land use in their research on four-lane suburban arterial. They identified the median presence and roadside development as having a significant impact on vehicle speeds on curve segments. Figueroa Medina and Tarko (2005) estimated free-flow speed on tangent and curve segments on two-lane rural highways. They identified the number of driveways per mile, shoulder width, sight distance, lateral clearance, etc. as independent variables; however, they did not consider land use.

The NCHRP report 504 (Fitzpatrick et al. 2003) explores the relationship between operating speed, design speed, and posted speed limits. The authors recommend that future studies incorporate all of the geometric and roadside variables (such as: posted speed limits, functional class, access density, pedestrian activity, centreline marking, edge line marking, parking, median type, the distance between controls, shoulder width, signal density, total pavement width, median width, roadside environment, lane width, and roadside development) as independent variables and adopt SEM as their research methodology.

Many studies have been conducted to model operating speed in two-lane highways, and a variety of variables such as grade, segment length, posted speed limits, land use, traffic flow, etc., have been considered by such studies. However, not all the variables have been considered simultaneously by any of

the studies. Moreover, only a few researchers, such as Esposito et al. (2011) and Bassani et al. (2015) have presented a general model of tangent and curve.

On the other hand, while previous studies have suggested that latent variables play a crucial role in speed prediction (Washington et al. 2010; Sadia et al. 2017; Bamdad Mehrabani and Mirbaha, 2019; Bhowmik et al., 2019), the simultaneous impact of latent geometrical and roadside variables on operating speed has been overlooked by previous speed prediction models. For instance, Sadia et al. (2017) have demonstrated the impact of latent variables on speed while Washington et al. (2010) have highlighted the consideration of such variables as essential to transportation studies. The effects of new roadside variables, such as land-use type, on speed prediction have also been overlooked.

Therefore, the current state of the art will be advanced by 1- Evaluating the effects of new roadside variables on operating speed 2- Incorporating both roadside and geometric factors simultaneously for operating speed modelling and 3- Evaluating the effects of latent variables on operating speed. To do so, the authors employ SEM, in addition to LRM, to analyse and predict the operating speed.

3. Data Description

Two highways from the main arterial rural roads of Markazi and Lorestan Provinces in Iran were selected as the case study. As shown in Figure 1, these are two-way four-lane highways (two lanes in each direction) in mountainous terrain that connects Kole Jub to Borujerd city and Borujerd to Khorramabad city. The length of the entire path is 135 km (in both directions). The width of each lane (3.75 m) and the lighting condition were constant along the whole road.

Figure 1: Kole jub – Khoramabad path

To collect speed data, the road should be divided into homogeneous segments. The following characteristics were used as segmentation criteria:

1- Curvature

2- Highway grade

3- Dominant land use

4- Posted speed limit

A new segment was defined where there is a change in at least one of the mentioned characteristics. Considering the specified criteria, the road was divided into 138 segments (108 in tangent and 30 in curve). Following segmentation, a laser gun was employed to collect speed data from segments. The person collecting the data was inside a car safely parked on the roadside without being noticed by the drivers. To avoid distracting the drivers, the data collector person attempted to collect speed data before the vehicles reached the data collection point. In this process, the spot speeds of 100 passenger cars were collected from the middle of each segment. The measured speed is directly related to the cosine of the angle between the laser gun and travel direction. Thus, since the location of the data collector vehicle is fixed, an average cosine effect angle is assumed at each segment, and all of the measured speed values are converted to actual speed values based on the cosine effect angle. The 85^{th} percentile of the spot speeds was considered as operating speed in each segment. Data collection was performed during daylight and under appropriate weather conditions (not rainy, snowy, and cloudy weather condition). Traffic flow was less than 1000 vehicles per hour, and no queue or congestion was observed during the entire data collection process. In total, more than 13800 (100*138) observations were collected.

After collecting spot speed data, the geometrical and roadside factors from all segments in a field survey were gathered using a recording camera and a GPS device. The GPS data was imported into Google Earth, which was used to measure highway grade. Moreover, other geometric factors, such as paved width, were measured in the field. Roadside factors such as land-use types were also extracted by matching video images and GPS data.

The variables of each segment that affect the operating speed of drivers are presented in Table 1. Due to their definition, the variables are divided into continuous and categorical groups.

Table 1. Descriptive Statistics of Variables

4. Modelling Approach

Since most previous studies have used a linear regression model to estimate operating speed, this study employs LRM as one of the approaches to modelling. Moreover, due to traditional regression limitations on considering latent variables and assessing their direct and indirect effects (Washington et al. 2010), and the recommendation of simultaneous equations by previous studies (Fitzpatrick et al. 2003; Hassan and Sarhan 2011), this research employs SEM to predict operating speed. These methods are presented in what follows.

4.1. Linear Regression Model (LRM)

Multiple linear regression modelling was employed to predict operating speed as the dependent variable and roadside and geometric factors as independent variables. The form of multiple regression models is presented by the following equation:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \quad (1)$$

4.2. Structural Equation Model (SEM)

SEM has numerous advantages over regression models. Some of these advantages are as follows (Xiao 2013):

- The simultaneous use of a given parameter as both a predicted and an explanatory variable.
- Use of latent variables: A latent variable is a representative variable that cannot be directly measured or observed.
- Dealing with time-series data.

SEM is a multivariate regression model that allows the researcher to simultaneously measure a set of relations between the measured and latent variables. In other words, this method is a combination of principal factor analysis (PFA) and multivariate regression (Hair et al. 2010). This hybrid model is a combination of the measurement and structural models. Measurement models identify the latent variable measured by observed variables, while structural models identify correlating independent and dependent variables.

The basic equation of the structural model is defined as (De Arcangelis 1993):

$$\eta = B\eta + \Gamma\xi + \zeta \quad (2)$$

The basic equation of the measurement model is as follows:

$$X = \Lambda_x\xi + \delta \quad (3)$$

$$Y = \Lambda_y\eta + \varepsilon \quad (4)$$

The SEM elements are summarized in a measurement model, structural model and covariance matrix. Table 2 defines the SEM elements.

Table 2. Definition of SEM elements

There are several fit indices in previous studies for evaluating the performance of SEM. These indices are grouped into three categories: 1-Absolute fit indices 2- Incremental fit indices and 3- Parsimony fit indices. Brown (2014) recommends that at least one indicator from each category should be reported in order to have a fair evaluation of the SEM. Table 3 presents the indices of these three categories, as well as their cut-off point values.

Table 3. SEM Fit Indices and Their Thresholds (Schumacker and Lomax 2004; Hooper et al. 2008; Hair et al. 2010)

5. Results and Discussion

In this section, the result of both LRM and SEM are presented and discussed:

5.1. LRM Results

Several model specifications (with different roadside and geometric factors) were tested to achieve better fitting results using SPSS 24**.** The best fitted linear regression model is presented in Table 4 for predicting operating speed in tangents and curves. The mean residual for this model is equal to zero, indicating that the data is normal and accurate and that the models are calibrated. In the presented model, various geometric and roadside features are evaluated. All of the variables are statistically significant, and this model has an acceptable R-square (R-square= 0.828).

Table 4. Multivariable Linear Regression Models Results

Considering the model presented in Table 4, segment length [LNM] is statistically significant and has a positive coefficient; as such, it can be concluded that the operating speed is higher in longer segments. This result is consistent with results reported by H. Hashim et al. (2016) and numerous other studies (Esposito et al. 2011; Gargoum and El-Basyouny 2016; Eboli et al. 2017). Other significant variables identified by the present research and other studies are the highway grade [GRD] and posted speed limits [PSL] (Fitzpatrick et al. 2003; Gong and Stamatiadis 2008; Esposito et al. 2011; Bassani et al. 2015; Boroujerdian et al. 2016; Shallam and Ahmed 2016). The negative coefficient of highway grade indicates a decrease in vehicle speeds uphill, and an increase in vehicle speeds downhill. Furthermore, operating speeds increase when posted speed limits are increased.

Moreover, the type of median barrier and roadside slope type are among the factors affecting operating speed. To evaluate the effects of such variables, the dummy variables related to the median (guardrail and no barrier) and the roadside slope type (flat, fore slope, and back slope) were separately incorporated into the model and identified as statistically insignificant. Thus, the dummy variable, which is the multiplication of the median barrier type and roadside slope type dummy variables (the presence of a guardrail and flat roadside slope), was used in the model [MTRSC]. It should be mentioned that Fitzpatrick et al. (2001) found that the median barrier type is a significant variable in predicting operating speed.

According to Fitzpatrick et al. (1997), one of the important roadside variables used in this model is the number of access points in each segment [N.ACCESS]. Because the coefficient of N.ACCESS is negative in the presented LRM, it can be concluded that an increase in the number of access points results in decreased operating speeds in each segment.

A normal Q-Q plot for model 1 is shown in Figure 2. This plot reveals disturbances that lie on the dotted straight line, which, as Washington et al. (2010) indicate, demonstrates that the operating speed data is normally distributed.

Figure 2: Q-Q plot of residuals

5.2. SEM Results

The authors employed SEM in order to investigate the direct and indirect effects of variables, assess the impact of all variables on operating speed, take effective latent variables into account, and achieve superior fit indices. The primary motivation for the adoption of SEM was to investigate the effects of latent variables and determine the impact of all variables on operating speed, on the one hand, and reduce the number of independent variables used in predicting operating speed, on the other hand.

Geometric and roadside factors were regarded as independent variables while the operating speeds of tangents and curves were considered as the dependent variable. A two-steps approach was used to estimate SEM. The first step (measurement model) is concerned with the extent to which various measured exogenous variables measure latent variables. Factor analysis was employed to determine whether there are latent variables that best explain or account for the correlations among the observed variables. Factor analysis was employed to reveal the true nature of the data. Moreover, there was a theoretically motivated reason to suspect that some variables are measuring the same underlying phenomenon, with a subsequent examination of whether the data support this expected underlying measurement model.

Therefore, a variety of variables were considered in order to select the most suitable combination of latent variables. However, variables were kept in the model only if they were statistically significant. it was assumed that roadside variables (namely land-use type [LUR], Land use density [LUD], and access per segment [N.ACCESS]) have a statistically significant correlation, and the linear composition of all of them represents a latent variable (“roadside effects”). A similar assumption was made for geometric variables (segment length [LOGLN], highway grade [GRD], curvature [C], the presence of guardrail and flat roadside slope [MTRSC], and posted speed limits [PSL]). The results of the confirmatory factor analysis revealed that the assumptions are true.

Other roadside and geometric variables (Paved width, paved shoulder width, etc.) do not exist in the composition of latent variables because these variables do not have a statistically significant correlation with other variables (that already exists in the linear composition of latent variables). For instance, paved

width does not measure the same underlying phenomenon that other geometric variables (LOGLN, GRD, C, MTRSC, PSL) are measuring.

Consequently, in the second step (structural model), different model specifications of how these latent variables affect operating speed were investigated. The second step was to identify the latent construct that exerted the highest influence on operating speed (the structural model is concerned with how the model variables are related to one another). Moreover, goodness-of-fit indices (classified into three subcategories) were considered in order to achieve the best-fit model. The process of SEM for the prediction of operating speed is presented in figure 3.

Figure 3: structural equation modelling process for operating speed prediction

The results of SEM are presented in Figure 4 and Table 5. The Goodness of Fit Index (GFI) and Adjusted Goodness of Fit Index (AGFI) are among the evaluated absolute fit indices, which were respectively measured as 0.909 and 0.843, values that exceed acceptable thresholds (0.8). The acceptable range for the Root Mean Square Error of Approximation (RMSEA) index is 0.08–0.1, while in this research, the value was determined to be 0.087. For the incremental fit indices, the Tucker-Lewis index (TLI) and Comparative Fit Index (CFI) were investigated. The closer these values are to 1, the better their fit. The values of these indices for the model proposed in this work are acceptable. The recommended value of the Parsimonious Comparative Fit Index (PCFI) and Parsimonious Normed Fit Index (PNFI) is around 0.5, which are among parsimony fit indices. The obtained values of these indices are also acceptable for the presented model. As can be seen, all of the variables and paths are statistically significant, and the descriptive fit statistics for this model are acceptable based on the fitting thresholds.

Figure 4: Structural equation model for operating speed

Table 5. Structural Equation Model

According to the proposed model, the relationships between the variables are written as follows:

Measurement model of roadside effects:

$$LUR = 1 * roadside\ effects + e2 \quad (5)$$

$$LUD = 0.420 * roadside\ effects + e3 \quad (6)$$

$$N.ACCESS = 0.473 * roadside\ effects + e4 \quad (7)$$

Measurement model of geometric effects:

$$LOGLN = 1 * geometric\ effects + e5 \quad (8)$$

$$GRD = -7.635 * geometric\ effects + e6 \quad (9)$$

$$C = -0.005 * geometric\ effects + e7 \quad (10)$$

$$MTRSC = 1.837 * geometric\ effects + e8 \quad (11)$$

$$PSL = 56.690 * geometric\ effects + e9 \quad (12)$$

Confirmatory factor analysis is used in testing the measurement model, and the hypothesized factors are referred to as latent variables. Latent variables (roadside effects and geometric effects) are operationalized in a reflective way. In other words, the unknown parameters (coefficients and latent variables value) in equations 5-7 and 8-12 are simultaneously estimated using the Maximum Likelihood (ML) estimator. Also, note that the factor loadings for one indicator of each latent variable are set to 1, which is done for two reasons: so that the model is identifiable and so that the latent variable is on the same statistical scale as the observed variables. The amount of latent variables are then fed to the structural model in order to predict the operating speed:

*V85 = -3.618*roadside effects + 52.630*geometric effects + e1* *(13)*

The latent variables, their indicators, and their relative weights are discussed as follows:

1-Roadside Effects: This latent variable represents a combination of three roadside factors, namely land-use type [LUR], Land use density [LUD], and access per segment [N.ACCESS]. Land use is a roadside factor that affects operating speed (Yagar and Van Aerde 1983; Fitzpatrick et al. 2001). The presence of different land-use types (residential, commercial, school, etc.) was considered as categorical variables (1 if present; 0 other). Finally, only the presence of residential land use was found to be statically significant. Thus, the land-use type index (1 if the land-use type is residential, 0 otherwise) is used in the model. Not

only the effect of land use type is considered in this study, but also its density is used. Land use density is defined by equation 14.

$$\text{land use density(LUD)} = \frac{\text{length of adjacent land use(km)}}{\text{segment length(km)}} \quad (14)$$

LUD is the most significant factor that positively affects the “roadside effects” (standard regression weight = 0.924) followed by LUR (standard regression weight = 0.767) and N.ACCESS (standard regression weight = 0.221). The standard regression weight of “roadside effects” is -0.123, suggesting that the presence of residential land use in a segment, incremental increases in land use density, and access per segment lower operating speed.

2-geometric effects: This variable is extracted from segment length [LOGLN], highway grade [GRD], curvature [C], the presence of guardrail, and flat roadside slope [MTRSC], and posted speed limits [PSL]. Segment length is the main factor that influences “geometric effects” (standard regression weight = 0.866), followed by the posted speed limit with the standard regression weight of 0.686. The positive sign of LOGLN, MTRSC, and PSL indicates that as these variables increases, the “geometric effects” also increases. In comparison, GRD and C influence the “geometric effects” negatively. This latent variable affects the operating speed with a standard regression weight of 0.944. Considering the exogenous relationship between operating speed and “geometric effects” it can be seen that highway grade and curvature have an adverse effect, whereas segment length, the presence of a guardrail and flat roadside slope, and posted speed limits have a positive impact on operating speed.

As it can be seen, the standard regression weight of “geometric effects” (0.944) is more than that of “roadside effects” (-0.123), which means that “geometric effects” is more important than “roadside effects” in predicting operating speed. Also, “geometric effects” positively affects operating speed, while “roadside effects” affects it negatively. To investigate how each variable affects operating speed, the sensitivity analysis is illustrated for all the latent variables’ indicators in Figure 5. The sensitivity analysis results are in line with the above discussion about the importance of each indicator.

Figure 5. Sensitivity analysis

Thus, based on the above-mentioned latent variables, most geometric and roadside features are simultaneously considered as independent variables in order to evaluate operating speed in a multi-lane highway.

5.3. Comparison of SEM and LRM

The regression model proposed in this research identifies segment length, highway grade, the presence of guardrail and flat roadside slope, posted speed limits, and access per segment as statistically significant. On the other hand, two latent variables, including “roadside effects” and “geometric effects” were incorporated into SEM. The latent “roadside effects” variable is a combination of land-use type, land use density, and access per segment, whereas the “geometric effects” are extracted from segment length, highway grade, curvature, the presence of guardrail and flat roadside slope, and posted speed limits. In SEM, all variables related to roadside factors were significant, whereas the number of access points was the only roadside factor considered by LRM. The use of latent variables as well as a more significant number of roadside factors such as land-use type and land use density, and geometric factors such as curvature, resulted in better fit indices by SEM compared to LRM. These variables, however, were not significant when used separately, suggesting that they affect operating speed when combined with other variables and in the form of latent variables. Thus, SEM and latent variables make it possible to simultaneously employ a more significant number of roadside and geometric factors and consider only two latent variables to predict operating speed successfully.

To compare the results of SEM with LRM, the range of residuals returned from the difference between the observed and predicted operating speed for SEM and LRM was assessed (Figure 6). According to Dell’Acqua and Russo (2011), Esposito et al., (2011) and P. Garcia and T. Nodari (2013), if the residuals between the predicted and observed operating speed values assume limited values, the presented model is statistically significant and shows improvements compared to past predictions. Figure 6 shows that the values of the residuals for SEM are lower than those for LRM and that the distribution of residuals around the mean is low for SEM, which implies that both of the models are statistically significant and SEM has superior prediction capabilities.

Figure 6: Residual plot

Also, for a better comparison between these two modelling approaches, the observed and predicted value of operating speed for some real-life examples (other than the records used for modelling) and their relevant predictor variables' value are shown in Table 6.

Table 6. SEM and LRM real-life example

A cumulative standardised residual analysis is carried out (Figure 7). The closer the cumulative standardised residual curve stays to the x-axis, the more appropriate the model form is. Figure 7 shows the appropriateness of SEM in comparison with LRM regarding cumulative standardised residual curves. On the other hand, the value of R-squared for SEM is 0.940, while the value is 0.828 for LRM.

Figure 7: Cumulative standardised residual plot

Thus, based on the residual analysis, fit indices, and the significance of different variables, it can be concluded that SEM provides better fitting performance compared to LRM. The advantages of the presented SEM in modelling operating speed are summarised in Table 7.

Table 7. Advantages of SEM over LRM in the current study

6. Conclusion

This research employs two approaches to modelling, namely LRM and SEM, to identify the significant factors that affect operating speeds in rural multi-lane highways. A laser gun was utilised to collect more than 13,800 samples from 138 segments in Kole Jub-Borujerd and Borujerd-Khorramabad multi-lane highways in Iran. LRM considers roadside factors (access per segment) in addition to geometric factors (segment length, highway grade, the presence of guardrail and flat roadside slope, and posted speed limits). All these variables had a positive coefficient, except highway grade and access per segment. It should be mentioned that this research is among the few studies that have considered both geometric and roadside effects. To identify all of the effective variables and investigate their impact, two latent variables ("roadside effects" and "geometric effects") were defined in SEM. The latent "roadside effects" variable is

the combination of land-use type, land use density, and access per segment, whereas “geometric effects” are extracted from segment length, highway grade, curvature, the presence of guardrail, flat roadside slope, and posted speed limits. Therefore, the impacts of all variables on operating speed are assessed by SEM using latent variables. However, some of these variables, such as curvature, land-use type, and land use density, were not statistically significant in the regression model.

On the other hand, residual analysis reveals a lower value of residual for SEM compared to LRM. The value of R-squared for LRM is lower than that for SEM. Therefore, it can be concluded that SEM provides better fitting performance compared to LRM.

For further research, it is recommended to replicate this study using other roadside factors (such as pavement conditions), traffic factors (such as traffic flow), environmental factors (such as weather conditions and time of day), and driver characteristics (such as age, gender, etc.) simultaneously. A more extensive database is also recommended, which could lead to more reliable results.

References

Abbas S.K.S, Adnan M.A. and Endut I.R (2011) Exploration of 85th percentile operating speed model on horizontal curve: a case study for two-lane rural highways. Procedia - Social and Behavioural Sciences 16: 352-363, https://doi.org/10.1016/j.sbspro.2011.04.456.

Bamdad Mehrabani B. and Mirbaha B. (2019) Modeling the Operating Speed in Tangents and Curves of Four-lane Highways Based on Geometric and Roadside Factors, International Journal of Transportation Engineering 6(4): 355-366, https://doi.org/10.22119/IJTE.2018.113361.1385.

Bassani M, Cirillo C, Molinari S. and Tremblay J.M (2015) Fixed and random effect models to predict operating speed percentiles on rural two-lane highways (15-2307), Presented at the 94th Annual Meeting of the Transportation Research Board, Washington, D,C., https://trid.trb.org/view/1337552.

Bella F 2013. Driver perception of roadside configurations on two-lane rural roads: Effects on speed and lateral placement. Accident Analysis and Prevention 50:251-62, https://doi.org/10.1016/j.aap.2012.04.015.

Bhowmik T, Yasmin S and Eluru N (2019). A multilevel generalized ordered probit fractional split model for analysing vehicle speed. Analytic Methods in Accident Research 21: 13-31, https://doi.org/10.1016/j.amar.2018.12.001.

Boroujerdian AM, Seyedabrishami E and Akbarpour H (2016) Analysis of Geometric Design Impacts on Vehicle Operating Speed on Two-Lane Rural Roads. Procedia Engineering 161: 1144-1151, https://doi.org/10.1016/j.proeng.2016.08.529.

Brown TA (2014) Confirmatory factor analysis for applied research. Guilford Publications.

Cheng W, Gill GS, Sakrani T, Ralls D and Jia X (2018) Modelling the endogeneity of lane-mean speeds and lane-speed deviations using a Bayesian structural equations approach with spatial correlation. Transportation Research Part A: Policy and Practice 116: 220-231, https://doi.org/10.1016/j.tra.2018.06.014.

Cruzado I and Donnell E (2011) Models of vehicle operating speeds along two-lane rural highway transition zones: panel and multilevel modelling approaches. Transportation Letters 3(4): 265-278, https://doi.org/10.3328/TL.2011.03.04.265-278.

De Arcangelis G (1993) Structural equations with latent variables kenneth a. bollen. john wiley and sons, inc., new york, chichester, brisbane, toronto, singapore. Journal of Applied Econometrics 8(1): 111-113, https://doi.org/10.1002/jae.3950080111.

Dell'Acqua G and Russo F (2011) Road performance evaluation using geometric consistency and pavement distress data. Transportation Research Record 2203(1): 194-202, https://doi.org/10.3141/2203-24.

Eboli L, Guido G, Mazzulla G and Pungillo G (2017) Experimental relationships between operating speeds of successive road design elements in two-lane rural highways. Transport 32(2): 138-145, https://doi.org/10.3846/16484142.2015.1110831.

Echaveguren T, Díaz Á and Vargas-Tejeda S (2015) Operating speed models for horizontal reverse curves. In Proceedings of the Institution of Civil Engineers-Transport 168(6):510-522, https://doi.org/10.1680/jtran.13.00016

Esposito T, Mauro R, Russo F and Dell'Acqua G (2011) Speed prediction models for sustainable road safety management. Procedia - Social and Behavioural Sciences 20: 568-576, https://doi.org/10.1016/j.sbspro.2011.08.063.

Figueroa Medina AM and Tarko AP (2005) Speed factors on two-lane rural highways in free-flow conditions. Transportation Research Record 1912(1): 39-46, https://doi.org/10.3141/1912-05.

Fitzpatrick K, Carlson P, Brewer M and Wooldridge M (2001) Design factors that affect driver speed on suburban streets. Transportation Research Record 1751(1): 18-25, https://doi.org/10.3141/1751-03.

Fitzpatrick K, Paul C, Marcus A.B, Mark D.W, and Shaw-Pin M (2003) Design Speed, Operating Speed, and Posted Speed Practices (NCHRP report 504), Transportation Research Board, National Research Council, Washington, D.C.

Fitzpatrick K, Shamburger CB, Krammes RA and Fambro DB (1997) Operating speed on suburban arterial curves. Transportation Research Record 1579(1): 89-96, https://doi.org/10.3141/1579-11.

Garcia DS, Lindau LA, and Nodari CT (2013) Multivariate operating speed forecasting model based on the geometric elements of two-lane highways. Journal of Traffic and Logistics Engineering 1(2): 97-103, https://doi.org/10.12720/jtle.1.2.97-103.

Gargoum SA and El-Basyouny K (2016) Exploring the association between speed and safety: A path analysis approach. Accident Analysis & Prevention 93: 32-40. https://doi.org/10.1016/j.aap.2016.04.029.

Gong H and Stamatiadis N (2008) Operating speed prediction models for horizontal curves on rural four-lane highways. Transportation Research Record 2075(1): 1-7, https://doi.org/10.3141/2075-01.

Hair JFJ, Black WC, Babin BJ, and Anderson RE (2010) Multivariate Data Analysis Seventh Edition Prentice Hall.

Hashim IH, Abdel-Wahed TA and Moustafa Y (2016) Toward an operating speed profile model for rural two-lane roads in Egypt. Journal of Traffic and Transportation Engineering (English Edition) 3(1): 82-88, https://doi.org/10.1016/j.jtte.2015.09.005.

Hassan Y and Sarhan M (2011) Modelling Operating Speed: Synthesis Report. Chapter 1: Introduction. Transportation Research E-Circular E-C151.

Himes SC and Donnell ET (2010) Speed prediction models for multilane highways: simultaneous equations approach. Journal of Transportation Engineering 136(10): 855-862, https://doi.org/10.1061/(ASCE)TE.1943-5436.0000149.

Hooper D, Coughlan J and Mullen MR (2008) Structural equation modelling: Guidelines for determining model fit. Electronic Journal of Business Research Methods 6(1): 53-60, https://arrow.dit.ie/libart/4/.

Ke R, Li W, Cui Z and Wang Y (2019) Two-stream multi-channel convolutional neural network (TM-CNN) for multi-lane traffic speed prediction considering traffic volume impact. arXiv preprint arXiv:1903.01678.

Park YJ and Saccomanno FF (2006). Evaluating speed consistency between successive elements of a two-lane rural highway. Transportation Research Part A: Policy and Practice 40(5): 375-385. https://doi.org/10.1016/j.tra.2005.08.003.

Sadia R, Bekhor S and Polus A (2017) Structural equations modelling of drivers' speed selection using environmental, driver, and risk factors. Accident Analysis and Prevention 116: 21-29. https://doi.org/10.1016/j.aap.2017.08.034.

Schumacker RE and Lomax RG (2004) A beginner's guide to structural equation modelling. Psychology Press.

Semeida AM (2013) Impact of highway geometry and posted speed on operating speed at multi-lane highways in Egypt. Journal of Advanced Research 4(6):515-523, https://doi.org/10.1016/j.jare.2012.08.014.

Shallam RDK and Ahmed AM (2016). Operating Speed Models on Horizontal Curves for Two-lane Highways. Transportation Research Procedia, 17: 445-451, https://doi.org/10.1016/j.trpro.2016.11.086.

Shankar V and Mannering F (1998) Modelling the endogeneity of lane-mean speeds and lane-speed deviations: a structural equations approach. Transportation Research Part A: Policy and Practice 32(5): 311-322, https://doi.org/10.1016/S0965-8564(98)00003-2.

Sil G, Nama S, Maji A. and Maurya AK (2020) Modelling Operating Speeds for Multilane Divided Highways. In Transportation Research 367-375 Springer Singapore, https://doi.org/10.1007/978-981-32-9042-6_29.

Washington SP, Karlaftis MG and Mannering F (2010) Statistical and econometric methods for transportation data analysis. CRC press.

Xiao X (2013) Structural equation modelling compared with ordinary least squares in simulations and life insurers' data. http://hdl.handle.net/2152/22516.

Yagar S and Van Aerde M (1983) Geometric and environmental effects on speeds of 2-lane highways. Transportation Research Part A: General 17(4): 315-325, https://doi.org/10.1016/0191-2607(83)90094-8.

Table captions

Table 1. Descriptive Statistics of Variables
Table 2. Definitions of SEM elements
Table 3. Fit Indices and Their Thresholds (Hu and Bentler 1999; Schumacker and Lomax 2004; Hooper et al. 2008; Hair et al. 2010)
Table 4. Multivariable Linear Regression Models Results
Table 5. Structural Equation Model
Table 6. SEM and LRM real-life example
Table 7. Advantages of SEM over LRM in the current study

10 Table 1. Descriptive Statistics of Variables

Variable	**Continuous Variables**				
	Variable Symbol	Min	max	Mean	Standard Deviation
Operating speed(km/h)	V85	62.00	110.00	94.05	10.12
Segment length (m)	LNM	200	2700	1192.80	503.36
Natural logarithm of the Segment length	LOGLN	2.30	3.83	3.04	0.20
Highway grade (%)	GRD	-7.80	12.00	0.21	3.88
Curvature ($\frac{1}{radius}$)	C	0.00	0.0115	0.0008	0.0018
Paved Shoulder width (m)	SHP	0.50	2.50	1.68	0.71
Unpaved shoulder width[1] (gravel and untreated) (m)	SHD	2.00	10.00	3.55	2.06
Posted speed limit (km/h)	PSL	40.00	110.00	89.28	11.99
Number of access per segment	N.ACCESS	0.00	3.00	1.02	0.73
Access density (number of access / segment length)	AD	0.00	6.66	1.08	1.14
Adjacent Land use length (km)	LULN	0.00	0.45	0.15	0.14
Land use density	LUD	0.00	0.62	0.14	0.15
Paved width (m)	PW	7.80	10.80	8.56	0.89
Variable	**Categorical Variables**				
	Variable Symbol				
LU (1 if land use exists in segment; 0 otherwise)	LU				
Land use type (1 if residential; 0 otherwise)	LUR				
Median barrier type (1 if guardrail; 0 other wise)	MT				
Road side configuration type (1 if flat; 0 otherwise)	RSC				
Paved shoulder width (1 if > 2 (m); 0 otherwise)	SHP01				
Unpaved shoulder width (1 if > 3 (m); 0 otherwise)	SHD01				
Median barrier and roadside slope type (1 if guardrail and flat; otherwise)	MTRSC				

1- beyond the paved width.

11

Table 2. Definition of SEM elements

Model	element	Definition
Measurement Model	X	$q \times 1$ column vector of observed exogenous variables
	Y	$p \times 1$ column vector of observed endogenous variables
	ξ	$n \times 1$ column vector of observed exogenous variables
	η	$m \times 1$ column vector of observed endogenous variables
	δ	$q \times 1$ column vector of measurement error terms for observed variables x
	ε	$p \times 1$ column vector of measurement error terms for observed variables y
	Λ_x	Matrix $(q \times n)$ of structural coefficients for latent exogenous variables based on their observed indicator variables
	Λ_y	Matrix $(p \times m)$ of structural coefficients for latent endogenous variables based on their observed indicator variables
Structural model	Γ	Matrix $(m \times n)$ of direct effects among endogenous latent variables
	B	$m \times 1$ column vector of the error terms
	ξ	Vector of exogenous latent variables
	η	Vector of endogenous latent variables
	ζ	Estimation error associated with the endogenous variables

Table 3. SEM Fit Indices and Their Thresholds (Schumacker and Lomax 2004; Hooper et al. 2008; Hair et al. 2010)

Absolute fit indices "Absolute fit indices determine how well a certain model fits the sample data and demonstrates which proposed model has the most superior fit". (Hooper et al. 2008)				
Fit index	**Threshold levels**			
	perfect	**good**	**Acceptable**	**fair**
Chi-Square	Low chi-square relative to degrees of freedom with an insignificant p-value ($p < 0.05$)			
Chi-Square / degree of freedom		< 3	< 5	
RMSEA	< 0.06	0.06 ≤ ≤ 0.08	0.08 < ≤ 0.1	> 0.1
GFI	> 0.95	0.9 ≤ ≤ 0.95	> 0.8	
AGFI	> 0.95	0.9 ≤ ≤ 0.95	> 0.8	
RMR	Good models have small RMR			
SRMR	> 0.08			
Incremental fit indices Incremental fit indices assess how well the estimated model fits relative to some baseline model for rejecting the null hypothesis that all variables are uncorrelated**.**				
Fit index	**Threshold levels**			
	perfect	**good**	**Acceptable**	**fair**
NFI	> 0.95			
TLI	> 0.95	0.9 ≤ ≤ 0.95	> 0.9	> 0.85
CFI	> 0.95	0.9 ≤ ≤ 0.95	> 0.9	> 0.85
Parsimony fit indices Parsimony fir indices measures of overall goodness-of-fit representing the degree of model fit per estimated coefficient. (Hair et al. 2010).				
Fit index	**Threshold levels**			
	perfect	**good**	**Acceptable**	**fair**
PCFI		> 0.5		

17 Table 4. Multivariable Linear Regression Models Results

Variables	Coefficient	P
Constant	67.829	<0.001
Geometric variables		
LNM	0.011	<0.001
GRD	-0.380	<0.001
MTRSC	3.147	0.002
PSL	0.166	<0.001
Roadside variables		
N.ACCESS	-2.790	<0.001
Model fit indices		
R-Square	0.828	
F	113.112	
Durbin- Watson	1.654	
Mean residual	0.00	
MAD	3.05	
MSE	17.76	
RMSEA	0.873	
K-S test significance for operating speed = 0.058		

18

19 Table 5. Structural Equation Model

Measurement model					
Latent variable (exogenous variable)	**Observed variable**	**Regression weight**	**Standard error**	**p**	**Standard regression weight**
Roadside effects	LUR	1	Reference variable		0.767
	LUD	0.420	0.062	<0.001	0.924
	N.ACCESS	0.473	0.193	0.014	0.221
Geometric effects	LOGLN	1	Reference variable		0.866
	GRD	-7.635	1.759	<0.001	-0.357
	C	-0.005	0.001	<0.001	-0.550
	MTRSC	1.837	0.199	<0.001	0.666
	PSL	56.690	5.885	<0.001	0.686
Structural model					
Observed variable (endogenous variable)	**latent variable (exogenous variable)**	**Regression weight**	**Standard error**	**p**	**Standard regression weight**
Operating speed	Roadside effects	-3.618	1.269	0.004	-0.123
	Geometric effects	52.630	2.823	<0.001	0.944
CMIN/DF=1.947, AGFI=0.843, GFI=0.909, RMSEA=0.087, CFI=0.783, TLI=0.700, PNFI=0.478, PCFI=0.566, R-Square= 0.940					

20

Table 6. SEM and LRM real-life example

No	Predictor Variables									LRM		SEM			observed V85
	LUR	LUD	N.ACCESS	LOGLN	LNM	GRD	C	MTRSC	PSL	constant	predicted V85	roadside effects	geometric effects	predicted V85	
1	0	0.12	2	3.40	2500	0.23	0	1	90	67.83	107.74	0.02	1.99	104.66	105
2	0	0.00	3	2.95	900	5.78	0	0	70		78.78	-0.05	1.34	70.71	70
3	0	0.00	1	3.36	2300	-0.13	0	1	100		110.13	-0.19	1.99	105.42	106
4	0	0.00	1	3.08	1200	-1.67	0	1	90		96.95	-0.16	1.79	94.79	95
5	0	0.36	1	2.85	700	1.14	0	0	80		85.58	0.43	1.62	83.70	84
6	0	0.19	0	3.20	1600	2.38	0	1	100		104.26	0.11	1.9	99.60	100
7	1	0.23	1	3.11	1300	3.38	0	1	90		96.13	0.36	1.86	96.59	97
8	0	0.00	0	3.26	1800	-2.17	0	1	90		106.53	-0.22	2.01	106.58	107
9	0	0.25	1	3.11	1300	0.15	0	1	90		97.36	0.24	1.83	95.44	96
10	0	0.00	0	3.00	1000	2.40	0	0	90		92.86	-0.2	1.8	95.46	96

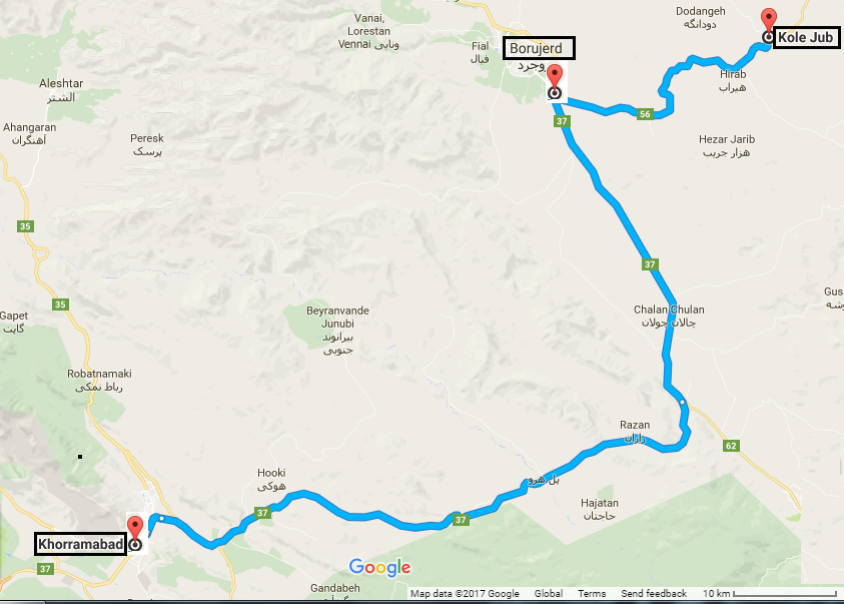
23 Table 7. Advantages of SEM over LRM in the current study

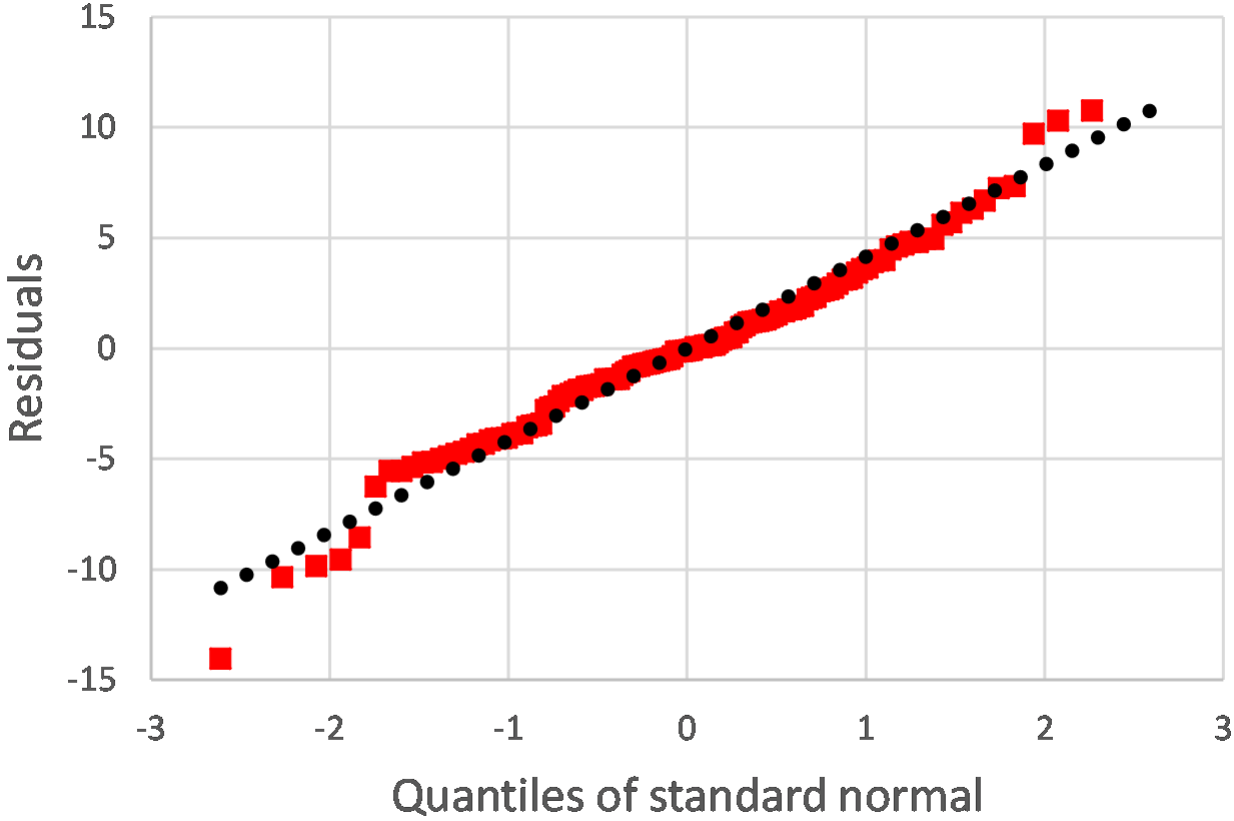
Feature	Modelling approach	SEM	LRM
Consideration of Geometric features	Segments length	✓	✓
	Highway grade	✓	✓
	Median barrier and roadside slope type	✓	✓
	Posted speed limit	✓	✓
	Curvature	✓	×
Consideration of Roadside features	Number of Access	✓	✓
	land use density	✓	×
	Land use type	✓	×
Assessing the direct effects of variables		✓	✓
Assessing the indirect effects of variables		✓	×
Considering latent variables		✓	×
Simultaneous equations		✓	×
Fewer variables which has direct effect on operating speed		✓	×
Lower residuals value		✓	×
Fitting indices	R-Square	0.940	0.828
	RMSEA	0.087	0.873

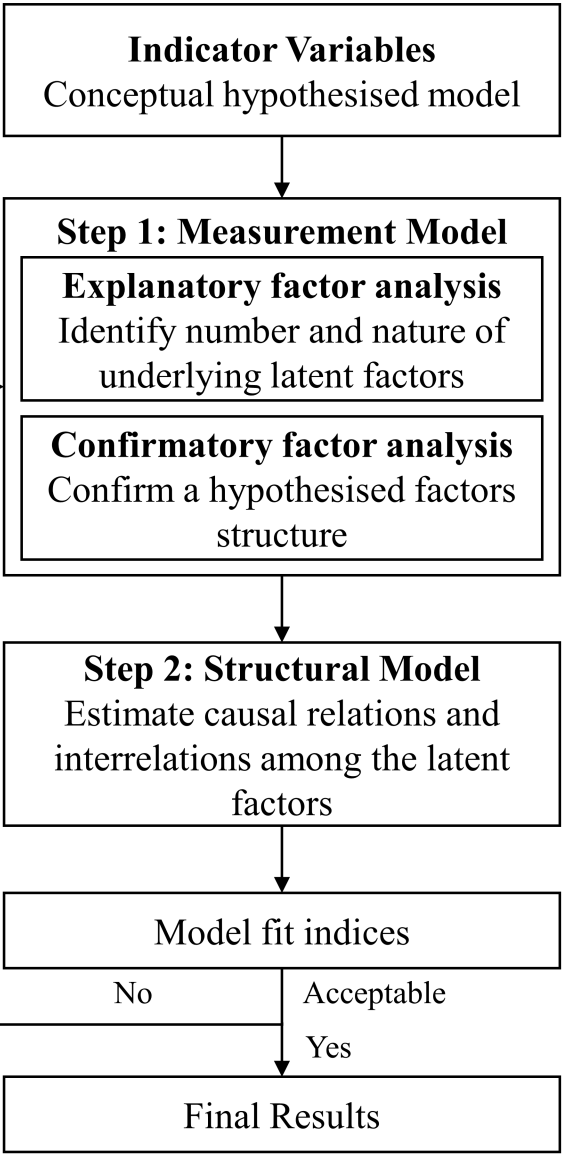
24

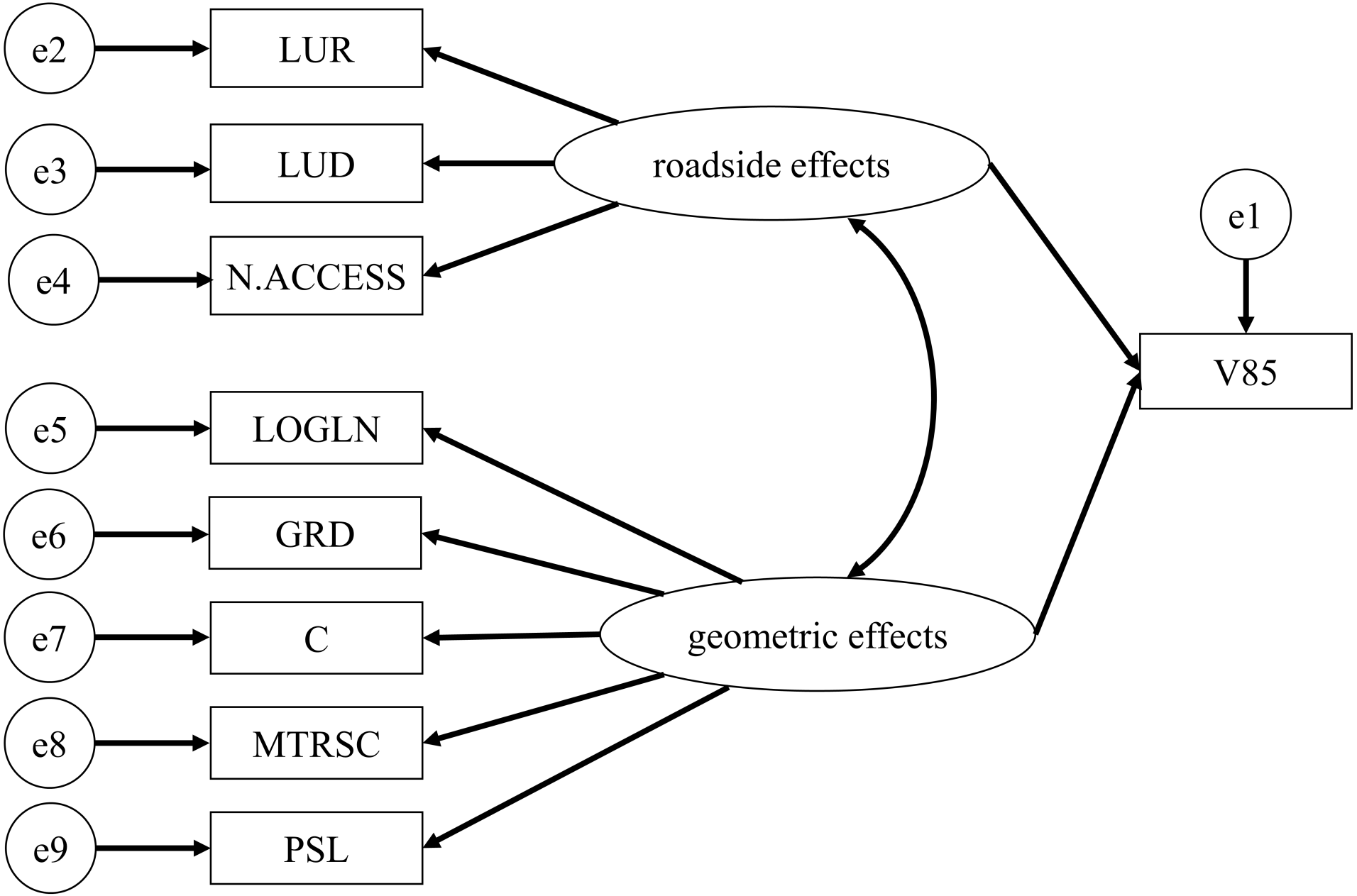
Figure Captions

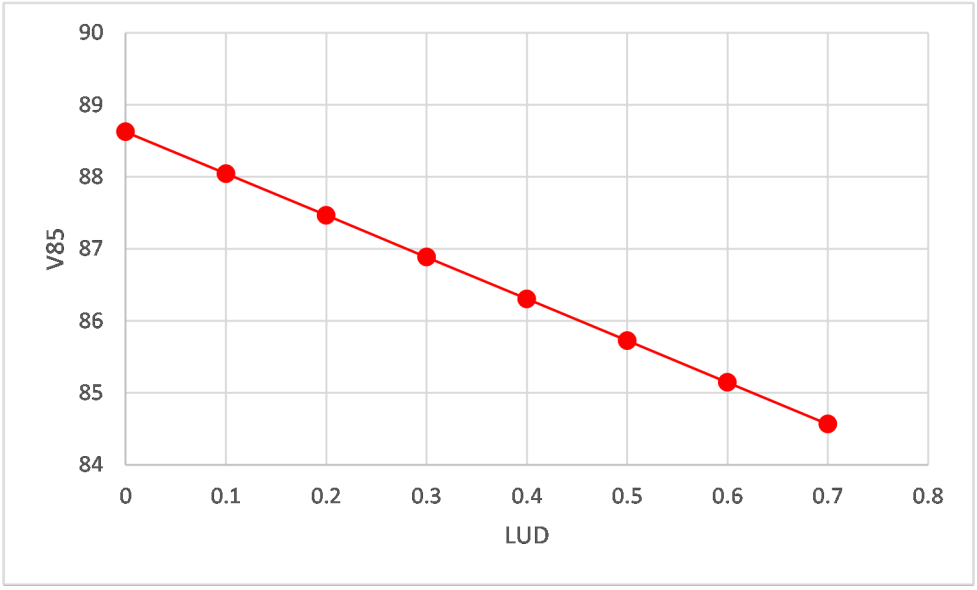
Figure 1: Kole jub – Khoramabad path
Figure 2: Q-Q plot of residuals
Figure 3: structural equation modelling process for operating speed prediction
Figure 4: Structural equation model for operating speed
Figure 5: Sensitivity analysis
Figure 6: Residual plot
Figure 7: Cumulative standardised residual plot

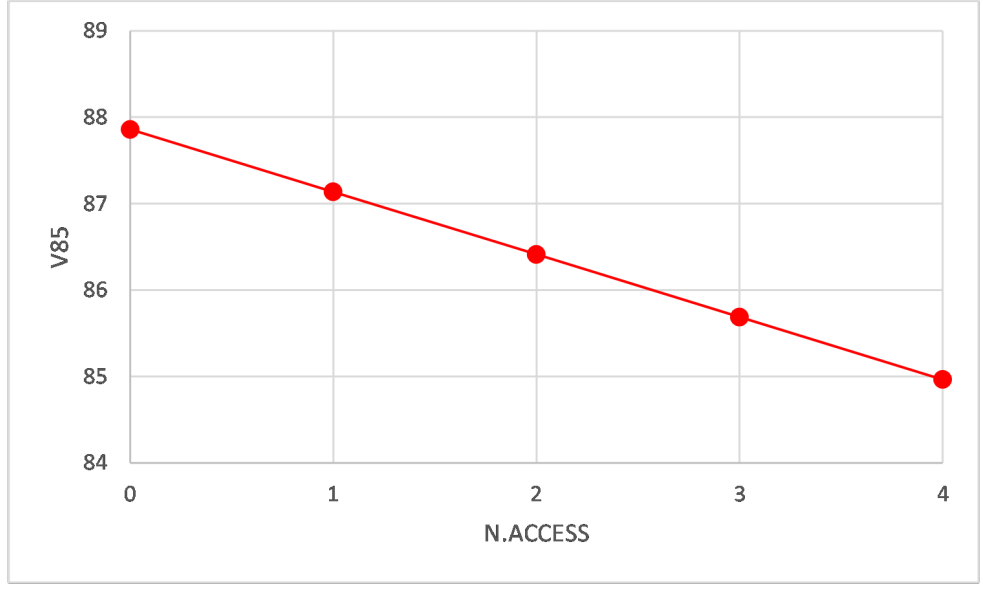

Kole Jub
Dodangeh
دودانگه
Borujerd
بروجرد
Fial
فیال
Vanai,
Lorestan
Vennai
وانایی
Hirab
هیراب
Aleshtar
الشتر
Ahangaran
اهنگران
Peresk
پرسک
Hezar Jarib
هزار جریب
Chalan Chulan
چالان چولان
Beyranvande
Junubi
بیرانوند
جنوبی
Gapet
گاپت
Robatnamaki
رباط نمکی
Razan
رازان
Hooki
هوکی
Hajatan
حاجتان
Khorramabad
Gandabeh
Google
Map data ©2017 Google
Global
Terms
Send feedback
10 km

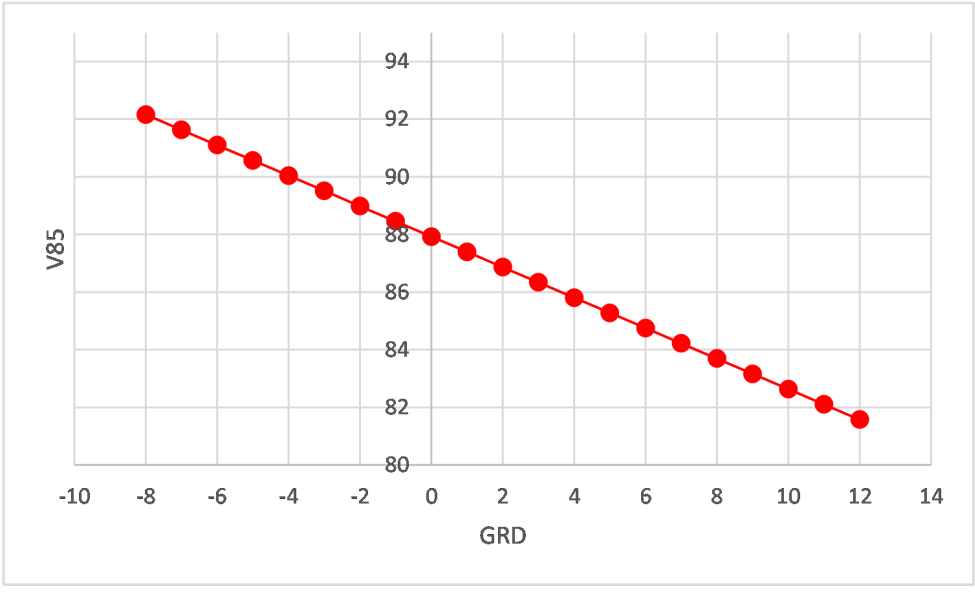

15
10
5
0
-5
-10
-15
Residuals
-3
-2
-1
0
1
2
3
Quantiles of standard normal

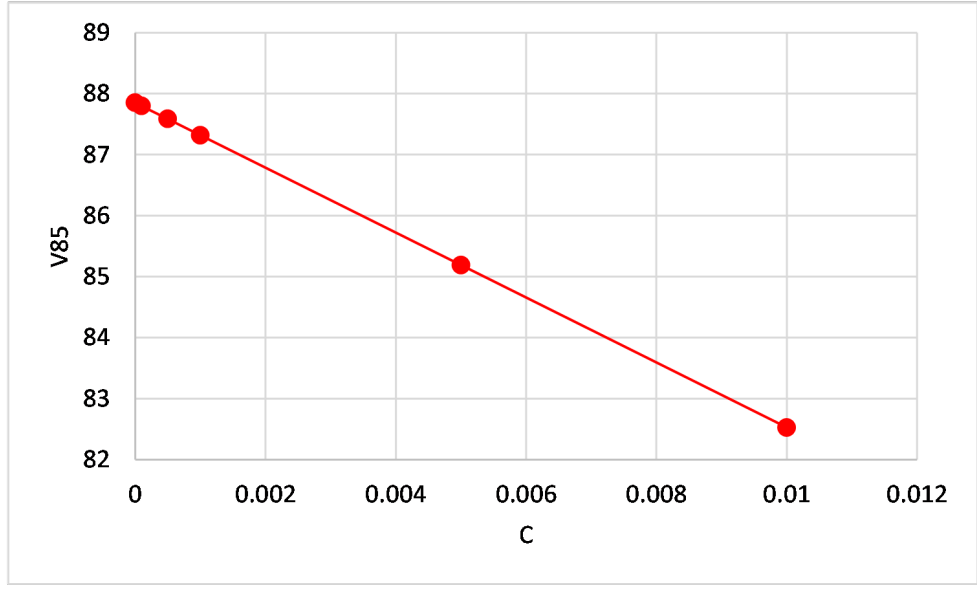

Indicator Variables
Conceptual hypothesised model
Step 1: Measurement Model
Explanatory factor analysis
Identify number and nature of underlying latent factors
Confirmatory factor analysis
Confirm a hypothesised factors structure
Step 2: Structural Model
Estimate causal relations and interrelations among the latent factors
Model fit indices
No
Acceptable
Yes
Final Results

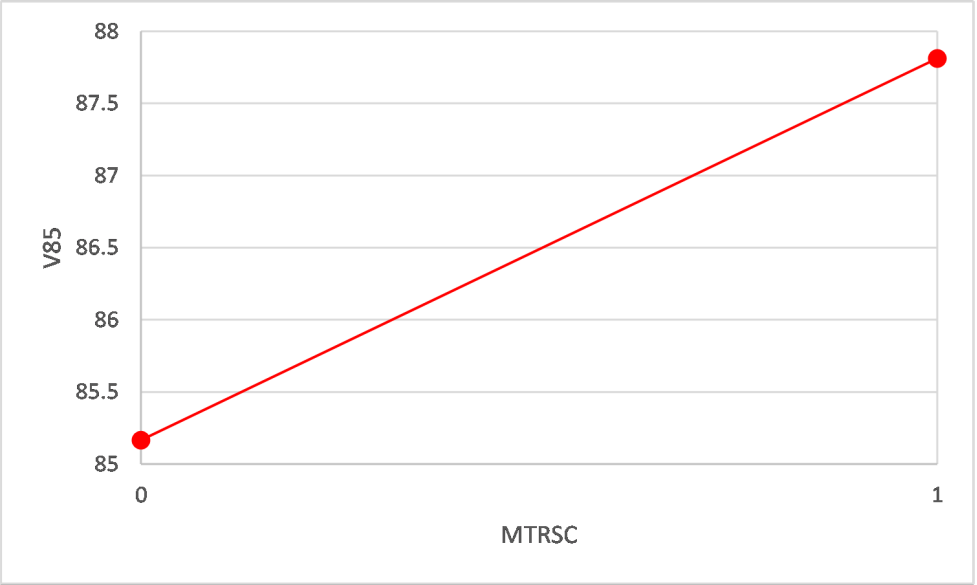

e2
LUR
e3
LUD
e4
N.ACCESS
roadside effects
e1
V85
e5
LOGLN
e6
GRD
e7
C
geometric effects
e8
MTRSC
e9
PSL

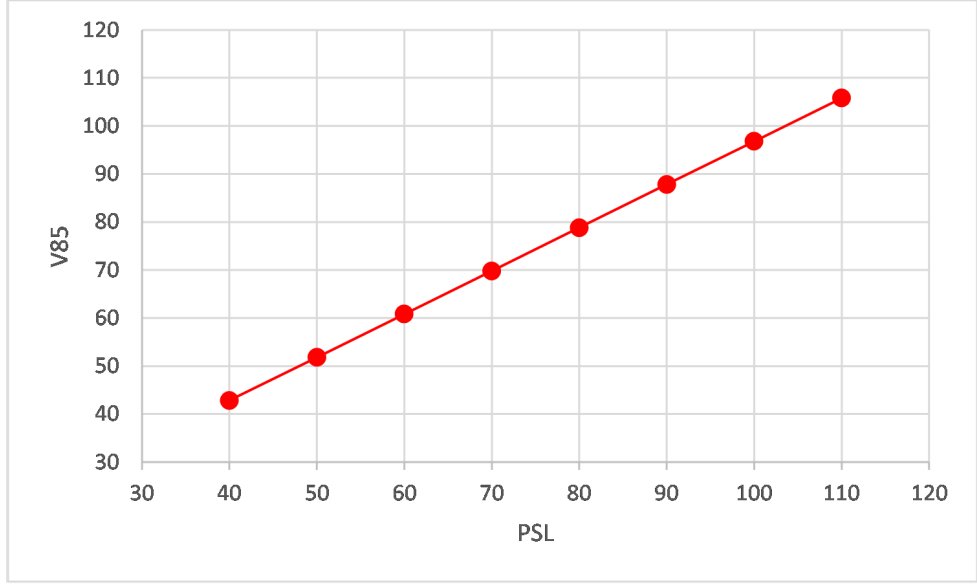

V85
LUD
90
89
88
87
86
85
84
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

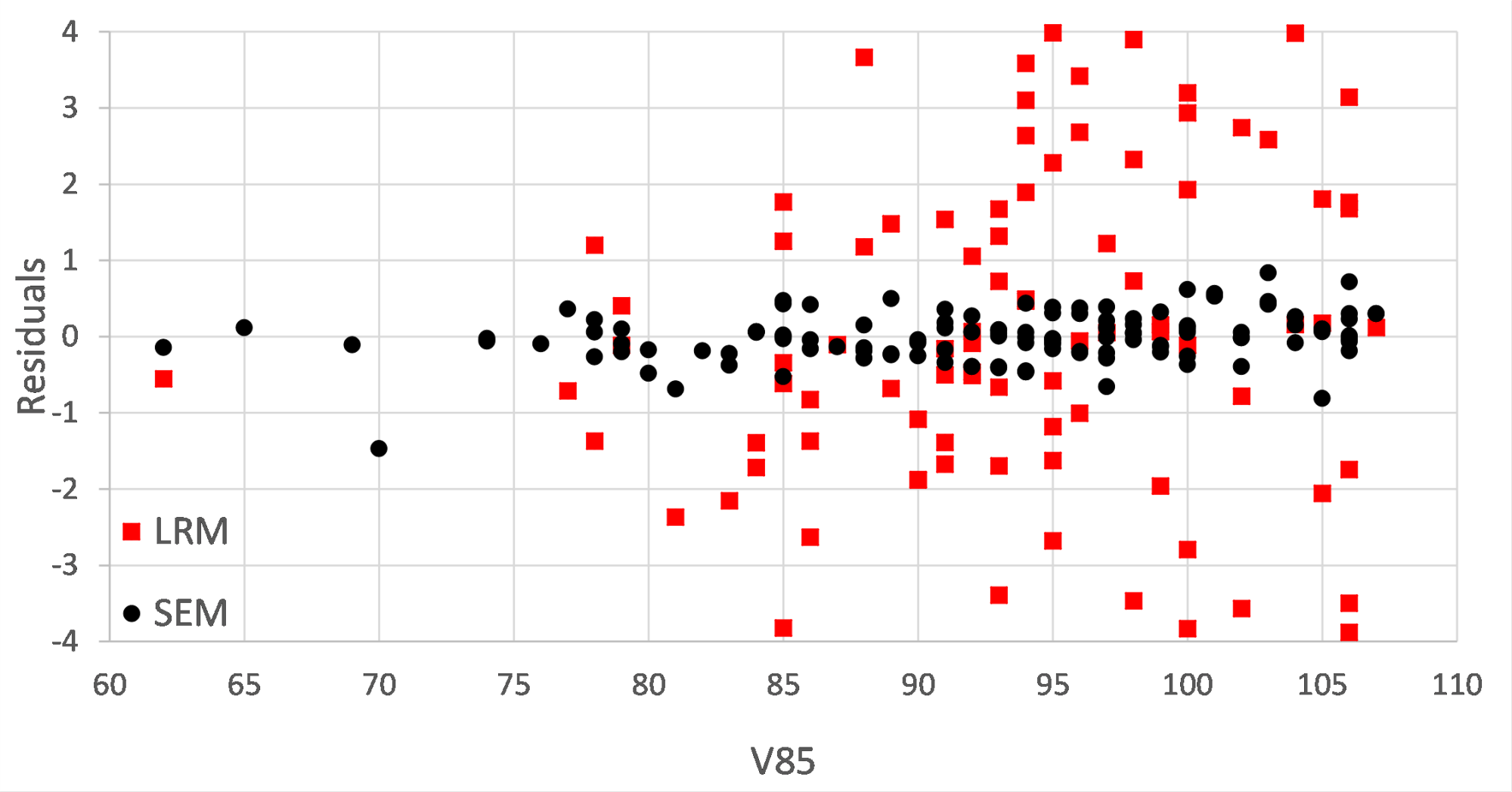

V85
N.ACCESS
89
88
87
86
85
84
0
1
2
3
4

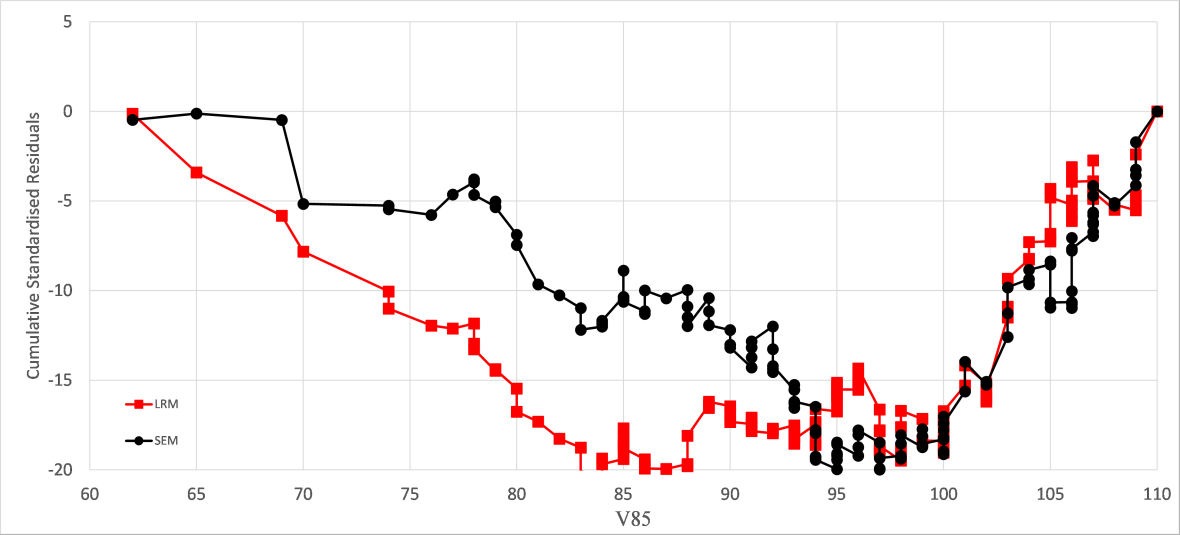

V85
GRD
94
92
90
88
86
84
82
80
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
14


V85
C
89
88
87
86
85
84
83
82
0
0.002
0.004
0.006
0.008
0.01
0.012


V85
MTRSC
88
87.5
87
86.5
86
85.5
85
0
1


V85
PSL
120
110
100
90
80
70
60
50
40
30
30
40
50
60
70
80
90
100
110
120


Residuals
4
3
2
1
0
-1
-2
-3
-4
LRM
SEM
60
65
70
75
80
85
90
95
100
105
110
V85


5
0
-5
-10
-15
-20
Cumulative Standardised Residuals
LRM
SEM
60
65
70
75
80
85
90
95
100
105
110
V85