UNIVERSITY OF PARDUBICE
FACULTY OF TRANSPORT ENGINEERING

THE FORMATION AND USAGE OF A REGIONAL PUBLIC TRANSPORT SERVICES MODEL

DOCTORAL THESIS PRECIS

2019 Martin Trpišovský

Author: Martin Trpišovský

Doctoral study programme:
P3710 Technique and Technology in Transport and Communications

Field of study:
3708V024 Technology and Management in Transport and Telecommunications

Doctoral thesis name:
The formation and usage of a regional public transport services model

Supervisor: Petr Průša
Supervisor specialist:

Doctoral thesis has arisen at the supervising:
Department of Transport Management, Marketing and Logistics

TABLE OF CONTENTS

INTRODUCTION

Human society has very close relationship to transport since ever. Without the existence of quality transport, the historical development would never be possible in its full extent and complexity. It its modes from individual one to the public one transport is the engine and an essential condition of the economic performance of contemporary civilization. To provide public transportation services is one of the basic indicators of the life standard in certain locality, thus public transport services shall be seen as important issue in each region. It is essential to understand local circumstances and transport needs for its proper provision, this is the only way how to achieve the aim of high-quality service offer fulfilling contemporary sustainable mobility requirements.

Public transportation services are natural part of public services provided by the self-governed regions and central government in the Czech Republic as well as former Czechoslovakia, the development of recent years brings up new issues and challenges though. Dynamic development of recent three decades brought extraordinary amount of changes that had to affect the public transport services too. The massive outbreak of individual motorism, growing regional disparities in development and economic efficiency, weakening of local industrial concentration, political and social integration into the structures in politically-western-European area leading to the growth of the standard of living, suburbanization and urban sprawl, cultural and social megatrends, integration of public transport subsystems and cost optimization pressure lead to new requirements.

Public transport services and their provision by public transport contracts are part of complicated multidisciplinary issue. This complex issue has strong impact on the citizens, which very often leads to the fact that the solution are more political compromises (on the local, regional, national as well as international level) than the outcomes of scientific approach.

The evaluation of the public transport services quality level, the influence of certain factors determining its extent and how anticipatively it reacts to the transport needs of local inhabitants with respect to the financial limitations of public budgets are very complicated, it is influenced by many social, political, demographical and economic factors.

The subject of this thesis is to design a methodology of the evaluation of the extent of public transport services in certain municipality units within a region with the usage of a mathematical-statistical model of public transport services connection count prediction. The main aim is accomplished by using appropriate scientific methods based on detailed analysis of present-time knowledge of the issue.

1 ANALYSIS OF CURRENT STATE IN THE FIELD OF THE DOCTORAL THESIS

The first chapter of the thesis covers the analysis of the existing situation in the field of public transportation services. The chapter is based on thorough review of wide variety of domestic and foreign literature. The analysis consists of the topics of public transportation services and their undisputable benefits to contemporary society as it is widely accepted as the only sustainable alternative to individual motorism. The benefits can be recognized in many fields, in particular the benefits in social field are crucial.

Under certain circumstances the public transportation services can bring the economical profit, but in the overwhelming majority of the cases public transportation services aren't viable in purely economic point of view. The evaluation of the public transportation services shall be shifted from purely economical to social-economical point and accepted as services provided in public interest. This approach also transforms the economic model, the services are provided not only to meet the demand, but also to meet other needs than a priori demand. Including the social aspects in public transportation services planning and organizing is essential part of sustainable mobility. Public interest is represented by public transport services contracts. The subsidies involved aren't primarily the support to the operator, but the payment for services provided for the region and its inhabitants.

The public transportation services system in the Czech Republic is well known for its long tradition, considerably high quality and high usage by commuting, most of the connections are provided on the base of public transportation contracts. Annual subsidy to the system reaches approximately 20 billion of CZK.

The emphasis was put on public transportation services determining factors, the importance of certain factors varies, but authors of public literature agree on crucial importance of the population. The problem of discontinuity of public transport services quality in different regions is identified, the extent of public transport services varies among regions. Despite this fact no methodology of the evaluation of the extent of public transport services in certain municipality units within a region was found, public transport services standards define only minimum level of service.

Chosen competent authorities (regions, ministry of transport) come up with standard and above-standard extent of public transport services provided and also stronger emphasis on planning and its obligatory character appears recently. These aspects are other reasons to come up with the methodology of the evaluation of the extent of public transport services.

2 AIMS OF THE DOCTORAL THESIS

Based on the detailed literature review of the current state in the field of doctoral thesis in Czech Republic and abroad the aims of the thesis are set. The main aim of this doctoral thesis is to design a methodology of the evaluation of the extent of public transport services in certain municipality units within a region with the usage of a mathematical-statistical model of public transport services connection count prediction.

The model will create predictions of the public transportation services connection count within defined time period in certain municipal units of analyzed regions with high statistical credibility. The starting point for reaching the aim will be the complex analysis of particular factors determining public transportation services in municipal units. The outcome will not be only the count prediction, but also the influence rate of particular factors. The model will cover wide variety of specific aspects determining the final count of public transport service connections serving particular municipal unit, it will allow transparent comparison of the predicted count and reference count. This makes it a useful decision support tool when deciding about public transport services. Also, there is vast potential of wide usage of the whole methodology in order to understand deeper this issue following by the benefits in the sector of education.

The main aim will be reached by accomplishing a serie of sequential sub-aims:

1. To form a mathematical-statistical model predicting the count of public transportation service connections in particular municipal units of a region. The sub-aim is accomplished by using appropriate scientific methods based on detailed analysis of present-time knowledge of the issue. The main requirement to be met by the model is its ability to predict the level of public transport services in assessed municipal units based on factors that are identified as the ones that determine the extent of public transportation services, the prediction shall be done with high statistical credibility.
2. To form a methodology of the evaluation of the extent of public transport services in certain municipality units with the usage of a model created according to the previous sub-aim.
3. Use the methodology and create the model.
4. To apply the methodology and model created within on certain region, to adjust the model based on the real data from that region and assess its credibility. The outcome should undergo a critical analysis that will be the last step of the methodology.

3 LIST OF USED METHODS

Throughout all the doctoral thesis only methods and techniques based on rational logics in compliance with science methodology are used to reach the aims. The choice of scientific methods originates from the principles of scientific research, the emphasis is put on systematic and organized character. Scientific research is realized in three standard steps – the exploration, the prediction and the explanation (Olecká and Ivanová, 2010):

- The exploration is the introductory part of the scientific research, basic elements of the research scope are identified and characterized.
- The prediction follows, during this phase certain qualitative and quantitative characteristics are identified and links between them are described.
- Final research part is the explanation, during this phase the final information is presented and issue explained.

3.1 Basic Explanatory Methods

The cornerstones are basic explanatory methods that are universally applicable at all steps of scientific research. Among methods used in this thesis literature review, abstraction and concretization, deduction and induction, analysis and synthesis, analogy and comparing and expert estimation.

3.2 Prognostic Methods

Fundamental methods in this thesis are prognostic methods, exact prognostic methods according to Jirsák, Mervart and Vinš (2012) are used. The model created in the methodology is based in regression, that is basic statistical method when exploring links between numerical signs (Blatná, 2008).

3.2.1 Linear regression

Regression function is expected value of the dependent (stochastic; output) variable derived from changeable combinations of independent (deterministic; input) variables, thus regression can be understood in the way presented in the formula (1) according to van Wieringen (2019); Blatná (2008) and Hebák (1998).

$$Y = f(\mathrm{x}, \beta_1, \dots, \beta_k) + \varepsilon = E(Y|x) \tag{1}$$

where:

Y .... output variable [-],

x .... input variable [-],

$\beta_1, \dots, \beta_k$ .... regression parameters [-],

k .... regression parameters count, $k \geq 1 \wedge k \in \mathbb{N}$ [-],

$f(\mathrm{x}, \beta_1, \dots, \beta_k)$.........regression function [-],
ε...............................error variable caused by the influences not included in the model [-],
$E(Y|\mathrm{x})$.....................expected value of the output variable Y derived from x [-].

Basic vector view on vector regression based on van Wieringen (2019), Tvrdík (2013), Bremer (2012), Anděl (2005), Zvára (2002), Hebák (1998). Rousseeuw and Yohai (1984), Golub, Heath and Wahba (1979) and Hoerl and Kennard (1970) is presented in formula (2).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

where:

$\mathbf{y}$ ....vector of real empirically found values of dependent variable consisting of n dimensions [-],
$\mathbf{X}$..............................matrix of independent variables of the $(n \times p)$ dimensions [-],
$\boldsymbol{\beta}$..............................vector of regression coefficients consisting of p dimensions [-],
$\boldsymbol{\varepsilon}$...............................vector of error variable consisting of n dimensions [-].

The linear regression is based on ordinary least squares method introduced by Gauß (1809). The interpretations by Bondell and Stefanski (2013), Bremer (2012) and Franc (2011) are presented in the formula (3).

$$\min_{\widehat{\boldsymbol{\beta}}} \mathbf{e}^{\mathrm{T}}\mathbf{e} = \min_{\widehat{\boldsymbol{\beta}}} \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right)^{\mathrm{T}} \left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) = \min_{\widehat{\boldsymbol{\beta}}} \left(\mathbf{y} - \sum_{j=1}^{k} \mathbf{x}_j{}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right)^2 \tag{3}$$

where:

$\mathbf{e}$ ....residual vector of differences between real empirically found values of y variable and their predicated estimation consisting of n dimensions [-],
$\widehat{\boldsymbol{\beta}}$..............................vector of the estimations of the regression coefficient β consisting of p dimensions [-],
$\mathbf{y}$ ....vector of real empirically found values of dependent variable consisting of n dimensions [-],
$\mathbf{X}$..............................matrix of independent variables of the $(n \times p)$ dimensions [-],
j ....identifier of particular input valuable, $j \in \langle 1; k \rangle \wedge j \in \mathbb{N}$ [-],
k ....regression parameters count, $k \geq 1 \wedge k \in \mathbb{N}$ [-],
$\mathbf{x}_j$.............................vector of the values of jth independent variable consisting of k dimensions [-].

The outcome of the linear regression model is shown in the formula (4) according to Neubauer (2016), Bremer (2012) and Hebák and Svobodová (2001).

$$\hat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \tag{4}$$

where:

$\hat{\mathbf{y}}$ .... vector of predicted values of independent variable consisting of n dimensions [-],

$\mathbf{X}$ .... matrix of independent variables of the $(n \times p)$ dimensions [-],

$\hat{\boldsymbol{\beta}}$ .... vector of the estimations of the regression coefficient β consisting of p dimensions [-].

3.2.2 Robust regression using LTS

According to Hebák and Svobodová (2001) any systematic pattern or nonrandomness of the residuals indicates to shortcomings of the regression model. The linear regression model based on the ordinary least squares method is very sensitive to outliers. Robust regression is involved in order to deal with outliers. Based on comparison of different robust regression methods LTS (Least trimmed squares method) is chosen to be used in the model. LTS is based on the iteration principle, it has four main steps (Doornik, 2011):

1. Setting the trimming constant of r.
2. All possible subsets γ consisting of r observations are generated. For each subset γ the regression coefficients β estimations are counted by standard ordinary least squares method.
3. The residuals $e^{m}_{LTS,\gamma}$ are counted for each vector $\hat{\boldsymbol{\beta}}^{\mathbf{LTS}}_{\boldsymbol{\gamma}}$ and all n observations. Resiudals vector $\mathbf{e}_{\boldsymbol{\gamma}}$ is assembled for each vector $\hat{\boldsymbol{\beta}}^{\mathbf{LTS}}_{\boldsymbol{\gamma}}$. The dimensions of vector $\mathbf{e}_{\boldsymbol{\gamma}}$ are ordered ordinaly according to the size of the second powers of their residuals.
4. Particular vector $\hat{\boldsymbol{\beta}}^{\mathbf{LTS}}_{\boldsymbol{\gamma}}$ is chosen, it is the vector that leads to the lowest values of the minimization criterion. The criterion is introduced in the formula (5) (Alfons, Croux and Gelper, 2013).

$$\begin{aligned}\min_{\gamma} \min_{\hat{\boldsymbol{\beta}}^{\mathbf{LTS}}_{\boldsymbol{\gamma}}} {\mathbf{e}^{*}_{\boldsymbol{\gamma}}}^{2} &= \min_{\gamma} \min_{\hat{\boldsymbol{\beta}}^{\mathbf{LTS}}_{\boldsymbol{\gamma}}} \left(\mathbf{y}^{*}_{\boldsymbol{\gamma}} - \mathbf{X}^{*}_{\boldsymbol{\gamma}} \hat{\boldsymbol{\beta}}^{\mathbf{LTS}}_{\boldsymbol{\gamma}}\right)^{2} = \min_{\gamma} \min_{\hat{\beta}_0^{LTS,\gamma}, \hat{\beta}_j^{LTS,\gamma}} \sum_{[m_\gamma]=1}^{r} {e_{\gamma}^{[m_\gamma]}}^{2} = \\ &= \min_{\gamma} \min_{\hat{\beta}_0^{LTS,\gamma}, \hat{\beta}_j^{LTS,\gamma}} \sum_{[m_\gamma]=1}^{r} \left(y^{[m_\gamma]} - \hat{\beta}_0^{LTS,\gamma} - \sum_{j=1}^{k} \hat{\beta}_j^{LTS,\gamma} \times x_j^{[m_\gamma]} \right)^{2}\end{aligned} \tag{5}$$

where:

γ .... identifier of particular subset consisting of r observations, $\gamma \in \langle 1; \Gamma \rangle \wedge \gamma \in \mathbb{N}$ [-],

$\hat{\boldsymbol{\beta}}^{\mathbf{LTS}}_{\boldsymbol{\gamma}}$ .... vector of the estimations of the regression coefficient β numerated by LTS method for the subset γ consisting of p dimensions [-],

$\mathbf{e}^{*}_{\boldsymbol{\gamma}}$ .... residual vector of differences between real empirically found values of y variable and their predicated estimation consisting of r dimensions representing the observations included in subset γ, the order of particular dimensions $e_{\gamma}^{[m_{\gamma}]}$ is set ordinally according to the size of their second powers [-],

$\mathbf{y}_{\boldsymbol{\gamma}}^{*}$............................vector of real empirically found values of dependent variable consisting of r dimensions representing the observations included in subset γ, the order of particular dimenstions $y_{\gamma}^{[m_{\gamma}]}$ is set ordinally according to the size of the second powers of their residuals $e_{\gamma}^{[m_{\gamma}]}$ counted for assessed vector $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}^{\mathbf{LTS}}$ [-],

$\mathbf{X}_{\boldsymbol{\gamma}}^{*}$............................matrix of independent variables of the $(r \times p)$ dimensions, the order of particular rows $[m_{\gamma}]$ is set ordinally according to the size of the second powers of their residuals $e_{\gamma}^{[m_{\gamma}]}$ counted for assessed vector $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}^{\mathbf{LTS}}$ [-],

$\hat{\beta}_{0}^{LTS,\gamma}$.......................the estimation of the intercept term counted by the LTS method for subset γ [-],

$\hat{\beta}_{j}^{LTS,\gamma}$.......................the estimation of the regression coefficient of jth independent variable counted by the LTS method for subset γ [-],

$[m_{\gamma}]$........................identifier of particular observation in the adjusted ordinal order according to the size of the second powers of their residuals $e_{\gamma}^{[m_{\gamma}]}$ counted for asessed vector $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}^{\mathbf{LTS}}$, $[m_{\gamma}] \in \langle 1; n \rangle \wedge [m_{\gamma}] \in \mathbb{N}$ [-],

r..............................trimming constant, the number of observations included in particular subsets γ of the minimization criterion of LTS method, $r \in \langle 1; n \rangle \wedge r \in \mathbb{N}$ [-],

$e_{\gamma}^{[m_{\gamma}]}$........................the residuum of the observation $[m_{\gamma}]$ counted with the use of assessed estimations of regression coefficients β for subset γ [-],

$y^{[m_{\gamma}]}$........................true empirically identified value of the independent variable for the observation $[m_{\gamma}]$ [-],

j..............................identifier of particular input valuable, $j \in \langle 1; k \rangle \wedge j \in \mathbb{N}$ [-],

k..............................regression parameters count, $k \geq 1 \wedge k \in \mathbb{N}$ [-],

$x_{j}^{[m_{\gamma}]}$........................value of jth independent variable for the observation $[m_{\gamma}]$ [-].

3.2.3 Ridge regression

By the multiple regression analysis growing number of variables k leads to a drop in the credibility of the ordinary least squares method, the vectors of input variables aren’t orthogonal, the absence of orthogonality means that multicollinearity appears in the model (Bremer, 2012; Hoerl and Kennard, 1970). Reaching clear orthogonality between all input variables isn’t possible in real world. Then multicollinearity leads to the distortion of values of regression coefficients, instability of regression model a very low model credibility (Dorugade, 2018; Buonaccorsi, 1996).

Ridge regression deflects lightly the regression coefficients estimations, the principle is penalization of such regression coefficients β that evince the marks of multicollinearity (NCSS Statistical Software, 2019; Oleszak, 2019; Breheny, 2011; Golub, Heath and Wahba, 1979). The real effect of the ridge regression is strengthening the diagonal of the $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ matrix as described by Zvára (2008). The mathematical form is provided in the formula (6) according to Xiao, Coots and Ye (2017) and Zvára (2008).

$$\widehat{\boldsymbol{\beta}}_{\mathbf{RR}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}_{\boldsymbol{n}})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \tag{6}$$

where:

$\widehat{\boldsymbol{\beta}}_{\mathbf{RR}}$.......................... vector of the estimations of the regression coefficient β numerated by ridge regression for the subset γ consisting of p dimensions [-],

$\mathbf{X}$.............................. matrix of independent variables of the $(n \times p)$ dimensions [-],

λ.............................. ridge regression constant [-],

$\mathbf{I}_{\boldsymbol{n}}$............................ identity matrix of the $(n \times n)$ dimensions [-],

$\mathbf{y}$.............................. vector of real empirically found values of dependent variable consisting of n dimensions [-].

Van Wieringen (2019) shows that the estimation of vector $\widehat{\boldsymbol{\beta}}_{\mathbf{RR}}$ allows the user to compute the final vector of predicted values $\hat{\mathbf{y}}_{\mathbf{HR}}$ as demonstrated in formula (7).

$$\hat{\mathbf{y}}_{\mathbf{RR}} = \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathbf{RR}} \tag{7}$$

where:

$\hat{\mathbf{y}}_{\mathbf{HR}}$.......................... vector of predicted values of independent variable consisting of n dimensions numerated by ridge regression [-],

$\mathbf{X}$.............................. matrix of independent variables of the $(n \times p)$ dimensions [-],

$\widehat{\boldsymbol{\beta}}_{\mathbf{RR}}$.......................... vector of the estimations of the regression coefficient β numerated by ridge regression for the subset γ consisting of p dimensions [-].

4 PROBLEMS SOLVING

This Chapter contains the description of the execution of the subject of the thesis.

4.1 Model inputs

The model quantifies the influence of particular factors, that were identified in the literature review as factors that determine the public transportation services. Each factor is represented by an appropriate input value. The overview of model inputs is introduced in the Table 1.

Table 1 The overview of all potentially entering inputs to the mathematical-statistical model

Input identifier *j*	**Name and description of the factor (always related to the assessed municipality unit)**	**Binary variable**
1	Inhabitants count	no
2	The women share on total inhabitants count	no
3	The share of economically active population on total inhabitants count	no
4	Unemployment rate	no
5	The count of microbusinesses	no
6	The count of little businesses	no
7	The count of middle businesses	no
8	The count of large businesses	no
9	The count of elementary schools	no
10	The count of secondary schools	no
11	The count of tertiary schools and universities	no
12	The count of municipal units with the inhabitant count within the $\langle A; B)$ range remoted no more then defined distance of D	no
13	The count of municipal units with the inhabitant count within the $\langle B;\infty)$ range remoted no more than defined distance of D	no
14	The distance to the regional center, in specific cases of bicentric regions the sum of distances to both centers can be used	no
15	Administrative part of the regional center	yes
16	Regional border remoted no more than defined distance of E	yes
17	National border remoted no more than defined distance of G	yes
18	The existence of important cultural sight or social-cultural facilities	yes
19	The existence of important natural sight or location within the area of natural protected area	yes
20	The location at the road with the transportation density within the $\langle J; K)$ range	yes
21	The location at the road with the transportation density within the $\langle K; \infty)$ range	yes
22	The location by the crossing of important roads both exceeding the transport density of L, or at the railway crossing where both lines are regularly served by passenger transport	yes
23	The existence of a railway station or stop on the track regularly served by passenger transport	yes
24	The altitude higher than the limit of M	yes
25	Specific factor defined in the critical analysis of previous model version	yes /no

Source: author

Particular limit values used in the Table 1 A, B, D, E, G, J, K, L and M shall be defined in more variants and all of them will be tested during the process as limit values ζ.

4.2 Methodology of the evaluation of the extent of public transport services

The methodology of the evaluation of the extent of public transport services in certain municipality units within a region with the usage of a mathematical-statistical model of public transport services connection count prediction is introduced in this chapter. The scientific methods described in chapter 3 are used in order to create the described methodology.

The methodology has 6 steps that will be shortly described in the subchapters.

4.2.1 Initialization – a decision to create the model

There are two basic types of reasons that lead to the decision to create the model:

1. The evaluation of the reference state, which can be present-time situation or potential alternative as s decision support during public transport optimization.
2. Theoretical point of view on this issue as part of scientific research in order to get better understanding of this area.

The decision about the model creation must be followed by exact specification of evaluated region.

4.2.2 Specification of the municipal units and the public transport stops and stations distribution

During the second step of the methodology the municipal units of the region are specified, their area is precisely defined, and the public transport stops and stations are assigned to the units with the emphasis on real importance of these stops and stations for certain municipal units.

When the public transport stops and stations serve more municipal units, it can lead the solver to the respecification of the municipal units structure by merging these municipal units together for the model.

4.2.3 Data collection

Two data spheres are identified and included to the model.

First data needed for the model are the data about reference public transport services within the analyzed region. The most usual type of the analyzed reference public transport services state is the contemporary situation of its provision. The data is structured as the count of public transport service connections in each single municipal unit in defined time unit, typically the amount of public transport service connections within 24 hours of a standard weekday. The difference between public transport service connections of different character aren’t considered as all of them lead to the same outcome for analyzed municipal unit, this covers the different transport modes as well as different financial background of the connections (provided by a transport company on a commercial basis or as an obligation given by public transport services contract).

The outcome is structured into a vector $\mathbf{y}$, the vector is described in the formula (8).

$$\mathbf{y} = \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix} \tag{8}$$

where:

$\mathbf{y}$ ....vector of reference values of public transport service connections in all n included municipal units within a defined time unit [-],

y^1 ....reference value of public transport service connections in the municipal unit $m = 1$ within a defined time unit [-],

y^n ....reference value of public transport service connections in the municipal unit $m = n$ within a defined time unit [-],

n ....the count of municipal units included to the model, $n \geq 1 \wedge n \in \mathbb{N}$ [-].

The other data needed are the numerical representations of all the factors that determine public transport service connections count, their overview is in Table 1. The data is structured in an $\mathbf{X}_i$ matrix, its structure is introduced in formula (9). The $\mathbf{X}_i$ matrix has N versions, one for each input combinations i. The input combinations are derived from the limit values introduced in subchapter 4.1.

$$\mathbf{X}_i = \begin{pmatrix} 1 & x_1^1 & \cdots & x_k^1 \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_1^n & \cdots & x_k^n \end{pmatrix} \tag{9}$$

where:

$\mathbf{X}_i$.............................matrix of independent variables of the $(n \times p)$ dimensions for the ith combination of inputs consisting of n dimensions [-],

i ....identifier of certain input combination, $i \in \langle 1; N \rangle \wedge i \in \mathbb{N}$ [-],

$x_1^1, \dots, x_k^n$ ....particular independent variables [-],

k ....regression parameters count, $k \geq 1 \wedge k \in \mathbb{N}$ [-],

n ....the count of municipal units included to the model, $n \geq 1 \wedge n \in \mathbb{N}$ [-].

4.2.4 Model construction

Fundamental step of the methodology is the construction of mathematical-statistical model. The main inputs for the model are vector $\mathbf{y}$ and matrices $\mathbf{X}_i$ both prepared in the subchapter 4.2.3. The model construction has three phases:

1. Initial construction of the regression model using the multiple linear regression – standard ordinary least squares method.

2. The second phase specifies the prediction from the first phase by implementing the LTS method of robust regression. This method is implemented in order to manage the existence of outlier in the input data (observations).
3. Third phase than deals with the multicollinearity by using ridge regression.

First phase is based on formulas (3) and (4). The output dimensions of vector $\hat{\mathbf{y}}$ are used to count the residuals, as shown in formula (10).

$$e_{OLS,i}^{m} = y^{m} - \hat{y}_{OLS,i}^{m} \tag{10}$$

where:

$e_{OLS,i}^{m}$ .... the residual in the regression model counted by ordinary least squares regression for the ith combination of inputs for municipal unit m [-],

y^{m} .... reference value of public transport service connections in the municipal unit m within a defined time unit [-],

$\hat{y}_{OLS,i}^{m}$ .... predicted value of public transport service connection in municipal unit m within defined time unit counted by the ordinary least squares regression for the ith combination of inputs [-].

The next phase is including the robust regression which deals in outliers in data inputs. The robust regression has several methods, the LTS method will be used in this case. The LTS method needs the information about the outlier count in the observation data. This count is represented by the trimming constant r. The outliers count is obtained by testing of studentized residuals of each observation. Afterwards the trimming constant r is set as accrding to the formula (11).

$$r_i = n - \rho_i \tag{11}$$

where:

r_i .... trimming constant, the number of municipal units included in particular subsets γ of the minimization criterion of LTS method for the ith combination of inputs, $r \in \langle 1; n \rangle \wedge r \in \mathbb{N}$ [-],

n .... the count of municipal units included to the model, $n \geq 1 \wedge n \in \mathbb{N}$ [-],

ρ_i .... the count of outliers obtained by the studentized residuals for the ith combination of inputs, $\rho_i \in \langle 1; n-2 \rangle \wedge \rho_i \in \mathbb{N}$ [-].

Afterwards the subsets γ are created for each ith combination of inputs. For each combination of subset γ and input combination i the vector of the estimations of the regression coefficient β is counted according to the formula (12).

$$\widehat{\boldsymbol{\beta}}_{\gamma}^{i} = \left(\mathbf{X}_{\gamma}^{i\,\mathrm{T}}\mathbf{X}_{\gamma}^{i}\right)^{-1}\mathbf{X}_{\gamma}^{i\,\mathrm{T}}\mathbf{y}_{\gamma} \tag{12}$$

where:

$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}^{\boldsymbol{i}}$ ....vector of the estimations of the regression coefficient β consisting of p dimensions numerated by ridge regression for the subset γ for the ith combination of inputs [-],

$\mathbf{X}_{\boldsymbol{\gamma}}^{\boldsymbol{i}}$ ....matrix of independent variables of the $(r_i \times p)$ dimensions for the ith combination of inputs containing only such municipal units m that were included in the subset γ [-],

$\mathbf{y}_{\boldsymbol{\gamma}}$ ....vector of reference values of public transport service connections within a defined time unit consisting of r_i dimensions containing only such municipal units m that were included in the subset γ [-].

Based on vector $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}^{\boldsymbol{i}}$ the values of $\hat{y}_{\gamma,i}^{m}$ and $e_{\gamma,i}^{m}$ with the usage of analogies to formulas (4) and (10). For each subset γ the ordinal order of second powers of the resiudals $e_{\gamma,i}^{[m_{\gamma}^{i}]}$ for each municipal unit included in such subset γ is set. The order is then used in LTS method as described in the subchapter 3.2.2. Once the optimal subset γ_{opt} is identified, matching vector $\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{\mathbf{LTS}}$ is found and this vector is used to count the vector $\hat{\mathbf{y}}_{\boldsymbol{i}}^{\mathbf{LTS}}$ as described in formula (13).

$$\hat{\mathbf{y}}_{\boldsymbol{i}}^{\mathbf{LTS}} = \mathbf{X}_{\boldsymbol{i}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{\mathbf{LTS}} \tag{13}$$

where:

$\hat{\mathbf{y}}_{\boldsymbol{i}}^{\mathbf{LTS}}$ ....vector of the predicted values of public transport service connections within defined time unit consisting of n dimensions counted by the LTS method of robust regression for the ith combination of inputs [-],

$\mathbf{X}_{\boldsymbol{i}}$ ....matrix of independent variables of the $(n \times p)$ dimensions for the ith combination of inputs consisting of n dimensions [-],

$\widehat{\boldsymbol{\beta}}_{\boldsymbol{i}}^{\mathbf{LTS}}$ ....vector of the estimations of the regression coefficient β consisting of p dimensions counted by the LTS method of robust regression for the ith combination of inputs [-].

The valuables of $\hat{y}_{LTS,i}^{m}$ and $e_{LTS,i}^{m}$ can be counted afterwards with the usage of analogies to formulas (4) and (10). The calculation follows by calculating the residual sum of squares for each ith combination of inputs as shown in formula (14).

$$RSS_i^{LTS} = \sum_{\substack{m=1 \\ m \in U_i}}^{n} {e_{LTS,i}^{m}}^2 = \sum_{\substack{m=1 \\ m \in U_i}}^{n} \left(y^m - \hat{y}_{LTS,i}^{m}\right)^2 \tag{14}$$

where:

RSS_i^{LTS} ....residual sum of squares for the ith combination of inputs constructed only for such municipal units m that are included in the optimum subset γ_{opt} [-],

m .... identifier of particular municipal unit, $m \in \langle 1; n \rangle \wedge m \in \mathbb{N}$ [-],

U_i .... set of municipal units m that are included in the optimum subset γ_{opt} of the robust regression LTS method for the ith combination of inputs [-],

n .... total count of municipal units in the analyzed region, $n \geq 1 \wedge n \in \mathbb{N}$ [-],

$e_{LTS,i}^{m}$ .... the residual for the municipal unit m in the model counted only for such municipal units m that were included in the optimum subset γ_{opt} of the robust regression LTS method for the ith combination of inputs [-],

y^m .... reference value of public transport service connections in the municipal unit m within a defined time unit [-],

$\hat{y}_{LTS,i}^{m}$ .... predicted value of public transport service connections in municipal unit m within defined time unit counted by the robust regression LTS method for the ith combination of inputs [-].

Once all the RSS_i^{LTS} counted the minimum value of RSS_i^{LTS} for each set θ_ζ can be found. The sets θ_ζ are defined for each defined limit value ζ. Finding the ith combination with lowest RSS_i^{LTS} in each set θ_ζ allows the user to introduce the final optimum combination of limit values ζ_{opt}. The LTS method is counted once again for the optimum limit values ζ_{opt} that create the matrix $\mathbf{X}_{\zeta_{opt}}$, the outcome of the calculation is new vector $\hat{\boldsymbol{\beta}}_{\zeta_{opt}}^{\mathbf{LTS}}$.

The statistical t test follows up to identify the variables with statistically significant influence. After carrying out the t tests the set of V that includes v values which represent all the variables that their statistically significant influence wasn't proven. Afterwards new matrix $\mathbf{X}^{\mathbf{opt}}$ that replaces the matrix $\mathbf{X}_{\zeta_{opt}}$ in upcoming formulas is created, the rules for $\mathbf{X}^{\mathbf{opt}}$ are described clearly in the formula (15).

$$\mathbf{X}^{\mathbf{opt}} = \left(x_j^m\right) \text{ for } \forall\, j \in Q \wedge\, j \notin V \wedge \text{ for } \forall\, m \tag{15}$$

where:

$\mathbf{X}^{\mathbf{opt}}$ .... matrix of the final values of input independent variables with proven statistically significant influence on the final count of the public transport service connections within a defined time unit of the $[n \times (p - v)]$ [-],

x_j^m .... the value of jth input for the municipal unit m [-],

j .... identifier of particular input independent variable, $j \in \langle 1; k \rangle \wedge j \in \mathbb{N}$ [-],

Q .... the set of input variables j that are included in the matrix $\mathbf{X}_{\zeta_{opt}}$ that is compliant with optimum limit values ζ_{opt} [-],

V .... the set of input variables j that were proven to have no statistically significant influence on the final count of the public transport service connections within a defined time unit [-],

m ....identifier of particular municipal unit, $m \in \langle 1; n \rangle \wedge m \in \mathbb{N}$, where n ($n \geq 1 \wedge n \in \mathbb{N}$) is total count of municipal units in the analyzed region [-].

Analogically to formula (15) the matrix $\mathbf{X}_{\mathbf{LTS}}^{\mathbf{opt}}$ is created. This matrix has same characteristics, but only $m \in U_{\zeta_{opt}}$ are included. $U_{\zeta_{opt}}$ is the set of municipal units that are included in the optimum subset γ_{opt}.

The ridge regression is included in the third phase. The ridge regression manages the problem of multicollinearity that occurs in the input data vectors. The ridge regression constant λ is counted by the leave-one-out version of generalized cross-validation. The ridge regression is then carried out using the formulas (6) and (7). Final values of the predicted values of public transport service connections in municipal units m within defined time unit is counted according to formula (16).

$$\hat{y}_{RR}^{m} = \hat{\beta}_{0}^{RR} + \sum_{\substack{j=1 \\ j \notin V}}^{k} \hat{\beta}_{j}^{RR} \times x_{j}^{m} \tag{16}$$

where:

$\hat{y}_{RR}^{m}$ ....predicted value of public transport service connections in municipal unit m within defined time unit counted by the ridge regression [-],

$\hat{\beta}_{0}^{RR}$ ....the estimation of the intercept term counted by the ridge regression [-],

j ....identifier of particular input independent variable, $j \in \langle 1; k \rangle \wedge j \in \mathbb{N}$ [-],

V ....the set of input variables j that were proven to have no statistically significant influence on the final count of the public transport service connections within a defined time unit [-],

k ....regression parameters count, $k \geq 1 \wedge k \in \mathbb{N}$ [-],

$\hat{\beta}_{j}^{RR}$ ....the estimation of the regression coefficient of jth independent variable counted by the ridge regression [-],

x_{j}^{m} ....the value of jth input for the municipal unit m [-].

The value $\hat{y}_{RR}^{m}$ as the result of formula (16) represents the main outcome from the model. The residual e_{RR}^{m} shall be counted afterwards by using the analogy of formula (10).

The model optimization is based on the multiply linear regression and ordinary least squares method, however it is adapted to the need to reflect the real state of different factors that determine public transportation services in analyzed region. The model is robust enough to prevent outliers from spoiling the outcome data and ridge regression removes ominous influence of the unavoidable multicollinearity.

4.2.5 *Model assessment*

The statistical model is subject to tests of statistical credibility afterwards, the adjusted index of determination is counted as well as t tests and F test are carried out.

The second phase of this step is critical assessment of the model and models comparison with the assumptions and experience. The critical assessment may lead to identification of new factor that was omitted when the first set of public transport determining factors were firstly set.

In case that statistical verification is negative or the logical assessment leads to such discrepancies that cause serious doubts about the models credibility it can lead to rejecting all the models outcome.

4.2.6 *Outcomes interpretation*

Final part of the methodology is the outcomes interpretation, the verbal explanation of values from the model leads to understanding the real importance of the outcomes and also it sketches the logical bonds in the model. The degree of influence of each factor can be counted for each municipal unit. Also, the residuals shall be discussed, and the larger residuals explained.

4.2.7 *Methodology scheme*

The schematic representation of the methodology of the evaluation of the extent of public transport services in certain municipality units within a region with the usage of a mathematical-statistical model of public transport services connection count prediction is provided in the Figure 1. The scheme shows basic structure of the methodology, it doesn't cover all the details, only main steps. The scheme includes the connection to the subchapters that describe particular step of the methodology in wider details.

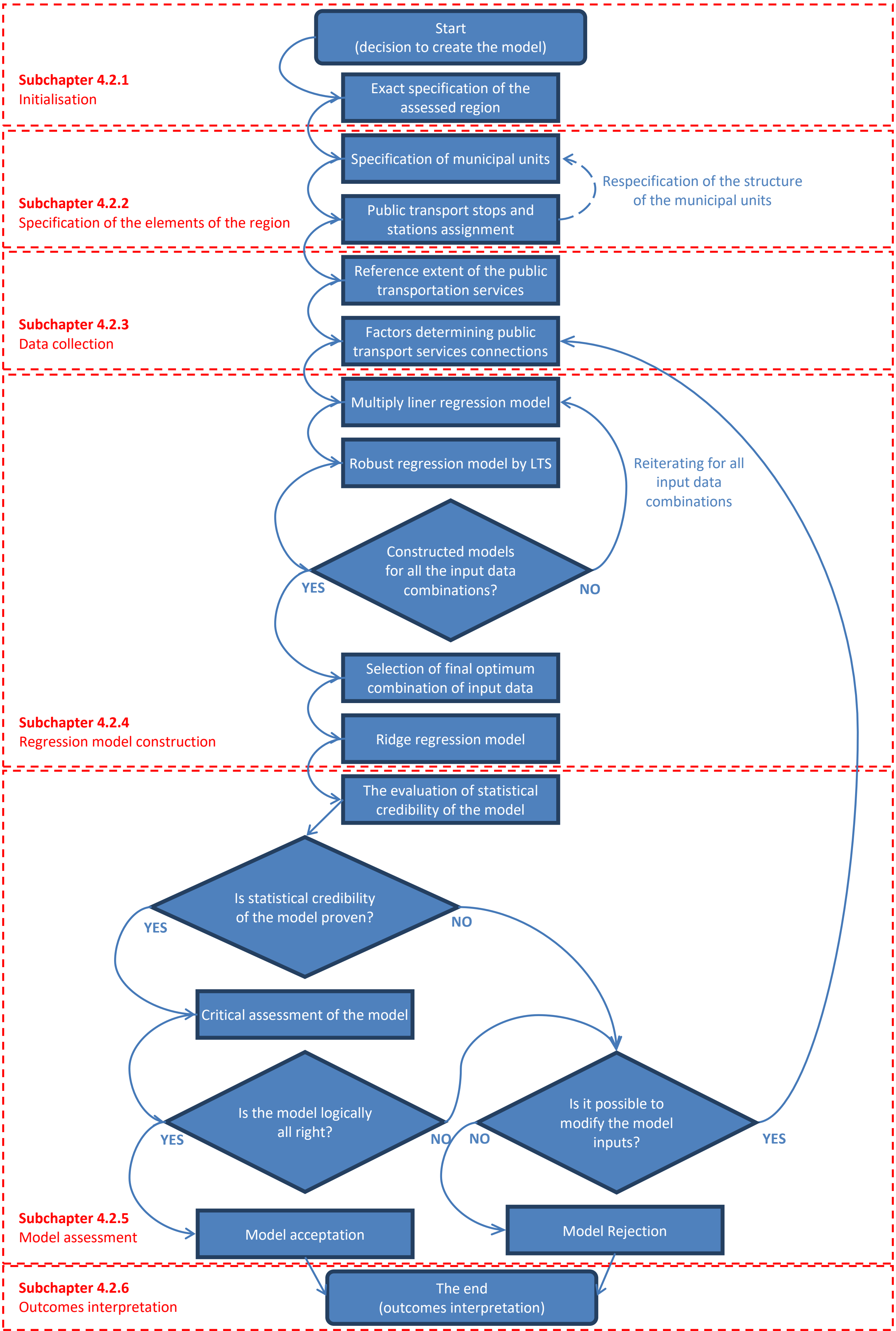

Figure 1 Schematic representation of the methodology (author)

4.3 Methodology application

Within the scope of the doctoral thesis the methodology was applied on a real data from chosen region of the Czech Republic. The application was carried out on the data from the district of Ústí nad Labem in northwest part of the Czech Republic in the Ústecký region. The methodology used the data updated to 1 August 2019. The acceptable walking distance was set to 1 km. According to Czech Statistical Office 207 basic municipal units were identified, out of them 97 were parts of the regional center of Ústí nad Labem. After the analysis 95 municipal units enter the model. Out of the 95 municipal units, the outcomes for 15 chosen municipal units is presented in Table 2.

Table 2 Final estimations of the public transport services connection counts in chosen municipal units m and the residuald of these estimations including the comparison with reference state

m	**Municipal unit**	**Reference count of public transport services connection y^m**	**The estimation of the public transport services connection count $\hat{y}_{RR}^m$**	**The residual e_{RR}^m**
2	Arnultovice	25,0	22,54	2,46
9	Čeřeniště	5,0	22,42	-17,42
11	Český Újezd	130,0	108,67	21,33
12	Dolní Zálezly	46,0	33,63	12,37
14	Dubice	18,0	27,00	-9,00
26	Knínice	42,0	37,04	4,96
32	Libouchec	113,0	75,60	37,40
37	Lysá	4,0	7,50	-3,50
52	Ostrov	0,0	7,54	-7,54
53	Petrovice	40,0	37,53	2,47
66	Řehlovice	44,0	53,29	-9,29
67	Řetouň	12,0	7,85	4,15
68	Sebuzín	68,0	83,48	-15,48
80	Telnice	29,0	24,39	4,61
84	Ústí nad Labem	2.498,4	2.374,14	124,26

Source: Dopravní podnik města Ústí nad Labem (2019), CHAPS and Ministerstvo dopravy ČR (2019), SŽDC (2019), author

The adjusted determination index reaches the value of 88,71 %, that is acceptable. Statistical tests approve the credibility of the model.

5 RESULTS AND DISCUSSION

In this thesis a methodology of the evaluation of the extent of public transport services in certain municipality units within a region with the usage of a mathematical-statistical model of public transport services connection count prediction has been proposed. This methodology has been proposed based on the results of the analysis of the current situation in the researched area in the Czech Republic and abroad in compliance with the findings of the analysis. Public transport services are essential for contemporary society, they bring crucial benefits especially in the social area and at the same time it’s the only competitive alternative to individual motorism that is widely understood as environmentally friendly. Its importance increases in suburb and rural areas, where often it is the only possibility to access civic amenities. Public transport services aren’t viable on commercial basis in most cases, they have to subsidized through public service obligations contracts. The system has long history and in general it’s rather quality system, the problem is its regional inconsistency. The issue of the thesis is very actual.

There are three general areas of usage of the proposed methodology. The first one is the practical application on particular regions according to the needs of the competent authorities managing public transport services, especially the organizers of integrated transport systems and transport responsible at the regions. The usage of such methodology can be very good decision supporting tool when deciding about public transport services and their extent. Also, it can be useful when defining standard and above-standard level of public transport services.

The second are of usage is the field of theoretical transport research for science, it is first tool that can help to understand which factors really influence the final extent of public transport services. The third area is strongly connected to the second one, it’s the field of education.

The overview of pros and cons of the methodology follows in the Table 3.

Table 3 Pros and cons of the methodology

Pros	Cons
High statistical credibility of the prediction	Limits given by the regression function
Stability and logical integrity of the prediction	The necessity to discard fully linearly dependent input variables
Models modularity	Extreme computation demanding
Universal applicability	Model can’t include all the existing factors, such as political ones
Simplicity of the construction	The accessibility and recency of input data

Source: author

Several ways of future development of the model are to be mentioned:

1. First area is the implementation of detailed characteristics of included public transport connections. This could cover their routing (towards the regional center,

from the regional center, round connection) as well as more accurate time description (peak/off-peak hour).

2. The second area leads to transforming the model into the normative one. The universal descriptive model can be easily transformed into normative one that would enable the competent authorities to model easily public transport services structure. Normative parameters would have to be introduced and implemented.
3. The third area covers the implementation of socio-economical characteristics that would enable the weighted comparison of particular regions based on their economical productivity.

In compliance with these three areas of future development there is still vast potential to implement new factors to the model and to find mathematical ways of quantification to include the factors that are known to determine the public transport service, but there is no known way how to transform them into values that could be used as inputs in the model. Also, minor simplifications can be done in the model after wider user experience, when it’s obvious that certain changes can lead to reducing the complicative manners of the model with only weak influence on the results if any influence at all.

6 AUTHOR'S OWN CONTRIBUTION

The doctoral thesis brings three major areas of author’s own contribution. First of them is the thorough and detailed literature review analyzing both on Czech and foreign literature sources. The literature review focuses on public transportation services and their position in context of transport in general. The emphasis is put on the importance of public transport services in contemporary society, especially from the point of view of social aspects. The stronger accent is kept on the area of factors determining public transport services, that is cornerstone of the literature review. Also transport policy and the actual political situation in the Czech Republic is covered. The literature review is widely used as a base for other research in this scientific field.

Second major area of author’s contribution is the analysis and choice of particular prognostic methods derived from basic liner regression. During the process of methodology creation several scientific methods based on regression were tested on particular regions. As the result of this analysis author introduces the issue of linear regression, it’s mathematical formulas and systematics and its application into practical problem. Considering the characteristic of input data the higher penetration of outliers would spoil the outcomes and also refuse the essential condition of normal statistical distribution of the residuals, thus robust regression is implemented. Several robust regression methods were analyzed and finally the LTS method was chosen. The outliers were followed by another problematic issue, the multicollinearity. Few methods were tested, among Partial Least Squares, Least Absolute Shrinkage and Selection Operator, Best Subset Regression and Ridge Regression the ridge regression was chosen because of higher fidelity and credibility, but still ability to interpret the influence of particular factors.

Third major area of author’s contribution is crucial one from the point of view of this doctoral thesis and it is designing a methodology of the evaluation of the extent of public transport services in certain municipality units within a region with the usage of a mathematical-statistical model of public transport services connection count prediction. The methodology leads to designing a model, that predicts the estimations of public transport service connection counts in particular municipality units in the analyzed region within defined time unit, with high credibility. Several control mechanisms are included in the methodology too, these mechanisms focus not only on mathematically correct way, but also on the logical connections and critical analysis of the outcome. The model is universal decision supporting tool for all public transport services organizers.

7 CONCLUSION

The public transport in the Czech Republic haven’t fully adapted yet to the complex major changes that happened in last 30 years in the structure of national economy and political and social life, however dynamical development is already setting new challenges of the social-cultural as well as economical kind. Its character as social service is emphasized recently, it leads to the essential mobility, whole this point is gaining more importance in the future prospect in compliance with society aging, living standard growth and paradigm change in the relationship to the environment and its harming by the society.

The aim of this doctoral thesis was designing a methodology of the evaluation of the extent of public transport services in certain municipality units within a region with the usage of a mathematical-statistical model of public transport services connection count prediction. The methodology was designed by the author, described in very detailed way and applied on real data from the Ústí nad Labem district. The model which is an outcome from the methodology reaches high statistical credibility and it is universally usable. The aim was thoroughly fulfilled.

Introduced methodology is widely usable as a decision supporting tool for the competent authorities responsible for public transport services in each region. At the same time constructed model brings valuable information for the theoretical research in the area of transport sciences. The scientific work can focus on further development of the model.

Authors ambition is to use this doctoral thesis as a starting point for wider discussion about the importance of public transport services provided in regions of the Czech Republic and complex point of view on transport in general, regional disparities and reasons that lead to them, extent of public transport services in particular regions and general purpose of public transport services supply and its priorities in future development.

8 REFERENCES

ALFONS, Andreas, Christophe CROUX and Sarah GELPER, 2013. Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *The Annals of Applied Statistics*. Vol. 7, no. 1, s. 226-248. ISSN 1932-6157.

ANDĚL, Jiří, 2005. *Základy matematické statistiky*. Praha: Matfyzpress. ISBN 80-86732-40-1.

BLATNÁ, Dagmar, 2008. Metody statistické analýzy. Vyd. 3. Praha: Bankovní institut vysoká škola. ISBN 978-80-7265-129-0.

BONDELL, Howard D. and Leonard A. STEFANSKI, 2013. Efficient Robust Regression via Two-Stage Generalized Empirical Likelihood. *Journal of the American Statistical Association*. Vol. 108, no. 502, s. 644-655. ISSN 0162-1459.

BREHENY, Patrick, 2011. Ridge Regression. *College of Arts & Science – University of Kentucky* [online]. [citováno 2016-12-15]. Dostupné z: http://web.as.uky.edu/statistics/users/pbreheny/764-f11/notes/9-1.pdf

BREMER, Martina, 2012. Multiple Linear Regression. *Jason Mezey Lab – Cornel University* [online]. [citováno 2017-06-22]. Dostupné z: http://mezeylab.cb.bscb.cornell.edu/labmembers/documents/supplement%205%20-%20multiple%20regression.pdf

BUONACCORSI, John P., 1996. A modified estimating equation approach to correcting for measurement error in regression. *Biometrika*. Vol. 83, no. 2, s. 433-440. ISSN 0006-3444.

DOORNIK, Jurgen A., 2011. Robust Estimation Using Least Trimmed Squares. *Department of Economics and Business Economics, School of Business and Social Sciences | Aarhus University* [online]. [citováno 2018-11-03]. Dostupné z: http://econ.au.dk/fileadmin/site_files/filer_oekonomi/subsites/creates/Seminar_Papers/2011/ELTS.pdf

DOPRAVNÍ PODNIK MĚSTA ÚSTÍ NAD LABEM, 2019. Jízdní řády. *Dopravní podnik města Ústí nad Labem* [online]. [citováno 2019-08-02]. Dostupné z: https://dpmul.cz/?page=jizdni-rady

DORUGADE, Ashok Vithoba, 2018. New ridge parameters for ridge regression. *Journal of the Association of Arab Universities for Basic and Applied Sciences.* Vol. 15, no. 1, s. 94-99. ISSN 1815-3852.

FRANC, Jiří, 2011. Robust regression - Robust estimation of regression coefficients in linear regression model when orthogonality condition is breaking. *Univerzita Karlova* [online]. [citováno 2017-06-20]. Dostupné z: http://ai.ms.mff.cuni.cz/~sui/franc.pdf

GAUß, Carl Friedrich, 1809. Theoria motus corporum coelestium in sectionibus conicis solem ambientium. *ETH Bibliothek Zürich* [online]. [citováno 2016-05-11]. Dostupné z: https://www.e-rara.ch/zut/doi/10.3931/e-rara-522

GOLUB, Gene H., Michael HEATH and Grace WAHBA, 1979. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics.* Vol. 21, no. 2, s. 215-223. ISSN 0040-1706.

HEBÁK, Petr, 1998. *Regrese. Část 1*. Praha: Vysoká škola ekonomická v Praze. ISBN 80-7079-909-9.

HEBÁK, Petr and Alžběta SVOBODOVÁ, 2001. *Regrese - II. část*. Praha: Vysoká škola ekonomická. ISBN 80-245-0134-1.

HOERL, Arthur E. and Robert W. KENNARD, 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics.* Vol. 12, no. 1, s. 55-67. ISSN 0040-1706.

CHAPS and MINISTERSTVO DOPRAVY ČR, 2019. Celostátní informační systém o jízdních řádech. *CHAPS spol. s r.o. – Portál CIS JŘ* [online]. [citováno 2019-06-12]. Dostupné z: http://portal.cisjr.cz

JIRSÁK, Petr, Michal MERVART and Marek VINŠ, 2012. *Logistika pro ekonomy: vstupní logistika.* Praha: Wolters Kluwer ČR. ISBN 978-80-7357-958-6.

NCSS STATISTICAL SOFTWARE, 2019. Chapter 335: Ridge Regression. *NCSS Statistical Software* [online]. [citováno 2019-04-10]. Dostupné z: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Ridge_Regression.pdf

NEUBAUER, Jiří, 2016. Lineární modely | Ekonometrie. *Katedra ekonometrie FVL UO Brno* [online]. [citováno 2018-10-21]. Dostupné z: https://k101.unob.cz/~neubauer/pdf/ekon_linearni_modely.pdf

OLECKÁ, Ivana and Kateřina IVANOVÁ, 2010. *Metodologie vědecko-výzkumné činnosti*. Olomouc: Moravská vysoká škola Olomouc. ISBN 978-80-87240-33-5.

OLESZAK, Michał, 2019. A Comparison of Shrinkage and Selection Methods for Linear Regression. *Towards Data Science* [online]. [citováno 2019-06-07]. Dostupné z: https://towardsdatascience.com/a-comparison-of-shrinkage-and-selection-methods-for-linear-regression-ee4dd3a71f16

ROUSSEEUW, Peter J. and Victor J. YOHAI, 1984. Robust Regression by Means of S-Estimators. In: FRANKE, Jürgen, Wolfgang HÄRDLE a Douglas MARTIN, eds. *Robust and Nonlinear Time Series Analysis.* New York, NY: Springer US, s. 256-272. ISBN 978-0-387-96102-6.

SŽDC, 2019. Jízdní řád. *Správa železniční dopravní cesty* [online]. [citováno 2019-08-03]. Dostupné z: https://www.szdc.cz/cestujici/jizdni-rad

TVRDÍK, Josef, 2013. Analýza vícerozměrných dat. *Ostravská univerzita v Ostravě* [online]. [citováno 2017-10-30]. Dostupné z: http://www1.osu.cz/~bujok/files/avdat.pdf

VAN WIERINGEN, Wessel N., 2019. Lecture notes on ridge regression. *Cornell University* [online]. [citováno 2019-08-05]. Dostupné z: https://arxiv.org/pdf/1509.09169;Lecture

XIAO, Ruikun, Reed COOTS and Yuzhe YE, 2017. Ridge Regression | R. *JB Hender* [online]. [citováno 2019-04-12]. Dostupné z: https://jbhender.github.io/Stats506/F17/Projects/G13/R.html

ZVÁRA, Karel, 2002. Regrese. *Univerzita Karlova* [online]. [citováno 2017-06-20]. Dostupné z: http://www.karlin.mff.cuni.cz/~zvara/regrese/PREDN01.pdf

ZVÁRA, Karel, 2008. *Regrese*. Praha: Matfyzpress. ISBN 978-80-7378-041-8.

9 LIST OF AUTHOR'S PUBLICATIONS RELATED TO THE FIELD OF THE DOCTORAL THESIS

TRPIŠOVSKÝ, Martin and Petr PRŮŠA, 2012. Užití CBA pro hodnocení investic do dopravy. In: *LOGI 2012 – Conference Proceeding*. Brno: Tribun EU, s. 340-350. ISBN 978-80-263-0336-7.

ČÁP, Jiří, Jindřich JEŽEK and Martin TRPIŠOVSKÝ, 2013. Plány dopravní obslužnosti krajů ČR. In: *Horizons of Railway Transport 2013*. Žilina: EDIS - vydavateľstvo Žilinskej univerzity, s. 93-102. ISBN 978-80-554-0764-7.

TRPIŠOVSKÝ, Martin, Jan CHOCHOLÁČ, Jindřich JEŽEK and Petr PRŮŠA, 2013. Rovnoběžné trasy MHD obsluhující sídliště. In: *Current trends in transport and economy in 2013 (Aktuální trendy v dopravě a ekonomice 2013)*. Pardubice: Univerzita Pardubice, s. 203-217. ISBN 978-80-86530-90-1.

TRPIŠOVSKÝ, Martin, Jindřich JEŽEK and Petr PRŮŠA, 2013. Faktory determinující dopravní obslužnost v městské zástavbě. In: *Current trends in transport and economy in 2013 (Aktuální trendy v dopravě a ekonomice 2013)*. Pardubice: Univerzita Pardubice, s. 47-56. ISBN 978-80-86530-90-1.

TRPIŠOVSKÝ, Martin, Radhika JADE and Petr PRŮŠA, 2013. Public transport services in the Czech Republic. *Focus: The International Journal of Management Digest*. Vol. 9, no. 2, s. 6-14. ISSN 0973-9165.

TRPIŠOVSKÝ, Martin, Tomáš RÝC a Petr PRŮŠA, 2013. Modelování dopravní obslužnosti obcí. In: *Sborník příspěvků z mezinárodní Masarykovy konference pro doktorandy a mladé vědecké pracovníky 2013*. Hradec Králové: MAGNANIMITAS, s. 765-774. ISBN 978-80-87952-00-9.

TRPIŠOVSKÝ, Martin, Jan CHOCHOLÁČ and Darko BABIĆ, 2013. City public transport infrastructure investments determining factors. In: *ZIRP 2013*. Zagreb: University of Zagreb, s. 111-122. ISBN 978-953-243-064-6.

TRPIŠOVSKÝ, Martin, Jan CHOCHOLÁČ and Petr PRŮŠA, 2013. Plány dopravní obslužnosti krajů ČR. In: *Dopravní obslužnost měst a krajů*. Praha: B.I.D. services, s. 17-35. ISBN 978-80-87534-51-9.

GĄSKA, Damian, Martin TRPIŠOVSKÝ and Maria CIEŚLA, 2013. Comparison of public transport services organization in the Prague and Warsaw metropolitan regions. In: *II International Symposium of Young Researchers TRANSPORT PROBLEMS 2013*. Katowice: Silesian University of Technology, s. 87-96. ISBN 978-83-935232-1-4.

TRPIŠOVSKÝ, Martin and Petr PRŮŠA, 2014. Regional Public Transportation Services Modelling. *Naše more: Journal of Marine Science*. Vol. 61, no. 3-4, s. 77-82. ISSN 0469-6255.

HUSÁK, Jiří, Dalibor GOTTWALD and Martin TRPIŠOVSKÝ, 2014. Rozvoj dopravní infrastruktury ve vybraných regionech České republiky. In: *Sborník příspěvků Mezinárodní Masarykovy konference pro doktorandy a mladé vědecké pracovníky 2014*. Hradec Králové: MAGNANIMITAS, s. 1213-1222. ISBN 978-80-87952-07-8.

GĄSKA, Damian, Martin TRPIŠOVSKÝ and Maria CIEŚLA, 2015. Comparison of public transport services organization in the Prague and Warsaw metropolitan regions. *Zeszyty Naukowe Politechniki Śląskiej: Transport*. Vol. 86, no. 1, s. 21-32. ISSN 0209-3324.

TRPIŠOVSKÝ, Martin, 2015. Public transportation services determining factors in the urban areas. In: *ZIRP 2015*. Zagreb: University of Zagreb, s. 219-224. ISBN 978-953-243-073-8.

CHOCHOLÁČ, Jan, Martin TRPIŠOVSKÝ and Dana SOMMERAUEROVÁ, 2016. Komparace marketingového mixu železničních dopravců v souvislosti s dopravní obslužností na trase Praha - Ostravsko v souladu s principy udržitelného rozvoje. In: *Aktuální trendy v dopravě a ekonomice.* Pardubice: Univerzita Pardubice, s. 246-251. ISBN 978-80-8653-095-6.

CHOCHOLÁČ, Jan, Martin TRPIŠOVSKÝ and Nina KUDLÁČKOVÁ, 2018. The evaluation of the service quality performed by the rail passenger transport carriers on the Prague – Ostrava region route: primary marketing research. In: *Transport Means: proceedings of the international scientific conference.* Kaunas: Kaunas University of Technology., s. 91-106. ISSN 1822-296X.