

RĪGAS TEHNISKĀ UNIVERSITĀTE

Transporta un mašīnzinību fakultāte

Diāna SANTALOVA

Transporta sistēmas tehniskā nodrošinājuma doktora programmas doktorante

**PUSPARAMETRISKIE REGRESIJAS
MODEĻI KRAVU UN PASAŽIERU
PĀRVADĀJUMU APJOMU
ANALĪZEI UN PROGNOZĒŠANAI**

Promocijas darba kopsavilkums

Zinātniskais vadītājs
Dr.hab.sc.ing., profesors
A. ANDRONOVS

**RTU Izdevniecība
Rīga 2009**

UDK [656.025+519.233.5](043.2)

Sa 611 p

Santalova D.

Sa 611 p Pusparametriskie regresijas modeļi kravu un pasažieru pārvadājumu apjomu analīzei un prognozēšanai. Promocijas darba kopsavilkums. – R.: RTU, 2009. – 43 p.

Iespiests saskaņā ar “RTU P-20” promocijas padomes 2009. gada 30.septembra lēmumu, protokols Nr. 01/2009.

Šis darbs izstrādāts ar Eiropas Sociālā Fonda atbalstu Nacionālās programmas “Atbalsts doktorantūras programmu īstenošanai un pēcdoktorantūras pētījumiem” projekta “Atbalsts RTU doktorantūras attīstībai”.

Šis darbs izstrādāts ar Eiropas Sociālā Fonda atbalstu projektā „Atbalsts RTU doktora studiju īstenošanai”.

ISBN XXX

**PROMOCIJAS DARBS
IZVIRZĪTS RĪGAS TEHNISKAJĀ UNIVERSITĀTĒ
INŽENIERZINĀTŅU DOKTORA GRĀDA IEGŪŠANAI**

Promocijas darbs inženierzinātņu doktora grāda iegūšanai tiek publiski aizstāvēts 2009. gada 7. decembrī plkst. 14:30, Rīgas Tehniskās universitātes Transportmašīnu tehnoloģiju institūtā Lomonosova ielā 1, korpusā V, 218. auditorijā.

OFICIĀLIE OPONENTI:

Profesors, *Dr.sc.ing.* Pēteris Balckars
Rīgas Tehniskā universitāte, Latvija

As. profesors, *Dr.sc.ing.* Viktors Ļumkis
Transporta un telekomunikāciju institūts, Latvija

Profesors, *Dr.phys.math.* Nikolajs Ušakovs
Norvēģijas Zinātņu un tehnoloģijas universitāte, Trondheima, Norvēģija

APSTIPRINĀJUMS

Apstiprinu, ka esmu izstrādājusi doto promocijas darbu, kas iesniegts izskatīšanai Rīgas Tehniskajā universitātē inženierzinātņu doktora grāda iegūšanai. Promocijas darbs nav iesniegts nevienā citā universitātē zinātniskā grāda iegūšanai.

Diāna Santalova

Datums: 2009. gada 8. oktobris

Promocijas darbs ir uzrakstīts angļu valodā, satur ievadu, 7 nodaļas, nobeigumu, literatūras sarakstu, 2 pielikumus un glosāriju. Darbā ietilpst 147 formulas, 60 tabulas, 30 zīmējumi, kopā 162 teksta lappuses. Literatūras sarakstā ir 128 nosaukumi.

ANOTĀCIJA

Promocijas darbu «Pusparametriskie regresijas modeļi kravu un pasažieru pārvadājumu apjomu analīzei un prognozēšanai» izstrādājusi Diāna Santalova inženierzinātņu doktora zinātniskā grāda iegūšanai. Darba zinātniskais vadītājs ir *Dr. hab. sc. ing.*, profesors Aleksandrs Andronovs.

Darba mērķis ir dažādu transporta veidu pasažieru un kravu pārvadājumu apjomu prognozēšana ar mūsdienu statistisko metožu pielietošanas palīdzību. Patlaban visā pasaulē novērojama gan ražošanas decentralizācija, gan arī iedzīvotāju mobilitātes pieaugums. Tā rezultātā novērojams visu transporta veidu pasažieru un kravu pārvadājumu apjomu nepārtraukts pieaugums. Šādi globālie procesi nosaka izvēlētā pētnieciskā darba virziena perspektivitāti, tātad darbs ir aktuāls.

Par pirmo pētījuma virzienu kļuva grupu modeļu izstrādāšana un verifikācija pārvadājumu apjomu prognozēšanai ES valstīm. Šajā ziņā bija jāatrisina četri uzdevumi:

- summāro kravu aviopārvadājumu apjomu prognozēšana uz parametriskā daudzdimensiju modeļa bāzes;
- dzelzceļa pasažieru nosūtījumu korespondenču matricas novērtēšana ar modificētā parametriskā gravitācijas modeļa palīdzību;
- dzelzceļa kravu apgrozījuma novērtēšana un prognozēšana uz pusparametriskā viena indeksa modeļa bāzes;
- summāro pasažieru aviopārvadājumu apjomu prognozēšana datu nepilnības gadījumā ar parametriskā SURE-modeļa izmantošanu.

Otro virzienu veidoja grupu pusparametrisko modeļu izstrādāšana un izvērtēšana pasažieru dzelzceļa nosūtījumu analīzei un prognozēšanai Latvijas reģioniem.

Tika piedāvātas jaunas modeļu novērtēšanas metodes, izstrādāti veidojamo modeļu kvalitātes verifikācijas kritēriji. Parādītas viena indeksa modeļu pielietošanas priekšrocības salīdzinājumā ar lineārās regresijas modeļiem.

SATURS

1. PROBLĒMAS AKTUALITĀTE.....	6
2. PĒTĪJUMA MĒRĶI UN UZDEVUMI.....	7
3. TĒMAS PAŠREIZĒJĀS IZPĒTES LĪMENIS.....	7
4. METODOLOĢIJA UN PĒTNIECISKĀS METODES	13
5. ZINĀTNISKĀ NOVITĀTE.....	13
6. PRAKTISKĀ NOZĪME UN REALIZĀCIJA	13
7. PROMOCIJAS DARBA STRUKTŪRA	14
8. PĒTĪJUMU REZULTĀTU KOPSAVILKUMS	15
8.1. STARPTAUTISKO KRAVU AVIOPĀRVADĀJUMU ANALĪZE UN PROGNOZĒŠANA ES VALSTĪM UZ DAUDZDIMENSIJU MODEĻA BĀZES	15
8.2. DZELZCEĻA PASAŽIERU KORESPONDENČU IZVĒRTĒŠANA STARP ES VALSTĪM UZ MODIFICĒTĀ GRAVITĀCIJAS MODEĻA BĀZES.....	19
8.3. VIENA INDEKSA MODEĻA PIELIETOŠANA PĀRVADĀJUMU APJOMU IZPĒTEI.....	23
8.3.1. VIENA INDEKSA MODEĻA NOVĒRTĒŠANA	23
8.3.2. STARPTAUTISKĀ KRAVU DZELZCEĻA TRANSPORTA APGROZĪJUMA ANALĪZE UN PROGNOZĒŠANA ES VALSTĪM.....	26
8.3.3. IEKŠĒJO DZELZCEĻA PASAŽIERU PĀRVADĀJUMU ANALĪZE UN PROGNOZĒŠANA LATVIJAS REĢIONIEM	29
8.4. STARPTAUTISKO PASAŽIERU AVIOPĀRVADĀJUMU PROGNOZĒŠANA ES VALSTĪM UZ SURE- MODEĻA BĀZES.....	34
NOBEIGUMS	39
AUTORES PUBLIKĀCIJAS.....	41
DALĪBA KONFERENCĒS	42

1. PROBLĒMAS AKTUALITĀTE

Izstrādātais promocijas darbs ir veltīts dažādu transporta veidu pasažieru un kravu pārvadājumu apjomu prognozēšanai Eiropas Savienībā ar mūsdienu statistisko metožu palīdzību, īpaši akcentējot Latvijas valsti un balstoties uz parametriskiem, kā arī pusparametriskiem regresijas modeļiem, ar šo rezultātu pārbaudi praksē un to attiecīgu izvērtēšanu.

Transportēšanas apjomi veido to aktuālo informāciju, uz kuras pamata ir iespējams izstrādāt nākotnes plānus un prognozes attiecībā uz transporta nozaru funkcionēšanu un attīstības perspektīvām. Konkrēti, transportēšanas apjomi tiek izmantoti sekojošos nolūkos:

- 1) veidojot perspektīvo transporta līdzekļu parku,
- 2) plānojot transporta tīklus,
- 3) nosakot tendences prasībās pēc jaudām un investīcijām transporta līdzekļu un to komponentu attīstībā, utt.

Latvijas gadījumā ģeogrāfiskais stāvoklis veido vienu no tās nacionālajām bagātībām. Latvijas teritoriju šķērso liels kravu un pasažieru pārvadājumu apjoms. Latvijas transporta kompānijas darbojas stipras ārzemju kompāniju konkurences apstākļos. Attiecīgi, tikai pareiza un pārdomāta ekonomiskā politika spēj nodrošināt vietējo kompāniju konkurences spējas un efektivitāti. Līdz ar to kvalitatīvai transporta apjomu prognozēšanai jābūt balstītai uz moderniem matemātiskiem modeļiem.

Īpaši atzīmēsim faktu, ka līdz šim pusparametriskie modeļi un metodes Latvijā netika izmantotas. Savukārt pasaules praksē pamata pētījumi neparametriskās un pusparametriskās regresijas laukā tika uzsākti divdesmitā gadsimta beigās un patlaban jau tiek plaši pielietoti praksē. Līdz ar to aplūkojamā zinību nozare uzskatāma par jaunu un vienu no perspektīvākajām.

Iegūtie modeļi dos iespēju definēt, kādi faktori un uzskaitāmie lielumi nosaka transporta plūsmu intensitāti un virzienus. Kvalitatīvos faktoros (piemēram, politisko situāciju Eiropas Savienībā) zināmā mērā var aprakstīt skaitliski caur kvantitatīvajiem faktoriem (piemēram, nacionālo valūtu maiņas kursu, naftas cenām u.c.). Šāda pieeja palīdzēs izstrādāt pasākumu kopumu, kas Latvijai nodrošinās efektīvu transporta sistēmas attīstību.

Piedāvājamais pētījums var tikt iedalīts divās daļās, atbilstoši pielietojuma jomai:

- 1) Transportēšanas apjomu aprakstošo modeļu izstrādāšana un novērtēšana ES dalībvalstīm, t.i.:
 - a) summāro kravu aviopārvadājumu apjomu analīze un prognozēšana, balstoties uz parametrisko daudzdimensiju modeli;
 - b) dzelzceļa pasažieru pārvadājumu korespondenču matricas novērtēšana, izmantojot modificētu gravitācijas modeli;
 - c) dzelzceļa kravu pārvadājumu apgrozījuma novērtēšana un prognozēšana, balstoties uz pusparametrisko viena indeksa modeli;

- d) pasažieru summāro aviopārvadājumu apjomu analīze un prognozēšana nepilnīgu datu gadījumā, izmantojot SURE-modeļi;
- 2) Pusparametrisko viena indeksa modeļu izstrādāšana un novērtēšana, lai analizētu un prognozētu dzelzceļa pasažieru pārvadājumus no dažādiem Latvijas reģioniem.

2. PĒTĪJUMA MĒRĶI UN UZDEVUMI

Promocijas pētījuma darba galvenie mērķi bija:

- 1) Izstrādāt matemātiskos modeļus nolūkā analizēt un prognozēt pasažieru un kravu pārvadājumu apjomus.
- 2) Rūpīgi izstrādāt metodes, ar kuru palīdzību būtu iespējams novērtēt piedāvātos modeļus.
- 3) Izpētīt izstrādāto novērtēšanas metožu efektivitāti.

Lai realizētu formulētos mērķus, bija jārisina sekojoši uzdevumi:

- 1) Aplūkot problēmas saistībā ar transportēšanas apjomu prognozēšanu.
- 2) Izpētīt mūsdienās izplatītākos modeļus un metodes, ko izmanto prognozēšanā.
- 3) Izvērst to faktoru, kas ietekmē pasažieru pārvadājumu un kravu transportēšanu apjomu, analīzi.
- 4) Izveidot pētījuma informācijas bāzi, veicot statistisko datu uzkrāšanu gan par Eiropas Savienības dalībvalstīm, gan par Latvijas reģioniem.
- 5) Izstrādāt daudzfaktoru lineāros regresijas modeļus un pusparametriskos regresijas modeļus, ar kuru palīdzību būtu iespējams prognozēt pasažieru pārvadājumu un kravu transportēšanas apjomus.
- 6) Izstrādāt metodes un algoritmus, ar kuru palīdzību varētu veikt piedāvāto modeļu novērtēšanu.
- 7) Veikt izstrādāto modeļu verificēšanu un efektivitātes novērtēšanu.
- 8) Uzskatāmi parādīt pusparametriskā modeļa priekšrocības salīdzinājumā ar daudzfaktoru lineāro modeli.
- 9) Modificēt gravitācijas modeli ar atbilstošās metodes un algoritma attīstību, lai veiktu šī modeļa novērtējumu.
- 10) Pielietot piedāvāto modeli, lai izvērtētu pasažieru nosūtījumu korespondenču matricu.
- 11) Veikt SURE-modeļa formalizēšanu ar atbilstošās metodes un algoritma attīstību, lai veiktu piedāvātā modeļa novērtējumu.
- 12) Uzskatāmi parādīt SURE-modeļa priekšrocības salīdzinājumā ar daudzdimensiju modeli nepilnīgu datu gadījumā.

3. TĒMAS PAŠREIZĒJĀS IZPĒTES LĪMENIS

Tendences transporta veikumā seko ekonomikas attīstībai. Atbilstoši *EuroSTAT Yearbook 2008*, kamēr IKP periodā no 1995. līdz 2007. gadam caurmērā gadā audzis par 2,4%, preču pārvadājumu pieaugums bija līdz 2,85%,

bet pasažieru pārvadājumi pieauga ik gadu par 1,7%. Pēdējā pusotra gada laikā, pastāvot nestabilai ekonomiskajai situācijai pasaulē, nevar apgalvot par nepārtrauktu pārvadājumu apjomu izaugsmi. Sevišķi, rēķinoties ar IKP pastāvīgu samazināšanos un bezdarbnieku skaita pieaugumu, jo šie faktori runā pretī iespējamai transportēšanas izaugsmei.

Atbilstoši preses relīzēm, kuras publicētas *EuroSTAT*, redzams, ka Eiropas Savienībā IKP 2009. gada I ceturksnī krities par 2,5% salīdzinājumā ar iepriekšējo ceturksni. ES 2009. gada jūnijā bezdarbnieki sastādīja 9,4%, salīdzinot ar 9,3% 2009. gada maijā. ES dalībvalstu vidū visaugstākais bezdarbnieku procents bija Spānijā (18,1%), Latvijā (17,2%) un Igaunijā (17,0%). Gada inflācija ES tika prognozēta 1,2% 2009. gada februārī un 0,6% 2009. gada jūlijā.

Augstākminētie dati kalpo par pārliecinošu argumentu tam, ka šādos ekonomiski nestabilos apstākļos īpaši svarīga kļūst pārvadājumu apjomu prognozēšana. Lai iegūtu kvalitatīvu prognozi par pārvadājumu apjomu, tika izstrādāta un pilnveidota virkne matemātisko modeļu:

- 1) Parametriskie modeļi:
 - a) lineārie modeļi,
 - i) daudzfaktoru regresijas modelis,
 - ii) daudzzimensiju regresijas modelis,
 - iii) SURE-modelis;
 - b) Nelineārie modeļi:
 - i) vispārinātais lineārais modelis,
 - ii) gravitācijas modelis;
- 2) Neparametriskie modeļi:
 - a) Nadaraya-Watson novērtētājs,
 - b) viena indeksa modelis.

Raksturojot darbus, kas veikti daudzzimensiju statistikas un parametriskās regresijas jomā, būtu jāatzīmē sekojoši darbi, proti, *N.Draper* un *H.Smith* grāmata *Applied Regression Analysis*, *M.Srivastava* monogrāfija *Methods of Multivariate statistics*, *E.L.Lehman* grāmatas *Theory of Point Estimation* un *Statistical Hypothesis Testing* un *S.Weisberg* darbs *Applied Linear Regression*.

Tāpat autore nevar nepieminēt arī savu pasniedzēju: *A.Andronova*, *E.Kopitova* un *L.Gringlaza* darbu *Varbūtības teorija un matemātiskā statistika* (krievu valodā), kā arī *I.Jackivas* darbu *Intensīvās datormetodes statistisko datu apstrādei* (krievu valodā).

Vispārējā regresijas modeļa forma ir sekojoša:

$$Y_i = m(\mathbf{x}_i) + \varepsilon_i, \quad (1)$$

kur Y_i ir atkarīgais mainīgais i -jā novērojumā, $m(\mathbf{x}_i)$ ir nezināma regresijas funkcija, \mathbf{x}_i ir neatkarīgo mainīgo d -dimensionāls vektors, ε_i ir kļūda

(gadījumlīkums). Turpmāk mūsu rīcībā ir neatkarīgo un nekorelēto novērojumu \mathbb{R} , x_i virkne, $i = 1, 2, \dots, n$. Nezināmajai funkcijai $m(\cdot)$ jābūt novērtētai uz to bāzes.

Vienkāršā gadījumā tiek izmantots *daudzfaktoru regresijas modelis*:

$$Y = X\beta + \varepsilon, \quad (2)$$

$$\text{kur } Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_d \end{pmatrix} \text{ un } X' = \begin{pmatrix} x_{1,1} & \dots & x_{1,d} \\ \vdots & & \vdots \\ x_{n,1} & \dots & x_{n,d} \end{pmatrix},$$

turklāt $x_i = (x_{i,1} \dots x_{i,d})$. Galvenie pieņēmumi $E(\varepsilon) = 0$, $Cov(\varepsilon) = \sigma^2 I$ ir pazīstami kā Gausa-Markova nosacījumi. Vektora β mazāko kvadrātu novērtējums ir sekojošs:

$$\hat{\beta} = (X'X)^{-1} X'Y. \quad (3)$$

Daudzdimensiju regresijas modelis tiek pierakstīts šādā veidā:

$$Y = XB + E, \quad (4)$$

kur Y ir atkarīgo mainīgo novērojumu $\mathbb{R} \times p$ -matrica, X ir neatkarīgo mainīgo $\mathbb{R} \times (d+1)$ -matrica, B ir nezināmo regresijas koeficientu $\mathbb{R} \times p$ -matrica, E ir gadījuma locekļu $\mathbb{R} \times p$ -matrica, p ir atkarīgo mainīgo skaits, d ir neatkarīgo mainīgo skaits un n ir novērojumu skaits. Pieņemts, ka kļūdas ir savstarpēji neatkarīgas. Nezināmo koeficientu vektoru mazāko kvadrātu novērtējumi ir šādi:

$$\tilde{\beta}^{(k)} = (X'X)^{-1} X'Y^{(k)}, \quad k = 1, \dots, p, \quad (5)$$

kur $\tilde{\beta}^{(k)}$ ir nezināmo koeficientu novērtējums k -m atkarīgajam mainīgajam.

Šķietami nesaistīto regresijas modeli (*seemingly unrelated regression equation model*, turpmāk *SURE-modelis*) var apskatīt kā daudzdimensiju modeļa (4) turpinājumu gadījumā, kad daļa ietekmējošo faktoru atbilst visiem atkarīgajiem mainīgajiem, un otrā daļa neatbilst. Turklāt novērojumu skaits katram atkarīgajam mainīgajam var atšķirties. Šajā jomā būtu jāatzīmē tādu zinātnieku ka A.Zellner, R.Velu un J.Richards, un A.Andronova ieguldījums.

Tātad tiek aplūkota G objektu grupa ar numuriem $i = 1, 2, \dots, G$. Katrs i -s objekts ir pārbaudīts n_i reizes, laika momentos $t_{i,1} < t_{i,2} < \dots < t_{i,n_i}$. Katrā j -jā laika momentā $t_{i,j}$ tiek reģistrēts neatkarīgo mainīgo vektors $x_{i,j} = (x_{i,j}^{(1)}, x_{i,j}^{(2)}, \dots, x_{i,j}^{(m_i)})$, kur $m_i < n_i$, un atkarīgā mainīgā vērtība ir $Y_{i,j}$. Ir pieņemts, ka atkarīgais mainīgais $Y_{i,j}$ tiek konstruēts ar sekojošā lineārā regresijas vienādojuma palīdzību:

$$Y_{i,j} = \sum_{v=1}^{m_i} \beta_{i,v} x_{i,j}^{(v)} + Z_{i,j}, \quad (6)$$

kur $\beta_{i,v}$ ir koeficients i -m objektam un v -m neatkarīgajam mainīgajam, $Z_{i,j}$ ir pēc normālā sadalījuma līkuma sadalīts gadījumloceklis ar nulles vidējo vērtību un dispersiju σ_i^2 .

Tālāk, ja diviem dažādiem objektiem i un i' laika momenti $t_{i,j}$ un $t_{i',j'}$ sakrīt, tad gadījuma locekļi $Z_{i,j}$ un $Z_{i',j'}$ ($Y_{i,j}$ un $Y_{i',j'}$ arī) ir korelējošie gadījumlielumi ar kovariāciju $c_{i,i'}$, turklāt vienlaikus dažādiem laika momentiem ir pieņemts, ka tie ir neatkarīgi ($Z_{i,j}$ un $Z_{i',j'}$ arī ir neatkarīgi, ja $j \neq j'$). Tātad, kļūdas $Z_{i,j}$ ir vienlaicīgi savstarpēji korelējošas.

Parasti, tiek pieņemts, ka visiem $i = 1, 2, \dots, G, j = 1, 2, \dots, n_i$ $x_{i,j} = (x_{i,j}^{(1)}, x_{i,j}^{(2)}, \dots, x_{i,j}^{(m_i)})$ ir zināms konstantu vektors, $Y_{i,j}$ ir fiksēta noteikta vērtība. Pamatojoties uz visu augstāk izklāstīto, tiek novērtēti regresijas modeļa nezināmie parametri $\beta_{i,v}$ un nezināmā kovariācija $c_{i,i'}$, kur $\sigma_i^2 = c_{i,i}$.

Gravitācijas modelim (*gravity model*, pēc A. Wellington un E. Lille) piemīt šāda forma:

$$T_{ij} = K \frac{P_i P_j}{D_{ij}^2}, \quad (7)$$

kur T_{ij} ir pasažieru pārvadājumu apjoms starp ģeogrāfiskiem punktiem i un j ; P_i un P_j ir iedzīvotāju skaits punktos i un j , attiecīgi; D_{ij} ir attālums starp punktiem i un j ; K ir konstante.

Gravitācijas modelis turpmāk tika attīstīts pasažieru aviopārvadājumu apjomu prognozēšanai gan esošām, gan plānojamām aviolīnijām. Ievērojamu ieguldījumu pasažieru aviācijas plūsmu prognozēšanā, izmantojot dažādus modificētus gravitācijas modeļus, veica J. Doganis un A. Andronov et al. pagājušā gadsimta sešdesmitajos un astoņdesmitajos gados. Savukārt OD-matricu novērtēšanas jomā jāatzīmē J. de D. Ortuzar un L. G. Willumsen grāmatu *Modelling Transport* un A. Andronova pēdējos pētījumus.

Pusparametrisko un neparametrisko modeļu apskats

Tā kā parametriskie modeļi ne vienmēr nodrošina pietiekami labas prognozes, to trūkuma novēršanai tiek pielietoti *ne-* un *pusparametriskie modeļi*.

Neparametriskās regresijas jomā ir vērts atzīmēt W. Härdle, M. Müller, S. Sperlich and A. Werwatz grāmatu *Nonparametric and Semiparametric Models*, B. W. Silverman darbu *Density Estimation for Statistics and Data Analysis*, W. Härdle monogrāfiju *Applied Nonparametric Regression*, A. Bowman and A. Azzalini grāmatu *Applied Smoothing Techniques for Data Analysis*, J. Simonoff monogrāfiju *Smoothing Methods in Statistics* un A. Pagan un A. Ullah darbu

Nonparametric Econometrics.

Pusparametriskās regresijas vispārējos aspektus mēs rekomendējam aplūkot *A.Yatchew* pētījumā *Semiparametric Regression for the Applied Econometrician* un *D.Ruppert, M.P.Wand and R.J.Carrol* darbā *Semiparametric Regression*.

Vispārinātais lineārais modelis (generalized linear model, turpmāk GLM) ir visu pusparametrisko modeļu pamats (skat. *P.McCullagh* un *J.A.Nelder* fundamentālo darbu *Generalized Linear Models*):

$$m(x_i) = G(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_d x_{i,d}) = G(\beta^T x_i) \tag{8}$$

kur ar $G(\cdot)$ tiek apzīmēta zināma nepārtraukta funkcija, tas ir, viendimensionālā mainīgā $\beta^T x_i$ saites funkcija, turklāt $\beta^T x_i$ saucas par indeksa funkciju vai par indeksu. Vispār eksistē sekojoši ne- un pusparametrisko modeļu veidi: aditīvais modelis (AM), daļēji lineārais modelis (PLM), vispārinātais aditīvais modelis (GAM), vispārinātais daļēji lineārais modelis (GPLM), vispārinātais aditīvais daļēji lineārais modelis (GAPLM).

Šajā promocijas darbā lielākoties tiek pētīts viena indeksa modelis (turpmāk VIM). Viena indeksa modelis ir viens no GLM vispārinājumiem un apvieno ietekmējošo faktoru $x = (x_1, x_2, \dots, x_d)$ efektus vienīgajā mainīgajā, ko sauc par indeksu. Viena indeksa modelis tiek izteikts sekojošā veidā:

$$E(y | x) = m(\beta^T x) = g(\eta) \tag{9}$$

Pieņemts, ka nezināma funkcija $m(\cdot)$ ir gluda funkcija. Funkcija $g(\cdot)$ ir nezināmā saites funkcija no viendimensionālā mainīgā η , ko sauc par indeksu. Par indeksa funkciju var būt paņemta jebkura funkcija. *W.Härdle* izvirzīja divas metodes VIM novērtēšanai, t.i., *pusparametrisko mazāko kvadrātu metodi* un *pseido maksimālās ticamības metodi*.

Nadaraya-Watson kodola novērtētājs, kuru ieviesa *Nadaraya* un *Watson* 1964. gadā:

$$\hat{m}_h(\eta) = \frac{\sum_{i=1}^n K_h(\eta - X_i) Y_i}{\sum_{j=1}^n K_h(\eta - X_j)} \tag{10}$$

ir vispārēja kodola novērtētāja loģiskais turpinājums nezināmās regresijas funkcijas novērtēšanas problēmai. Izlīdzināšanas parametrs h (*bandwidth*) noteica funkcijas \hat{m}_h gluduma līmeni.

Nobeigumā, *individuālie prognozēšanas modeļi* aplūko katru prognozēšanas objektu atsevišķi. *Grupu modeļu* pamatā ir pētāmo objektu vienlaicīgā analīze.

Neskatoties uz to, ka VIM ir viens no vispopulārākajiem un labāk izpētītajiem pusparametriskajiem modeļiem, tā pētījumi turpinās. Šajā jomā būtu

jāatzīmē arī tādus zinātniekus kā *J.S.Simonoff* un *C.-L.Tsai*, *M.Hristache*, *A.Juditski* un *V.Spokoiny*, *L.Xue*, *Y.Xia*, *H.Tong*, *E.Kong*, *W.K.Li*, *L.Zhu*, *H.Wong*, *W.C.Ip* un *R.Zha*.

Vairākās pārvadājumu apjomu prognozēšanai veltītās publikācijās ir aprakstīti vienkāršie prognozēšanas modeļi uz laika rindu metožu bāzes, vai lineārie regresijas modeļi, kas satur 2-3 ietekmējošus faktorus (kā piemērus varam izcelt *A.Baublys*, *J.Butkevičius*, *A.Cok asova*, *N.R.Farnum* un *L.W.Staton*, *Ü.Hunt*, *T.Šliupas*, *E.Spissu*, *N.Taneja*, *S.Wheatcroft* and *G.Lipman*, *S.Makridakis*, *S.Wheelwright* un *R.Hyndman* pētījumus). Vienlaikus šajos darbos nav pietiekoši precīzi un detalizēti aprakstīta nedz vispārējā prognozēšanas tehnika, nedz nav izvēlēti prognozēšanas metodes kvalitātes kritēriji.

Pretstatītajā pētījumā autore vadījās no sekojošas pārvadājumu struktūras izkārtojuma, iedalot to:

- 1) Pēc pārvadājuma tipa:
 - a) pasažieru;
 - b) kravu.
- 2) Pēc pārvadājuma veida:
 - a) starptautiskie;
 - b) iekšējie.
- 3) Pēc transporta veida:
 - a) dzelzceļa pārvadājumi;
 - b) aviopārvadājumi.
- 4) Pēc prognozēšanas objekta:
 - a) valstis vai valsts reģioni;
 - b) OD-pāris.
- 5) Pēc virziena:
 - a) ienākošie;
 - b) izejošie.
- 6) Pēc prognozējamā indikatora:
 - a) Pārvadājumu vai nosūtījumu apjomi, tūkst. pasažieru vai tonnu;
 - b) Apgrozījums, tonnas/km vai pasažieri/km.

Atkarībā no aprakstāmās informācijas tipa faktori, kas ietekmē prognozējamus indikatorus, tiek iedalīti *kvantitatīvajos* un *kvalitatīvajos*. Faktori, kuru lielumi mainās no novērojuma uz novērojumu, tiek attiecināti uz *kvantitatīvajiem* faktoriem. *Kvalitatīvajiem* faktoriem saglabājas noteiktas vērtības novērojumu grupām, turklāt šīs vērtības var būt diskrētie lielumi – atbilstoši iepriekš noteiktam principam. Atkarībā no uzrādītā informācijas rakstura visus ietekmējošos faktorus var iedalīt sekojošās pamatgrupās: *ekonomiskajās*, *sociālajās* un *struktūras*.

Piedāvātā pētījuma ietvaros tika izstrādātas divas statistiskās datu bāzes.

Pirmā statistiskā datu bāze tika ģenerēta viscaur uz *The Statistical Office of the European Communities (EuroSTAT)* materiālu pamata un satur statistiskās ziņas gan par pasažieru, gan kravas pārvadājumiem katrai valstij – Eiropas Savienības dalībnieci laika periodā no 1995. līdz 2008. gadam. Datu bāze

saturēja arī dažādus faktoros pēc ES dalībvalstīm, kuri savukārt ietekmē abus pārvadājumu veidus.

Otrā statistiskā datu bāze satur statistiskos datus par dzelzceļa pasažieru pārvadājumiem Latvijā (no reģioniem un pilsētām). Šī bāze savukārt tika ģenerēta uz materiāliem, kas iegūti no Latvijas Republikas Centrālās statistikas pārvaldes (*LR CSP*) un AS „Latvijas dzelzceļš” Gada grāmatas.

Atkarībā no veiktajiem eksperimentiem un no prognozējama lieluma, modeļos arī tika izmantoti vairāki kvalitatīvie faktori.

4. METODOLOĢIJA UN PĒTNIECISKĀS METODES

Veiktais pētījums ir balstīts uz:

- 1) Mūsdienu regresijas analīzes teoriju, turklāt īpaša uzmanība tika pievērsta tādiem vispārinātās regresijas modeļu veidiem kā pusparametriskais regresijas modelis; SURE-modelis un modificētais gravitācijas modelis; nelineārās optimizācijas metodes tika izmantotas kā palīgīdzekļi.
- 2) Statistikas datiem, kas iegūti no Eiropas Savienības Statistikas dienesta EuroSTAT datu bāzes, Latvijas Republikas Centrālās Statistikas pārvaldes un AS “Latvijas dzelzceļš” Gada grāmatas.
- 3) Zinātniskajām publikācijām, preses relīzēm, publikācijām internetā, kas veltītas apskatāmajām problēmām.
- 4) Datorpakotnes Statistica 6.0 un MathCad 13 vides nodrošinātām iespējām aprēķinu veikšanai un izstrādāto algoritmu programmēšanai.

5. ZINĀTNISKĀ NOVITĀTE

Pētījuma novitāti veido:

- 1) Daudzdimensiju regresijas modelis summāro kravu aviopārvadājumu apjomu prognozēšanai.
- 2) Metode un programmatūra viena indeksa modeļa novērtēšanai un tā adekvātuma un efektivitātes verifikācijai pārvadājumu apjomu analīzē un prognozēšanā.
- 3) Oriģināls nelineārās regresijas modelis (uz gravitācijas modeļa bāzes) un programmatūra transporta tīkla korespondenču matricas novērtēšanai.
- 4) SURE-modelis un atbilstošā novērtēšanas procedūra summāro pasažieru aviopārvadājumu apjomu prognozēšanai.

6. PRAKTISKĀ NOZĪME UN REALIZĀCIJA

- 1) Uz iegūto rezultātu pamata ir sagatavota un nolasīta kursa „Mathematical Methods of Traffic Flow Analysis and Forecasting” daļa RTU 2. kursa ārvalstu bakalaura programmas studentiem *Mechanical Engineering* specialitātē.
- 2) Autores izstrādātie modeļi un metodes dzelzceļa kravu pārvadājumu no

ES valstīm apjomu prognozēšanai tika izmantoti zinātniskajā projektā U7107 „Matemātisko modeļu un to novērtēšanas metodes izstrādāšana Baltijas reģiona transporta plūsmu analīzei un prognozēšanai”, kas veidoja daļu no zinātniskā projekta II IZM-RTU „Zinātniskās darbības attīstība augstskolās”, kas notika laikā no 01.06.06. līdz 31.12.06.

- 3) Autores izstrādātie modeļi un metodes pasažieru dzelzceļa pārvadājumu apjomu prognozēšanai no Latvijas reģioniem tika izmantoti zinātniskajā projektā U1212 „Matemātisko modeļu, algoritmu un datorprogrammu izstrādāšana Latvijas transporta sistēmas analīzei, attīstības prognozēšanai un optimizācijai”, kas veidoja daļu no zinātniskā projekta IZM-RTU „Zinātniskās darbības attīstība augstskolās”, kas notika laikā no 01.06.08. līdz 31.12.08.
- 4) Iegūtās prognozes var izmantot transporta kompānijas – savas darbības optimālai plānošanai, t.i., izstrādājot lidojumu sarakstu, veidojot biļešu cenas utt.; Satiksmes ministrija – kapitālo ieguldījumu optimālajam sadalījumam ceļu būvei, dzelzceļu remontam, termināļu rekonstrukcijai u.c.

7. PROMOCIJAS DARBA STRUKTŪRA

1. nodaļa. Prognozēšanas problēmas pašreizējā stāvokļa apskats. Tiek raksturota pašreizējā situācija Eiropas Savienībā kravu un pasažieru pārvadājumu jomā. Tiek aprakstīta darba izpildes gaitā izmantotā literatūra, īsi aprakstīti pēdējos gados veiktie pētījumi neparametriskās un pusparametriskās regresijas novērtēšanas jomā.

2. nodaļa. Prognozēšanas problēmas informācijas nodrošinājums. Šī nodaļa ir veltīta šajā darbā izmantoto statistisko datu vākšanas procesa analīzei. Pirmkārt tiek aprakstīti pārvadājumu tipi kā prognozējamie lielumi. Tiek izklāstīta pārvadājumus ietekmējošo faktoru klasifikācija. Tiek aprakstītas divas statistiskās datubāzes, kuras izveidotas darba gaitā. Turklāt, ir raksturoti statistiskās informācijas avoti, t.i., EuroSTAT, LR Centrālā Statistikas pārvalde un AS “Latvijas dzelzceļš” Gada grāmata.

3. nodaļa. Matemātiskie modeļi prognozēšanai un to novērtēšanas metodes. Tiek aprakstīts darbā izmantoto regresijas modeļu teorētiskais pamatojums. Pirmkārt, sniegts individuālo un grupu modeļu raksturojums; ir iztīrīts tas, kuros gadījumos šie modeļi ir pielietojami. Otrkārt, ir pastāstīts par parametriskajiem modeļiem. Šajā nolūkā ir aprakstīts plaši pazīstams un izmantojams daudzfaktoru (vai lineārais) modelis. Ir izklāstīti arī parametriskie, bet lietojumā daudz sarežģītākie daudzdimensiju un SURE-modeļi. Viena no apakšnodaļām ir pilnībā veltīta gravitācijas modeļiem. Speciāla uzmanība ir pievērsta neparametriskajiem un pusparametriskajiem modeļiem. Tiek piedāvāts dažādu pusparametrisko modeļu apskats. Tā kā viens no darbā akcentētiem aspektiem saistīts ar viena indeksa modeļa verifikāciju, tiek rūpīgi apskatīts Nadaraya-Watson kodola novērtētājs un tā īpašības.

4. nodaļa. *Starptautisko kravu aviopārvadājumu analīze un prognozēšana ES valstīm uz daudzdimensiju modeļa bāzes.* Tiek uzskaitīti summāro kravu aviopārvadājumu (t.i., iekšējo un ārējo attiecībā pret ES robežām) apjomu analīzes un prognozēšanas rezultāti Eiropas Savienības valstīm. Atbilstošais daudzdimensiju regresijas modelis satur svarīgākos ekonomiskos faktoros, kas ietekmē iekšējos un ārējos pārvadājumus katrai analizējamajai valstij. Ir parādīts ārējo un iekšējo pārvadājumu summārās prognozes iegūšanas process noteiktam gadam. Tika aprēķināti ticamības intervāli iegūtajām prognozēm (pie dažādām varbūtībām).

5. nodaļa. *Dzelzceļa pasažieru korespondenču izvērtēšana starp ES valstīm uz modificētā gravitācijas modeļa bāzes.* Šeit tiek stāstīts par pasažieru dzelzceļa nosūtījumu korespondenču matricas novērtēšanu starp ES valstīm uz modificētā gravitācijas modeļa bāzes. Ir izklāstīts efektīvais algoritms šī modeļa nezināmo parametru novērtēšanai. Tika novērtētas pasažieru dzelzceļa korespondences starp 23 ES valstīm 2008. gadam.

6. nodaļa. *Vienindeksa modeļa pielietošana pārvadājumu apjomu izpētei.* Šī nodaļa ir veltīta viena indeksa modeļa efektivitātes izpētei. Pirmkārt, tika izstrādāts algoritms viena indeksa modeļa nezināmo parametru novērtēšanai. Otrkārt, tika izveidotas procedūras visefektīvākā viena indeksa un lineārā modeļu izvēlei. Piedāvātā kross-pārbaudes pieeja ļauj izvērtēt pētāmo modeļu kvalitāti ne tikai datu izlīdzināšanā, bet arī prognozēšanas gadījumā. Ir pretstatīti vairāku eksperimentu rezultāti, kuros ir uzskatāmi redzamas viena indeksa modeļa neapšaubāmās priekšrocības salīdzinājumā ar parametriskajiem modeļiem.

7. nodaļa. *Starptautisko pasažieru aviopārvadājumu prognozēšana ES valstīm uz SURE-modeļa bāzes.* Šajā nodaļā ir apskatīts viens no SURE-modeļa vispārinājumiem. Uzskatīts, ka individuālajos novērojumos nav jābūt informācijai par visiem atkarīgajiem mainīgajiem. Iegūts modeļa kovariācijas matricas nenovirzīts novērtējums. SURE-modeļa pielietošanas priekšrocība salīdzinājumā ar daudzdimensiju modeli statistisko datu nepilnības gadījumā ir parādīta, balstoties uz summāro pasažieru aviopārvadājumu apjomu prognozēšanas piemēru ES valstīm.

8. PĒTĪJUMU REZULTĀTU KOPSAVILKUMS

8.1. STARPTAUTISKO KRAVU AVIOPĀRVADĀJUMU ANALĪZE UN PROGNOZĒŠANA ES VALSTĪM UZ DAUDZDIMENSIJU MODEĻA BĀZES

Uzdevuma nostādne

Kravu un pasta starptautiskie aviopārvadājumi ir sadalīti divās daļās atbilstoši Eiropas Savienības robežām: iekšējie un ārējie starptautiskie kravu un pasta aviopārvadājumi.

Uzdevums ir regresijas modeļu izstrāde minēto pārvadājumu prognozēšanai. Tiek aplūkotas sekojošas problēmas:

1. Iekšējo un ārējo aviopārvadājumi *atsevišķā* prognozēšana.
2. Iekšējo un ārējo aviopārvadājumi *kopējā* prognozēšana.

Pirmās problēmas atrisināšanai tiek izmantots daudzfaktoru regresijas modelis (2), otrai problēmai – daudzdimensiju modelis (4). Tiek izvirzīts pieņēmums, ka apskatāmie ietekmējošie faktori uz iekšējiem un ārējiem pārvadājumiem iedarbojas dažādi. Pētījuma mērķis ir kopējās prognozēšanas nozīmīguma pierādījums.

Pētījuma *objekts* ir kāda Eiropas Savienības *valsts*. Par *novērojumu* tiek uzskatīti dati par objektu konkrētajā laika momentā, t.i., ārējie un iekšējie pārvadājumi, tūkst. tonnu. Ietekmējošie faktori ir:

- t_1 – laika faktors (*YEAR*);
- t_2 – tirdzniecības līmenis (*TI*), % no IKP;
- t_3 – gada vidējā EUR likme pret nacionālo valūtu (*EURrates*);
- t_4 – gada vidējais inflācijas līmenis, patērētāju cenu harmonizētajos indeksos (*IR*);
- t_5 – cenas uz naftas produktiem (*PP*), eirās par tonnu;
- t_6 – IKP uz vienu iedzīvotāju, cenu standartos (*GDP_PPS*);
- t_7 – valsts gradācija pēc prognozējamā lieluma (*GradY*), ir ievadīta modelī ar nolūku izdalīt valstis ar lieliem īpatnējiem (komentārus skatīt zemāk) pārvadājumiem (Beļģija, Vācija un Lielbritānija) atsevišķā grupā;
- t_8 – valsts gradācija pēc populācijas blīvuma (*GradPD*);
- t_9 – valsts gradācija pēc platības (*GradArea*);
- t_{10} – valsts gradācija pēc līdzdalības ilguma Eiropas Savienībā (*GradMember*);
- t_{11} – iedzīvotāju skaits, tūkst. cilvēku (*TP*).

Sekojoši faktori ir prognozējamie lielumi, tūkst. tonnu:

- t_{12} – iekšējie kravu un pasta starptautiskie aviopārvadājumi (*FrM_Intra*);
- t_{13} – ārējie kravu un pasta starptautiskie aviopārvadājumi (*FrM_Extra*).

Apskatāmo modeļu vispārējā struktūra un novērtēšanas procedūra

Kovariācijas novērtējums starp diviem novērtētiem vektoriem:

$$Cov^* \left(\tilde{\beta}^{(k)}, \tilde{\beta}^{(l)} \right) = \tilde{\sigma}_{kl} \left(X \right), \quad k, l = 1, \dots, p. \quad (11)$$

Vektori $\tilde{\beta}^{(k)}$ un $\tilde{\beta}^{(l)}$ tiek novērtēti atsevišķi pēc formulas (5), un $\tilde{\sigma}_{kl}$ ir atbilstošais elements no kovariācijas matricas $\tilde{\Sigma}$ nenovirzītā novērtējuma. Tiek iegūtas divas atsevišķas prognozes $\tilde{Y}^{(1)} = X\tilde{\beta}^{(1)}$ un $\tilde{Y}^{(2)} = X\tilde{\beta}^{(2)}$. Prognožu summa ir $\tilde{S} = \tilde{Y}^{(1)} + \tilde{Y}^{(2)}$. Šīs summas vidējā vērtība un dispersija ir:

$$E \left(\tilde{S} \right) = E \left(\tilde{Y}^{(1)} \right) + E \left(\tilde{Y}^{(2)} \right) = X \left(\tilde{\beta}^{(1)} + \tilde{\beta}^{(2)} \right), \quad (12)$$

$$D \left(\tilde{S} \right) = D \left(\tilde{Y}^{(1)} \right) + D \left(\tilde{Y}^{(2)} \right) + 2Cov \left(\tilde{Y}^{(1)}, \tilde{Y}^{(2)} \right). \quad (13)$$

Pirmais un otrais formulas (13) loceklis ir:

$$D\left(\tilde{\gamma}^{(k)}\right) = D\left(\tilde{\beta}^{(k)}\right) = X \cdot Cov\left(\tilde{\beta}^{(k)}\right) \cdot X' \quad (14)$$

kur $Cov\left(\tilde{\beta}^{(k)}\right)$, $k=1,2$, ir vektoru $\tilde{\beta}^{(1)}$ un $\tilde{\beta}^{(2)}$ kovariācijas matricas. Trešo formulu (13) locekli var noteikt sekojoši:

$$\begin{aligned} Cov\left(\tilde{\gamma}^{(1)}, \tilde{\gamma}^{(2)}\right) &= Cov\left(\tilde{\beta}^{(1)}, \tilde{\beta}^{(2)}\right) = E\left(\left(\tilde{\beta}^{(1)} - X\beta^{(1)}\right) \cdot \left(\tilde{\beta}^{(2)} - X\beta^{(2)}\right)\right) = \\ &= X \cdot E\left(\left(\tilde{\beta}^{(1)} - \beta^{(1)}\right) \cdot \left(\tilde{\beta}^{(2)} - \beta^{(2)}\right)\right) \cdot X. \end{aligned} \quad (15)$$

Pēdējais loceklis ir divu koeficientu vektoru kopīgā kovariācijas matrica (11). Ticamības intervāla augšējā robeža summārajai prognozei $\bar{E}\left(\tilde{\gamma}\right) = E\left(\tilde{\gamma}\right)$ atbilstoši varbūtībai γ ir $\left(\bar{S}_\gamma\right)$, un

$$\bar{S}_\gamma = E\left(\tilde{\gamma}\right) + \sqrt{D\left(\tilde{\gamma}\right)} \cdot \Phi^{-1}\left(\gamma\right), \quad (16)$$

kur $\Phi^{-1}\left(\gamma\right)$ ir standarta normālā sadalījuma γ -kvantīls.

Pētāmie modeļi pārvadājumu prognozēšanai

Visi piedāvātie modeļi ir *grupu* modeļi. Eksperimentiem tika atlasītas gan ES vecās dalībnieces, gan arī jaunās: Austrija, Beļģija, Čehija, Vācija, Dānija, Igaunija, Spānija, Somija, Francija, Grieķija, Ungārija, Īrija, Itālija, Lietuva, Holande, Polija, Portugāle, Slovākija un Lielbritānija. Analizējamais laika periods ir no 2001. līdz 2006.gadam. Turklāt, daži objekti nesatur novērojumus par visu apskatāmo periodu.

Pirmais un otrais modelis ir daudzfaktoru regresijas modelis (2).

Atkarīgais mainīgais *pirmajā modelī* $Y^{\text{C}} = \frac{t_{12}}{t_{11}}$ ir iekšējie kravu un pasta starptautiskie pārvadājumi (tonnās), dalīti ar iedzīvotāju skaitu (tūkst. cilvēku), t.i., īpatnējie pārvadājumi. Ietekmējošie mainīgie ir $x_1 = t_1$, $x_2 = t_2$, $x_3 = t_3$, $x_4 = t_4$, $x_5 = t_5$, $x_6 = t_6$, $x_7 = t_7$, $x_8 = t_8$, $x_9 = t_9$, $x_{10} = t_{10}$.

Atkarīgais mainīgais *otrajā modelī* $Y^{\text{C}} = \frac{t_{13}}{t_{11}}$ ir ārējie kravu un pasta starptautiskie pārvadājumi (tonnās), dalīti ar iedzīvotāju skaitu (tūkst. cilvēku).

Trešais modelis ir daudzdimensiju modelis (4), kas satur divus atkarīgus mainīgos, kuri ir aprakstīti ar vektoru $Y^{\text{C}} = \left(Y_1^{(3)} = \frac{t_{12}}{t_{11}}, Y_2^{(3)} = \frac{t_{13}}{t_{11}}\right)$.

Ietekmējošo faktoru kopas visiem modeļiem sakrīt.

Iekšējo un ārējo pārvadājumu atsevišķs novērtējums

Ir apstrādāti 94 novērojumi. Novērtētie modeļi ir:

$$\begin{aligned} \tilde{E}\left(\tilde{\gamma}^{(1)}\right) &= -1193 + 0.6x_1 + 0.09x_2 - 0.001x_3 - 0.17x_4 - 0.01x_5 + \\ &+ 1.99x_7 - 7.1x_8 + 16.47x_9, \end{aligned} \quad (17)$$

$R_0 = 2\,202$, $R^2 = 0.82$, $F^l = 38$, $F(10,83) = 1.95$, $\alpha = 5\%$.

$$\tilde{E} \left(\gamma^{(2)} \right) = -3394 + 1.7x_1 - 0.001x_3 - 0.02x_5 + 0.07x_6 - 52.5x_7 - 6.6x_8 + 9.7x_9 + 69.3x_{10}. \quad (18)$$

$R_0 = 21\,989$, $R^2 = 0.95$, $F = 151$, $F(10,83) = 1.95$, $\alpha = 5\%$.

Kopējo pārvadājumu novērtējums

Tika aprēķināta kļūdu kovariācijas matrica:

$$\tilde{\Sigma} = \begin{bmatrix} 9.12 & 5.24 \\ 5.24 & 22.45 \end{bmatrix}. \quad (19)$$

Tālāk uz formulas (19) bāzes tiek iegūtas kovariācijas matricas pēc formulas (11) novērtētiem regresijas koeficientiem $\tilde{\beta}^{(1)}$ un $\tilde{\beta}^{(2)}$.

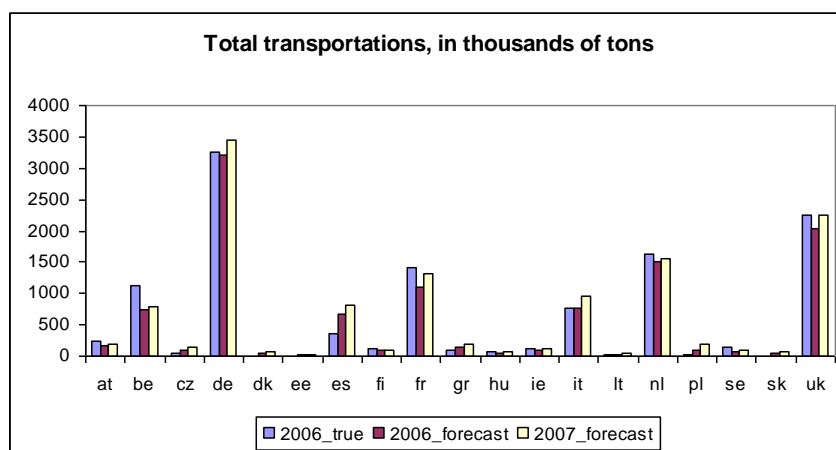
1. tabula satur summārās prognozes 2007. un 2006. gadam (kolonnas 2007. un 2006.), patiesus summārus pārvadājumus 2006.gadam (kolonna 2006*), un ticamības intervāla augšējo robežu (16) pie dažādām varbūtībām γ summārajai prognozei 2007.gadam. 1. att. vizualizē rezultātus, pretstatītus 1. tabulai.

1. tabula

Summārās pārvadājumu prognozes, tūkst. tonnu

Valsts	2006*	2006	2007	Ticamības intervāla augšējā robeža			
				60%	70%	80%	90%
Austrija	227.92	174.21	196.60	204.77	213.73	224.09	238.47
Beļģija	1124.95	742.95	781.78	788.92	796.75	805.80	818.35
Čehija	57.05	90.63	133.22	136.51	140.12	144.30	150.09
Vācija	3267.91	3213.93	3444.38	3 498.02	3556.90	3624.96	3719.35
Dānija	6.69	55.71	72.64	73.45	74.35	75.38	76.82
Igaunija	10.05	16.08	20.96	24.03	27.40	31.29	36.69
Spānija	368.97	673.48	811.22	817.36	824.11	831.91	842.72
Somija	118.43	84.17	101.16	126.16	153.60	185.32	229.31
Francija	1421.41	1105.82	1307.02	1 340.66	1377.59	1420.27	1479.47
Grieķija	90.60	154.94	197.92	243.74	294.03	352.17	432.80
Ungārija	64.89	41.21	81.90	84.04	86.39	89.10	92.87
Īrija	111.64	86.28	109.44	114.88	120.86	127.77	137.35
Itālija	764.05	762.71	949.88	963.28	977.97	994.96	1018.52
Lietuva	12.67	21.94	36.12	40.69	45.70	51.49	59.53
Holande	1621.35	1499.07	1545.27	1 577.28	1612.40	1653.00	1709.32
Polija	32.02	104.63	193.87	197.25	200.97	205.26	211.22
Zviedrija	139.48	66.41	94.66	96.93	99.41	102.29	106.28
Slovākija	5.36	50.44	61.37	68.46	76.25	85.25	97.74
Lielbritānija	2248.41	2044.67	2248.39	2 289.97	2335.60	2388.36	2461.52

¹ Fišera kritērija aprēķina vērtība



1.att. Summārie novērotie un prognozētie pārvadājumi

Iegūtie rezultāti liecina par piedāvātā paņēmiena efektivitāti. Tomēr būtu labāk izmantot individuālus prognozēšanas modeļus tām atsevišķām valstīm, kurām ir stabilas pārvadājumu tendences. Īpaši svarīgs, ja atsevišķām valstīm trūkst novērojumu. Nav rekomendēts izmantot summāro pārvadājumu apjomu prognozēšanai laika rindās. Šajā ziņā būtu vēlams turpināt esošo pētījumu ar SURE-modeļa palīdzību.

8.2. DZELZCEĻA PASAŽIERU KORESPONDENČU IZVĒRTĒŠANA STARP ES VALSTĪM UZ MODIFICĒTĀ GRAVITĀCIJAS MODEĻA BĀZES

Uzdevuma nostādne un izvērtēšanas procedūra

Ir doti n ģeogrāfiskie punkti ar numuriem $i = 1, 2, \dots, n$. Katram punktam i ir zināms iedzīvotāju skaits h_i un m faktori $c_{i,j}$, $j = 1, 2, \dots, m$, kuri tāpat ir zināmās konstantes. Ir zināms attālums $d_{i,l}$ starp punktiem i un l visiem punktu pāriem \mathcal{L} . Ir zināms nosūtīto pasažieru skaits Y_i no punkta i par noteiktu laika periodu, turklāt tas ir gadījuma lielums.

Ir nepieciešams novērtēt korespondenču $Y_{i,l}$ lielumus visiem punktu pāriem \mathcal{L} . Lielumu $Y_{i,l}$ matrica saucas par *korespondenču matricu*. Korespondences $Y_{i,l}$ novērtējums tiek apzīmēts ar $Y_{i,l}^*$. Tiek izvirzītas sekojošas prasības: 1) $Y_{i,l}^* > 0$ visiem $i \neq l$; 2) $Y_{i,i}^* = 0$; 3) $Y_{i,l}^* = Y_{l,i}^*$.

Matemātiskais modelis atsevišķai korespondencei \mathcal{L} ir:

$$Y_{i,l} = \frac{h_i h_l}{d_{i,l}^\tau} \exp \left(a + \alpha_{(i)} + c_{(l)} \right) g_{(i,l)} \beta + V_{i,l} \quad (20)$$

kur a , $\alpha = \alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_m$ un $\beta = \beta_1 \quad \beta_2 \quad \dots \quad \beta_m$ ir nezināmie regresijas parametri, τ un θ ir nezināmie formas parametri, $c_{(i)} = c_{i,1} \quad \dots \quad c_{i,m}$ un $g_{(i,l)} = c_{i,1} c_{l,1} \quad \dots \quad c_{i,m} c_{l,m}$ ir m -vektori-rindas, $V_{i,l}$ ir neatkarīgi identiski

sadalītie gadījumlīkumi ar nulles vidējo un nezināmo dispersiju σ^2 . Ir pieņemts, ka $V_{i,l}$ atbilst normālajam sadalījumam. Tad $Z_{i,l} = \exp\left(\frac{h_{i,l}}{\tau}\right)$ ir sadalīts pēc log-normālā līkuma. Atzīmēsim, ka pie $\theta = 1$ un $\tau = 2$ atbilst tā saucamais gravitācijas modelis.

Izteiksme, kas apraksta pasažieru nosūtījumu lielumu no punkta i , izriet no modeļa (20):

$$Y_i = \sum_{\substack{l=1 \\ i \neq l}}^n Y_{i,l} = \sum_{\substack{l=1 \\ i \neq l}}^n \frac{h_i h_l}{h_{i,l}} \exp\left(\frac{h_{i,l}}{\tau}\right) \exp\left(\mu + \mu_{(i)} + c_{(l)}\right) + g_{(i,l)}\beta + V_{i,l} \quad (21)$$

Sekojošie parametri ir jānovērtē uz zināmo nosūtījumu vērtību \mathcal{H} bāzes: $\theta, \tau, a, \alpha, \beta$ un σ^2 . Tā kā vienlaicīgi identificēt parametrus a un σ^2 nav iespējams, pētījumā tika ieviests apvienotais parametrs $\tilde{a} = a + \frac{1}{2}\sigma^2$. Kā novērtējumu efektivitātes kritērijs tiek izmantota mazāko kvadrātu svērtā summa:

$$R(\gamma, w) = \sum_{i=1}^n w_i \left(\sum_{l=1, l \neq i}^n \frac{h_i h_l}{h_{i,l}} \exp\left(\frac{h_{i,l}}{\tau}\right) \exp\left(\mu + \mu_{(i)} + c_{(l)}\right) + g_{(i,l)}\beta - Y_i \right)^2, \quad (22)$$

kur $\gamma = \left(\mu, \tau, \tilde{a}, \alpha^T, \beta^T \right)^T$ un $w = (w_1, w_2, \dots, w_n)^T$ ir svaru vektors. Kritērija (22) minimizācijai tiek izmantota gradienta metode, un

$$\nabla R(\gamma, w) = \left(\frac{\partial R}{\partial \theta}, \frac{\partial R}{\partial \tau}, \frac{\partial R}{\partial a}, \frac{\partial R}{\partial \alpha}, \frac{\partial R}{\partial \beta} \right)^T. \quad (23)$$

Ja $w_i = \text{const}$, gradienta metode ātri dod novērtējumus $\theta^*, \tau^*, \tilde{a}^*, \alpha^*, \beta^*$. Turklāt,

$$\sigma^{2*} = \ln \left\{ 1 + \left(2 \sum_{i=1}^{n-1} \sum_{l=i+2}^n \frac{h_i h_l}{h_{i,l}} \exp\left(\frac{h_{i,l}}{\tau^*}\right) \right)^{-1} \sum_{i=1}^n \left(\sum_{l=1, l \neq i}^n \frac{h_i h_l}{h_{i,l}} \exp\left(\frac{h_{i,l}}{\tau^*}\right) \exp\left(\mu^* + \mu_{(i)}^* + c_{(l)}^*\right) + g_{(i,l)}\beta^* - Y_i \right)^2 \right\}. \quad (24)$$

Tagad parametra a novērtējums tiek aprēķināts kā $a^* = \tilde{a}^* - \frac{1}{2}\sigma^{2*}$.

Statistiskajiem datiem par nosūtījumiem \mathcal{H} un novērtētām korespondencēm \mathcal{H}_i^* jābūt līdzsvarotiem:

$$\sum_{l=1}^n Y_{i,l}^* = Y_i, \quad i = 1, \dots, n. \quad (25)$$

Šim nolūkam katram punktam i tika ieviests korekcijas koeficients $\delta_i > 0$. Tātad novērtējuma $Y_{i,l}$ korekcija ir:

$$\tilde{Y}_{i,l} = \delta_i Y_{i,l}^* \delta_l, \quad i, l = 1, \dots, n. \quad (26)$$

Skaitliskais piemērs

Mēs pielietojām piedāvāto pieeju pasažieru dzelzceļa pārvadājumu novērtējumam starp 23 valstīm – ES dalībniecēm, turpmāk *valstis*. Par faktoriem tika pieņemti sekojošie katras valsts parametri:

- c_1 vidējā mēnešalga, EUR;
- c_2 valsts gradācija pēc aviācijas transporta izmantošanas intensitātes;
- c_3 valsts gradācija pēc dzelzceļa transporta izmantošanas intensitātes;
- c_4 valsts gradācija pēc jūras transporta izmantošanas intensitātes;
- c_5 valsts gradācija pēc tūristiskās pievilcības;
- c_6 valsts gradācija pēc līdzdalības ilguma Eiropas Savienībā.

Visi gradācijas mainīgie var pieņemt jebkuras pozitīvas integras vērtības, tikai c_6 pieņem divas vērtības: 0 valstīm – Eiropas Savienības vecajām dalībniecēm, un 1 – jaunajām dalībniecēm. Gradācijas mainīgo vērtības tiek noteiktas atbilstoši ekspertu vērtējumam. Piemēram, mainīgo vērtības Dānijai ir sekojošas: $c_1 = 4.34$, $c_2 = 4$, $c_3 = 3$, $c_4 = 4$, $c_5 = 2$, $c_6 = 0$; Latvijai ir $c_1 = 0.68$, $c_2 = c_3 = c_4 = 1$, $c_5 = 0$, $c_6 = 1$. Gradācijas mainīgo vērtību iegūšanai tika analizēti statistikas dati apskatāmajām valstīm par starptautiskajiem jūras, dzelzceļa un aviopārvadājumiem 2007. gadā.

Par attālumiem starp ģeogrāfiskajiem punktiem, tika ņemti attālumi starp valstu galvaspilsētām. Kā viens no iespējamajiem attāluma mēriem, var tikt pieņemts attālums starp valstu vidējām svērtām koordinātēm.

Tiek novērtētas korespondences starp sākumpunktiem (pacelšanās valstis) un beigu punktiem (nosēšanās valstis) uz zināmo nosūtījumu bāzes no sākumpunktiem. Tātad Y_i vērtības ir starptautiskie izejošie dzelzceļa pārvadājumi, tūkst. pasažieru, no nosēšanās valstīm (skat. 2. tabulu). Tāpēc ir nepieciešams novērtēt 15 parametrus. Aprakstītā novērtēšanas procedūra sniedz sekojošus parametru novērtējumus:

$$\theta^* = 0.788, \quad \tau^* = 2.786, \quad \tilde{a}^* = -10.01,$$

$$\alpha^* = \begin{pmatrix} 0.074 & 0.061 & 0.077 & 0.063 & 0.078 & 0.33 \end{pmatrix}^T,$$

$$\beta^* = \begin{pmatrix} 0.197 & 0.194 & 0.152 & 0.054 & 0.097 & 0.104 \end{pmatrix}^T.$$

Novērtētais parametrs $\sigma^{2*} = 0.065$. Izmantojot novērtējumu $\tilde{a}^* = -10.01$, atradām

$$a^* = \tilde{a}^* - \frac{1}{2} \sigma^{2*} = -10.01 - \frac{1}{2} 0.065 = -10.04.$$

Novērotie Y_i , novērtētie Y_i^* nosūtījumi no katras valsts, tūkst. pasažieru, un korekcijas koeficienti δ_i ir pārstāvēti 2. tabulā.

3. tabula satur novērtētās un patiesās korespondences vairākām valstīm. Patieso novērojumu būtiskā daļa nav pieejama. Pārim “Francija-Vācija” novērtētā korespondence ir 153 tūkst. pasažieru. Apgrieztā korespondence ir 155

tūkst. pasažieru. Diemžēl iepazīties ar šā pāra patieso korespondenci nav iespējams. Bet ir nenoliedzams, ka tie novērtējumi, kas iegūti ar piedāvātā modeļa palīdzību, nav pretrunā ar augstāk izklāstītiem teorētiskiem pieņēmumiem, un ir ļoti tuvu patiesām korespondencēm.

2. tabula

Novērtēšanas rezultāti

Valsts	Y_i	Y_i^*	δ_i	Valsts	Y_i	Y_i^*	δ_i
ES	44 270	41 210	-	Lietuva	7	0.99	7.062
Beļģija	3 187	3 741	0.705	Luksemburga	2 333	905	2.882
Bulgārija	68	23	1.040	Ungārija	631	164	3.424
Čehija	1015	864	1.212	Holande	2 617	2 321	1.211
Dānija	4 974	5 557	0.806	Austrija	1 575	2 017	0.362
Vācija	5 072	5 279	0.956	Polija	353	202	1.487
Īrija	347	178	1.909	Portugāle	98	41	1.629
Grieķija	18	6	2.050	Rumānija	197	37	3.888
Spānija	329	164	1.554	Slovēnija	97	37	2.064
Francija	5 184	5 240	0.948	Slovākija	1 459	1 081	2.939
Itālija	1 600	890	1.654	Zviedrija	5 023	4 591	1.226
Latvija	2	0.85	2.354	Lielbritānija	8 082	7 875	1.003

3. tabula

Daži korespondenču novērtējumi

Pacelšanās valsts	Vācija		Francija		Latvija		Lietuva		Holande	
	$Y_{i,l}$	$Y_{i,l}^*$	$Y_{i,l}$	$Y_{i,l}^*$	$Y_{i,l}$	$Y_{i,l}^*$	$Y_{i,l}$	$Y_{i,l}^*$	$Y_{i,l}$	$Y_{i,l}^*$
Beļģija	:	113	1852	601	:	0	:	0	1280	548
Bulgārija	:	1	:	0.1	:	0	:	0	:	0
Čehija	:	581	0	5	:	0	0	0	:	3
Dānija	:	1683	0	37	:	0	:	0	:	51
Vācija	5072	0	545	155	:	0.1	0	0.1	807	153
Īrija	:	9	0	15	:	0	:	0	:	6
Grieķija	:	1	:	0.4	:	0	:	0	:	0
Spānija	:	15	315	27	:	0	:	0	:	4
Francija	:	153	3	0.00	:	0	:	0	520	197
Itālija	:	324	777	109	:	0	:	0	:	24
Latvija	:	0	:	0	:	0	2	0.1	:	0
Lietuva	:	1	:	0	2	0.2	0	0	:	0
Luksemburga	:	223	238	572	:	0	:	0	:	113
Ungārija	:	39	:	2	:	0	:	0	0	1
Holande	:	193	265	251	:	0	:	0	0	0
Austrija	:	136	8	6	:	0	0	0	10	2
Polija	:	120	0	1	:	0.1	5	0.5	:	1
Portugāle	:	1	0	1	:	0	:	0	:	0
Rumānija	:	12	0	0	:	0	:	0	:	0
Slovēnija	:	7	0	1	:	0	0	0	:	0
Slovākija	:	18	0	0	:	0	0	0	:	0
Zviedrija	:	1156	0	61	:	0.3	:	0.1	:	49
Lielbritānija	:	512	1181	3615	:	0	:	0	:	1007

Novērtēta nosūtīšana no Vācijas uz Vāciju ir vienāda ar nulli, un atbilstoši fiksētam modelim (20), tādām korespondencēm $Y_{i,i}^*$ patiesi ir jābūt vienādām ar nulli. Patiesā korespondence ir vienāda ar 5072 tūkst. pasažieru, kas nonāk pretrunā ar šo pieņēmumu. Analizējot 2. tabulu, var redzēt, ka kopējā novērtētā nosūtīšana no Vācijas ir ļoti tuva īstai. Šis fakts pierāda bilances nosacījuma (26) patiesumu.

Analizējot rezultātus, ir redzams, ka labākie korespondenču novērtējumi atbilst valstīm – ES vecajām dalībniecēm. Iespējams, novērtējumu precizitātes uzlabošanai par jaunajām ES dalībvalstīm, tās ir jāanalizē atsevišķi.

Ir piedāvāts oriģināls nelineārais regresijas modelis konkrētās korespondences $Y_{i,l}$ novērtēšanai starp diviem ģeogrāfiskajiem punktiem. Modeļa nezināmie parametri ir novērtēti ar speciāli izveidoto procedūru uz gradienta metodes bāzes. Nepieciešamā programmatūra ir izstrādāta MathCad 13 vidē. Piedāvātās pieejas verifikācija pasažieru dzelzceļa korespondenču noteikšanai starp Eiropas Savienības valstīm norādīja uz samērā labiem rezultātiem.

8.3. VIENA INDEKSA MODEĻA PIELIETOŠANA PĀRVADĀJUMU APJOMU IZPĒTEI

8.3.1. VIENA INDEKSA MODEĻA NOVĒRTĒŠANA

Viena indeksa modeļa vispārējā formula ir sekojoša:

$$E \left(Y_{i,l} \mid x \right) = m \left(x \right) = g \left(v_{\beta} \left(x \right) \right). \quad (27)$$

Funkcija $g \left(\cdot \right)$ ir *saites nezināmā funkcija*. Kā indeksa funkciju izmanto jebkuru piemēroto gludo funkciju. Šajā pētījumā tika izvēlēta lineārā kombinācija:

$$m(x_i) = g \left(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_d x_{i,d} \right) = g \left(\tau_i \right), \quad (28)$$

kur $\tau_i = \beta^T x_i$ un ir nosaukts par *indeksu*.

Viena indeksa modeļa novērtēšanas procedūrā ietilpst divi soļi. Vispirms tiek novērtēts nezināmo koeficientu vektors β . Otrā soļa gaitā, izmantojot iegūtās indeksa vērtības, tiek novērtēta funkcija g kā Y vienfaktora neparametriskā regresija no $v_{\beta} \left(x \right)$. Ir jāveic divas būtiskas lietas: jāizvēlas aizvietotājs nezināmai saites funkcijai g un jānosaka piemērotā mērķa funkcija nezināmo koeficientu vektora β novērtēšanai. Turklāt ir iespējamas vairākas metodes izvēlētajās mērķa funkcijas optimizācijai.

Ņemot vērā vienīgo pieņēmumu attiecībā uz nezināmo funkciju, proti, tai ir jābūt gludai, kā nezināmo funkciju $m \left(x \right)$ var izmantot *Nadaraya-Watson kodola novērtētāju*:

$$\tilde{g}(\boldsymbol{\tau}) = \frac{1}{\sum_{i=1}^n K_h(\boldsymbol{\tau}_i)} \sum_{i=1}^n K_h(\boldsymbol{\tau}_i) \tilde{Y}_i, \quad (29)$$

kur $\tau_i = \boldsymbol{\tau} - x_i$ ir indeksa vērtība i -jā novērojumā, Y_i ir atkarīgā mainīgā vērtība i -jā novērojumā un $K_h(\boldsymbol{\tau})$ ir kodola funkcija. Šajā pētījumā kā $K_h(\boldsymbol{\tau})$ tika izmantota tikai Gausa kodola funkcija:

$$K_h(\boldsymbol{\tau}) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\boldsymbol{\tau}}{h}\right)^2\right), \quad -\infty < \boldsymbol{\tau} < \infty, \quad (30)$$

kur h ir izlīdzināšanas parametrs (vai josla). Nezināmo koeficientu vektoru β novērtē pēc mazāko kvadrātu kritērija:

$$R(\beta) = \sum_{i=1}^n \left(Y_i - \tilde{g}(\boldsymbol{\tau}_i) \right)^2 \rightarrow \min_{\beta}. \quad (31)$$

Kritērija minimizācijai ir izmantota *gradianta metode*. Atbilstošā gradianta formula ir sekojoša:

$$\nabla R(\beta) = -2 \sum_{i=1}^n \left(Y_i - \frac{\sum_{i=1}^n K_h(\boldsymbol{\tau}_i) \tilde{Y}_i}{\sum_{i=1}^n K_h(\boldsymbol{\tau}_i)} \right) \left(\sum_{i=1}^n K_h(\boldsymbol{\tau}_i) \right)^{-2} \left(\frac{1}{h} \sum_{i=1}^n \frac{\partial}{\partial \tau_i} K_h(\boldsymbol{\tau}_i) \left(Y_i \sum_{i=1}^n K_h(\boldsymbol{\tau}_i) - \tilde{Y} \right) x_i \right), \quad (32)$$

kur

$$\tilde{Y} = \sum_{i=1}^n K_h(\boldsymbol{\tau}_i) \tilde{Y}_i \quad (33)$$

un

$$\frac{\partial}{\partial \tau_i} K_h(\boldsymbol{\tau}_i) = -\frac{\tau_i}{h^2 \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\boldsymbol{\tau}_i}{h}\right)^2\right) \quad (34)$$

ir Gausa kodola funkcijas (30) pirmais atvasinājums.

Visnozīmīgākā viena indeksa modeļa atrašanas procedūras vienkāršošanai, kā arī parametrisko un pusparametrisko modeļu efektivitātes salīdzināšanai ir piedāvāts izmantot *atlikumu kvadrātu summu* R_0 , kura tiek aprēķināta sekojošā veidā:

$$R_0 = \frac{1}{n-d} \sum_{i=1}^n \left(Y_i - \hat{Y}_i \right)^2 \quad (35)$$

kur n ir novērojumu skaits, d ir novērtējamo koeficientu skaits, Y_i ir novērotā vērtība un \hat{Y}_i ir novērtētā gan ar lineāro, gan ar viena indeksa modeļa palīdzību. Tātad \hat{Y}_i var tikt attēlots ar $\tilde{g}(\boldsymbol{\tau}_i)$ vai ar $\hat{\beta}^T x_i$ palīdzību.

Zemāk ir izklāstīta visatbilstīgākā viena indeksa modeļa atrašanas *I. Procedūra*.

I. Procedūras sākums.

1. Iekļaut modelī visnozīmīgākos faktoros pēc atbilstošā labākā lineārā modeļa izvērtēšanas rezultātiem (skat. atbilstošo *II. Procedūru* zemāk). Noteikt modeļa kārtas numuru $n = 0$.
2. $n = n + 1$. Izvēlēties sākuma vērtību izlīdzināšanas parametram $h^{(n)}$ uz pārliecinošo argumentu bāzes. Vērtību var izvēlēties uz *Mahalanobisa attāluma* bāzes, vai izmantojot *AMISE* formulu.
3. Izvēlēties adekvātu precizitātes līmeni ε kritērija (31) optimālās vērtības aprēķinam, un precizitātes līmeni η gradienta (32) optimālās vērtības aprēķinam.
4. Piešķirt $i = 0$ vērtību. Izvēlēties sākuma vērtību nezināmajiem koeficientiem $\beta^{(i)}$ uz pārliecinošo argumentu bāzes.
5. Aprēķināt gradienta $\nabla R(\beta^{(i)})$ vērtību.
6. Aprēķināt pārvietojuma virziena $\theta^{(i)}$ pašreizējo vērtību uz gradienta $\nabla R(\beta^{(i)})$ bāzes:

$$\theta^{(i)} = \frac{\nabla R(\beta^{(i)})}{\sqrt{\nabla R(\beta^{(i)})^T \cdot \nabla R(\beta^{(i)})}}. \quad (36)$$

7. Aprēķināt pārvietojuma soļa $v^{(i)}$ pašreizējo vērtību ar jebkuru piemēroto viendimensijas optimizācijas metodi (dihotomijas metode, zelta griezuma metode u.c.).
8. Noteikt $i = i + 1$. Aprēķināt nezināmo koeficientu pašreizējās vērtības $\beta^{(i)}$ pēc *visstraujākās samazināšanas metodes* (viens no *gradienta metodes* variantiem):

$$\beta^{(i)} = \beta^{(i-1)} - v^{(i)} \cdot \theta^{(i)}. \quad (37)$$

9. Apstādināšanas tests. Ja $|R(\beta^{(i-1)}) - R(\beta^{(i)})| > \varepsilon$, tad noteikt $n = n + 1$ un pāriet pie 6. soļa.
10. Gradienta vienādības nullei tests. Ja $|\nabla R(\beta^{(i)})| > \eta$, tad noteikt $n = n + 1$ un pāriet pie 5. soļa.
11. Atzīt pašreizējo novērtējumu $\beta^{(i)}$ par optimālu β^* .
12. Izveidot labākā izvēlētā viena indeksa modeļa vienādojumu, ar tā palīdzību aprēķināt prognozējamā lieluma novērtējumus.
13. Aprēķināt atlikumu kvadrātu summu $R_0^{(i)}$.
14. Ja $n = 1$, tad pāriet pie 15. soļa. Citos gadījumos pāriet pie 16. soļa.
15. Atzīt pašreizējo $R_0^{(i)}$ par atlikumu kvadrātu summas optimālo vērtību (t.s. *rekordu* $RE^{(i)}$) un pāriet pie 2. soļa.
16. Ja pašreizējais $R_0^{(i)}$ ir mazāks par pēdējo rekordu $RE^{(i-1)}$, tad atzīt to par jaunu rekordu, noteikt $RE^{(i-1)} \leftarrow R_0^{(i)}$.
17. Ja $h^{(i)}$ joprojām var mainīt, tad pāriet pie 2. soļa.

18. Optimālais *VIM* ir atrasts.

I. Procedūras beigas.

Tā kā viena indeksa modeļa efektivitātes izpēte iekļauj sevī salīdzināšanu ar lineāro modeli (ar tādiem pašiem faktoriem), visnozīmīgākā lineārā modeļa sameklēšanai ir izstrādāta *II. Procedūra*. Šī procedūra ir izklāstīta promocijas darba pilnā tekstā.

Ir jāuzsver tāds fakts, ka augstāk izklāstītās procedūras ļauj novērtēt apskatāmo modeļu efektivitāti tikai esošo datu izlīdzināšanai. Tādēļ jāizšķir *izlīdzināšanas* kritēriji un īsti *prognozēšanas*, t.i., *kross-pārbaudes* (turpmāk *KP*), kritēriji.

Izlīdzināšanas procedūra modeļa koeficientu novērtēšanai tiek veikta uz visu retrospektīvo datu pamata. Pēc tam uz novērtēto koeficientu bāzes aprēķina atkarīgā mainīga Y vērtības \hat{Y} atbilstoši esošiem novērojumiem. *Izlīdzināšanas* kvalitāti nosaka novērtējumu \hat{Y} novirzes mērs no atkarīgā mainīga Y patiesajām vērtībām.

Kross-pārbaudes elementārajā variantā datu izlasi dala divās daļās. Datus no pirmās daļas izmanto apskatāmā modeļa nezināmo koeficientu novērtēšanai. Tālāk iegūtos koeficientu novērtējumus izmanto atkarīgā mainīgā vērtību aprēķinam atbilstoši datiem no otrās izlases daļas. Tādā veidā aprēķināto prognožu vērtību un novēroto vērtību salīdzināšana ļauj verificēt apskatāmā modeļa prognozēšanas iespējas.

Paredzēta piedāvāto modeļu efektivitātes pārbaude divos gadījumos: esošo datu izlīdzināšanai un prognozēšanai. Ir iespējams pielāgot agrāk izklāstītās procedūras arī *kross-pārbaudes* gadījumam.

8.3.2. STARPTAUTISKĀ KRAVU DZELZCEĻA TRANSPORTA APGROZĪJUMA ANALĪZE UN PROGNOZĒŠANA ES VALSTĪM

Uzdevuma nostādne

Aplūkojamais pētījuma *objekts* ir Eiropas Savienības *valsts*. Par *novērojumu* tiek uzskatīti dati par objektu konkrētajā laika momentā.

Pētījuma mērķis ir izveidot dažādus regresijas modeļus un izvēlēties no tiem visefektīvākos dzelzceļa kravu transporta apgrozījuma prognozēšanai.

Prognozējamais lielums, apzīmēts ar t_0 , ir starptautiskais dzelzceļa kravu transporta apgrozījums, miljonos t/km. Statistiskās analīzes priekšmets ir 75 novērojumi par 15 valstīm ES no 1996. gada līdz 2000. gadam. Ietekmējošie faktori ir sekojoši:

- t_1 – valsts platība, tūkst. km² (*SQUARE*);
- t_2 – IKP uz vienu iedzīvotāju (*GDP_PPS*);
- t_3 – salīdzinošais cenu līmenis (*CPL*);
- t_4 – dzelzceļa līniju garums, tūkst. km, (*TOTLEN*);
- t_5 – lokomotīvu skaits, tūkst. (*LOKOM*);
- t_6 – preču vagonu skaits, tūkst. (*WAGONS*);
- t_7 – valsts platības gradācija (*GradAREA*).

Izvirzītie modeļi

Pirmie divi modeļi ir lineārie, apzīmēti ar $L1$ un $L2$. Pārējie divi modeļi ir vienindeksa modeļi, apzīmēti ar $VIM1$ un $VIM2$. Atkarīgais mainīgais $Y^{(1)} = t_0$. Ietekmējošie mainīgie ir $x_1 = t_2$, $x_2 = t_3$, $x_3 = t_2/t_3$, $x_4 = t_4$, $x_5 = t_5$, $x_6 = t_6$. Atkarīgais mainīgais $Y^{(2)} = t_0/\sqrt{t_1}$. Ietekmējošie mainīgie ir tādi paši, turklāt apskatā tika ievadīta gradācija pēc valsts platības, $x_7 = t_7$. Atkarīgais mainīgais $Y^{(IM1)} = t_0/t_1$. Atkarīgais mainīgais $Y^{(IM2)} = t_0/\sqrt{t_1}$. Ietekmējošo mainīgo kopa modeļiem $VIM1$ un $VIM2$ sakrīt ar mainīgo kopu modelim $L1$. Visi augstāk minētie modeļi ir grupu modeļi.

Izvirzīto modeļu izvērtēšanai un verifikācijai tiek izmantotas gan I . un II . *Procedūras*, gan kross-pārbaudes (KP) paņēmieni, kas aprakstīts agrāk. Speciāli viena indeksa modelim tika veikta eksperimentu sērija, lai noteiktu optimālās vērtības izlīdzināšanas parametru h .

Lineāro modeļu novērtēšana

Novērtējamo modeļu $L1$ un $L2$ vienādojumi ir sekojoši:

$$\hat{E}^{(1)} = -3713 + 118x_1 + 26x_2 - 11769x_3 + 879x_4 + 549x_5 + 158x_6. \quad (38)$$

$R^2 = 0.98, F = 384.$

$$\hat{E}^{(2)} = -120.4 - 1.2x_1 + 1.4x_2 + 110.2x_3 + 0.2x_4 + 5.9x_5 + 0.3x_6 + 29.2x_7. \quad (39)$$

$R^2 = 0.98, F = 314.$

Vienindeksa modeļa novērtēšana

Koeficientu β novērtējumi tika iegūti pie optimālajām joslām $h^* = 7$ modelim $VIM1$ un $h^* = 6$ modelim $VIM2$. Novērtētais modelis tiek pierakstīts sekojošā veidā:

$$\hat{E} = \frac{\sum_{i=1}^n Y_i K_h(x_i) \hat{\beta}}{\sum_{i=1}^n K_h(x_i) \hat{\beta}} \quad (40)$$

kur Y var pieņemt vērtības $Y^{(IM1)}$ vai $Y^{(IM2)}$. Novērtētie koeficientu vektori ir $\hat{\beta}_{SIM1}^T = [10 \quad -1 \times 10^3 \quad 18 \times 10^{-5} \quad 758 \quad 155 \quad -2 \times 10^3]$ un $\hat{\beta}_{SIM2}^T = [16 \quad -1 \times 10^3 \quad -4 \quad 853 \quad 62 \quad -871]$.

Kross-pārbaudes analīze

Koeficienti β tiek novērtēti uz apskatāmā perioda (no 1996.g. līdz 1999.g.) bāzes, pēc tam tiek iegūtas prognozes vērtības analizējamajam 2000.gadam. Modeļu $L1$ un $L2$ vienādojumi ir sekojošie:

$$\hat{E}^{(1)} = 4154 + 198x_1 + 45x_2 - 20556x_3 + 899x_4 + 532x_5 + 148x_6. \quad (41)$$

$$\hat{E}^{(2)} = -120 + 28x_1 - 1.2x_2 + 1.4x_3 + 110x_4 + 0.2x_5 + 6x_6 + 0.3x_7. \quad (42)$$

Iegūtie koeficientu β novērtējumi atbilst optimālajām joslām $h^* = 7$ un $h_0 = 8$ modelim *VIM1* un $h^* = 6$ modelim *VIM2*. Novērtētie koeficientu vektori ir

$$\hat{\beta}_{SIM1}^T = \begin{pmatrix} 3160 & -22420 & 566 & 19870 & 3996 & 56310 \end{pmatrix} \quad \text{un}$$

$$\hat{\beta}_{SIM1}^T = \begin{pmatrix} 45 & -96 & 4 & 121 & 25 & 357 \end{pmatrix}; \quad \text{un}$$

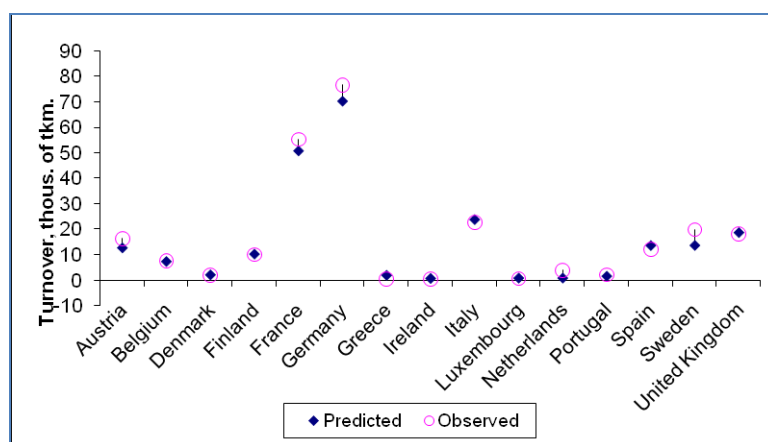
$$\hat{\beta}_{SIM2}^T = \begin{pmatrix} 63 & 1 \times 10^3 & -2 & 604 & 49 & 690 \end{pmatrix}.$$

4. tabula satur kritērija R_0 vērtības apskatāmajiem modeļiem. Novērotie un prognozētie dzelzceļa kravu apgrozījuma apjomi tiek parādīti 2. un 3. attēlā.

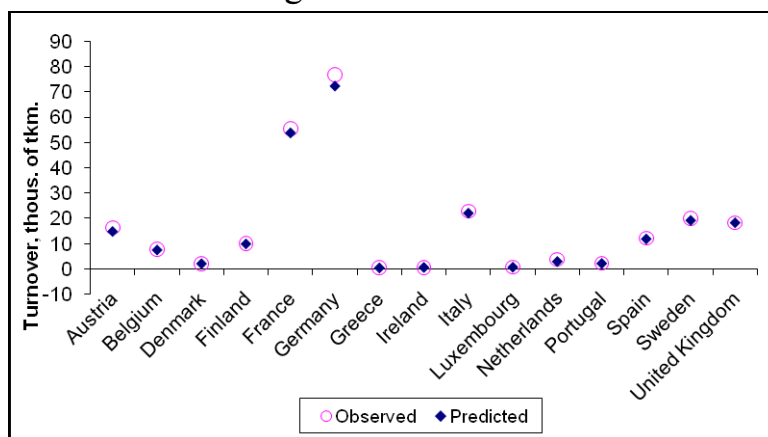
4. tabula

Kritērija R_0 vērtības

Modelis	<i>L1</i>	<i>L2</i>	<i>VIM1</i>	<i>VIM2</i>
Izlīdzināšana	11 543 065	4 830 576	894 265	565 407
KP	18 509 464	8 941 875	1 894 237	1 896 287



2.att. Prognozēšana ar modeli *L2*



3.att. Prognozēšana ar modeli *VIM2*

Bija izveidoti un izvērtēti četri regresijas modeļi, divi no tiem lineārie un divi – viena indeksa modeļi. Modeļu efektivitāte tika pētīta, balstoties uz datiem par kravu apgrozījumu 15 Eiropas Savienības valstīs. Pētījuma rezultāti liecina par to, ka viena indeksa modelis sniedz precīzākas prognozes, nekā lineārais, atbilstoši kritērija R_0 vērtībām. Turklāt, eksperimentāli ir iegūta izlīdzināšanas parametra h optimālā vērtība analizējamajam viena indeksa modelim.

8.3.3. IEKŠĒJO DZELZCEĻA PASAŽIERU PĀRVADĀJUMU ANALĪZE UN PROGNOZĒŠANA LATVIJAS REĢIONIEM

Pētījuma mērķis ir grupu parametrisko un pusparametrisko modeļu konstruēšana un pilnveidošana, kas dotu iespēju vienādi labi aprakstīt gan mazus, gan lielus dzelzceļa pasažieru pārvadājumus Latvijas reģionos.

Galvenā problēma ir pārvadājumu nevienmērīgs sadalījums. Pārvadājumi lielākiem reģioniem vairākās kārtās pārsniedz pārvadājumus pārējiem reģioniem. Viens no iespējamiem iemesliem ir tas, ka iedzīvotāji izmanto dzelzceļa transportu atšķirīgos nolūkos. Lielu reģionu iedzīvotāji izmanto elektrovilcienus kā parasto pilsētas transportu, bet mazo reģionu iedzīvotāji izmanto vilcienus tikai tālajiem braucieniem.

Sakarā ar to tika veikti vairāki pētījumi, kuros esošā statistika tika iedalīta divās grupās atkarībā no pārvadājumu lieluma. Bija izstrādāts modelis, kas domāts tikai mazu pārvadājumu aprakstam. Labākie prognozēšanas rezultāti bija iegūti pēc izlecošo novērojumu izslēgšanas atbilstoši Mahalanobisa attālumam.

Uzdevuma nostādne

Apskata *objekts* ir Latvijas reģions, kurā tiek veikti noteikti dzelzceļa pasažieru pārvadājumi, turpmāk *reģions*. Par *novērojumu* tiek uzskatīti dati par objektu konkrētajā laika momentā.

Prognozējamais lielums, apzīmēts ar t_0 , ir iekšējo dzelzceļa pasažieru pārvadājumu apjoms, tūkst. pasažieru, Latvijas zināmajam reģionam. Statistiskās analīzes priekšmets ir 113 novērojumi pēc 33 Latvijas reģioniem par periodu no 2000.g. līdz 2003.g. Ietekmējošie faktori ir sekojoši:

t_1 – iedzīvotāju blīvums, tūkst. uz platības vienību, $TP/1000*SQUARE$;

t_2 – uzņēmumu blīvums, uz platības vienību, $NE/SQUARE$;

t_3 – uzņēmumu blīvums, uz tūkst. iedzīvotāju, $NE/1000*TP$;

t_4 – nestrādājošo iedzīvotāju blīvums, tūkst. uz platības vienību, $UP/SQUARE$;

t_5 – vispārizglītojošo iestāžu blīvums, uz platības vienību, $NGEI/SQUARE$;

t_6 – autobusu blīvums, uz platības vienību, $NB/SQUARE$;

t_7 – autobusu blīvums, uz tūkst. iedzīvotāju, $NB/1000*TP$;

t_8 – dzelzceļa staciju skaits, NRS .

Abos modeļos atkarīgie mainīgie $Y^{(1)}$ un $Y^{(2)}$ ir t_0 . Ietekmējošie mainīgie abos modeļos ir visi 8 augstākminētie.

Katrs eksperiments sastāv no abu modeļu novērtēšanas izlīdzināšanas procesā. Pēc tam abi modeļi ir izvērtēti prognozēšanai, izmantojot kross-pārbaudes paņēmieni: nezināmie koeficienti β tiek vērtēti uz apskatāmā perioda

bāzes – no 2000. gada līdz 2002. gadam. Izmantojot iegūtos novērtējumus, tiek prognozēti pārvadājumi analizējamajam periodam – 2003. gadam. Tālāk tiks parādīti galvenie rezultāti bez liekiem komentāriem.

Pilnās statistikas izlases analīzes rezultāti

Tika apstrādāti visi 113 novērojumi visos 33 reģionos. Turklāt daļai reģionu statistikas par visu apskatāmo periodu trūkst.

Izlīdzināšana. Labākajam lineārajam modelim atbilst sekojošais vienādojums:

$$\hat{E} = -75 - 4x_1 + 323x_2 + 168x_3 - 476x_7 + 146x_8. \quad (43)$$

$$R^2 = 0.94, F = 358.$$

Koeficientu β novērtēšana vienindeksa modelim bija veikta pie joslas h dažādām vērtībām. Labākais izvēlētais vienindeksa modelis atbilst optimālajai joslai $h^* = 1$ un var tikt prezentēts formā (40), pie novērtētā koeficientu vektora $\hat{\beta}^T = [183 \quad -395 \quad 7375 \quad -3523 \quad 0.1]$.

Kross-pārbaude. Vienādojums izvēlētajam lineārajam modelim ir sekojošs:

$$\hat{E}_{cv} = -167 - 3.9x_1 + 311x_2 + 144x_3 - 406x_7 + 140x_8. \quad (44)$$

$R^2 = 0.94, F = 245$. Šis modelis sniedz negatīvas prognozes 7 reģioniem, t.i., apmēram 30% novērojumu.

Labākais izvēlētais vienindeksa modelis atbilst optimālajai joslai $h^* = 30$, pie novērtētā koeficientu vektora $\hat{\beta}_{cv}^T = [34500 \quad 3975 \quad 8273 \quad -1287 \quad 24270]$.

Kritērija R_0 vērtības abiem analizētajiem modeļiem izlīdzināšanas un kross-pārbaudes (KP) gadījumā ir apkopoti 5. tabulā. 6. tabulā ietilpst ar abu modeļu palīdzību novērotie un novērtētie pārvadājumu apjomi KP gadījumā analizējamam periodam, un relatīvu kļūdu δ_i katram reģionam, %.

5. tabula

Kritērija R_0 vērtības		
	Izlīdzināšana	KP
LM	884 141	1 181 086
VIM	133 556	466 016

Nepilnīgu datu analīzes rezultāti

Tika apstrādāti 88 novērojumi no 22 reģioniem. Tika ņemti vērā reģioni, kuriem pieejama statistika par visu apskatāmo periodu.

Izlīdzināšana. Labākā izvēlēta lineārā modeļa vienādojums ir:

$$\hat{E} = -2094 - 19x_1 + 389x_2 + 83x_3 + 75995x_5 - 886x_6 + 155x_8. \quad (45)$$

$R^2 = 0.95, F = 306$. Labākais izvēlētais viena indeksa modelis atbilst optimālajai joslai $h^* = 1$, pie novērtētā koeficientu vektora

$$\hat{\beta}^T = \begin{pmatrix} 48 & -73 & 226 & -0.1 & -188 & 0.1 \end{pmatrix}.$$

Kross-pārbaude. Labākā novērtētā lineārā modeļa vienādojums ir:

$$\hat{E}_{cv} = -2135 - 21x_1 + 380x_2 + 75x_3 + 86963x_5 - 928x_6 + 146x_8. \quad (46)$$

$R^2 = 0.96$, $F = 225$. Labākais izvēlētais viena indeksa modelis atbilst optimālajai joslai $h^* = 30$, pie novērtētā koeficientu vektora $\hat{\beta}_{cv}^T = \begin{pmatrix} 1620 & 235 & 2216 & 3 & 132 & 7522 \end{pmatrix}$. Kritērija R_0 vērtības ir apkopotas 7. tabulā.

6. tabula

Kross-pārbaudes rezultāti

	Novērotie pārvadājumi	Prognozes		Relatīvā kļūda, %	
		LM	VIM	LM	VIM
Rīga	23323.05	20067.43	22370.00	13.96	4.09
Daugavpils	259.85	979.57	301.83	-276.98	-16.16
Jelgava	1676.11	1611.31	1412.00	3.87	15.76
Jūrmala	6436.22	3360.92	4682.00	47.78	27.26
Rēzekne	192.50	1532.21	193.77	-695.95	-0.66
Aizkraukles r.	795.41	427.85	120.98	46.21	84.79
Cēsu r.	272.15	332.88	174.91	-22.31	35.73
Daugavpils r.	47.71	-618.34	72.59	1395.97	-52.14
Dobeles r.	43.10	78.08	142.96	-81.17	-231.72
Gulbenes r.	10.25	183.36	94.47	-1689.53	-822.01
Jelgavas r.	267.40	315.14	132.57	-17.85	50.42
Jēkabpils r.	251.74	522.49	126.90	-107.55	49.59
Limbažu r.	96.13	-71.44	85.70	174.32	10.85
Ludzas r.	90.78	-152.40	84.19	267.88	7.26
Madonas r.	66.30	-165.41	130.82	349.50	-97.33
Ogres r.	4108.32	2269.00	2149.00	44.77	47.69
Preiļu r.	104.37	155.95	100.04	-49.43	4.14
Rēzeknes r.	53.34	-119.31	81.32	323.67	-52.46
Rīgas r.	6219.64	5660.02	4813.00	9.00	22.62
Saldus r.	9.36	-249.69	115.56	2768.52	-1134.99
Tukuma r.	698.92	332.84	149.75	52.38	78.57
Valkas r.	74.60	-97.24	89.15	230.34	-19.50
Valmieras r.	86.74	62.98	185.40	27.39	-113.74

7. tabula

Kritērija R_0 vērtības

	Izlīdzināšana	KP
LM	823 311	1 335 774
VIM	171 435	909 120

Diemžēl lineārais modelis sniedz negatīvas prognozes 8 reģioniem, t.i. apmēram 36% novērojumu. Vienindeksa modelis apraksta mazos pārvadājumus efektīvāk nekā lineārais. Šis fakts rada priekšnoteikumu veidot atšķirīgus

modeļus mazo un lielo pārvadājumu apjomu analīzei un prognozēšanai.

Mazo pārvadājumu atsevišķa analīze un prognozēšana

Tika apstrādāti 95 novērojumi no 28 reģioniem, t.i., mazie pārvadājumi. *Izlīdzināšana.* Vienādojums labākajam novērtētajam lineārajam modelim ir:

$$\hat{E} = -255 - 82x_2 + 16x_3 - 34x_4 + 13033x_5 + 18x_8. \quad (47)$$

$R^2 = 0.48, F = 17$. Labākais izvēlētais viena indeksa modelis atbilst optimālajai joslai $h^* = 0.5$ pie novērtētā koeficientu vektora $\hat{\beta}^T = (2 \quad 265 \quad 78 \quad 0.3 \quad 0.1)$.

Kross-pārbaude. Labākais lineārās regresijas vienādojums ir sekojošs:

$$\hat{E}_{cv} = -223 - 74x_2 + 14x_3 - 28x_4 + 11258x_5 + 16x_8. \quad (48)$$

$R^2 = 0.42, F = 10$. Labākais izvēlētais viena indeksa modelis atbilst optimālajai joslai $h^* = 2$, pie novērtētā koeficientu vektora $\hat{\beta}_{cv}^T = (-38 \quad 2555 \quad 180 \quad 0.1 \quad 1561)$. Kritērija R_0 vērtības ir apkopotas 8. tabulā.

8. tabula

Kritērija R_0 vērtības

	Izlīdzināšana	KP
LM	42 430	79 029
VIM	8 877	40 992

Izlecošo novērojumu izslēgšana atbilstoši Mahalanobisa attālumam

Tika apstrādāts 91 novērojums. Tika izdarīti sekojoši esošā modeļa uzlabojumi:

- 1) kvalitatīvā mainīgā ievadīšana modelī;
- 2) izlecošo novērojumu izslēgšana atbilstoši Mahalanobisa attālumam.

Izlīdzināšana. Vienādojums labākajam novērtētajam lineārajam modelim ir:

$$\hat{E} = -20x_1 + 436x_2 + 163x_3 + 64418x_5 - 734x_6 - 398x_7 + 147x_8. \quad (49)$$

$R^2 = 0.96, F = 258$.

Tagad kvalitatīvais mainīgais GRAD ir ievadīts modelī. Tas pieņem vērtību 1 tiem reģioniem, kuriem apriori raksturīgi lieli pārvadājumi, un vērtību 0 – visiem pārējiem reģioniem. Pēc uzlabošanas lineārā modeļa vienādojums var tikt pierakstīts sekojošā formā:

$$\hat{E} = -20x_1 + 618x_2 + 81x_3 + 122x_4 + 38525x_5 - 921x_6 + 63x_8 + 2743x_9. \quad (50)$$

$R^2 = 0.97, F = 339$. Labākais izvēlētais viena indeksa modelis atbilst optimālajai joslai $h^* = 1$, pie novērtētā koeficientu vektora

$$\hat{\beta}^T = \begin{pmatrix} 959 & -386 & 7195 & 0.1 & -3437 & 0.1 \end{pmatrix}.$$

Regresijas vienādojums pēc izlecošo novērojumu izslēgšanas ir:

$$\hat{E}_{LM} = 2307x_2 + 59x_3 - 393x_4 - 1207x_6 - 361x_7 + 128x_8. \quad (51)$$

$R^2 = 0.91$, $F = 91$. Labākais izvēlētais viena indeksa modelis atbilst optimālajai joslai $h^* = 1$, pie koeficientu vektora $\hat{\beta}_M^T = \begin{pmatrix} 6 & 820 & 24 & 4 & 114 & 0.1 \end{pmatrix}$.

Kross-pārbaude. Lineārās regresijas vienādojumam ir sekojošā formā:

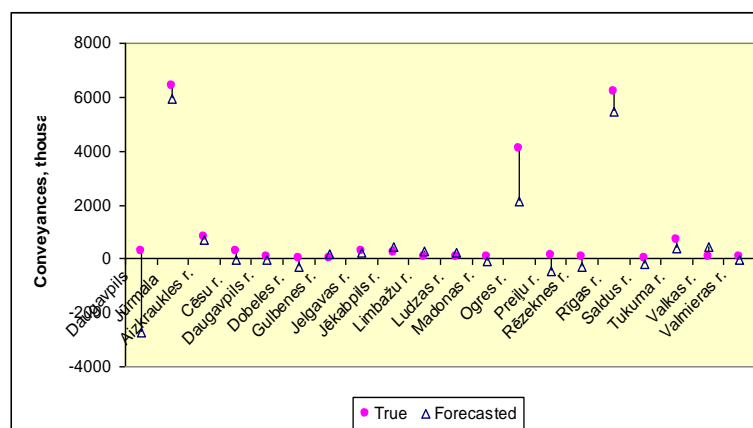
$$\hat{E}_{CV} = 2692 + 3863x_2 - 525x_4 - 117031x_5 - 416x_7 + 110x_8. \quad (52)$$

$R^2 = 0.91$, $F = 94$. Labākais izvēlētais viena indeksa modelis atbilst optimālajai joslai $h^* = 1$, pie novērtētā koeficientu vektora $\hat{\beta}_{CV}^T = \begin{pmatrix} 83 & -380 & 4 & 2709 & 0.1 \end{pmatrix}$. Kritērija R_0 vērtības ir savāktas 9. tabulā. Novērotie un prognozētie pārvadājumu apjomi ir pretstatīti 4. un 5. attēlā.

9. tabula

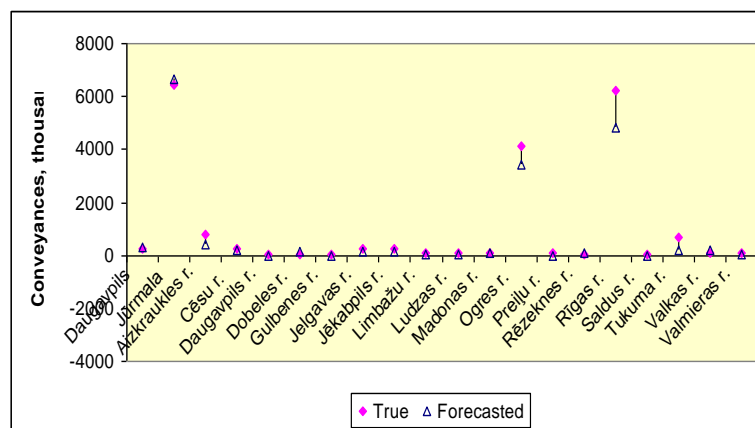
Kritērija R_0 vērtības

Modelis	Izlīdzināšana		KP
	pirms	pēc	
	izlecošo novērojumu izslēgšanas		
LM	801 111 / 546 895 ²	259 096	739 151
VIM	165 843	10 053	147 810



4.att. Prognozēšana ar Lineāro Modeli

² ieskaitot mainīgo GRAD



5.att. Prognozēšana ar viena indeksa modeli

Tika iegūti regresijas modeļi, kuri ļauj noteikt svarīgo ekonomisko un sociālo faktoru ietekmi uz pasažieru dzelzceļa pārvadājumu apjomiem Latvijas reģionos. Rezultātā ir salīdzināti divi grupu modeļi: daudzfaktoru lineārais regresijas modelis un vienindeksa modelis. Izlecošo novērojumu izslēgšana no statistikas datiem atbilstoši Mahalanobisa attālumam samazina izlīdzināšanas kļūdu, un savukārt palielina prognozēšanas precizitāti. Analīzes rezultāti liecina par vienindeksa modeļa priekšrocībām izlīdzināšanas un prognozēšanas gadījumos.

8.4. STARPTAUTISKO PASAŽIERU AVIOPĀRVADĀJUMU PROGNOZĒŠANA ES VALSTĪM UZ SURE-MODEĻA BĀZES

SURE-modelis, kuru piedāvāja *A.Zellner* 1962. gadā, ir piemērots un lietderīgs līdzeklis plašam uzdevumu klāstam ekonometrikā, loģistikā u.c. Šajā darbā tiek pētīts viens no iespējamajiem SURE-modeļa vispārinājumiem. Pieņem, ka atsevišķie novērojumi nesatur informāciju par visiem atkarīgajiem mainīgajiem. Ir iegūts kovariācijas matricas nenovirzītais novērtējums. Pierādīta SURE-modeļa pielietošanas priekšrocība salīdzinājumā ar daudzdimensiju modeli, statistikas datu nepilnības gadījumā, uz kopējo pasažieru aviopārvadājumu prognozēšanas piemēra, kā arī uz *S.Bogdanovas* maģistra darba praktisko rezultātu bāzes, izpildīta kandidātes vadībā.

Uzdevuma nostādne

SURE-modelis (6) un tā svarīgākās atšķirības no daudzdimensiju modeļa (4) ir aprakstītas esošā kopsavilkuma 3. nodaļā. Tātad mērķis ir iegūt prognozi summas

$$W(t) = \sum_{i=1}^G Y_{i,j(t)} = \sum_{i=1}^G \left(\sum_{v=1}^{m_i} \beta_{i,v} x_{i,j(t)}^{(v)} + Z_{i,j(t)} \right) \quad (53)$$

matemātiskajai cerībai, nākamajam laika momentam t . Nezināmie koeficienti $\beta_{i,v}$ tiek novērtēti pēc zināmās formulas

$$\beta_i^* = \left(X_i^T X_i \right)^{-1} X_i^T Y_i, \quad (54)$$

kur $\beta_i^* = \left(\beta_{i,1}^*, \beta_{i,2}^*, \dots, \beta_{i,m_i}^* \right)^T$ un $Y_i = \left(Y_{i,1}, Y_{i,2}, \dots, Y_{i,n_i} \right)^T$ ir regresijas koeficientu novērtējumu vektors un atkarīgā mainīgā vērtību vektors. Nenovirzītais novērtējums summas $W(t)$ (53) matemātiskajai cerībai:

$$W^*(t) = \sum_{i=1}^G Y_{i,j(t)}^* = \sum_{i=1}^G x_{i,j(t)} \beta_i^* = \sum_{i=1}^G \left(\sum_{v=1}^{m_i} \beta_{i,v}^* x_{i,j(t)}^{(v)} \right). \quad (55)$$

Atbilstoši formulai (54), divu vektoru β_i^* un β_l^* kovariācijas matrica tiek aprēķināta pēc formulas

$$\text{Cov}(\beta_i^*, \beta_l^*) = \left(X_i^T X_i \right)^{-1} X_i^T \text{Cov}(Y_i, Y_l) X_l \left(X_l^T X_l \right)^{-1}. \quad (56)$$

Kovariācijai $\text{Cov}(Y_i, Y_l)$ arī jābūt noteiktai. Lai $D^{(i,l)}$ ir $(n_i \times n_l)$ -matrica, kuras elements $D_{j,f}^{(i,l)} = 1$ ja $t_{i,j} = t_{l,f}$ un $D_{j,f}^{(i,l)} = 0$ - pretējā gadījumā. Divu atkarīgo mainīgo $Y_{i,j}$ un $Y_{l,f}$ kovariācija vienam un tam pašam laika momentam $t_{i,j} = t_{l,f}$ ir vienāda ar $c_{i,l}$. Tādēļ $\text{Cov}(Y_i, Y_l) = c_{i,l} D^{(i,l)}$. Pārrakstīsim formulu (56) sekojošā veidā:

$$\text{Cov}(\beta_i^*, \beta_l^*) = c_{i,l} \left(X_i^T X_i \right)^{-1} X_i^T D^{(i,l)} X_l \left(X_l^T X_l \right)^{-1}. \quad (57)$$

Nezināmās kovariācijas $\{c_{i,l}\}$ novērtēšanai tiek izmantots mazāko kvadrātu novērtētājs:

$$c_{i,l}^* = \frac{1}{v_{i,l}} \left(Y_i - X_i \beta_i^* \right)^T D^{(i,l)} \left(Y_l - X_l \beta_l^* \right), \quad (58)$$

kur $v_{i,l}$ ir konstante, definēta pēc nenovirzītā novērtētāja nosacījumiem. Konstantes $v_{i,l}$ noteikšanai ir vajadzīga kovariācijas novērtējuma $c_{i,l}^*$ matemātiskā cerība.

Gadījumā, ja $R_j(i)^T$ ir matricas $\left(Y_i - X_i (X_i^T X_i)^{-1} X_i^T \right)$ j -ā rinda, tad $f(l, i, j)$ ir l -tā objekta novērojuma numurs, kuram laika moments sakrīt ar $t_{i,j}$, vai ir vienāds ar nulli, ja novērojuma ar tādu numuru vispār nav:

$$f(l, i, j) = \sum_{v=1}^{n_l} v D_{j,v}^{(i,l)}. \quad (59)$$

Tātad

$$E(c_{i,l}^*) = \frac{1}{v_{i,l}} \sum_j \sum_f E \left(R_j(i)^T D^{(i,l)} R_f(l) Z_{l,f} \right) \\ = \frac{1}{v_{i,l}} \sum_j R_j(i)^T D^{(i,l)} R_{f(l,i,j)}(l) E(Z_{i,j} Z_{l,f(l,i,j)}) = \frac{1}{v_{i,l}} c_{i,l} \sum_j R_j(i)^T D^{(i,l)} R_{f(l,i,j)}(l).$$

Nemot vērā normalizējošās konstantes $v_{i,l}$ sekojošo vērtību,

$$v_{i,l} = \sum_j R_j(i)^T D^{(i,l)} R_{f(l,i,j)}(l), \quad (60)$$

formula (58) sniedz kovariācijas $c_{i,l}$ nenovirzīto novērtējumu. Tagad summas (55) dispersija tiek rēķināta kā parasti:

$$\text{Var} \left(\sum_{i=1}^G x_{i,j(t)} \beta_i^* \right) = \sum_{i=1}^G x_{i,j(t)} \text{Cov}(\beta_i^*) x_{i,j(t)}^T + \\ + 2 \sum_{i=1}^{G-1} \sum_{l=i+1}^G x_{i,j(t)} \text{Cov}(\beta_i^*, \beta_l^*) x_{l,j(t)}^T. \quad (61)$$

Šeit kovariācija $\text{Cov}(\beta_i^*, \beta_l^*)$ tiek aprēķināta pēc formulas (57), un kovariācijas matricas katram vektoram β_i^* tiek aprēķinātas pēc zināmās formulas

$$\text{Cov}(\beta_i^*) = c_{i,i} \left(X_i^T X_i \right)^{-1} = \sigma_i^2 \left(X_i^T X_i \right)^{-1}. \quad (62)$$

Kopējo pasažieru aviopārvadājumu prognozēšana ES valstīm

Piedāvātā pieeja tiek demonstrēta uz kopējo pasažieru aviopārvadājumu prognozēšanas piemēra ES valstīm. Mūsu rīcībā ir divi prognozējamie lielumi: iekšējie un ārējie starptautiskie pasažieru aviopārvadājumi, apzīmēti kā Intra-EU un Extra-EU, tūkstošos pasažieru. Abi šie lielumi tiek prognozēti vienlaikus, pateicoties pastāvošai stiprai savstarpējai korelācijai, t.i., tiek iegūta divu apskatāmu atkarīgo mainīgo *kopējā prognoze*.

Radās nepieciešamība iegūt kopējo pārvadājumu prognozes arī statistikas datu nepilnības gadījumā. Atbilstošais regresijas modelis saturēja svarīgus ekonomiskus un sociālus faktoros, kas ietekmējuši Intra- un Extra-EU aviopārvadājumus katrai valstij. Šajā pētījumā bija izanalizētas 24 valstis, t.i., gandrīz visas ES dalībnieces, izņemot Bulgāriju, Zviedriju un Somiju.

Tātad apskatā ir divi objekti ($G = 2$) ar numuriem $i = 1, 2$. Katram objektam ir viens atkarīgais mainīgais: pirmā objekta atkarīgais mainīgais Y_1 ir starptautisko Intra-EU pasažieru aviopārvadājumu apjoms, otrā objekta atkarīgais mainīgais Y_2 ir Extra-EU aviopārvadājumu apjoms. Modelis satur 7 neatkarīgus mainīgos. Pirmie trīs mainīgie abiem objektiem ir kopēji:

t_1 – valsts platība, tūkst. km², *SQUARE*;

t_2 – nodarbinātības norma, t.i., kopējais strādājošo cilvēku skaits valstī, izmaiņu

procents attiecībā pret iepriekšējo gadu, *EGAP*;

t_3 – valsts iedzīvotāju skaita izmaiņas, algebriskā dabiskā pieauguma un tīkla migrācijas summa, tūkst. iedzīvotāju, *PC*.

Nākamie divi attiecas tikai uz pirmo objektu:

t_4 – populācijas apjoms, tūkst. iedzīvotāju, *TP*;

t_5 – iekšzemes kopprodukta pieauguma tempi, procentos, attiecībā pret iepriekšējo gadu, *GDPrate*.

Pēdējie divi apraksta tikai otro objektu:

t_6 – galīgais enerģijas patēriņš pēc visiem starptautiskā un iekšējā transporta veidiem, tūkst. tonnu, *FECT*;

t_7 – galīgais elektroenerģijas patēriņš, kurš tiek patērēts ražošanā, transportā, mājsaimniecībās, *GWt*, *CEITH*.

Analizējamais periods ir no 2000. līdz 2007. gadam. Turklāt vairāki statistiskie dati abiem objektiem trūkst. Šajā gadījumā parasti pielieto divus paņēmienus:

- novērojumi pēc atbilstošiem laika momentiem tiek izslēgti no visiem objektiem;
- trūkstošās vērtības tiek aizvietotas ar piemērotiem novērtējumiem.

Abi šie paņēmieni kropļo statistikas datus un prognozēšanas rezultātus. Piedāvātās pieejas efektivitātes analīzei reģistrēsim laika momentus, kuros statistikas datu trūkst. Tātad pirmajā objektā ir 122 novērojumi, otrajā – 100. Tiek pētīts, kādā mērā iztrūkstošo novērojumu skaits ietekmē pāravadājumu summas dispersiju.

Datu apstrādei tiek izmantotas divas pieejas:

1. zināmie novērojumi trūkst un statistikas dati tiek izslēgti no abiem objektiem pēc visiem atbilstošiem laika momentiem, pēc kā katram objektam paliek 99 novērojumi;
2. zināmie novērojumi trūkst un tiek izmantota piedāvātā pieeja.

Kā efektivitātes kritērijs tiek izmantots *kopējās prognozes dispersijas novērtējums*. Tiek apstrādāti dati par analizējamo periodu un tiek iegūtas prognozes 2008. gadam uz novērtēto koeficientu bāzes. Ir nepieciešams iegūt novērtētu dispersiju gadījumā, ja lielumi $Y_{1,j}$ un $Y_{2,j}$ ir atkarīgi. Šim nolūkam ir iespējams izmantot formulu (61). Šajā eksperimentā mēs rīkojamies ar divām matricām R_1 un R_2 , vienu matricu $D^{(1,2)}$ un vienu normalizējošo konstanti $v_{1,2}$.

Normalizējošās konstantes aprēķins aizņem daudz laika, jo ir nepieciešams veikt virkni operāciju ar lielās dimensijas matricām. Novērtētie regresijas koeficienti un kovariācijas matrica izskatās sekojoši.

Iegūtie ar pirmās pieejas palīdzību:

$$\beta_1^* = \begin{pmatrix} 0.02 \\ 49.59 \\ -21.32 \\ 4.15 \\ -0.66 \end{pmatrix}, \beta_2^* = \begin{pmatrix} 8.79 \\ -28.35 \\ 28.82 \\ 0.22 \\ 164.29 \end{pmatrix}, \text{ un } \tilde{\Sigma} = \begin{pmatrix} 1.684 \cdot 10^8 & 1.687 \cdot 10^7 \\ 1.687 \cdot 10^7 & 8.234 \cdot 10^6 \end{pmatrix}.$$

Iegūtie ar otrās pieejas palīdzību:

$$\beta_1^* = \begin{pmatrix} -0.19 \\ 49.12 \\ -21.09 \\ 4.14 \\ -0.66 \end{pmatrix}, \beta_2^* = \begin{pmatrix} 11.93 \\ -30.78 \\ 29.68 \\ 0.21 \\ 102.48 \end{pmatrix}, \text{ un } \tilde{\Sigma} = \begin{pmatrix} 1.668 \cdot 10^8 & 8.538 \cdot 10^6 \\ 8.538 \cdot 10^6 & 8.994 \cdot 10^6 \end{pmatrix}.$$

Normalizējošā konstante ir $v_{1,2} = 92.976$.

Kopējo pasažieru aviopārvadājumu apjomu prognozēšanas rezultāti ir salīdzināti 10. un 11. tabulā. 10. tabulas pirmā kolonna satur kopējo pasažieru aviopārvadājumu novērotās vērtības 2008. gadam; nākamās divas kolonnas satur prognozes vērtības, iegūtas ar abu pieeju palīdzību; kārtējās divas kolonnas satur dispersijas; un pēdējās divas kolonnas reprezentē 95% ticamības intervālu augšējās robežas, arī atbilstoši abām pieejām. 11. tabulā ir apkopotas R_0 vērtības izlīdzināšanas un prognozēšanas gadījumos.

10. tabula

Prognozēšanas rezultāti

ES valsts	2008	2008*, pieejas Nr.		Dispersija, tūkst., pieejas Nr.		95% augš. ticamības intervāls, pieejas Nr.	
		1	2	1	2	1	2
Beļģija	21 982	19 283	19 499	6 197	5 572	35 495	36 061
Čehija	13 429	14 548	14 454	2 952	2 602	26 676	26 350
Dānija	24 629	15 217	15 377	10 600	8 764	30 073	30 295
Vācija	165 759	129 700	129 710	42 180	36 090	222 560	223 376
Igaunija	1 804	3 829	3 692	3 585	3 134	8 958	9 385
Īrija	30 016	18 900	18 916	3 968	3 475	34 053	34 289
Griekija	34 790	29 426	29 539	4 111	3 625	51 566	51 584
Spānija	161 401	114 130	115 750	30 750	26 100	196 267	198 208
Francija	122 724	142 710	141 520	18 080	16 200	240 645	239 066
Itālija	106 300	105 940	106 630	22 630	20 780	181 543	182 349
Kipra	7 218	5 209	5 400	4 798	4 197	11 903	12 448
Latvija	3 687	5 717	5 656	3 513	2 999	12 450	12 116
Lietuva	2 552	6 439	6 349	3 262	2 820	13 522	13 166
Luksemburga	1 713	10 221	9 945	3 951	3 642	20 022	19 440
Ungārija	8 429	15 116	15 050	3 002	2 649	27 632	27 351
Malta	3 125	2 013	2 203	3 064	2 710	6 001	6 484
Holande	50 419	41 197	41 496	5 523	4 843	71 172	71 908
Austrija	23 900	14 515	14 624	3 736	3 222	26 748	27 153
Polija	18 727	33 708	33 346	12 030	10 450	60 583	60 376
Portugāle	25 047	23 317	23 344	3 034	2 648	40 909	41 141
Rumānija	8 031	11 260	10 960	6 872	5 842	22 430	22 274
Slovēnija	1 649	2 407	2 181	4 189	3 724	7 304	6 742
Slovākija	2 596	2 779	2 684	2 930	2 619	7 364	7 056
Lielbritānija	213 888	183 880	184 090	48 430	46 640	312 976	313 108

Kritērija R_0 vērtības

	Pieejas Nr	
	1	2
Izlīdzināšana	386 427 788	243 593 829
Prognozēšana	235 066 477	224 884 525

Ir redzams, ka prognozes, iegūtas ar mūsu piedāvātās pieejas palīdzību, ir tuvākas reāliem kopējiem pasažieru aviopārvadājumiem, nekā iegūtas ar agrāk izmantoto pieeju. Tas uzskatāmi demonstrē mūsu piedāvātās pieejas priekšrocību kopējo pārvadājumu prognožu iegūšanai datu nepilnības gadījumā.

NOBEIGUMS

1. Promocijas darbs ir veltīts mūsdienīgu statistisko metožu izstrādāšanai un pielietošanai pārvadājumu apjomu analīzei un prognozēšanai valstīm – Eiropas Savienības dalībniecēm, īpaši izceļot Latviju. Tā kā patlaban visā pasaulē vērojams gan kravu, gan pasažieru pārvadājumu apjomu pieaugums, izvēlētais virziens ir perspektīvs, un iesniegtais darbs ir visai aktuāls.
2. Pētījumā ietilpst sekojošie pārvadājumi: pasažieru un kravu; starptautiskie un iekšējie; dzelzceļa un aviācijas; prognozēšanas objekts ir valsts (reģions) vai OD-pāris; prognozēšanas indikators ir pārvadājuma apjoms vai apgrozījums. Augstākminēto pārvadājumu analīze un prognozēšana tika izvērstas uz parametrisko un pusparametrisko modeļu bāzes. No parametriskajiem modeļiem tika izmantoti gan lineārie, respektīvi, daudzfaktoru, daudzdimensiju un SURE-modeļi, gan arī nelineārais modificētais gravitācijas modelis; tāpat darbā tika plaši izmantots pusparametriskais viena indeksa modelis.
3. Izpētīts plašs jautājumu klāsts, kas attiecas uz veiktā pētījuma uzdevumiem: pārvadājumu tipu klasifikācija; faktori, kas ietekmē pārvadājumu apjomus; dažādas prognozēšanas pieejas un metodes.
4. Veikta tāda literatūras avotu, kas veltīti gan parametriskajiem, gan pusparametriskajiem pārvadājumu apjomu prognozēšanas metodēm un modeļiem, kā arī korespondenču matricas novērtēšanas metodēm, analīze.
5. Par sākotnējo modeli, prognozējot gan pasažieru un kravu pārvadājumu apjomus, gan pasažieru un kravu apgrozījumu, tika izmantots klasiskais daudzfaktoru lineārās regresijas modelis. Tika izveidoti vairāki lineārās regresijas modeļi. Atšķirībā no lineārās regresijas modeļiem, kas izklāstīti vairumā skatīto zinātnisko publikāciju, visi darbā apskatītie modeļi ir grupu modeļi, un modeļos tika izmantots daudz lielāks ietekmējošo faktoru un to kombināciju skaits.
6. Apkoposim uzdevumus, kas atrisināti saskaņā ar izmantoto modeļu klasifikāciju.

- Pusparametrisko modeļu izstrādāšana un izvērtēšana dzelzceļa kravu apgrozījuma apjomu prognozēšanai ES valstīm un dzelzceļa pasažieru pārvadājumu apjomu prognozēšanai no Latvijas reģioniem. Šajā ziņā tika komplektētas atbilstošās statistiskas datu bāzes, kas saturēja informāciju par prognozējamās indikatorus ietekmējošiem faktoriem. Tika piedāvāta viena indeksa modeļa novērtēšanas metodika, izstrādātas visefektīvāko modeļu izvēles procedūras. Vairāku eksperimentu gaitā tika pierādītas viena indeksa modeļa priekšrocības salīdzinājumā ar lineāro regresijas modeli kā datu izlīdzināšanas gadījumā, tā arī prognozēšanas nolūkos. Pēdējā gadījumā tika izmantota kross-pārbaudes pieeja.
- Nelineārā parametriskā modeļa izstrādāšana korespondenču matricas novērtēšanai dzelzceļa pasažieru nosūtījumiem starp ES valstīm. Piedāvātais modelis ir gravitācijas modeļa oriģinālā modifikācija. Izstrādātā modeļa novērtēšanas procedūra tika realizēta un verificēta. Iegūto novērtējumu salīdzināšanas rezultāti ar reālām korespondencēm liecina par piedāvātā izstrādātā modeļa augsto efektivitāti.
- Summāro kravu aviopārvadājumu un summāro pasažieru aviopārvadājumu apjomu prognozēšana ES valstīm. Šim nolūkam tika izmantots daudzdimensiju regresijas modelis un SURE-modelis. Pēdējam modelim tika izstrādāta oriģināla procedūra summārās prognozes dispersijas aprēķinam. Eksperimentāli parādītas SURE-modeļa izmantošanas priekšrocības prognozēšanas nolūkā datu nepilnības gadījumā, salīdzinot prognožu dispersijas, kas iegūtas ar abu pieeju palīdzību.

Visi aprēķini tika veikti ar Statistica 6.0 lietojumprogrammas palīdzību. Visas procedūras, kas izstrādātas darba izpildes gaitā, tika programmētas MathCad13 vidē.

7. Autores izstrādātie modeļi un metodes kravu dzelzceļa pārvadājumu apjomu prognozēšanai no ES valstīm tika izmantoti zinātniskajā projektā U7107 „Matemātisko modeļu un to novērtēšanas metodes izstrādāšana Baltijas reģiona transporta plūsmu analīzei un prognozēšanai”, kas veidoja daļu no zinātniskā projekta II IZM-RTU „Zinātniskās darbības attīstība augstskolās” un kas notika laikā no 01.06.06. līdz 31.12.06.
8. Autores izstrādātie modeļi un metodes pasažieru dzelzceļa pārvadājumu no Latvijas reģioniem apjomu prognozēšanai tika izmantoti zinātniskajā projektā U1212 „Matemātisko modeļu, algoritmu un datorprogrammu izstrādāšana Latvijas transporta sistēmas analīzei, attīstības prognozēšanai un optimizācijai”, kas veidoja daļu no zinātniskā projekta IZM-RTU „Zinātniskās darbības attīstība augstskolās” un kas notika laikā no 01.06.08. līdz 31.12.08.
9. Uz iegūto rezultātu pamata ir sagatavota un nolasīta kursa „Mathematical Methods of Traffic Flow Analysis and Forecasting” daļa RTU 2. kursa ārvalstu bakalaura programmas studentiem *Mechanical Engineering*

specialitātē.

10. Būtiskākie pētījuma rezultāti ir publicēti deviņos rakstos un prezentēti vienpadsmit starptautiskās zinātniskās konferencēs Latvijā, Lietuvā, Igaunijā, Vācijā, Grieķijā un Šveicē, kur autore nolasīja vienpadsmit ziņojumus par promocijas darba tēmu.

AUTORES PUBLIKĀCIJAS

1. Andronov A., Zhukovskaya C. and Santalova D. On Mathematical Models for Analysis and Forecasting of the Europe Countries Conveyances. RTU Zinātniskie raksti, Datorzinātne. – 5. sērija, Informācijas tehnoloģija un vadības zinātne, 28. sējums (2007), pp. 96 – 106.
2. Andronov A., Santalova D. On Nonlinear Regression Model for Correspondence Matrix of Transport Network. The XIIIth International Conference on Applied Stochastic Models and Data Analysis. 2009, June 30 – July 3, Lithuania, Vilnius. – Vilnius: Vilnius University, Applied Stochastic Models and Data Analysis International Society, L.Sakalauskas, C.Skiadas and E.K.Zavadskas (Eds.), 2009, pp. 90-94.
3. Jackiva I., Santalova D. Faktoru analīzes izmantošana transporta nozares attīstības tendenču noteikšanai ES valstīs. RTU Zinātniskie raksti, Mašīnzinātne un transports. – 6. sērija, Intelektuālas transporta sistēmas, 18. sējums (2005), 139. – 147. lpp.
4. Kopytov E, Santalova D. Application of the Single Index Model for Forecasting of the Inland Conveyances. Recent Advances in Stochastic Modelling and Data Analysis. Editor: Christos H. Skiadas. – Singapore: World Scientific Publishing Co Pte Ltd., 2007, pp. 268 – 276.
5. Santalova D. Vairumtirdzniecības noliktavas pārdošanu lieluma regresijas modelis. RTU Zinātniskie raksti, Mašīnzinātne un transports. – 6. sērija, Intelektuālas transporta sistēmas, 18. sējums (2005), 67. – 73. lpp.
6. Santalova D. Regression Model of Sales Volume from Wholesale Warehouse. 13th International Conference on Analytical and Stochastic Modelling Techniques and Applications. 2006, May 28 – 31, Germany, Bonn, Sankt Augustin.– Bonn: Bonn-Rhein-Sieg University of Applied Science, Khalid Al-Begain (Ed.), 2006, pp. 133 – 137.
7. Santalova D. Forecasting of Rail Freight Conveyances in EU Countries on the Base of the Single Index Model. Computer Modelling and New Technologies. – Vol. 11, No. 1. (2007), pp. 73 – 83.
8. Santalova D. The Use of Multivariate Regression Model to Forecast the Freight Air Transportations in the Members Countries of the Europe Union. The International Conference „Modelling of Business, Industrial and Transport Systems – 2008”. 2008, May 7 – 10, Latvia, Riga. – Riga: TSI, 2008, pp. 258 – 266.
9. Santalova D. Forecasting of Passenger Conveyances in Latvian Regions

Applying Semiparametric Regression Models. RTU Zinātniskie raksti, Mašīnzinātne un transports. – 6. sērija, Intelektuālas transporta sistēmas, 18. sējums (2008), drukā.

DALĪBA KONFERENCĒS

1. The 46th Scientific Conference of Riga Technical University. Riga, Latvia, October 13 – 15, 2005. Report: „Vairumtirdzniecības noliktavas pārdošanu lieluma regresijas modelis”, author: Santalova D.
2. International Scientific Conference: ECMS 2006, ASMTA 2006. Bonn-Rhein-Sieg University of Applied Science, Bonn, Sankt Augustin, Germany, May 28 – 31, 2006. Report: „Regression Model of Sales Volume from Wholesale Warehouse”, author: Santalova D.
3. The 47th Scientific Conference of Riga Technical University. Riga, Latvia, October 12 – 14, 2006. Report: „Aim and problems of the project “Mathematical models and their estimation methods elaboration for analysis and forecasting of the Baltic region passenger and freight flows”, authors: Santalova D.
4. The 6th International Conference on Reliability and Statistics in Transportation and Communication (RealStat'06). Riga, Latvia, October 25 – 28, 2006. Report: „Forecasting of rail freight conveyances in EU countries on the base of the Single Index Model”, author: Santalova D.
5. The XIIth International Conference on Applied Stochastic Models and Data Analysis (ASMDA-2007). Applied Stochastic Models and Data Analysis International Society, Agronomic Institute of Chania – Greece, Crete, Chania, May 28 – June 2, 2007. Report: „Application of the Single Index Model for Forecasting of the Inland Conveyances”, authors: Kopytov E., Santalova D.
6. The 8th Tartu Conference on Multivariate Statistics and The 6th Conference on Multivariate Distributions with Fixed Marginals. Tartu, Estonia, June 25 – 31, 2007. Report: „Single Index Model for Railway Passenger Conveyances Forecasting in Regions of Latvia”, author: Santalova D.
7. The 48th Scientific Conference of Riga Technical University. Riga, Latvia, October 11 – 12, 2007. Report: „Investigation of classical and semiparametric regression for the forecasting of rail conveyances in Latvia”, author: Santalova D.
8. The International Conference “Modelling of Business, Industrial and Transport Systems” (MBITS'08). Riga, Latvia, May 7 – 10, 2008. Report: „The Use of Multivariate Regression Model to Forecast the Freight Air Transportations in the Members Countries of the European Union”, author: Santalova D.
9. The 2nd International Workshop on Computational and Financial Econometrics (CFE-2008). University of Neuchatel, Neuchatel,

- Switzerland, June 19 – 21, 2008. Report: “On Some Generalization of Seemingly Unrelated Regression Equation Models”, authors: Andronov A., Santalova D., Svirchenkov A.
10. The 49th Scientific Conference of Riga Technical University. Riga, Latvia, October 13 – 15, 2008. Report: „Forecasting of passenger conveyances in Latvian regions applying semiparametric regression models”, author: Santalova D.
 11. The XIIIth International Conference on Applied Stochastic Models and Data Analysis (ASMDA-2009). Applied Stochastic Models and Data Analysis International Society, Vilnius University, Vilnius, Lithuania, June 30 – July 3, 2009. Report: „On Nonlinear Regression Model for Correspondence Matrix of Transport Network”, authors: Andronov A., Santalova D.