

LATVIJAS UNIVERSITĀTE  
EKSAKTO ZINĀTŅU UN TEHNOLOĢIJU FAKULTĀTE  
MATEMĀTIKAS NODAĻA

**EMPĪRISKĀS TICAMĪBAS METODES  
LIETOJUMS INSTRUMENTĀLO MAINĪGO  
REGRESIJĀ SKOLĒNU MATEMĀTIKAS  
ZINĀŠANU KAUZĀLAJĀ ANALĪZĒ**

BAKALaura DARBS

Autors:  
Studenta apliecības nr.:  
Darba vadītājs:

RĪGA 2025

## ANOTĀCIJA

Instrumentālo mainīgo regresija ir metode, kas novērš endogenitāti, tādējādi nodrošina nenovirzītus modeļa novērtētājus. Darbā tika skatīts empīriskās ticamības pielietojums instrumentālo mainīgo regresijā, salīdzinot to ar mazāko kvadrātu metodēm, kā arī, pielietojot Hausmena testu par endogenitāti empīriskās ticamības novērtētājiem. Tika veiktas simulācijas novērtēšanas metožu salīdzinājumam un Hausmena testa jaudas analīzei. Turpinot, teorija tika pielietota skolēnu matemātikas zināšanu kauzālā analīzē, pētot digitālo ierīču lietošanas politikas skolā ietekmi uz disciplināro klimatu matemātikas stundās.

**Atslēgas vārdi:** Instrumentālo mainīgo regresija, empīriskās ticamības metode, starptautiskā skolēnu novērtēšanas programma (PISA), Hausmena tests.

## ABSTRACT

Instrumental variable regression is a method that eliminates endogeneity, thus providing unbiased model estimators. This thesis examined the application of empirical likelihood in instrumental variable regression, comparing it with least squares methods, as well as applying the Hausman test of endogeneity to empirical likelihood estimators. Simulations were performed to compare the estimation methods and assess the power of the Hausman test. The theory was applied to the causal analysis of students' mathematics knowledge, studying the impact of digital device use policies in schools on the disciplinary climate in mathematics lessons, which in turn affects the level of mathematics knowledge.

**Keywords:** Instrumental variable regression, empirical likelihood, Programme for International Student Assessment (PISA), Hausman test.

# SATURS

<b>Apzīmējumu saraksts</b>	<b>5</b>
<b>Ievads</b>	<b>6</b>
<b>1. Regresija kauzālā analīzē un endogenitātes problēma</b>	<b>8</b>
1.1. Endogenitāte . . . . .	9
1.2. Instrumentālo mainīgo regresija . . . . .	10
1.2.1. Vāji instrumenti . . . . .	11
<b>2. Endogenitātes pārbaudīšana</b>	<b>12</b>
2.1. Teorētiskais formulējums parametriskām novērtēšanas metodēm . . . . .	12
2.2. Teorētiskais formulējums neparametriskām novērtēšanas metodēm . . . . .	13
<b>3. Parametru novērtēšanas metodes un to pielietojums Hausmena testā</b>	<b>14</b>
3.1. Mazāko kvadrātu metodes . . . . .	14
3.1.1. Lineārās regresijas mazāko kvadrātu metode . . . . .	14
3.1.2. Instrumentālo mainīgo regresijas divfāžu mazāko kvadrātu metode . . . . .	14
3.1.3. Hausmena testa pielietojums . . . . .	15
3.2. Uz empīrisku ticamību balstītas metodes . . . . .	16
3.2.1. Empīriskās ticamības metode parametriskajā regresijā . . . . .	16
3.2.2. Instrumentālo mainīgo regresija ar empīriskās ticamības metodi . . . . .	18
3.2.3. Hausmena tests empīriskās ticamības novērtētājiem . . . . .	20
<b>4. Simulācijas</b>	<b>23</b>
<b>5. Dati</b>	<b>26</b>
<b>6. Regresijas mainīgo izvēle</b>	<b>30</b>
6.1. Regresijas mainīgo izvēles metodes . . . . .	30
6.2. Prediktoru un to instrumentālo mainīgo izvēle . . . . .	31
6.3. Kontroles mainīgo izvēle . . . . .	32
6.4. Izvēlēto mainīgo apkopojums un datu analīzē skatītie modeļi . . . . .	35
<b>7. Datu analīze</b>	<b>36</b>

<b>8. Rezultāti</b>	<b>40</b>
<b>9. Secinājumi</b>	<b>41</b>
<b>Literatūras saraksts</b>	<b>42</b>
<b>Pielikums</b>	<b>46</b>
<b>A. Simulāciju kods</b>	<b>46</b>
A.1 Endogēnu un eksogēnu datu simulācijas . . . . .	46
A.2 Jaudas analīze . . . . .	51
<b>B. Praktiskās daļas kods</b>	<b>55</b>

## APZĪMĒJUMU SARAKSTS

OECD - Ekonomiskās sadarbības un attīstības organizācija (angl. *Organisation for Economic Co-operation and Development*)

PISA - starptautiskā skolēnu novērtēšanas programma (angl. *Programme for International Student Assessment*)

AIC - Akaika informācijas kritērijs

2SLS - divfāžu mazāko kvadrātu metode (angl. two stage least squares)

EL - empīriskā ticamība (angl. empirical likelihood)

IV - instrumentālie mainīgie (angl. instrumental variables)

## IEVADS

Ik pēc trīs gadiem Ekonomiskās sadarbības un attīstības organizācija OECD veic starptautisku skolēnu novērtēšanas programmu (PISA), kur tiek novērtētas lasīšanas, dabaszinātņu un matemātikas prasmes 15 gadus veciem skolēniem. Katrā programmas ciklā dalībvalstu skaits pieaug un 2025. gada ciklā tiek ziņots par līdz pat 90 valstu dalību.

Koncentrējoties uz skolēnu matemātikas sniegumu, Latvijas rādītāji ir bijuši nestabili un svārstīgi, kur ik ciklu vidējais rādītājs vai nu strauji pieaug, vai krasi sarūk. Šāda tendence aktualizē jautājumu par matemātikas sniegumu ietekmējošiem faktoriem. Viens no tiem ir disciplinārais klimats mācību stundās, īpaši matemātikas priekšmetā. 2022. gada PISA atskaitēs tika novērots, ka lielākoties disciplinārais klimats matemātikas stundās nav atbilstošs kvalitatīva mācību procesa nodrošināšanai [31]. Pastāv iespēja, ka disciplinārais klimats matemātikas stundās ir endogēns mainīgais, jo tas var būt saistīts ar faktoriem, kas netiek skatīti pētījumu ietvaros. Tad nepieciešams izmantot instrumentālo mainīgo regresijas metodi, kas novērš endogenitāti un nodrošina precīzākus kauzālus secinājumus. Tās ietvaros tiek skatīti instrumentālie mainīgie, kas ietekmē atbildes mainīgo tikai caur skaidrojošo mainīgo jeb tiem nav tiešas kauzālas ietekmes uz matemātikas zināšanu līmeni, bet tie ietekmē disciplināro klimatu.

Viens no šādiem instrumentāliem mainīgajiem jeb ietekmējošiem faktoriem ir skolēnu uzmanības novēršana mobilo tālrunu dēļ. 2023. gada globālā izglītības monitoringa atskaitē uzsvērtā ne tikai pozitīvā, bet arī negatīvā tehnoloģiju ietekme uz skolēna kognitīvajām spējām, kur digitālo ierīču klātbūtnē skolēniem tiek novērsta uzmanība, kas turpmāk negatīvi ietekmē to spēju koncentrēties un uztvert mācību vielu [16]. No 2025. gada 31. maija stāsies spēkā Izglītības likuma grozījums, kas liedz skolēniem no 1. līdz 6. klasei izmantot mobilos tālrunus izglītības iestādes telpās [34]. Tad nepieciešams izpētīt vai līdzīgi ierobežojumi ir nepieciešami arī pamatskolas līmeņa skolēniem. Pētāmās problēmas ietvaros tiks apskatīta skolas mobilo tālrunu izmantošanas politikas ietekme uz disciplināro klimatu matemātikas stundās, kas savukārt ietekmē skolēnu matemātikas zināšanu līmeni.

Empīriskās ticamības metode ir neparametriska koeficientu novērtēšanas metode, kas dažādos kontekstos ir parādījusi jaudīgākus rezultātus nekā parametriskās metodes. Tās ietvaros nav nepieciešami pieņēmumi par datu sadalījumu, tādējādi vienkāršojot un vispārinot tās praktisko pielietojumu. Tad darba ietvaros instrumentālo mainīgo regresija tiks realizēta, izmantojot empīriskās ticamības metodi parametru novērtēšanai. Turpinot tiks formulēts Hausmena tests par endogenitāti empīriskās ticamības novērtētājiem un iegūtie rezultāti tiks salīdzināti ar mazāko kvadrātu metodēm.

Tiek izvirzīta hipotēze, ka skolās, kur mobilo tālruņu lietošana ir aizliegta, ievērojama pozitīva ietekme uz disciplināro klimatu, sekojoši arī uz skolēnu matemātikas zināšanām.

Darba mērķis ir pielietot instrumentālo mainīgo regresiju skolēnu matemātikas zināšanu kausālā analīzei, izmantojot empīriskās ticamības metodi parametru novērtēšanai. Tā sasniegšanai tika izvirzīti vairāki uzdevumi:

- Instrumentālo mainīgo regresijas literatūras apskats.
- Empīriskās ticamības metodes pielietošana instrumentālo mainīgo regresijas novērtēšanai.
- Hausmena testa par endogenitāti korekcija instrumentālo mainīgo regresijas empīriskās ticamības metodes novērtētājiem.
- Izvirzītās hipotēzes pārbaude par skolas mobilo tālruņu lietošanas politikas pozitīvo ietekmi uz skolēnu matemātikas zināšanu līmeni.

Darbs sastāv no 8 nodaļām. Darba struktūra ir sekojoša: 1. nodaļā tiek sniegts lineārās regresijas formulējums un pieņēmumi, endogenitātes raksturojums un instrumentālo mainīgo regresijas teorētiskais apskats, 2. nodaļā izskatīta endogenitātes pārbaudīšana ar Hausmena testu, 3. nodaļā tiek dots teorētiskais apraksts regresijas parametru novērtēšanas metodēm un to pielietojums Hausmena testā, 4. nodaļā skatīti simulāciju rezultāti, 5. nodaļā tiek sniegts izmantoto datu raksturojums, 6. nodaļā izskata regresijas mainīgo izvēli izvēlētiem datiem, 7. tiek veikta datu analīze, 8. nodaļā skatīti rezultāti un diskusija un 9. nodaļā veikti secinājumi. Darba noslēgumā ir izmantotās literatūras saraksts, kā arī pielikums, kur redzams programmēšanas valodas R kods, praktiskās daļas izpildei.

# 1. REGRESIJA KAUZĀLĀ ANALĪZĒ UN ENDOGENITĀTES PROBLĒMA

Regresiju analīze ir metode mainīgo attiecību izpratnei, viens no šādu attiecību pielietojumiem ir kauzāla efekta pētījums jeb cēloņa meklēšana starp mainīgajiem [39]. Vienkāršākais regresijas modelis var tikt formulēts sekojoši [39]

$$y_i = \alpha + x_i\beta + \epsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

kur  $\alpha$  ir konstante,  $\beta$  ir prediktora koeficients,  $\epsilon$  ir troksnis jeb kļūda un  $n$  ir izlases apjoms [39]. Ar  $y$  un  $x$  apzīmē attiecīgi rezultatīvo pazīmi un prediktoru jeb skaidrojošo pazīmi [39]. Svarīgi noprast, ka  $x$  un  $y$  ir novēroti lielumi un atrodami datos, taču  $\alpha$  un  $\beta$  ir koeficienti un regresijas analīzes rezultātā šiem koeficientiem tiek aprēķināti novērtētāji [39]. Situācijas, kur pietiek ar vienu skaidrojošo mainīgo ir ļoti specifiskas un retas, tāpēc biežāk izmanto matricu formas pierakstu.  $X$  ir skaidrojošo mainīgo matrica, tās kolonnu skaits atbilst prediktoru skaitam, pievienojot pirmo kolonnu konstantes novērtēšanas nolūkiem [38]. Tad matricu formā modelis izskatās gandrīz identiski,

$$Y = X\beta + \epsilon, \quad (1.2)$$

kur  $X = (x_{ij})_{n \times (p+1)}$ ,  $x_{i0} = 1$ ,  $y = [y_i]_{n \times 1}$  un  $\epsilon = [\epsilon_i]_{n \times 1}$ ,  $n$  ir objektu skaits un  $p$  ir skaidrojošo mainīgo jeb prediktoru skaits [38]. Lineārās regresijas ietvaros ir formulēti septiņi pieņēmumi, pie kuriem modelis tiek uzskatīts par patiesu [38],[42].

- R1 Prediktori un rezultatīvā pazīme tiek novēroti bez mērījuma kļūdas [38],[42].
- R2 Modelis ir lineārs jeb  $Y$  attiecība ar katru  $X_i$  prediktoru ir lineāra [38],[42].
- R3 Kļūdas gadījuma lieluma  $\epsilon$  vidējā vērtība ir 0 [38],[42].
- R4 Pastāv homoskedasticitāte jeb dispersija ir konstanta [38],[42].
- R5 Kļūdas gadījuma lielumi ir nekorelēti [38],[42].
- R6 Prediktori  $X_i$  ir lineāri neatkarīgi savā starpā [38],[42].
- R7 Kļūdas gadījuma lielums  $\epsilon$  ir sadalīts pēc Normālā sadalījuma [38],[42].

Lineārā regresija savu popularitāti guva vieglās interpretācijas un pielietojuma dēļ [38]. Taču kā jau visiem modeļiem, tās lietošanai jābūt teorētiski un praktiski korektai, jo pretējā gadījumā tās rezultāti ir neuzticami, novirzīti vai pat kļūdaini. Nepieciešams skaidri noteikt datu

veidu kuriem tiks veikta analīze, jo eksperimentālu un novērotu datu problemātikas krasi atšķiras. Eksperimentālos datos tiek nodrošināti vienādi apstākļi, tādējādi tiek novērsta neparedzētu situāciju ietekme uz rezultātu, turpretim novērotos datos šāda kontrole ir neiespējama, secinot var rasties endogenitāte, kas ietver plašu klāstu ekonometrijas problēmu [35].

### 1.1. Endogenitāte

Lai veiktu kauzālus secinājumus novērojuma tipa nevis eksperimentāla tipa datiem, ir jāizpildās konkrētiem pieņēmumiem [19]. Viens no šiem pieņēmumiem ir regresoru eksogenitāte, tie nekorelē ar kļūdas gadījuma lielumu [5]. Standarta lineārās regresijas formulējumā,  $y = X\beta + \epsilon$ , eksogenitāte ir gadījumos, kad

$$E(\epsilon|X) = 0 \quad (1.3)$$

vai pie lieliem izlases apjomiem

$$plim \frac{1}{n} X^T \epsilon = 0, \quad (1.4)$$

kur *plim* apzīmē konvergenci pēc varbūtības [17]. Gadījumu, kad korelāciju ar kļūdas gadījuma lielumu nav nulle, sauc par endogenitāti un tās rezultātā modeļa novērtētāji ir novirzīti un iegūtie kauzālie secinājumi var būt nepatiesi [1].

Viens no endogenitātes cēloņiem ir skaidrojošo mainīgo izslēgšana no modeļa, precīzāk, neiekļaujot modelī skaidrojošo mainīgo, kas korelē gan ar atkarīgo mainīgo, gan ar citiem skaidrojošajiem mainīgajiem, tādējādi nekontrolējot kopīgos cēloņus modelī [33]. Izglītības datu ietvaros šādu endogenitāti var radīt, piemēram, sociālekonomiskā stāvokļa mainīgā neiekļaušana modelī, jo šim rādītājam ir korelācija gan ar skolēna zināšanām, gan ar kopējo skolas disciplināro klimatu un attieksmi pret mācībām.

Endogenitāte arī ietver gadījumus, kad skaidrojošo mainīgo ietekmei uz rezultatīvo mainīgo, seko rezultatīvā mainīgā ietekme uz skaidrojošajiem mainīgajiem, šādu situāciju sauc par vienlaicīgu endogenitāti [1]. Par piemēru var uzskatīt skolēna matemātikas zināšanu līmeni un iesaistīšanos mācību stundās, pie augstāka matemātikas zināšanu līmeņa skolēnam ir vairāk interese un vēlmes aktīvi piedalīties matemātikas stundās, pretēji, jo aktīvāk skolēns piedalās mācību stundās, jo augstāks ir tā matemātikas zināšanu līmenis.

Līdzvērtīgi endogenitāti izraisa arī apstākļi, kad vēsturiski atkarīgā mainīgā rezultāti ietekmē šī brīža skaidrojošos mainīgos, tādējādi radot dinamisko endogenitāti, kas atkarīga no laika [1]. Piemēram, matemātikas zināšanu līmenis un vispārīgā attieksme pret mācībām ir dinamiski endogēni mainīgie, jo iepriekšējie matemātikas zināšanu pārbaudes rezultāti ietekmē skolēna attieksmi pret mācībām, jāpiebilst, ka šeit ietekme var būt gan pozitīva, gan negatīva.

Secinot, endogenitāte ir viena no svarīgākajām problēmām regresiju analīzē, kas veikta kauzālu secinājumu rašanai. Neņemot vērā visus rezultātu ietekmējošus mainīgos, kļūda korelē ar skaidrojošajiem mainīgajiem, ieviešot novirzi un veicinot kļūdainus secinājumus[1].

## 1.2. Instrumentālo mainīgo regresija

Instrumentālo mainīgo regresija ir kauzālās analīzes metode, ko piemēro novērotiem datiem [37]. Tā nodrošina endogenitātes novēršanu caur instrumentāliem mainīgiem, kuriem nav tieša efekta uz rezultējošo mainīgo [12]. Tā nosaka  $X$  efektu uz  $Y$ , caur instrumentālo prediktoru  $Z$ , kas ietekmē  $X$ , bet neietekmē  $Y$ , sekojoši tā saistība ar  $Y$  ir tikai caur  $X$  [37]. Lineārā instrumentālo mainīgo regresijā  $X$  ir potenciāli endogēnu prediktoru matrica ar dimensiju  $p \times n$ , kur  $p$  ir prediktoru skaits un  $n$  ir objektu skaits,  $Z$  ir instrumentālo mainīgo matrica, kuras izmēri ir  $k \times n$ , kur  $k$  ir instrumentālo mainīgo skaits,  $W$  ir kontroles mainīgo matrica ar izmēriem  $u \times n$  un  $Y_i$  ir atkarīgā mainīgā vektors garumā  $n$  [4]. Tad var iegūt lineāro instrumentālo mainīgo regresijas modeli

$$Y = X^T \beta + W^T \gamma + \epsilon, \quad (1.5)$$

$$X = \Pi Z + \Gamma W + V, \quad (1.6)$$

kur (1.5) ir strukturālais vienādojums, bet (1.6) ir prediktora  $X$  prognozētās vērtības, izmantojot instrumentālos mainīgos  $Z$  un kontrolējot mainīgos  $W$ ,  $\epsilon$  un  $V$  ir kļūdu vektori [4]. Nepieciešami arī regresijas modeļa parametru skaidrojumi, tad  $\beta = (\beta_1, \dots, \beta_p)^T$  ir endogēno mainīgo koeficientu vektors,  $\gamma = (\gamma_1, \dots, \gamma_u)^T$  ir kontroles mainīgo parametru vektors,  $\Pi$  ir instrumentālo mainīgo parametru matrica ar dimensiju  $p \times k$  un  $\Gamma$  ir pirmās fāzes kontroles mainīgo koeficientu matrica ar izmēriem  $p \times u$  [4].

Kauzālai analīzei šāds modelis ir nepieciešams, lai korekti interpretētu rezultātu, jo kontroles mainīgajiem reti ir kauzāla interpretācija [20] un galvenais jautājums tomēr ir par instrumentālo mainīgo spēju skaidrot skaidrojošo mainīgo. Taču, teorētisko secinājumu veikšanai un novērtēšanas metožu pielietošanai, kontroles mainīgo atdalīšana no instrumentāliem mainīgiem nav pamatota, tādēļ izmanto vienkāršotu modeli. Tā ietvaros pirmās fāzes skaidrojošos mainīgos savieno vienā matricā

$$\begin{cases} Y = X^T \beta + W^T \gamma + \epsilon, \\ X = K^T \Phi + V, \end{cases} \quad (1.7)$$

$$\text{kur } K = \begin{pmatrix} Z \\ W \end{pmatrix}, \text{ bet } \Phi = \begin{pmatrix} \Pi \\ \Gamma \end{pmatrix}.$$

Instrumentālo mainīgo regresijas ietvaros ir jāizpildās trīs pieņēmumi par instrumenta mainīgajiem, korektu un nenovirzītu rezultātu iegūšanai.

IV1 Saistības pieņēmums: Instrumentam  $Z$  ir kauzāls efekts uz  $X$  [26], [18].

IV2 Izslēgšanas pieņēmums: Instrumentam  $Z$  ir efekts uz  $Y$  tikai caur  $X$  [26], [18].

IV3 Neatkarības pieņēmums:  $Z$  un  $Y$  nav kopīgu cēloņu [26], [18].

Ja neizpildās šie pieņēmumi, tad novērtējais kļūs novirzīts un tā kauzālā interpretācija būs nekorekta [18]. Turklāt, saistības pieņēmumam IV1 ir ne tikai jāizpildās, bet starp instrumenta mainīgo un skaidrojošo mainīgo ir jābūt stiprai korelācijai, pretēji, pat, ja instrumentam ir kauzāls efekts uz  $X$ , bet tas ir ļoti vājš, modelis uzrādīs novirzītus un neuzticamus rezultātus [4].

### 1.2.1. Vāji instrumenti

Instrumentu mainīgo sauc par vāju, ja tā korelācija ar endogēno regresoru ir tuvu nullei [4]. Pie šādas situācijas ierastās metodes instrumentālo mainīgo regresijas novērtēšanai kļūst neuzticamas, konkrētāk, to rezultāti kļūst novirzīti un ticamības intervāli var neiekļaut īstā parametra vērtību [4].

Pie vienkārša modeļa, kur strukturālais vienādojums ir  $y = x\beta + \epsilon$  un pirmās fāzes vienādojums ir  $x = Z\pi + v$ . Konkretizējot, ka  $y$ ,  $x$ ,  $u$ ,  $Z$ ,  $v$  ir vektori garumā  $n$ , bet  $\beta$  un  $\pi$  ir regresijas koeficienti. Instrumentālo mainīgo uzskata par vāju, ja  $\pi$  ir tuvu nullei attiecībā pret  $\hat{\pi}$  izlases mainīgumu [4].

Tad testus, kas pārbauda instrumentu vājumu balsta uz pirmās fāzes koeficientu statistisko nozīmību [4]. Viens no tiem ir F-statistikas tests pirmajam instrumentālo mainīgo regresijas solim [4]. Nulles hipotēze šajā gadījumā ir instrumentālā mainīgā neatkarība no endogēnā regresora jeb  $H_0 : \pi = 0$ , to noraidot pie konkrēta nozīmības līmeņa var pieņemt, ka instrumentālie mainīgie nav vāji jeb instrumentālie mainīgie ir statistiski nozīmīgi [4].

Secinot, instrumentālo mainīgo regresija ir jaudīgs rīks kauzālās analīzes ietvaros, taču tās pilnvērtīgai izmantošanai ir jāņem vērā instrumentālo mainīgo pieņēmumi, kā arī instrumentālā mainīgā attiecība ar skaidrojošo mainīgo, jo šādu pieņēmumu neizpildīšanās rada novirzītus rezultātus, kas noved pie neuzticamiem kauzāliem secinājumiem.

## 2. ENDOGENITĀTES PĀRBAUDIŠANA

J. A. Hausmena 1978. gadā piedāvātais specififikācijas tests nodrošina instrumentālo mainīgo regresijas nepieciešamības pārbaudi [17]. Par nulles hipotēzi šādos testos uzskata gadījumu, kad modelī nav specififikācijas kļūdas [17]. Alternatīvā hipotēze ir nosacījuma 2.1 neizpildīšanās

$$plim \frac{1}{n} X^T \epsilon = 0, \quad (2.1)$$

kur  $plim$  apzīmē konverģenci pēc varbūtības,  $n$  ir izlases apjoms,  $X$  ir prediktors un  $\epsilon$  ir modeļa kļūdu vektors [17]. Tests balstās uz ideju, ka eksistē alternatīvs novērtētājs  $\hat{\beta}_1$ , kas ir būtisks, gan pie nulles hipotēzes, gan pie alternatīvās hipotēzes [17].

Tādējādi salīdzinot novērtētājus  $\hat{\beta}_0$  un  $\hat{\beta}_1$ , pie nulles hipotēzes novērtētājs  $\hat{\beta}_0$  ir būtisks un efektīvs, taču pie alternatīvās hipotēzes tas vairs nav būtisks un lietderīgāk izmantot novērtētāju  $\hat{\beta}_1$ , kurš ir būtisks arī pie alternatīvās hipotēzes [17].

Visbiežākais Hausmena testa pielietojums ir endogenitātes pārbaude, kur salīdzina mazāko kvadrātu metodi un divfāžu mazāko kvadrātu metodi, lai noteiktu instrumentālās regresijas nepieciešamību un endogenitātes kontroli [36]. Šādā kontekstā mazāko kvadrātu metodes novērtētājs ir būtisks un efektīvs pie endogenitātes nosacījuma, savukārt divfāžu mazāko kvadrātu metodes novērtētājs ir būtisks neatkarīgi no endogenitātes nosacījuma, taču nav efektīvs [17].

### 2.1. Teorētiskais formulējums parametriskām novērtēšanas metodēm

Pie nulles hipotēzes, kas paredz, ka modelī nav specififikācijas kļūdu, eksistē būtisks, asimptotiski normāli sadalīts un asimptotiski efektīvs novērtētājs, kur efektivitāte ir Krāmera-Rao robežas asimptotiskā sasniegšana [17]. Pie alternatīvās hipotēzes, kad pastāv modeļa specififikācijas kļūdas, šis novērtētājs būs novirzīts un nebūs būtisks [17]. Testa konstruēšanai atrod citu novērtētāju, kuru neietekmē specififikācijas kļūdas, bet zem nulles hipotēzes, tas nav asimptotiski efektīvs [17]. Par efektīvu novērtētāju pie nulles hipotēzes uzskata  $\hat{\beta}_0$ , savukārt  $\hat{\beta}_1$  ir būtisks pie alternatīvās hipotēzes [17].

Šī testa ietvaros svarīgi piebilst, ka novērtētāji  $\hat{\beta}_0$  un  $\hat{\beta}_1$  ir būtiski un ir asimptotiski normāli sadalīti, kur  $\hat{\beta}_0$  asimptotiski sasniedz Krāmera-Rao robežu [17]. Tad J. A. Hausmens piedāvā sekojošu testa statistiku

$$H = (\hat{\beta}_1 - \hat{\beta}_0)^T [Var(\hat{\beta}_1) - Var(\hat{\beta}_0)]^{-1} (\hat{\beta}_1 - \hat{\beta}_0), \quad (2.2)$$

un pierāda, ka pie nulles hipotēzes, šī statistika ir sadalīta pēc  $\chi_k^2$  sadalījuma, kur  $k$  ir parametru skaits [17].

Secinot, Hausmena testa pielietošanai nepieciešams atrast tādus divus novērtētājus, kas ir asimptotiski normāli sadalīti, būtiski pie nulles hipotēzes, viens no tiem zem nulles hipotēzes asimptotiski sasniedz Krāmiera-Rao robežu, bet otrs ir būtisks arī pie alternatīvās hipotēzes [17].

## 2.2. Teorētiskais formulējums neparametriskām novērtēšanas metodēm

Vispārīgāks Hausmena testa formulējums tiek balstīts uz paplašināto modeli un sekojoši dēvēts par paplašināto Hausmena testu. Tā ietvaros tiek pielietoti attiecīgie statistiskie testi parametra statistiskās nozīmības pārbaudei, tādējādi paplašinot iespējamo parametru novērtēšanas metožu klāstu. Parametrisku novērtēšanas metožu gadījumā, pielietojot F-testu, bet neparametrisku novērtēšanas metožu gadījumā to statistiskās nozīmības testu analogus.

Testa nostādne un novērtētāji nemainās, joprojām tiek uzskatīts, ka pie nulles hipotēzes novērtētājs  $\hat{\beta}_0$  ir efektīvs un būtisks, bet pie alternatīvās hipotēzes būtisks ir novērtētājs  $\hat{\beta}_1$  [17].

Uzskata, ka  $\hat{X} = K^T \hat{\Phi}$ , kur  $K$  ir instrumentālo un kontroles mainīgo matrica. Tad vienādojumā 1.5 ievieto vienādojumu  $X = \hat{X} + V$ . Parametru pie kļūdu vektora  $V$  apzīmē ar  $\tau$  un iegūst [21]

$$Y = (\hat{X} + V)^T \beta + W^T \gamma + \epsilon = \hat{X}^T \beta + V^T \beta + W^T \gamma + \epsilon = \hat{X}^T \beta + V^T \tau + W^T \gamma + \epsilon. \quad (2.3)$$

Turpinot vienādojuma pārveidojumus tam pieskaita un atņem izteiksmi  $V^T \beta$

$$Y = \hat{X}^T \beta + V^T \tau + W^T \gamma + \epsilon + V \beta - V \beta = (\hat{X} + V)^T \beta + V^T (\tau - \beta) + W^T \gamma + \epsilon, \quad (2.4)$$

tad apzīmējot izteiksmi  $\theta := \tau - \beta$ , iegūst paplašināto modeli, kas iekļauj pirmās fāzes kļūdas vektoru [21]

$$Y = X^T \beta + V^T \theta + W^T \gamma + \epsilon. \quad (2.5)$$

Endogenitātes pārbaudīšanas tests pārtop par  $\theta$  parametra statistiskās nozīmības testu. Kur ar atbilstošu statistisko testu, tiek pārbaudīta nulles hipotēze  $H_0 : \theta = 0$  [21]. Ja nulles hipotēzi noraidīt nevar, tad pirmās fāzes kļūdas nav statistiski nozīmīgas un instrumentālo mainīgo regresija nav nepieciešama. Ja nulles hipotēzi noraida, tad pirmās fāzes kļūda ir statistiski nozīmīga pie fiksēta nozīmības līmeņa un var apgalvot, ka  $X$  ir endogēns un instrumentālo mainīgo regresija ir nepieciešama nenovirzītu rezultātu iegūšanai.

Šī darba ietvaros Hausmena tests tiks pielietots lineārās regresijas un instrumentālo mainīgo empīriskās ticamības novērtētājiem, lai noteiktu mainīgo endogenitāti vai eksogenitāti. Tā kā empīriskās ticamības metode ir neparametriska, tad tiks izmantots otrais Hausmena testa formulēšanas veids, kas balstās uz pirmās fāzes atlikumu parametra statistisko nozīmību. Šie rezultāti tiks salīdzināti ar mazāko kvadrātu metodes rezultātiem.

### 3. PARAMETRU NOVĒRTĒŠANAS METODES UN TO PIELIETOJUMS HAUSMENA TESTĀ

Šajā nodaļā tiks izklāstītas četras parametru novērtēšanas metodes: mazāko kvadrātu metode, divfāžu mazāko kvadrātu metode, empīriskās ticamības metode un empīriskās ticamības metode instrumentālo mainīgo regresijas gadījumā. Sekojoši tiek apskatīts arī Hausmena testa pielietojums savstarpēji salīdzinot mazāko kvadrātu metodes un empīriskās ticamības metodes.

#### 3.1. Mazāko kvadrātu metodes

##### 3.1.1. Lineārās regresijas mazāko kvadrātu metode

Mazāko kvadrātu metode ir visvienkāršākā un visbiežāk izmantotā metode parametra  $\beta$  novērtēšanai. Pie modeļa  $y = X\beta + \epsilon$ , tiek uzskatīts, ka matricai  $X$  ir pilns rangs un tās kolonnas ir neatkarīgas, tad mazāko kvadrātu metodes mērķa funkcija sastāv no atlikumu kvadrātu summas un sekojoši to nepieciešams minimizēt [41]. Matricu formā to definē sekojoši

$$S(\hat{\beta}) = (y - X\hat{\beta})^T(y - X\hat{\beta}), \quad (3.1)$$

atvasinot šo funkciju pēc  $\hat{\beta}$  un pielīdzinot to nullei, iegūst mazāko kvadrātu metodes novērtētāju [10]

$$\hat{\beta} = (X^T X)^{-1} X^T y. \quad (3.2)$$

Izpildoties iepriekš minētiem regresijas nosacījumiem R1, ..., R6, spēkā ir Gausa-Markova teorēma, kas nodrošina šī novērtētāja būtiskumu [10].

##### 3.1.2. Instrumentālo mainīgo regresijas divfāžu mazāko kvadrātu metode

Instrumentālo mainīgo regresijas gadījumā ir nepieciešams novērtēt gan  $X$ , gan  $\beta$ , tāpēc arī mazāko kvadrātu metode tiek pielietota divreiz, sākotnēji novērtējot vienādojuma 1.6 parametru un tad novērtējot  $\beta$  [22]. Taču svarīgi piebilst, ka teorētiskā pielietojumā nav nepieciešams atdalīt kontroles mainīgos no skaidrojošajiem endogēniem mainīgajiem, jo ir vieglāk darboties matricu formā, tad par saistošo vienādojumu uzskata vienādojumu  $Y = X^T\beta + \epsilon$ , kur  $X$  satur gan endogēnos mainīgos, gan kontroles mainīgos un pirmā fāze ir  $X = K^T\Phi + V$ .

Pirmās fāzes mērķis ir noteikt parametra  $\Phi$  novērtētāju no vienādojuma  $X = K^T\Phi + V$ , pie pieņēmuma, ka kļūdas gadījuma lielums ir sadalīts pēc normālā sadalījuma,  $V \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_V^2 I)$ , tiek iegūts  $\Phi$  novērtētājs [22]:

$$\hat{\Phi} = (K K^T)^{-1} K X^T. \quad (3.3)$$

Tad izmantojot šo novērtētāju, iespējams novērtēt skaidrojošo mainīgo matricu  $\hat{X}$

$$\hat{X} = K^T \hat{\Phi} = K^T (K K^T)^{-1} K X^T = P_K X^T, \quad (3.4)$$

kur  $P_K = K^T (K K^T)^{-1} K$  [22].

Otrajā fāzē meklē vienādojuma  $Y = \hat{X}^T \beta + \epsilon$  parametra  $\beta$  novērtētāju. Tad pieņemot, ka  $X^T P_K X = \hat{X}^T \hat{X}$  ir invertējama matrica, divfāžu mazāko kvadrātu metodes novērtētājs ir [22]

$$\hat{\beta}_{2SLS} = (\hat{X}^T \hat{X})^{-1} \hat{X}^T Y. \quad (3.5)$$

### 3.1.3. Hausmena testa pielietojums

Hausmena testa pielietošanai mazāko kvadrātu metožu novērtētājiem  $\hat{\beta}_{LS}$  un  $\hat{\beta}_{2SLS}$  nepieciešams parādīt, ka novērtētāji ir asimptotiski normāli sadalīti. Pie nulles hipotēzes par eksogenitāti novērtētājs  $\hat{\beta}_{LS}$  ir būtisks un efektīvs jeb tas sasniedz Krāmēra-Rao robežu, bet novērtētājs  $\hat{\beta}_{2SLS}$  ir būtisks, bet nav efektīvs. Savukārt pie alternatīvās hipotēzes novērtētājs  $\hat{\beta}_{2SLS}$  joprojām ir būtisks jeb tā būtiskumu neietekmē endogenitātes klātbūtne, bet novērtētājs  $\hat{\beta}_{LS}$  kļūst novirzīts.

Pieņemot, ka  $\epsilon$  un  $V$  kovariācija ir nulle, mazāko kvadrātu metodes novērtētājs vienādojumam 1.5 ir asimptotiski efektīvs  $\beta_0$  novērtētājs, bet, kad kovariācija nav nulle, tad tas vairs nav būtisks [17]. Tad divfāžu mazāko kvadrātu metodes mainīgais ir būtisks [17]. Tā kā mazāko kvadrātu metodes ir parametriskas novērtēšanas metodes, tad arī variācijas aprēķins ir iespējams un vienkāršs. Turpinot Hausmena statistika var tikt aprēķināta sekojoši

$$H = (\hat{\beta}_{2SLS} - \hat{\beta}_{LS})^T (V(\hat{\beta}_{2SLS}) - V(\hat{\beta}_{LS}))^{-1} (\hat{\beta}_{2SLS} - \hat{\beta}_{LS}), \quad (3.6)$$

kur pie nulles hipotēzes tā ir sadalīta pēc  $\chi^2$  sadalījuma.

Šeit iespējams skatīt arī otro Hausmena testa formulēšanas veidu, kur no pirmās fāzes iegūst atlikumus un ievieto tos strukturālajā vienādojumā pārbauda to statistisko atšķirību no nulles, izmantojot F-testu.

## 3.2. Uz empīrisko ticamību balstītas metodes

### 3.2.1. Empīriskās ticamības metode parametriskajā regresijā

Vispārīgi par parametrisku regresijas modeli uzskata sekojošu reprezentāciju:

$$Y_i = m(X_i; \beta) + \epsilon_i \quad i = 1, \dots, n, \quad (3.7)$$

kur  $m(x; \beta)$  ir regresijas funkcija ar nezināmu parametru  $\beta \in R^p$  un kļūdām  $\epsilon_i$  [13]. Empīriskās ticamības formulējumam nepieciešams nosacījums, ka  $\epsilon_i$  piemīt galīga nosacītā dispersija [13]. Pieņemot, ka  $p_1, \dots, p_n$  ir varbūtību svāri, kas piemēroti datiem, empīriskās ticamības funkciju  $L_n(\beta)$  definē sekojoši [13]:

$$L_n(\beta) = \max \prod_{i=1}^n p_i, \quad (3.8)$$

ku maksimizē ar ierobežojumiem

$$\sum_{i=1}^n p_i = 1, \quad (3.9)$$

$$\sum_{i=1}^n p_i \frac{\partial m(X_i; \beta)}{\partial \beta} \{Y_i - m(X_i; \beta)\} = 0. \quad (3.10)$$

Tātad empīriskās ticamības novērtētājs ir argumenta vērtība, pie kuras  $L(\beta)$  tiek maksimizēta, to apzīmē ar  $\hat{\beta}_{EL}$  [13]. Turpinot ievieš apzīmējumu  $g(X, \beta) = \frac{\partial m(X; \beta)}{\partial \beta} \{Y_i - m(X; \beta)\}$ , kur tiek pieņemts, ka  $E[g(X, \beta)] = 0$  [13]. Empīriskās ticamības log-ticamības attiecība tiek definēta sekojoši [13]:

$$\begin{aligned} r_n(\beta) &= -2 \log \left\{ \frac{L_n(\beta)}{L_n(\hat{\beta})} \right\} = 2 \sum_{i=1}^n \log \{1 + \lambda^T g(X, \beta)\} \\ &= 2 \sum_{i=1}^n \log \left\{ 1 + \lambda^T \frac{\partial m(X_i; \beta)}{\partial \beta} (Y_i - m(X_i; \beta)) \right\}. \end{aligned} \quad (3.11)$$

Viena no svarīgākajām parametriskās ticamības īpašībām ir Vilka teorēma, kura izpildās arī empīriskās ticamības metodei

$$r_n(\beta_0) \xrightarrow[n \rightarrow \infty]{d} \chi_p^2, \quad (3.12)$$

kur  $\beta_0$  ir īstais  $\beta$  parametrs [13].

Šī darba ietvaros tiek apskatīts lineārs modelis jeb  $m(X; \beta) = X\beta$  un  $g(X, \beta) = X(Y - X\beta)$ . Tad empīriskās ticamības funkciju pieraksta

$$L(\beta) = \max \left\{ \prod_{i=1}^n p_i \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X(Y - X\beta) = 0 \right\}. \quad (3.13)$$

Empīriskās log-ticamības attiecība sekojoši ir

$$r_n(\beta) = 2 \sum_{i=1}^n \log \{1 + \lambda^T X(Y - X\beta)\}. \quad (3.14)$$

Pie konkrētiem regularitātes nosacījumiem var pierādīt, ka empīriskās ticamības novērtētājs ir asimptotiski normāli sadalīts. Šos nosacījumus uzskaita J. Kins un J. Laules [32]:

EL1  $E[g(x, \beta_0)g(x, \beta_0)^T]$  ir pozitīvi definīta matrica;

EL2  $\frac{\partial g(x, \beta)}{\partial \beta}$  ir nepārtraukta īstās vērtības  $\beta_0$  apkārtnē;

EL3  $\|\frac{\partial g(x, \beta)}{\partial \beta}\|$  un  $\|g(x, \beta)\|^3$  ierobežo integrējama funkcija  $G(x)$   $\beta_0$  apkārtnē;

EL4  $E[\frac{\partial g(x, \beta_0)}{\partial \beta}] = p$ ;

EL5  $\frac{\partial^2 g(x, \beta)}{\partial \beta \partial \beta^T}$  ir nepārtraukta pēc  $\beta$  īstās vērtības  $\beta_0$  apkārtnē;

EL6  $\|\frac{\partial^2 g(x, \beta)}{\partial \beta \partial \beta^T}\|$  ierobežo integrējama funkcija  $G(x)$  šajā apkārtnē.

**3.1. Teorēma.** [32] *Ja izpildās nosacījumi EL1 – EL6, tad empīriskās ticamības novērtētājs ir asimptotiski normāli sadalīts jeb*

$$\sqrt{n}(\hat{\beta}_{EL} - \beta_0) \xrightarrow[n \rightarrow \infty]{d} N(0, V), \quad (3.15)$$

$$V = \left[ \left\{ \sum_{i=1}^n \frac{1}{n} \frac{\partial g(x, \hat{\beta}_{EL})}{\partial \beta} \right\}^T \left\{ \sum_{i=1}^n \frac{1}{n} g(x, \hat{\beta}_{EL}) g^T(x, \hat{\beta}_{EL}) \right\}^{-1} \left\{ \sum_{i=1}^n \frac{1}{n} \frac{\partial g(x, \hat{\beta}_{EL})}{\partial \beta} \right\} \right]. \quad (3.16)$$

Pie papildus nosacījumiem EL7-EL11 var pierādīt, ka empīriskās ticamības novērtētājs ir būtisks [24].

EL7  $E[g(x, \beta_0)] = 0$  eksistē viens vienīgs atrisinājums  $\beta_0 \in B$  [24].

EL8  $E\|g(x, \beta)\|^3 < \infty$  jebkurai  $\beta \in B$  [24].

EL9 Jebkurai kompaktai kopai  $C \subseteq B$ , tādai, ka  $g(X, \beta)$  ir nepārtraukta Lipšitza nozīmē kopā  $C$ : eksistē funkcija  $G$ , tāda, ka  $\|g(x, \beta') - g(x, \beta)\| \leq G(x)\|\beta' - \beta\|$ , kur  $E(G^2(x)) < \infty$  visiem  $\beta, \beta' \in C$  [24].

EL10 Parametru kopa  $B$  ir slēgta  $R^q$  apakškopa [24].

EL11 Eksistē nepārtraukts vektors  $b(\beta)$ , tāds, ka  $h(X, \beta) = b^T(\beta)g(X, \beta)$  izpildās

(a)  $E\{\limsup \|h(x, \beta)^2\|\} < \infty$ ,

(b)  $\liminf \|E\{h(x, \beta)\}\| > 0$ ,

(c)  $E[\lim \Delta_h(X, r)] = 0$ , kur  $\Delta_h(X, r) = \sup \|h(x, \beta) - h(x, \beta')\|$  [24].

Rezultātu par novērtētāja būtiskumu formulē teorēmā 3.2 atbilstoši rakstam [24].

**3.2. Teorēma.** [24] Ja nosacījumi  $EL1$ ,  $EL7$ - $EL11$  izpildās, tad maksimālais empīriskās ticamības novērtētājs ir būtisks jeb

$$\hat{\beta}_{EL} \xrightarrow[n \rightarrow \infty]{} \beta_0 \quad (3.17)$$

Tāad empīriskās ticamības novērtētājs ir asimptotiski normāli sadalīti un būtisks pie minētiem nosacījumiem. Turklāt, tas ir efektīvs pie nulles hipotēzes, kur efektivitāti nosaka pēc Qin un Laules raksta [32].

### 3.2.2. Instrumentālo mainīgo regresija ar empīriskās ticamības metodi

Empīriskās ticamības metodes pielietojumu instrumentālo mainīgo regresijā pētīja Peixin Zhao un Liugen Xue [47], Liugen Xue un Lixing Zhu [45] un citi. Tās realizēšanai izmanto modeli 1.7 un papildus ir spēkā ir sekojošas izteiksmes par regresoru endogenitāti vai eksogenitāti [47].

$$E[\epsilon|X] \neq 0, \quad E[\epsilon|W] = 0, \quad E[\epsilon|Z] = 0. \quad (3.18)$$

Pēc Zhou aprakstītās metodikas nepieciešams ieviest palīgmainīgo  $\eta(\beta)$ , kas nodrošinās maksimizācijas funkcijas ierobežojumu, šeit  $\eta(\beta)$  ir analogisks nodaļā 3.2.1 apskatītai funkcijai  $g(X, \beta) = \frac{\partial m(X_i; \beta)}{\partial \beta} \{Y_i - m(X_i; \beta)\}$  [47]. Lai izpildītos  $E[\eta_i(\beta)] = 0$  pie endogenitātes nosacījumiem nepieciešams ieviest  $\eta(\beta)$  korekciju, kas balstās uz instrumentāliem mainīgajiem [47]. Tad izmantojot šo korekciju iespējams veidot log-ticamības attiecību un noteikt empīriskās ticamības instrumentālo mainīgo novērtētāju [47].

No nodaļas 3.2.1 var secināt, ka nepieciešams palīgmainīgais, kas iespējos endogēnu mainīgo iekļaušanu modelī. Šeit nereti izmanto momentu metodi, kur

$$E[X\epsilon] = E[X(Y - X^T\beta - W^T\gamma)] = 0. \quad (3.19)$$

Tad par palīgmainīgo uzskata funkciju  $\eta(\beta)$ , kas definēta [47]

$$\eta(\beta) = X(Y - X^T\beta - W\gamma). \quad (3.20)$$

Gadījumā, kad  $X$  ir eksogēns  $E[\eta(\beta)] = 0$  pie īstā  $\beta$  parametra [47]. Tad var konstruēt empīriskās ticamības attiecības funkciju parametram  $\beta$ , izmantojot  $\eta(\beta)$  [47]. Taču gadījumā, kad  $X_i$  ir endogēns, šis nosacījums neizpildās un nepieciešama  $\eta(\beta)$  korekcija, kas balstās uz instrumentālajiem mainīgajiem  $Z_i$  [47]. Tāad nepieciešams novērtēt koeficientu  $\Phi$  pēc momentu metodes, ņemot vērā, ka  $E[XZ^T] = \Phi E[ZZ^T]$  [47]:

$$\hat{\Phi} = \left( \sum_{i=1}^n X_i K_i^T \right) \left( \sum_{i=1}^n K_i K_i^T \right)^{-1}. \quad (3.21)$$

Pieņemot, ka  $\hat{X}_i = \hat{\Phi} K_i$  [47]. Var definēt  $\eta(\beta)$  korekciju [47]

$$\hat{\eta}(\beta) = \hat{X}(Y - \hat{X}^T \beta - W^T \gamma). \quad (3.22)$$

Nepieciešams pierādīt, ka šāda veida korekcijas matemātiskā cerība nav atkarīga no  $X$  endogenitātes jeb  $E[\hat{\eta}_i(\beta)] = 0$  [47].

$$E[\hat{\eta}_i(\beta)] = E[\hat{X}(Y - \hat{X}^T \beta - W^T \gamma)] = E[\hat{X} \beta K^T (\Phi^T - \hat{\Phi}^T) + \hat{\Phi} K \epsilon], \quad (3.23)$$

kur  $E[\hat{X} \beta K^T (\Phi^T - \hat{\Phi}^T)] = 0$ , jo  $\hat{\Phi}$  novērtētājs ir nenovirzīts, un  $E[\hat{\Phi} K \epsilon] = 0$  instrumentālo mainīgo un kontroles mainīgo nekorelētības ar atlikumiem dēļ [47]. Sekojoši, uz instrumentāliem mainīgiem balstītā empīriskās log-ticamības attiecība ir [47]

$$\hat{R}(\beta) = -2 \max \left\{ \sum_{i=1}^n \log(np_i) | p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\eta}_i(\beta) = 0 \right\}. \quad (3.24)$$

Pēc Lagranža reizinātāju metodes  $\hat{R}(\beta)$  var pierakstīt arī sekojoši [47]:

$$\hat{R}(\beta) = 2 \sum_{i=1}^n \log \{ 1 + \lambda^T \hat{\eta}_i(\beta) \}, \quad (3.25)$$

kur  $\lambda$  nosaka pēc vienādojuma [47]

$$\sum_{i=1}^n \frac{\hat{\eta}_i(\beta)}{1 + \lambda^T \hat{\eta}_i(\beta)} = 0. \quad (3.26)$$

Lai parādītu, ka  $\hat{R}(\beta)$  ir asimptotiski sadalīts pēc Hī-kvadrāta sadalījuma, kad  $\beta$  ir īstais parametrs, nepieciešams formulēt regularitātes pieņēmumus [47].

C1  $\exists \delta > 0$ , tāda, ka  $E[\epsilon^{2+\delta}] < \infty$  un  $E[\|\epsilon\|^{2+\delta}] < \infty$ . Turklāt, tiek pieņemts, ka  $E(\|X\|^4) < \infty$  un  $E(\|Z\|^4) < \infty$ , kur  $\|\cdot\|$  ir Eiklīda norma [47].

C2 Gadījuma lielumam  $W$  ir ierobežota nesējkopā un tā blīvuma funkcija ir ierobežota starp 0 un bezgalību un ir nepārtraukti diferencējama ierobežotā nesējkopā [47].

C3  $E(Z_r | W = w)$ ,  $r = 1, \dots, p$ , un  $W^T \gamma$  ir divas reizes nepārtraukti diferencējama kompaktā nesējkopā [47].

C4 Matrica  $\Sigma$  ir pozitīvi definīta un matrica  $E(ZZ^T)$  ir invertējama [47].

**3.3. Teorēma.** [47] Pieņemot, ka nosacījumi C1-C4 izpildās un  $\beta$  ir īstā parametra vērtība, tad

$$\hat{R}(\beta) \xrightarrow[n \rightarrow \infty]{d} \chi_p^2. \quad (3.27)$$

Maksimizējot  $\{-R(\beta)\}$  var iegūt  $\beta$  maksimālās empīriskās ticamības novērtētāju  $\hat{\beta}_{IVEL}$  [47]. Turpinot, var parādīt, ka  $\hat{\beta}_{IVEL}$  ir vienādojuma

$$\hat{\eta}(\beta) = 0 \quad (3.28)$$

atrisinājums [47]. Tad vienkāršu analītisku darbību rezultātā var pierādīt, ka [47]

$$\hat{\beta}_{IVEL} = (\hat{X}\hat{X}^T)^{-1}\hat{X}(Y - W^T\gamma). \quad (3.29)$$

Tad spēkā ir arī sekojoša teorēma par maksimālās empīriskās ticamības novērtētāja  $\hat{\beta}_{IVEL}$  asimptotisko normalitāti pie regularitātes nosacījumiem.

**3.4. Teorēma.** [47] Pieņemot, ka nosacījumi C1-C4 izpildās,

$$\sqrt{n}(\hat{\beta}_{IVEL} - \beta_0) \xrightarrow[n \rightarrow \infty]{d} N(0, \Sigma), \quad (3.30)$$

kur  $\Sigma = (\sigma_e^2 + \beta_0^T \Sigma_e \beta_0)(\Pi E[ZZ^T]\Pi^T)^{-1}$ .

### 3.2.3. Hausmena tests empīriskās ticamības novērtētājiem

Hausmena testa pielietošanai empīriskās ticamības novērtētājiem  $\hat{\beta}_{EL}$  un  $\hat{\beta}_{IVEL}$  nepieciešams pierādīt, ka novērtētāji ir asimptotiski normāli sadalīti. Pie nulles hipotēzes par eksogenitāti novērtētājs  $\hat{\beta}_{EL}$  ir būtisks un efektīvs, bet novērtētājs  $\hat{\beta}_{IVEL}$  ir būtisks, bet nav efektīvs [17]. Savukārt pie alternatīvās hipotēzes novērtētājs  $\hat{\beta}_{IVEL}$  joprojām ir būtisks jeb tā būtiskumu neietekmē endogenitātes klātbūtne, bet novērtētājs  $\hat{\beta}_{EL}$  kļūst novirzīts [17].

Pie nulles hipotēzes un regularitātes nosacījumiem EL1-EL6 var apgalvot, ka empīriskās ticamības novērtētājs  $\hat{\beta}_{EL}$  ir asimptotiski normāli sadalīts pēc teorēmas 3.1 [32]. Pie nulles hipotēzes un regularitātes nosacījumiem C1-C4 arī instrumentālo mainīgo empīriskās ticamības novērtētājs ir asimptotiski normāli sadalīts pēc teorēmas 3.4 [47].

Empīriskās ticamības novērtētāja būtiskums pie nulles hipotēzes ir formulēts teorēmā 3.2, alternatīvās hipotēzes gadījumā neizpildās nosacījums EL7 un empīriskās ticamības novērtētāju nevar uzskatīt par būtisku [24]. Tad nepieciešams pierādīt, ka instrumentālo mainīgo empīriskās ticamības novērtētājs

$$\hat{\beta}_{IVEL} = (\hat{X}\hat{X}^T)^{-1}\hat{X}(Y - W^T\gamma) \quad (3.31)$$

ir būtisks gan pie  $H_0$ , gan pie  $H_1$  jeb tas nav atkarīgs no  $X$  endogenitātes.

Novērtētāju sauc par būtisku, ja  $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|\hat{\beta}_n - \beta_0| > \varepsilon) = 0$  [14]. Tad, izmantojot Markova nevienādību

$$0 < P(|\hat{\beta}_n - \beta_0| > \varepsilon) \leq \frac{E[|\hat{\beta}_n - \beta_0|]}{|\varepsilon|}, \quad (3.32)$$

nepieciešams pierādīt, ka  $\lim_{n \rightarrow \infty} E[|\hat{\beta}_n - \beta|] = 0$  [14]. Robežas aprēķināšanai nepieciešams pielietot Lebesga dominējošās konverģences teorēmu.

**3.5. Teorēma.** [9] Ja  $\{f_n\}$  ir funkciju virkne uz kopas  $S$  un  $|f_n| < g$ , kur  $g$  ir integrējama funkcija uz kopas  $S$ , tad, ja  $f_n \rightarrow f$ , tad  $\int_S f_n \rightarrow \int_S f$ .

Skata Markova nevienādības skaitītāju vienādojumā 3.33. Sākotnēji izteiksmē ievieto vienādojumu  $Y = X^T \beta + W^T \gamma + \epsilon$ , to vienkāršo noīsinot kontroles mainīgos. Tad izteiksmi pārgrupē divos saskaitāmajos, pielietojot trijstūra nevienādību.

$$\begin{aligned}
& E|(\hat{X} \hat{X}^T)^{-1} \hat{X} (Y - W^T \gamma) - \beta| = \\
& = E|(\hat{X} \hat{X}^T)^{-1} \hat{X} (X^T \beta + W^T \gamma + \epsilon - W^T \gamma) - \beta| = \\
& = E|(\hat{X} \hat{X}^T)^{-1} \hat{X} (X^T \beta + \epsilon) - \beta| = \tag{3.33} \\
& = E|(\hat{X} \hat{X}^T)^{-1} \hat{X} X^T \beta + (\hat{X} \hat{X}^T)^{-1} \hat{X} \epsilon - \beta| \leq \\
& \leq E|((\hat{X} \hat{X}^T)^{-1} \hat{X} X^T - I) \beta| + E|(\hat{X} \hat{X}^T)^{-1} \hat{X} \epsilon|.
\end{aligned}$$

Pierādot, ka pēdējā izteiksme tiecas uz 0, kad  $n \rightarrow \infty$ , tiks pierādīts novērtētāja būtiskums. Tad, nepieciešams pierādīt, ka izteiksmes 3.34 un 3.35 ir patiesas.

$$\lim_{n \rightarrow \infty} E|((\hat{X} \hat{X}^T)^{-1} \hat{X} X^T - I) \beta| = 0 \tag{3.34}$$

$$\lim_{n \rightarrow \infty} E|(\hat{X} \hat{X}^T)^{-1} \hat{X} \epsilon| = 0 \tag{3.35}$$

Sāk ar izteiksmes 3.34 pierādījumu. Zināms, ka  $\hat{X} = \hat{\Phi} K$  un novērtētājs  $\hat{\Phi}$  ir būtisks, tātad  $\hat{\Phi} \rightarrow \Phi$ , kad  $n \rightarrow \infty$ . Tiek pieņemts, ka eksistē tāda integrējama funkcija  $H(X)$ , kas ierobežo funkciju  $|((\hat{X} \hat{X}^T)^{-1} \hat{X} X^T - I) \beta|$ . Sekojoši, izteiksmē 3.34 ievieto  $\hat{X}$ , tad mainot robežas un matemātiskās cerības operatorus iegūst robežas vērtību.

$$\begin{aligned}
& \lim_{n \rightarrow \infty} E|((\hat{X} \hat{X}^T)^{-1} \hat{X} X^T - I) \beta| = \\
& = \lim_{n \rightarrow \infty} E|((\hat{\Phi} K K^T \hat{\Phi}^T)^{-1} \hat{\Phi} K X^T - I) \beta| = \tag{3.36} \\
& = E|((\Phi K K^T \Phi^T)^{-1} \Phi K X^T - I) \beta|
\end{aligned}$$

Turpina pārveidojot robežas vērtību, lai novērtētu tās vienādību ar nulli. Ievieto vienādojumu  $X = \Phi K + V$ , atver iekavas un pielieto matricu īpašību. Pilna ranga matricas  $A$  reizinājums ar tās inverso matricu ir vienības matrica  $I$  jeb  $A^{-1} A = I$  [2].

$$\begin{aligned}
& E|(\Phi K K^T \Phi^T)^{-1} \Phi K (K^T \Phi^T + V^T) - I) \beta| = \\
& = E|(\Phi K K^T \Phi^T)^{-1} (\Phi K K^T \Phi^T + \Phi K V^T) - I) \beta| = \\
& = E|((\Phi K K^T \Phi^T)^{-1} \Phi K K^T \Phi^T + (\Phi K K^T \Phi^T)^{-1} \Phi K V^T - I) \beta| = \tag{3.37} \\
& = E|(I + (\Phi K K^T \Phi^T)^{-1} \Phi K V^T - I) \beta| \\
& = E|(\Phi K K^T \Phi^T)^{-1} \Phi K V^T \beta|.
\end{aligned}$$

Tā kā  $\Phi$  ir īsto parametru matrica, tad tie ir konstanti, tāpat mainīgie  $K$  ir fiksēti, tad  $|(\Phi K K^T \Phi^T)^{-1}|$  var iznest pirms matemātiskās cerības zīmes, jo šī izteiksme ir konstanta, sekojoši

$$E|(\Phi K K^T \Phi^T)^{-1} \Phi K V^T| \beta = |(\Phi K K^T \Phi^T)^{-1}| \cdot E|\Phi K V^T| \cdot |\beta| = 0, \quad (3.38)$$

jo pēc instrumentālo mainīgo regresijas pieņēmumiem matricas  $K$  elementi ir eksogēni jeb  $plim_n \frac{1}{n} K V = 0$ , kā arī eksistē integrējama funkcija  $H(X)$ , kas nodrošina absolūtās vērtības dominējošo konvergenci.

Turpina pierādījumu uzrādot, ka izteiksme 3.35 arī ir patiesa. Sākotnēji izteiksmē 3.35 ievieto  $\hat{X}$ , tad pieņemot, ka eksistē tāda integrējama funkcija  $H_1(X)$ , kas ierobežo izteiksmi  $|(\hat{X} \hat{X}^T)^{-1} \hat{X} \epsilon|$  aprēķina robežas vērtību, ņemot vērā, ka  $\hat{\Phi} \rightarrow \Phi$ , kad  $n \rightarrow \infty$ .

$$\lim_{n \rightarrow \infty} E|(\hat{X} \hat{X}^T)^{-1} \hat{X} \epsilon| = \lim_{n \rightarrow \infty} E|(\hat{\Phi} K K^T \hat{\Phi}^T)^{-1} (\hat{\Phi} K) \epsilon| = |(\Phi K K^T \Phi^T)^{-1}| \cdot E|\Phi K \epsilon|, \quad (3.39)$$

no instrumentālo mainīgo metodes nostādnes zināms, ka  $E[K \epsilon] = 0$  jeb  $plim_n \frac{1}{n} K \epsilon = 0$ , jo gan instrumentālie mainīgie, gan kontroles mainīgie ir eksogēni.  $\Phi$  ir īstais parametrs jeb konstante, kuru var iznest pirms matemātiskās cerības, tad otrais saskaitāmais ir vienāds ar nulli.

Tika pierādīts, ka abi saskaitāmie tiecas uz nulli, kad izlases apjoms tiecas uz bezgalību, sekojoši instrumentālo mainīgo empīriskās ticamības novērtētājs ir būtisks un nav atkarīgs no mainīgā  $X$  endogenitātes. Taču būtiskumam jānodrošina sekojoši papildus nosacījumi:

- eksistē tāda integrējama funkcija  $H(x)$ , kura ierobežo funkciju  $|((\hat{X} \hat{X}^T)^{-1} \hat{X} X^T - I) \beta|$ ,
- eksistē tāda integrējama funkcija  $H_1(x)$ , kura ierobežo funkciju  $|(\hat{X} \hat{X}^T)^{-1} \hat{X} \epsilon|$ .

Pie šādiem nosacījumiem Hausmena testu var veidot pēc tā neparametriskās specififikācijas. Tad nepieciešams iekļaut pirmās fāzes atlikumus strukturālajā vienādojumā

$$Y = X^T \beta + V^T \theta + W^T \gamma + \epsilon \quad (3.40)$$

un skatīt parametra  $\theta$  statistisko nozīmību ar testu piemērotu empīriskai ticamības novērtētājiem. Noraidot nulles hipotēzi, atlikumi ir statistiski būtiski, tātad instrumentālo mainīgo regresijas empīriskās ticamības novērtētājs ir būtisks. Savukārt empīriskās ticamības novērtētājs vairs nav būtisks.

## 4. SIMULĀCIJAS

Nodaļā apskata empīriskās ticamības novērtētāja Hausmena testa salīdzinājumu ar mazāko kvadrātu metodi. Simulācijas tiek realizētas programmēšanas valodā R, empīriskās ticamības metodi realizē ar pakotni `me1t`. Monte Karlo simulācijas veic balstoties uz modeli

$$\begin{cases} Y = X^T \beta + W^T \gamma + \epsilon, \\ X = \Pi Z + \Gamma W + V, \end{cases} \quad (4.1)$$

kur  $W \sim N(0,1)$ ,  $Z \sim N(1,2)$ ,  $\epsilon, V \sim N_2(0, \Sigma)$ , kur  $\Sigma = \begin{pmatrix} 40 & 12 \\ 12 & 20 \end{pmatrix}$ , kā arī tiek izmantoti sekojošie īstie parametri  $\beta = 2$ ,  $\gamma = 0.5$ ,  $\Gamma = 0.2$  un  $\Pi = 3$ . Šajā nostādņē mainīgais  $X$  ir endogēns jeb pastāv korelācija starp  $\epsilon$  un  $V$ . Tika apskatīti izlases apjomi  $n=100$ ,  $n=200$ ,  $n=300$  un  $n=400$ , simulācijas tika atkārtotas 1000 reizes. Simulāciju rezultāti redzami tabulā 1.

n	Mazāko kvadrātu metodes				Empīriskās ticamības metodes			
	$\hat{\beta}_{LS}$	$\hat{\beta}_{2SLS}$	H stat	p v.	$\hat{\beta}_{EL}$	$\hat{\beta}_{IVEL}$	H stat	p v.
100	2.16	1.99	3.22	0.03	2.16	1.99	11.74	0.02
200	2.16	2.00	4.56	0.00	2.16	1.99	21.21	0.00
300	2.16	2.00	5.54	0.00	2.16	2.00	29.76	0.00
400	2.16	2.00	6.39	0.00	2.16	2.00	38.97	0.00

1. tabula

**Endogēnu datu simulāciju rezultāti, mazāko kvadrātu metodes un empīriskās ticamības metodes novērtētāju vidējās vērtības, attiecīgās Hausmena testa statistikas un p-vērtības vidējās vērtības.**

Novērojams, ka gan mazāko kvadrātu metodes novērtētājs, gan empīriskās ticamības novērtētājs ir novirzīts endogēnu datu gadījumā, instrumentālo mainīgo novērtētājs abos gadījumos ir precīzāks un pie lielāka apjoma sakrīt ar īsto  $\beta$  parametra vērtību. Var secināt, ka arī Hausmena tests abām metodēm darbojas korekti, kur visiem izlases apjomiem tiek noraidīta nulles hipotēze par mainīgā eksogenitāti.

Apskata arī eksogēnu datu simulāciju, lai pārlicinātos, ka tests nenoraida nulles hipotēzi šādā gadījumā. Eksogēnu datu gadījumā nepieciešamas matricas  $\Sigma$  izmaiņas iepriekš raksturotajam modelim.  $\Sigma = \begin{pmatrix} 40 & 0 \\ 0 & 20 \end{pmatrix}$ , visi citi gadījuma lielumu sadalījumi, to parametri un regresijas koeficienti paliek nemainīgi. Eksogēnu datu simulāciju rezultātus apkopo tabulā 2

n	Mazāko kvadrātu metodes				Empīriskās ticamības metodes			
	$\hat{\beta}_{LS}$	$\hat{\beta}_{2SLS}$	H stat.	p v.	$\hat{\beta}_{EL}$	$\hat{\beta}_{IVEL}$	H stat.	p v.
100	2.00	2.00	0.03	0.51	2.00	1.99	1.07	0.50
200	2.00	2.00	0.02	0.50	2.00	2.00	1.08	0.49
300	2.00	2.00	-0.02	0.50	2.00	2.00	1.04	0.50
400	2.00	2.00	-0.03	0.51	2.00	2.00	1.00	0.51

2. tabula

**Eksogēnu datu simulāciju rezultāti, mazāko kvadrātu metodes un empīriskās ticamības metodes novērtētāju vidējās vērtības, attiecīgās Hausmena testa statistikas un p-vērtības vidējās vērtības.**

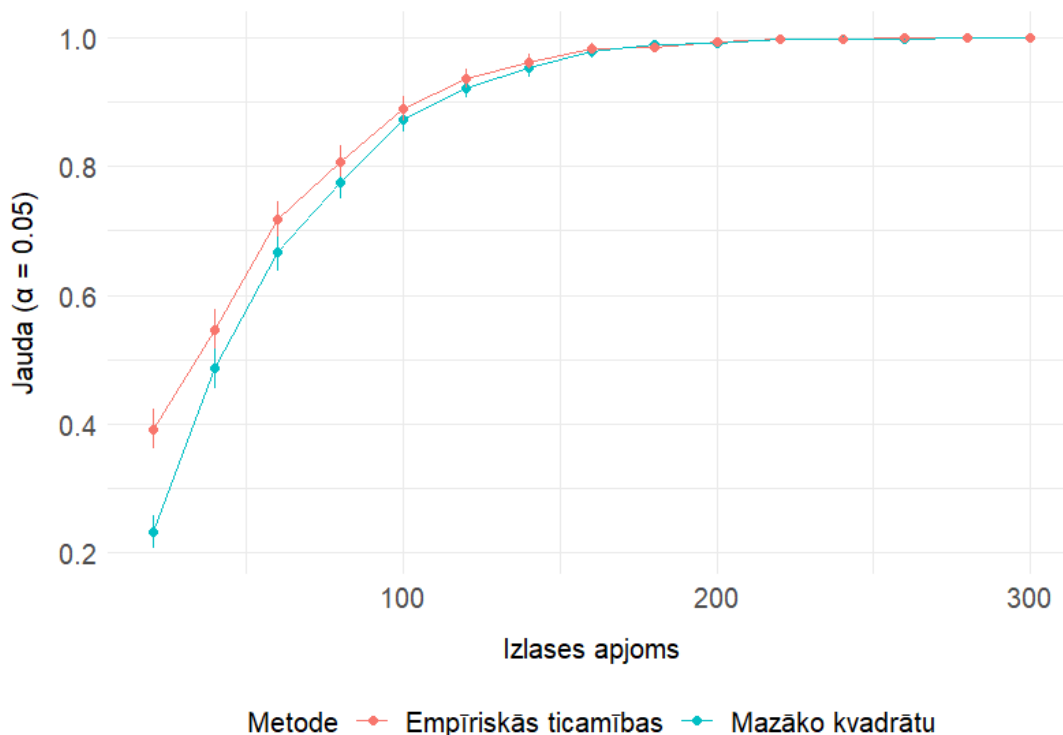
No tabulas 2. var secināt, ka Hausmena tests spēj atpazīt eksogēnus datus un pie visiem izlases apjomiem nulles hipotēzi par eksogenitāti nevar noraidīt pie nozīmības līmeņa 5%. Šim datu dizainam skata arī vidējās pirmā veida kļūdas rezultātus, dažādiem izlases apjomiem, lai novērtētu nozīmības līmeņa 5% derīgumu. Rezultātus apkopo tabulā 3. Var spriest, ka pie mazākiem izlases apjomiem mazāko kvadrātu metodei pirmā veida kļūda ir nedaudz zemāka par empīriskās ticamības metodi. Taču izlases apjomam palielinoties tā konverģē uz vēlamo nozīmības līmeni 5%.

n	Mazāko kvadrātu metodes	Empīriskās ticamības metodes
50	0.05	0.07
100	0.04	0.06
200	0.05	0.06
400	0.04	0.05
800	0.05	0.07
1000	0.05	0.04
1200	0.05	0.05

3. tabula

**Simulāciju pirmā veida kļūdas analīzes rezultāti.**

Skata arī Hausmena testa jaudas analīzi, iepriekš definētam endogēnam simulāciju dizainam, abām parametru novērtēšanas metodēm. Apskata izlases apjomus no 20 līdz 300 ar soli 20, simulācijas atkārtoti 1000 reizes, iegūtos rezultātus apkopo attēlā 1.



**1. att. Jauda atkarībā no izlases apjoma mazāko kvadrātu un empīriskās ticamības metodēm.**

Attēlā 1 redzams, ka pie izlases apjoma 20 starpība starp empīriskās ticamības un mazāko kvadrātu metodi ir gandrīz 0,2, kur empīriskās ticamības metode korekti noraida nulles hipotēzi ar varbūtību 0,39, bet mazāko kvadrātu metode ar varbūtību 0,23. Izlases apjomam palielinoties, starpība sarūk un pie izlases apjoma 160 jaudu starpība vairs nav būtiska. Tad empīriskās ticamības metodes Hausmena tests ir jaudīgāks par mazāko kvadrātu metodi pie maziem izlases apjomiem.

Secinot, simulāciju rezultātā tika noteikts, ka empīriskās ticamības metodes izmantošana parametru novērtēšanai instrumentālo mainīgo regresijā sniedz jaudīgus Hausmena testa rezultātus. Nodrošinot jaudu 0.9 pie izlases apjomiem, kas lielāki par 100. Salīdzinot ar mazāko kvadrātu metodi, līdzvērtīgi rezultāti netiek sasniegti. Pie maziem izlases apjomiem jauda ir krietni zemāka nekā empīriskās ticamības metodei, kur pie izlases apjoma 100 jauda ir 0.86. Tātad empīriskās ticamības novērtētāju lietojums instrumentālo mainīgo regresijā ir jaudīgāks par mazāko kvadrātu metodēm pie maziem izlases apjomiem.

## 5. DATI

Pētījumā tiek pielietoti starptautiskās skolēnu novērtēšanas programmas (PISA) 2022. gada dati. Tie ne tikai novērtē 15 gadus vecu skolēnu matemātikas, lasīšanas un dabaszinātņu prasmes, bet arī ietver studentu sociāldemogrāfisko izcelšanos, studenta pieredzi mācību procesā un tā mājas apstākļus, piemēram, elektronisko ierīču skaitu, grāmatu skaitu, vispārīgus dzīvošanas apstākļus. Pētījuma ietvaros arī skolas un vecāki aizpilda aptaujas, kas iekļauj informāciju par, piemēram, vecāku izglītības līmeni, skolas izmēru, atrašanās vietu, digitālo ierīču politiku un citiem aspektiem. Aptauju veidlapas, datu kopas un datu skaidrojumi atrodami PISA mājaslapā [29].

Liela mēroga novērtējumos tiek uzskatīts, ka vienkāršs testa punktu skaits jeb rezultāts no novērtēšanas darba nav precīzs skolēna spēju novērtējums, tāpēc tiek lietotas vairākas nejaušas vērtības no testa rezultātu sadalījuma funkcijas, šīs vērtības sauc par ticamajām vērtībām (angl. *plausible values*) [6]. Tad sekojoši katram studentam tā katra spēju aspekta novērtējumam tiek piekārtotas desmit ticamās vērtības. Šādu ticamo vērtību lietojums sekundārajā analīzē ir iespējams pēc sekojoša algoritma [6].

1. Novērtē interesējošo statistiku vai modeli desmit reizes, izmantojot katru ticamo vērtību, iegūstot desmit novērtējumus un to kļūdas [6].
2. Aprēķina šo novērtējumu vidējo vērtību [6].
3. Novērtē imputācijas kļūdas lielumu [6].
4. Aprēķina standartkļūdas vērtību, savienojot vidējo izlases kļūdu un imputācijas kļūdu [6].

Tad, pieņemot, ka katras ticamās vērtības novērtētā testa statistika ir  $\theta_i$ , kur PISA pētījumos  $i = 1, \dots, 10$  [28]. Populācijas testa statistiku aprēķina pēc vidējās vērtības formulas [28]

$$\theta = \frac{1}{10} \sum_{i=1}^{10} \theta_i. \quad (5.1)$$

Imputācijas kļūdu aprēķina pēc formulas

$$B_{10} = \frac{1}{9} \sum_{i=1}^{10} (\theta_i - \theta)^2, \quad (5.2)$$

sekojoši var iegūt standartkļūdas vērtību  $V$ , apvienojot izlases kļūdu  $U$  un imputācijas kļūdu  $B_{10}$  [28]

$$V = U + \left(1 + \frac{1}{10}\right) B_{10}. \quad (5.3)$$

Katram studentam ir ne tikai jāaizpilda trīs novērtēšanas testus, bet arī jāaizpilda sociāldemogrāfisko aptauju, kura satur informāciju par dažādiem skolēna dzīves aspektiem. Šajā aptaujā iespējams iegūt informāciju par skolēnu punktualitāti vai prombūtņu biežumu, trauksmi, kas saistīta ar matemātiku, skolotāju attieksmi pret studentiem matemātikas stundās, disciplinētību, skolas uzsākšanas vecumu, klases atkārtošānu, kas ir tikai īss uzskaitījums no plašā jautājumu loka. Tieši šīs aptaujas rezultāti nodrošina pētījumu ar nepieciešamajiem prediktoriem un rada iespējamo prediktoru bāzi.

Darbā tika izmantotas gan skolēnu, gan skolu aptaujas, to iekļautais jautājumu skaits ir plašs, tādēļ darba ietvaros tiks izklāstīts vien izmantoto mainīgo skaidrojums. Katram studentam, skolai un valstij ir savs identifikācijas kods, attiecīgi kolonnas ar nosaukumiem CNTSTUID, CNTSCHID un CNTRYID. Katra skolēna desmit ticamās vērtības atrodamas kolonnās PV1MATH, ..., PV10MATH, kuras matemātiskos modeļos turpmāk atzīmēs ar MATH.

Kontroles mainīgo ietvaros, tika lietoti dažādi skolēnu un skolu demogrāfiskie aspekti. Skolēna dzimums tika reģistrēts jautājumā ST004D01T, kur vērtība 1 apzīmē sievieti un vērtība 2 apzīmē vīrieti, turpmāk šo mainīgo dēvē par GENDER. Tiešais vecums (AGE) ir nepārtraukts mainīgais, kas aprēķināts kā starpība starp aptaujas veikšanas periodu ( gadu un mēnesi) un skolēna dzimšanas periodu ( gadu un mēnesi)[28]. Imigranta statuss (IMMIG) ir kopvērtējums no sešiem dažādiem aspektiem par skolēna un tā vecāku vēsturi, vērtība 1 norāda, ka students ir vietējs, 2 norāda uz pirmās pakāpes imigranta statusu un 3 norāda uz otrās pakāpes imigranta statusu [28]. Skolēna līmenī tika skatīts arī tā sociāldemogrāfiskais indekss (ESCS), kura aprēķins apvieno skolēnu vecāku izglītības un mantu, piemēram, grāmatu, aspektus [28]. Skolēna relatīvā klase GRADE ir indekss, kas norāda starpību starp skolēna klasi un tā vecumam atbilstošo klasi konkrētā valstī.

Skolas līmenī tika skatīti sekojoši rādītāji, skolas lielums (SCHSIZE), skolas atrašanās vieta (SC001Q01TA), vidējais studentu skaits matemātikas klasē (MCLSIZE), skolēnu-skolotāju attiecība (STRATIO) un skolēnu-matemātikas skolotāju attiecība (SMRATIO).

Skolas atrašanās vieta SC001Q01TA turpmāk dēvēta par LOCATION, ir skolas apkārtnes raksturojums tās populācijas ziņām iespējamās vērtības ir ciems ar populāciju zem 3000 (1), maza pilsēta no 3000 līdz 15 000 (2), pilsēta no 15000 līdz 100000 (3), lielpilsēta no 100000 līdz miljonam (4) un vēl divas kategorijas virs miliona (5) un virs desmit miljoniem (6), taču tās nav attiecināmas uz Latviju.

Datu kopā pieejamas ne tikai atbildes uz jautājumiem, bet arī dažādu aspektu kopvērtējumi, kas aprēķināti pēc Varda svērtā ticamības novērtējuma (angl. *Ward's weighted likelihoods estimate*). Visi kopvērtējuma indeksi tika standartizēti, to vidējā vērtība ir 0, bet standartnovirze

1. Galvenokārt pieturoties pie principa, ka pozitīvās vērtības ir vēlamākas un negatīvās vērtības norāda uz problēmām.

DISCLIM ir skolēnu aptaujas mainīgais, kas ietver septiņu dažādu jautājumu atbildes par disciplināro klimatu matemātikas stundās. Katram apgalvojumam iespējamas četras diskrētas atbildes vērtības: katru stundu (vērtība 1), lielāko daļu stundu (vērtība 2), dažās stundās (vērtība 3) un gandrīz nekad (vērtība 4). Šie jautājumi datu kopā marķēti ar kodu ST273 un apkopo informāciju par to cik bieži skolēni neklausā skolotājam, klasē ir troksnis un nekārtība, skolēni nespēj koncentrēties, kā arī pēdējie divi jautājumi attiecināmi uz skolēnu uzmanības novēršanu digitālo ierīču, piemēram, mobilo tālrunu, lietošanas dēļ, kur atdala divas situācijas, kad skolēns pats lieto digitālo ierīci vai kad to lieto kāds cits klasē. Tieši pēdējos divus jautājumus darbā skatīs arī atsevišķi, to jautājumu kodi datu bāzē ir attiecīgi ST273Q06JA un ST273Q07JA, taču uzskatāmības labad tos dēvē par DISCLIM6 un DISCLIM7.

DIGDVPOL ir skolas aptaujas kopvērtējuma mainīgais, kas balstās uz skolas digitālās politikas jautājumiem. Tas sastāv no 9 jautājumiem par skolas uzskatiem un noteikumiem šajā sakarā, taču šis mainīgais nav pieejams Latvijas datiem. Tāpēc tiks lietoti atsevišķi jautājumi, to formulējums un PISA datubāzes kods ir sekojoši:

1. SC190Q01JA - Skolā ir rakstiski noteikumi par digitālo ierīču lietošanu.
2. SC190Q02JA - Digitālo ierīču izmantošana ir aizliegta skolas teritorijā.
3. SC190Q05JA - Skolā ir vadlīnijas digitālo ierīču izmantošanai mācību procesā.
4. SC190Q06JA - Skolotāji nosaka noteikumus par digitālo ierīču izmantošanu.
5. SC190Q07JA - Skolotāji nosaka noteikumus sadarbībā ar skolēniem par digitālo ierīču izmantošanu.
6. SC190Q08JA - Skolai ir specifiska programma par atbildīgu interneta lietošanu.
7. SC190Q09JA - Skolā ir politika par sociālo mēdiju lietošanu mācību procesā.
8. SC190Q10JA - Skolā ir programma, kas veicina sadarbību digitālo ierīču izmantošanā.
9. SC190Q11JA - Skolā ir noteikts laiks, kad izvērtēt digitālo ierīču izmantošanu mācību procesā.

Turpmāk, tiks lietots apzīmējums DIGDVPOL1, ..., DIGDVPOL9 pieraksta saīsināšanai un skaidrībai. Katrs jautājums ir diskrēts mainīgais ar atbildes variantiem: jā (vērtība 1) un nē (vērtība 2).

Darbojoties ar liela izmēra izglītības pētījumiem nepieciešams ņemt vērā aptauju sarežģīto dizainu. Analīzes vienkāršošanas labad, OECD piedāvā speciālu pakotni programmēšanas valodā R `intsvy`, kas nodrošina standarta metožu pielietojumu, ņemot vērā pētījuma sarežģīto izlases dizainu un ticamo vērtību metodi.

Secinot, starptautiskās skolēnu novērtēšanas programmas dati ir visaptveroši un detalizēti. To korektai izmantošanai ir ne tikai jāizprot katra mainīgā uzbūve, bet arī jāņem vērā pētījuma dizains un jāpielieto attiecīgie svari. Šīs nodaļas ietvaros tik sniegts īss kopsavilkums par izmantotajām datu bāzes kolonnām, taču dziļākai izpētei var skatīt sekojošus materiālus [28], [30].

## 6. REGRESIJAS MAINĪGO IZVĒLE

Regresijas mainīgo izvēle ir ierobežota, jo iespējams izmantot tikai tos datus, kas tika ievākti PISA pētījuma ietvaros. Ņemot vērā likumu, kas stājas spēkā 2025. gada 31. maijā, kas paredz mobilo tālrunu izmantošanas ierobežošanas tiesības skolas pārvaldēm, tiks izskatīta šādu liegumu ietekme uz skolēnu matemātikas zināšanu līmeni.

Sākotnēji izskata pieejamo literatūru un nosaka galveno skaidrojošo mainīgo, tā instrumentālo mainīgo un iespējamo kontroles mainīgo kopu nodaļās 6.2, 6.3, tad šai iespējamo kontroles mainīgo kopai pielieto nodaļā 6.1 minētās metodes.

Prediktoru izvēlei tika apskatīti vairāki raksti [40], [15], [23], kas izmanto PISA datus matemātikas prasmju līmeņa kauzālā vai prediktīvā analizē vai apkopo vairāku pētījumu rezultātus vienviet. 2023. gada rakstā [40] tika apskatīti 156 raksti, kas pēta dažādu faktoru ietekmi uz matemātikas prasmēm. Sistemātiskā apskata rezultātā tika identificēti 57 visbiežāk izmantotie faktori un noteikta to ietekme [40].

### 6.1. Regresijas mainīgo izvēles metodes

Regresijas mainīgo izvēlei tiks pielietota pievienojošā mainīgo izvēles metode, kas izmanto informācijas kritērijus. Informācijas kritērijs ir funkcija, kas meklē balansu starp regresijas modeļa skaidrojošo funkciju un modeļa sarežģītību [25]. Visbiežāk izmantotie informācijas kritēriji lineārā regresijā ir  $R_{ADJ}^2$ , Akaika informācijas kritērijs (AIC), Akaika koriģētais informācijas kritērijs (AICC) un Beijesa informācijas kritērijs (BIC) [25]. Ņemot vērā to, ka mainīgo izvēle nav galvenais darba uzdevums, tiks izmantots tikai viens informācijas kritērijs - Akaika informācijas kritērijs jeb AIC.

Akaika informācijas kritērija interpretācija ir sekojoša, jo mazāka ir AIC vērtība, jo modelis kļūst vēlāmāks [25].

$$AIC = n \log \frac{SSR}{n} + 2k + n + n \log(2\pi), \quad (6.1)$$

kur  $SSR$  ir neizskaidrotā variācija,  $n$  objektu skaits un  $k$  prediktoru skaits modelī [25]. AIC ir arī koriģēta versija, kas nepieciešama, ja objektu skaits ir mazs un prediktoru skaits ir relatīvi liels [25], taču šī pētījuma ietvaros, šāda situācija nav iespējama, jo objektu skaits krasi pārsniedz prediktoru skaitu.

Pievienojošā izvēles metode ir iteratīva procedūra, kur sākotnējais stāvoklis ir modelis tikai ar konstantes vērtību [25]. Tad katrā iterācijā tiek pievienots tas mainīgais, kurš sasniedz optimālo informācijas kritēriju, to iekļaujot modelī [25]. Ja kādā solī vairs nav prediktoru, kas

pozitīvi maina informācijas kritērija rezultātu, tad algoritms beidz darbību un par optimālu tiek uzskatīts modelis no priekšpēdējās iterācijas [25].

## 6.2. Prediktoru un to instrumentālo mainīgo izvēle

Tiks izskatīti vairāki modeļi pētāmās problēmas ietvaros. Galvenokārt, tiks pētīta mobilo tālruņu ietekme uz skolēnu matemātikas prasmēm, izmantojot dažādus datu kopā pieejamos prediktorus un tiem atbilstošus instrumentālos mainīgos. Rezultējošais mainīgais šajā pētījumā būs nemainīgs, matemātikas zināšanu līmenis, tas datu kopā tiek reprezentēts caur desmit ticamajām vērtībām.

Par vienu no skaidrojošajiem mainīgajiem tiks skatīts disciplinārais klimats matemātikas stundās (DISCLIM). Šis mainīgais ir endogēns, jo to ietekmē gan vēsturiski apstākļi (slikta pieredze matemātikas stundās iepriekš), gan izlaista mainīgā situācijā, piemēram, skolēnu skolotāju attiecība vai vispārēja skolas disciplīnas nostāja. Disciplināro klimatu skata arī diviem konkrētiem jautājumiem, kas tieši saistīti ar uzmanības novēršanu digitālo ierīču dēļ, jautājumu formulējums ir sekojošs "Cik bieži skolēnu uzmanību novērš digitālo ierīču lietošana?" un "Cik bieži skolēnu uzmanību novērš tas, ka digitālās ierīces lieto kāds cits skolēns?".

Par vienu no instrumentālajiem mainīgajiem var uzskatīt mobilo tālruņu izmantošanas politiku skolā (DIGDVPOL). Taču Latvijas datiem nepieciešams lietot atsevišķos jautājumus DIGDVPOL1, ..., DIGDVPOL9 un pārbaudīt kuriem no tiem ir lielākā ietekme uz skaidrojošo mainīgo. Uzskata, ka digitālo ierīču politikai ir ietekme uz matemātikas prasmēm tikai caur disciplīnu matemātikas stundās, lai izpildītos instrumentālo mainīgo regresijas pieņēmumi. Turpinot, pie pieņēmuma, ka matemātikas klases lielums (MCLSIZE) jeb skolēnu skaits matemātikas stundās ietekmē matemātikas prasmes tikai caur disciplīnu, arī šādu mainīgo var skatīt kā potenciālo instrumentālo mainīgo.

Tabulā 4. atspoguļo potenciālo instrumentālo mainīgo pirmās fāzes nozīmības testa statistiku, lai noteiktu mainīgo derīgumu. No tabulas 4. var secināt, ka matemātikas klases lielums ir stiprs instrumentālais mainīgais visiem disciplīnu skaidrojošajiem mainīgajiem. Mainīgajam DISCLIM visi potenciālie instrumentālie mainīgie, kas saistīti ar tālruņu lietošanas politiku, ir vāji pie nozīmības līmeņa 5%. Skaidrojošajam mainīgajam DISCLIM6 par stipru instrumentu var uzskatīt digitālo ierīču lietošanas politikas 7 apgalvojumu jeb DIGDVPOL7, kas norāda uz sociālo mēdiju lietošanas nosacījumiem stundās, tad pie striktākiem nosacījumiem skolēna digitālo ierīču lietošana arī samazinās. Turpretim mainīgajam DISCLIM7 par stipriem var uzskatīt jautājumus DIGDVPOL4 un DIGDVPOL5, kas atbilst apgalvojumiem, ka skolotāji nosaka digi-

tālo ierīču lietošanas kārtību stundas laikā (ar vai bez skolēnu sadarbības lēmuma pieņemšanā).

Tad modelim, kur skaidrojošais mainīgais ir DISCLIM, par pamatotu instrumentālo mainīgo uzskata matemātikas klases lielumu. Modelim, kur neatkarīgais mainīgais ir DISCLIM6, atbilstoši instrumentālie mainīgie ir DIGDVPOL7 un MCLSIZE, bet DISCLIM7, skaidros ar instrumentāliem mainīgiem DIGDIVPOL4, DIGDIVPOL5 un MCSIZE.

IV	DISCLIM		DISCLIM6		DISCLIM7	
	t stat.	p vērt.	t stat.	p vērt.	t stat.	p vērt.
DIGDVPOL1	0.01	0.50	-0.55	0.29	0.75	0.23
DIGDVPOL2	-0.53	0.30	-1.01	0.16	0.11	0.46
DIGDVPOL3	-0.20	0.42	0.23	0.41	1.26	0.10
DIGDVPOL4	1.54	0.06	1.20	0.12	1.83	0.03
DIGDVPOL5	1.23	0.11	0.92	0.18	2.24	0.01
DIGDVPOL6	1.29	0.10	0.26	0.40	0.61	0.27
DIGDVPOL7	-1.38	0.08	-1.83	0.03	-1.03	0.15
DIGDVPOL8	1.33	0.09	0.39	0.35	0.39	0.35
DIGDVPOL9	0.37	0.36	0.45	0.32	0.62	0.27
MCLSIZE	-2.85	0.00	-2.19	0.01	-3.16	0.00

4. tabula

**Vāju instrumentu tests skolas digitālās politikas deviņiem apgalvojumiem, skolotāju-skolēnu attiecībai, matemātikas skolotāju-skolēnu attiecībai un matemātikas klases izmēram.**

### 6.3. Kontroles mainīgo izvēle

Novērotu nevis eksperimentālu datu gadījuma nav iespējams kontrolēt visus iespējamus ietekmējošos faktoros, tādēļ modeļos tiek iekļauti kontroles mainīgie, kas potenciāli ietekmē rezultējošo mainīgo [46]. Tos iekļauj regresijas analīzē, lai būtu iespējams novērtēt konkrēta mainīgā efektu uz rezultējošo mainīgo [20], galvenokārt, lai novērstu novērtētāja novirzi, taču to pārmērīga un nepamatota izmantošana var radīt pretēju efektu [44]. R. Jorks [46] secina, ka pārmērīga kontroles mainīgo izmantošana var rast līdzvērtīgu efektu kā nepietiekama to izmantošana, kā arī iesaka rast katram kontroles mainīgajam teorētisku pamatojumu. Turklāt nepieciešams identificēt kontroles mainīgo interpretāciju kauzālos modeļos. Šo problemātiku pētīja P. Hunermunds un B. Luvs [20] un secināja, ka kontroles mainīgajiem reti ir kauzāla interpretācija un to galvenā loma ir novirzes samazināšana [20].

Vairākos pētījumos tiek atklāts, ka demogrāfisku prediktoru iekļaušana modelī secina precīzākus rezultātus [8]. Par demogrāfiskiem mainīgajiem tiek uzskatīti sekojoši faktori: skolēna dzimums, skolēna sociālekonomiskais stāvoklis, dzimtā valoda un citi [8]. Skolas kontekstā šos demogrāfiskos mainīgos var pārveidot par dzimumu sadalījumu, skolas sociālekonomisko stāvokli, procentuālo dzimto valodu sastāvu. Piemēram, K. Bokhovs un G. Hampdens-Tomsons [11] savā pētījumā iekļāva sekojošus kontroles mainīgos, dzimums, izglītības līmenis, imigranta statuss, ESCS indekss, kopiena jeb apkārtnē, skolotāju trūkums, studentu-personāla attiecība.

Sekojoši R. Jorka ieteikumiem [46], visiem potenciālajiem kontroles mainīgajiem rada teorētisku pamatojumu, balstoties uz citu pētnieku atklājumiem. Sāk ar sociālekonomisko stāvokli, jo vadoties pēc X.S. Vanga [40] sistemātiskā apskata, tas tiek pielietots vairāk nekā 30 dažādos rakstos. Gamazo [15] pētījis matemātikas zināšanas ietekmējošos faktorus atklāj, ka skolas sociālekonomiskais indekss ir galvenais prediktors skolas matemātikas zināšanu modelēšanā. Šādu saisti atzīmē arī J. Verblovs [43], pētījis skolas lieluma ietekmi uz matemātikas sekmēm, tiek secināts, ka skolēna sociālekonomiskais stāvoklis daudz vairāk ietekmē rezultātu nekā skolas lielums. Galvenokārt, šo rādītāju veido vecāku izglītības līmenis, vecāku nodarbošanās, sadzīves mantas, piemēram, grāmatu, daudzums [7]. Šāda indeksa ietekme uz matemātikas zināšanām ir saistīta ar vairākiem faktoriem, piemēram, vecāku izglītības līmenis ietekmē to attieksmi pret mācībām un ieinteresētību skolēna zināšanās, to ienākumi un nodarbinātība veicina skolēnam pieejamo resursu daudzumu, bet sadzīves mantas kopskaits rada ietekmi uz skolēna ieinteresētību kultūrā un attīstībā [3]. Protams, šādus faktorus iespējams apskatīt atsevišķi, bet ņemot vērā, to, ka kontroles mainīgajiem nav kauzālas ietekmes uz modeli, sociālekonomiskā stāvokļa indeksu var izmantot kā šo faktoru apkopojumu. Var apskatīt arī skolas sociālekonomisko stāvokli, lai kontrolētu, ne tikai paša skolēna demogrāfisko fonu, bet arī vidi, kurā tas uzturas un mācās. Daudzos pētījumos tika konstatēts, ka skolas sociālekonomiskais stāvoklis daudz vairāk ietekmēja studentu rezultātus nekā individuālais indekss [23].

Dzimums ir viens no visbiežāk izmantotajiem prediktoriem [40], taču reti tiek pieminēts konkrēts iemesls vai skaidrojums šī mainīgā iekļaušanai. Daudzos rakstos tika novērots, ka vidēji vīriešu dzimuma pārstāvjiem ir labākas matemātikas zināšanas, taču šādus secinājumus veica vien globālā gadījumā, lokāli, piemēram, vienas valsts ietvaros, šādi rezultāti ne vienmēr ir novērojami [40]. Līdzvērtīgi, lielākā daļa pētījumu, kas norāda uz šādu sakarību tika veikti no 1990. gada līdz 2010. gadiem, jaunākie pētījumi liecina, ka mūsdienu pasaulē dzimumam nav stipras ietekmes uz skolēna matemātikas prasmēm [23].

Daudzos pētījumos izskata skolēna imigrācijas statusa korelāciju ar matemātikas zināšanām [23]. Pēc 2024. gada OECD datiem Latvijā dzīvo 13% imigrantu [27]. Protams, lielākā daļa

imigrē darba nolūkos, sekojoši tie nepiedalās PISA pētījumos [27]. Taču ņemot vērā to, ka 17 % no tiem ir ģimenes locekļi un 29 % no imigrantiem pārcēlās studiju dēļ, šī prediktora derīgumu nepieciešams izskatīt, jo matemātikas zināšanu līmeni var ietekmēt valodas barjera vai citāda veida iepriekšējā izglītība [27].

PISA datu kopā ir pieejama arī relatīvā klase, kurā atrodas skolēns. Šo rādītāju aprēķina salīdzinot skolēna klasi ar to klases līmeni, kurš tam būtu jāapmeklē pēc konkrētās valsts likumdošanas. Tādā veidā iespējams kontrolēt tos skolēnus, kuri atrodas ārpus ierastās klases, klasi zemāk vai augstāk. Tādējādi, kaut arī vecums visiem skolēniem ir līdzvērtīgs, tiem ir dažāds izglītības līmenis. Šeit redzama pozitīva korelācija, kur skolēniem, kas ir augstākās klasēs, bijusi plašāka pieredze un zināšanu klāsts, kā rezultātā arī matemātikas zināšanas ir labākas [40].

Tomēr skolēna vecuma ietekmi nevar izslēgt. Tā, arī, sistemātiskā apskata rakstā [40] 9 pētījumi atklāja pozitīvu asociāciju starp vecumu un matemātikas zināšanu līmeni, kas norāda, ka vecāki studenti uzrāda labākas matemātikas zināšanas.

Zināms, ka skolēna zināšanas ietekmē ne tikai tā individuālie aspekti, bet arī tā apkārtnē un vide, sekojoši nepieciešams apskatīt arī skolēna apmeklētās skolas rādītājus un aktuālākos no tiem iekļaut pētījumā.

Viens no šādiem rādītājiem būtu skolas izmērs, kuru nereti izmanto kā kontroles mainīgo, bet tā ietekme uz matemātikas zināšanām nav viennozīmīga [40]. Šāds rezultāts iespējams, jo skolas izmēram ir dažāda ietekme uz studentiem. Mazās skolās, kas nereti atrodas lauku reģionos, skolēnu skaits arī nav liels, sekojoši ir lielāka skolotāju-studentu attiecība un katra skolēna zināšanu līmenim tiek pievērsts vairāk uzmanības. Pretēji lielākās skolās arī studentu skaits ir lielāks un individuāla pieeja ir mazāk iespējama, taču šādā gadījumā lomu spēlē konkurence studentu vidū.

Turpinot nepieciešams apskatīt arī skolas atrašanās vietu, tās apkārtnes populācijas ziņā. Šis rādītājs PISA datu kopā ir apkopots vairākās kategorijās, kur Latvijai atbilst pirmās četras. Dažādos rakstos tiek atzīmēta pozitīva un statistiski nozīmīga atrašanās vietas ietekme uz matemātikas zināšanām, kur lielākās pilsētās nereti ir augstāks matemātikas zināšanu līmenis [40].

Skolēnu-skolotāju attiecība ir aspekts, kas ietekmē matemātikas prasmes neviennozīmīgi, dažos pētījumos atklāta pozitīva ietekme lielākam studentu skaitam uz vienu skolotāju, dažos pretēji negatīvu ietekmi, taču galvenokārt tā tomēr ir statistiski nozīmīga [40]. Neviennozīmīgie rezultāti iespējami, jo šeit rādītāju ietekmē, gan skolotāji, gan skolēni, un optimālā situācija vienmēr atkarīga no to tieksmēm.

Secinot, šajā nodaļā tika izskatīti un teorētiski pamatoti astoņi kontroles mainīgie: skolēna sociālekonomiskais stāvoklis, skolēna vecums, skolēna relatīvais izglītības līmenis, skolēna

imigranta statuss, skolas izmērs, skolas atrašanās vieta, skolēnu-skolotāju attiecība. Tad nepieciešams arī praktiski noteikt katra no šiem mainīgajiem spēju skaidrot kļūdu un novērst novirzi caur nodaļā 6. apskatītajām metodēm. Tā kā tiks skatīti dažādi pētāmie mainīgie, tad kontroles mainīgie tiks izvērtēti pie katra modeļa.

#### **6.4. Izvēlēto mainīgo apkopojums un datu analīzē skatītie modeļi**

Šajā nodaļā apkopo skaidrojošo, kontroles un instrumentālo mainīgo izvēli un formulē datu analīzē skatītos modeļus. Nodaļā 6.2. tika formulēti trīs skaidrojošie mainīgie, kas tiks skatīti trīs individuālos modeļos: DISCLIM, DISCLIM6, DISCLIM7. Attiecīgi katram no skaidrojošajiem mainīgajiem tika atrasti derīgi instrumentālie mainīgie, matemātikas klases lielums visiem trīs skaidrojošajiem mainīgajiem un digitālo ierīču politikas 7. apgalvojums skaidrojošajam mainīgajam DISCLIM6, digitālo ierīču politikas 4. un 5. apgalvojumi skaidrojošajam mainīgajam DISCLIM7. Tika noteikti arī astoņi kontroles mainīgie, kas tiks piemēroti katram modelim, izmantojot pievienojošo AIC algoritmu.

## 7. DATU ANALĪZE

Šajā nodaļā tiks apskatīti dažādi instrumentālo mainīgo regresijas modeļi, mobilo tālrunu ietekmei uz disciplināro klimatu skolā un sekojoši uz skolēnu matemātikas zināšanām. Katram modelim tiks pielietots pievienojošais AIC mainīgo izvēles algoritms, taču to pielietos izlasei, neņemot vērā izlases dizainu, tehnisku ierobežojumu dēļ. Regresijas modeļus veic izmantojot pakotnes `intsvy` un `melt`, kur empīriskajai ticamībai specificē svarus mainīgajā `weights`. Lai sekotu teorētiski aprakstītajam modelim, kontroles mainīgie tiek fiksēti pēc atbildes mainīgā jeb matemātikas zināšanu modeļa un netiek piemēroti katrai pirmās fāzes regresijai.

Pirmajā modelī disciplināro klimatu matemātikas stundās (*DISCLIM*) skata kā pētāmo mainīgo un matemātikas klases lielumu (*MCLSIZE*) kā instrumentālo mainīgo. Ievērojams ir tas, ka, pētot instrumentālo mainīgo vājumu nodaļā 6.2., neviens no digitālo ierīču izmantošanas politikas apgalvojumiem nevar tikt uzskatīts par stipru instrumentālo mainīgo. Tad sekojoši vispārīgais disciplinārais klimats nav tik ļoti atkarīgs no digitālām ierīcēm, bet to veido vairāku faktoru kopsumma. Pēc mainīgo izvēles algoritma par nozīmīgām tiek uzskatītas sekojošas kontroles: studenta sociāldemogrāfiskais indekss (*ESCS*), relatīvā klase (*GRADE*), skolas izmērs (*SCHSIZE*), skolēna dzimums (*GENDER*), skolas atrašanās vietas klasifikācija (*LOCATION*), skolēna tiešais vecums (*AGE*), skolēna imigranta statuss (*IMMIG*).

$$\begin{aligned} MATH = \beta DISCLIM + \gamma_1 ESCS + \gamma_2 GRADE + \gamma_3 SCHSIZE + \\ + \gamma_4 GENDER + \gamma_5 LOCATION + \gamma_6 AGE + \gamma_7 IMMIG \end{aligned} \quad (7.1)$$

Pirmās fāzes regresija ir sekojoša:

$$\begin{aligned} DISCLIM = \pi MCLSIZE + \kappa_1 ESCS + \kappa_2 GRADE + \kappa_3 SCHSIZE + \\ + \kappa_4 GENDER + \kappa_5 LOCATION + \kappa_6 AGE + \kappa_7 IMMIG \end{aligned} \quad (7.2)$$

Disciplinārā klimata ietekmi uz matemātikas zināšanām, kontrolējot iepriekš minētos mainīgos un skaidrojot klimatu ar matemātikas klases lielumu, apkopo tabulā 5. Var secināt, ka pastāv liela atšķirība starp vispārīgajiem un instrumentālo mainīgo modeļiem, neatkarīgi no novērtēšanas metodes. Pastāv pozitīva disciplinārā klimata ietekme uz matemātikas zināšanu līmeni, taču instrumentālo mainīgo regresijas novērtētāju ir ievērojami mazāks par mazāko kvadrātu metodes novērtētāju. Redzams arī, ka instrumentālo mainīgo novērtētāji abām metodēm nav statistiski nozīmīgi ( $p$ -vērtība  $> 0.05$ ). Hausmena testa rezultāti neatšķiras katrai no novērtēšanas metodēm, nulles hipotēzi par eksogenitāti noraidīt nevar pie nozīmības līmeņa 5%. Protams, formāli šāds secinājums ir korekts, taču empīriskās ticamības metodei  $p$ -vērtība ir 0.06, kas ir ļoti tuvu robežvērtībai, tas nozīmē, ka viennozīmīgi spriest par nulles hipotēzi nevar.

	Mazāko kvadrātu metodes			Empīriskās ticamības metodes		
	Novērtētājs	t stat.	p vērt.	Novērtētājs	$\chi^2$ stat.	p vērt.
$\hat{\beta}_1$	9.74	5.72	0.00	9.56	55.50	0.00
$\hat{\beta}_{IV}$	0.51	0.01	0.50	2.34	1.40	0.42
Hausmena tests		0.18	0.86		89.36	0.06

5. tabula

**Disciplinārā klimata rezultāti, kur ar  $\hat{\beta}_1$  apzīmē novērtētāju vispārīgajam modelim un ar  $\hat{\beta}_{IV}$  novērtētāju modelim ar instrumentālajiem mainīgajiem.**

Mainīgais DISCLIM ir disciplinārā klimata kopvērtējums un iekļauj ne tikai skolēnu uzmanības novēršanu digitālo ierīču dēļ, bet arī vispārīgo skolēnu uzvedību stundās. Tāpēc apskata modeļus, kur skaidrojošais mainīgais ir tieši skolēna uzmanības novēršana digitālo ierīču dēļ. Skolēnu aptaujā ir pieejami divi mainīgie, kas raksturo uzmanības novēršanu. Pirmais atbilst uzmanības novēršanai, jo skolēns pats izmanto digitālās ierīces. Otrais norāda uz uzmanības novēršanu, ja kāds cits izmanto digitālās ierīces. Tie šī darba ietvaros apzīmēti ar DISCLIM6 un DISCLIM7.

Modelim, kur skaidrojošais mainīgais ir DISCLIM6, pielietojot pievienojošo AIC modeli katrai ticamajai vērtība, iegūst, ka optimālais mainīgo skaits ir svārstīgs, starp 6, 7 un 8 kontroles mainīgajiem. Tātad šim mainīgajam izmanto visus kontroles mainīgos un iegūst sekojošu vispārīgo un pirmās fāzes modeļus

$$\begin{aligned}
 MATH &= \beta DISCLIM6 + \gamma_1 ESCS + \gamma_2 GRADE + \gamma_3 SCHSIZE + \\
 &+ \gamma_4 GENDER + \gamma_5 LOCATION + \gamma_6 AGE + \gamma_7 IMMIG + \gamma_8 STRATIO, \\
 DISCLIM6 &= \pi_1 MCLSIZE + \pi_2 DIGDVPOL7 + \kappa_1 ESCS + \kappa_2 GRADE + \\
 &+ \kappa_3 SCHSIZE + \kappa_4 GENDER + \kappa_5 LOCATION + \\
 &+ \kappa_6 AGE + \kappa_7 IMMIG + \kappa_9 STRATIO.
 \end{aligned}
 \tag{7.3}$$

Uzmanības novēršanas telefona izmantošanas dēļ ietekmi uz matemātikas zināšanām apkopo tabulā 6, kontrolējot iepriekš minētos mainīgos un izmantojot mobilo tālrunu lietošanas politiku par sociāliem mēdijiem, kā instrumentālo mainīgo. Var secināt, ka mazāko kvadrātu metodēm nevar noraidīt nulles hipotēzi par eksogenitāti pie nozīmības līmeņa 5%, bet empīriskās ticamības metodē gan eksogenitāte tiek noraidīta. Redzama krasa atšķirība instrumentālo mainīgo novērtētājā starp abām metodēm, kur empīriskās ticamības novērtētājs ir mazāks nekā mazāko kvadrātu metodes novērtētājs. Redzams, arī tas, ka instrumentālo mainīgo novērtētāji ir

negatīvi, kas norāda, ka pie sliktākas disciplīnas matemātikas zināšanu līmenis uzlabojas. Šāds secinājums norāda uz novirzi, kas var tikt radīta instrumentālo mainīgo pieņēmumu neizpildīšanās dēļ. Protams, politika par sociālo mēdiju lietošanu mācību procesā ietekmē uzmanības novēršanas biežumu, bet iespējams tas nav galvenais un kontekstuāli ir citi aspekti, kas ietekmē uzmanības novēršanu. Piemēram, skatot vispārējo stundu skaitu, kad skolēns izmanto mobilo telefonu izklaidei atrodoties skolā, šāda novirze krasi sarūk un tiek iegūti pozitīvi novērtētāji. Tādējādi instrumentālajam mainīgajam ir ne tikai jābūt stipram, bet arī kontekstuāli pietiekami jāietekmē skaidrojošo mainīgo, lai šī metode tiktu pielietota korekti. Šajā gadījumā kaut arī skolas nostāja par sociālo mēdiju lietošanu mācību procesā var tikt uzskatīts par stipru instrumentu, tas neietver pietiekami kontekstuālu cēloņu disciplīnas uzlabošanai, jo skolēni novēršas ne tikai sociālo mēdiju dēļ, bet arī dažādu spēļu, īsziņu un citu paziņojumu dēļ.

	Mazāko kvadrātu metodes			Empīriskās ticamības metodes		
	Novērtētājs	t stat.	p vērt.	Novērtētājs	$\chi^2$ stat.	p vērt.
$\hat{\beta}$	4.98	2.99	0.00	4.71	33.45	0.00
$\hat{\beta}_{IV}$	-29.98	-1.02	0.15	-45.96	2545.85	0.00
Hausmena tests		1.42	0.16		645.08	0.00

6. tabula

**DISCLIM6 rezultāti, kur ar  $\hat{\beta}$  apzīmē novērtētāju vispārīgajam modelim un ar  $\hat{\beta}_{IV}$  novērtētāju modelim ar instrumentālajiem mainīgajiem.**

Skatot modeli, kur pēta, uzmanības novēršanu citu skolēnu dēļ jeb skaidrojošo mainīgo DISCLIM7, visi kontroles mainīgie nav nepieciešami, jo, balstoties uz mainīgo izvēles algoritma rezultātiem, modelī var iekļaut vien 7 kontroles mainīgos.

$$\begin{aligned}
 MATH &= \beta DISCLIM7 + \gamma_1 ESCS + \gamma_2 GRADE + \gamma_3 SCHSIZE + \\
 &+ \gamma_4 GENDER + \gamma_5 LOCATION + \gamma_6 AGE + \gamma_7 IMMIG \\
 DISCLIM7 &= \pi_1 MCLSIZE + \pi_2 DIGDVPOL4 + \pi_3 DIGDVPOL5 + \\
 &+ \kappa_1 ESCS + \kappa_2 GRADE + \kappa_3 SCHSIZE + \\
 &+ \kappa_4 GENDER + \kappa_5 LOCATION + \kappa_6 AGE + \kappa_7 IMMIG
 \end{aligned} \tag{7.4}$$

Rezultātus apkopo tabulā 7, var secināt, ka ir krasa atšķirība starp vispārīgajiem un instrumentālo mainīgo modeļiem. Instrumentālo mainīgo novērtētājiem ir pozitīva ietekme uz matemātikas zināšanu līmeni, kas sakrīt ar kontekstuālo modeļa izpratni. Turpinot, redzams, ka šajā modelī instrumentālo mainīgo novērtētāji ir statistiski nozīmīgi pie nozīmības līmeņa 5%.

Līdzīgi, kā iepriekšējā modelī Hausmena testiem ir dažādi secinājumi, mazāko kvadrātu metode nenoraida nulles hipotēzi pie nozīmības līmeņa 5%, empīriskās ticamības metode noraida. Starp mazāko kvadrātu metodi un empīriskās ticamības metodes novērtētājiem atšķirība pastāv, bet ne tik ievērojama kā iepriekšējos modeļos, tas iespējams norāda uz labāku modeļa specifiskāciju un to, ka instrumentālie mainīgie labi skaidro disciplinārā klimata mainīgo.

	Mazāko kvadrātu metodes			Empīriskās ticamības metodes		
	Novērtētājs	t stat.	p vērt.	Novērtētājs	$\chi^2$ stat.	p vērt.
$\hat{\beta}$	10.56	6.55	0.00	10.68	432.11	0.00
$\hat{\beta}_{IV}$	51.07	1.92	0.03	63.74	3453.68	0.00
Hausmena tests		-1.62	0.11		667.47	0.00

7. tabula

**DISCLIM7 rezultāti, kur ar  $\hat{\beta}$  apzīmē novērtētāju vispārīgajam modelim un ar  $\hat{\beta}_{IV}$  novērtētāju modelim ar instrumentālajiem mainīgajiem.**

Salīdzinot visus trīs modeļus, var secināt, ka mazāko kvadrātu metodēm, Hausmena testu par endogenitāti noraidīt nevar. Tad nevar apgalvot, ka modelī pastāv specifiskācijas kļūda. Savukārt novērtējot modeļus ar empīriskās ticamības metodēm, diviem no trīs modeļiem nulles hipotēzi noraida pie nozīmības līmeņa 5%, tātad var uzskatīt, ka modeļos pastāv endogenitāte un instrumentālo mainīgo regresija ir nepieciešama. Neviennozīmīgi rezultāti iespējams rodas regresijas pieņēmumu neizpildīšanās dēļ, konkrētāk, homoskedasticitātes neizpildīšanās. Tika atklāts, ka visos modeļos pastāv heteroskedasticitāte, tad mazāko kvadrātu novērtētāji nevar tikt uzskatīti par nenovirzītiem. Savukārt, empīriskās ticamības novērtētāji ir robusti arī heteroskedasticitātes gadījumos, tādējādi tie nodrošina arī robustākus Hausmena testa rezultātus.

Secinot, tika skatīti trīs modeļi disciplinārā klimata ietekmei uz matemātikas zināšanu līmeni un tiem pielietota instrumentālo mainīgo regresija, izmantojot konkrētus aspektus no digitālās politikas skolā, kā arī matemātikas klases lielumu. Tiek secināts, ka empīriskās ticamības novērtētāji instrumentālo mainīgo regresijā ir efektīvāki, jo to korektai izpildei nav nepieciešami tik stingri pieņēmumi kā mazāko kvadrātu metodēm.

## 8. REZULTĀTI

Tika skatīti trīs dažādi modeļi par digitālo ierīču politikas ietekmi uz disciplināro klimatu, kas savukārt ietekmē skolēnu matemātikas zināšanas. Visos modeļos redzama būtiska skolēnu skaita ietekme uz disciplīnu, tādēļ, šo mainīgo arī iekļauj modeļos. Tam nerada kauzālu interpretāciju, izvairoties no nekorektiem secinājumiem, jo šādu attiecību ietekmē arī citi faktori, piemēram, konkrētā skolotāja uzvedība lielākās vai mazākās klasēs, konkurences uztvere skolēnu vidū un citi.

Pirmajam modelim, kur disciplinārais klimats matemātikas stundās skatīts kā kopvērtējuma indekss, netika novērota būtiska digitālo ierīču politikas ietekme pie nozīmības līmeņa 5%. Šādu rezultātu rada mainīgā vispārīgums, tā kā mainīgo ietekmē ne tikai skolēnu uzvedība saistībā ar digitālām ierīcēm, bet arī to nespēja klausīt skolotāja norādēm vai trokšņains klimats. Šo mainīgo daudz vairāk ietekmē, piemēram, skolēnu daudzums, skolotāja nostāja un metodika, skolēnu iepriekšējā pieredze un citi. Tādējādi tiek secināts, ka digitālo ierīču lietošana stundās ir tikai viens no faktoriem, kas ietekmē disciplīnu, bet tas nav noteicošais.

Tādēļ tika skatīti skaidrojošie mainīgie, kas tieši atbild par disciplinārā klimata aspektiem, kas iekļauj uzmanības novēršanu digitālo ierīču lietošanas dēļ. Skatot mainīgo, kas raksturo konkrētā studenta uzmanības novēršanu, jo tas pats lieto digitālās ierīces, tika novērota būtiska ietekme skolas politikai par sociālo mēdiju lietošanu mācību procesā. Taču instrumentālo mainīgo novērtētājs šai modelī ir negatīvs, kas ir pretrunā ar kontekstuālo izpratni (jo retāk lieto telefonu stundās jeb skaidrojošā mainīgā vērtības palielinās, jo labākus rezultātus iegūst, tātad atbildes mainīgā vērtības palielinās). Tas norāda uz iespējamu izlaistā mainīgā kļūdu vai nenovērstu endogenitāti. Tātad instrumentālajam mainīgajam ir jābūt ne tikai statistiski nozīmīgam, bet arī kontekstuāli jāietekmē skaidrojošo mainīgo, lai netiktu noraidīts instrumentālo mainīgo saistības pieņēmums.

Turpretim mainīgo, kur uzmanību novērš kāds cits skolēns, kurš lieto digitālās ierīces, būtiski ietekmē skolotāju spēja noteikt digitālo ierīču lietošanu mācību procesā. Turpinot, šis modelis uzrādīja ticamākos rezultātus endogenitātes novēršanas ietvaros, kur modelī redzama pozitīva disciplinārā klimata ietekme skolēnu matemātikas prasmēm jeb arī kontekstuāli, jo mazāk citi skolēni lieto digitālās ierīces, jo labāki matemātikas zināšanu rezultāti.

Secinot, pētot digitālo ierīču politikas ietekmi uz disciplināro klimatu matemātikas stundās, tādējādi novēršot endogenitāti un novirzi disciplīnas kauzālā ietekmē uz matemātikas zināšanu līmeni, svarīgi nodalīt tos disciplinārā klimata aspektus, kas tieši saistīti ar digitālo ierīču lietošanu, lai iegūtu precīzākus novērtējumus, kā arī, lai ievērotu korektu instrumentu principus.

## 9. SECINĀJUMI

Instrumentālo mainīgo regresija ir metode, kas nodrošina endogenitātes novēršanu, sekojoši rada iespēju veikt precīzākus secinājumus par atbildes mainīgā cēloņiem. Parametru novērtēšanu iespējams veikt pēc vairākām metodēm, taču šajā darbā tika skatīta divfāžu mazāko kvadrātu metode un empīriskās ticamības instrumentālo mainīgo metode. Tās tika savstarpēji salīdzinātas gan simulāciju, gan reālu datu ietvaros, kā arī tika pētīta instrumentālo mainīgo metodes nepieciešamība, izmantojot Hausmena testu endogenitātes noteikšanai.

Darba ietvaros tika izstrādāts teorētisks pamatojums Hausmena testa pielietojumam, izmantojot empīriskās ticamības metodi un veiktas jaudas simulācijas. To rezultātā tika noteikts, ka empīriskās ticamības Hausmena testa specifiskācija ir jaudīgāka maziem izlases apjomiem.

Praktiski tika pētīta disciplinārā klimata ietekme uz skolēnu matemātikas prasmēm, novēršot endogenitāti ar digitālo ierīču lietošanas politiku skolā. Galvenokārt tika secināts, ka disciplināro klimatu nevar pētīt kopsakarā un nepieciešams izvērtēt katru to ietekmējošo aspekta cēloņus, tādēļ tika izvēlēti mainīgie, kas raksturo uzmanības novēršanu digitālo ierīču lietošanas dēļ. Tādējādi tiek samazināta gan novirze, gan nodrošināta korekta instrumentālo mainīgo metodes lietošana. Izvirzītā nulles hipotēze neapstiprinājās, jo digitālo ierīču liegums nebija statistiski nozīmīgs nevienam no skatītajiem skaidrojošajiem mainīgajiem.

No optimālā modeļa var secināt, ka skolotāju spēja noteikt digitālo ierīču lietošanu mācību procesā būtiski ietekmē uzmanības novēršanas biežumu, kas rodas no citu skolēnu digitālo ierīču lietošanas. Tādējādi tiek iegūts, retāka uzmanības novēršana klasē, nodrošina augstāku matemātikas zināšanu līmeni. Tad ļaujot skolotājiem ierobežot vai pilnībā izslēgt digitālās ierīces no mācību procesa, var nodrošināt uzlabotu vienu no disciplinārā klimata aspektiem, sekojoši uzlabojot matemātikas zināšanu līmeni.

Tāpat disciplīnas uzlabošanai būtiska ir skolotāju rīcības brīvība, digitālo ierīču lietošanas kārtības ietvaros. Tieši skolotāju nostāja un skaidri noteikumi būtiski ietekmē skolēnu attieksmi pret disciplīnu un uzvedību. Ierobežojot uzmanības novēršanu, var uzlabot mācību vidi un veicināt labākus akadēmiskos rezultātus matemātikas priekšmetā.

Secinot, instrumentālo mainīgo regresija ir efektīvs rīks endogēnu situāciju risināšanai. Tās pielietojums ar empīriskās ticamības novērtētājiem ir jaudīgs modelis maziem izlases apjomiem, kā arī nodrošina precīzākus rezultātus gadījumos, kad regresijas pieņēmumi neizpildās. Šīs metodes lietojums matemātikas zināšanu līmeņa kauzālā analīzē apstiprina empīriskās ticamības metodes priekšrocības, tādējādi veicinot jomas izpēti arī citiem matemātikas zināšanu līmeņa aspektiem.

## LITERATŪRAS SARAKSTS

- [1] Wissam Abdallah, Marc Goergen un Noel O’Sullivan. Endogeneity: How failure to correct for it can cause wrong inferences and some remedies. *British Journal of Management*, 26(4):791—804, 2015.
- [2] PW Aitchison. Generalized inverse matrices and their applications. *International Journal of Mathematical Education in Science and Technology*, 13(1):99—109, 1982.
- [3] Ijtihadi Kamilia Amalina un Tibor Vidákovich. Cognitive and socioeconomic factors that influence the mathematical problem-solving skills of students. *Heliyon*, 9(9), 2023.
- [4] Isaiah Andrews, James H Stock un Liyang Sun. Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, 11(1):727—753, 2019.
- [5] Xiang Ao. Introduction to endogeneity-related methods. *vol*, 1:1—7, 2009.
- [6] J Aparicio, JM Cordero, L Ortiz un J Sancho. PLAUSIBLE VALUES IN THE PISA REPORT AND THEIR TREATMENT IN EFFICIENCY ANALYSIS USING EDUCATIONAL DATA. *INTED2020 Proceedings*, 765.—765. IAPPUSES. IATED, 2020.
- [7] Francesco Avvisati un Celine Wuyts. 'The Measurement of Socio-economic Status in PISA, 2024.
- [8] Ryan S Baker, Lief Esbenshade, Jonathan Vitale un Shamyia Karumbaiah. Using Demographic Data as Predictor Variables: A Questionable Choice. *Journal of Educational Data Mining*, 15(2):22—52, 2023.
- [9] Herbert Stanley Bear. *A primer of Lebesgue integration*. Academic Press, 2002.
- [10] Nathaniel Beck. OLS in matrix form. *Lect Notes*, 1:2, 2001.
- [11] Christian Bokhove un Gillian Hampden-Thompson. Country and school family composition’s effects on mathematics achievement. *School Effectiveness and School Improvement*, 33(2):280—302, 2022.
- [12] John Bound, David A Jaeger un Regina M Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430):443—450, 1995.

- [13] Song Xi Chen un Ingrid Van Keilegom. A review on empirical likelihood methods for regression. *Test*, 18:415—447, 2009.
- [14] Shailaja DESHMUKH. un Madhuri Kulkarni. *Asymptotic Statistical Inference*. Springer, 2021.
- [15] Adriana Gamazo un Fernando Martínez-Abad. An exploration of factors linked to academic performance in PISA 2018 through data mining techniques. *Frontiers in Psychology*, 11:575167, 2020.
- [16] GEM Report UNESCO. Global Education Monitoring Report 2023: Technology in education: A tool on whose terms?, 2023. gada jūlijs.
- [17] Jerry A Hausman. Specification tests in econometrics. *Econometrica: Journal of the econometric society*:1251—1271, 1978.
- [18] Miguel A Hernán un James M Robins. Instruments for causal inference: an epidemiologist’s dream? *Epidemiology*, 17(4):360—372, 2006.
- [19] Aaron D Hill, Scott G Johnson, Lindsey M Greco, Ernest H O’Boyle un Sheryl L Walter. Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management*, 47(1):105—143, 2021.
- [20] Paul Hünermund un Beyers Louw. On the nuisance of control variables in causal regression analysis. *Organizational Research Methods*, 28(1):138—151, 2025.
- [21] Alexandre Janot, Pierre-Olivier Vandanjon un Maxime Gautier. A revised Durbin-Wu-Hausman test for industrial robot identification. *Control Engineering Practice*, 48:52—62, 2016.
- [22] Alexandre Janot, Pierre-Olivier Vandanjon un Maxime Gautier. Identification of physical parameters and instrumental variables validation with two-stage least squares estimator. *IEEE Transactions on Control Systems Technology*, 21(4):1386—1393, 2012.
- [23] Anastasios Karakolidis, Vasiliki Pitsia un Anastassios Emvalotis. Examining students’ achievement in mathematics: A multilevel analysis of the Programme for International Student Assessment (PISA) 2012 data for Greece. *International Journal of Educational Research*, 79:106—115, 2016.
- [24] Haodi Liang un Jiahua Chen. Global Consistency of Empirical Likelihood. *arXiv preprint arXiv:2303.16410*, 2023.

- [25] Charles Lindsey un Simon Sheather. Variable selection in linear regression. *The Stata Journal*, 10(4):650—669, 2010.
- [26] Mette Lise Lousdal. An introduction to instrumental variable assumptions, validation and estimation. *Emerging themes in epidemiology*, 15(1):1, 2018.
- [27] OECD. International Migration Outlook 2024. [https://www.oecd.org/en/publications/international-migration-outlook-2024\\_50b0353e-en/full-report/latvia\\_dc364075.html](https://www.oecd.org/en/publications/international-migration-outlook-2024_50b0353e-en/full-report/latvia_dc364075.html), 2024. [Piekļūts 23-05-2025].
- [28] OECD. PISA Data Analysis Manual SPSS® SECOND EDITION. [https://www.oecd.org/content/dam/oecd/en/publications/reports/2009/03/pisa-data-analysis-manual-spss-second-edition\\_g1gh9e4d/9789264056275-en.pdf](https://www.oecd.org/content/dam/oecd/en/publications/reports/2009/03/pisa-data-analysis-manual-spss-second-edition_g1gh9e4d/9789264056275-en.pdf). [Piekļūts 23-5-2025].
- [29] PISA 2022 Database. <https://www.oecd.org/en/data/datasets/pisa-2022-database.html>. [Piekļūts 10-05-2025].
- [30] PISA 2022 Database — oecd.org. <https://www.oecd.org/en/data/datasets/pisa-2022-database.html>. [Piekļūts 23-05-2025].
- [31] PISA 2022 Results (Volume I and II) - Country Notes: Latvia, 2023. gada decembris. URL: [https://www.oecd.org/en/publications/2023/11/pisa-2022-results-volume-i-and-ii-country-notes\\_2fca04b9/latvia\\_fb915960.html#:~:text=In%20Latvia,%2078%%20of%20students,\(OECD%20average:%2069%\)..](https://www.oecd.org/en/publications/2023/11/pisa-2022-results-volume-i-and-ii-country-notes_2fca04b9/latvia_fb915960.html#:~:text=In%20Latvia,%2078%%20of%20students,(OECD%20average:%2069%)..)
- [32] Jin Qin un Jerry Lawless. Empirical likelihood and general estimating equations. *the Annals of Statistics*, 22(1):300—325, 1994.
- [33] Oliver J Rutz un George F Watson. Endogeneity and marketing strategy research: An overview. *Journal of the Academy of Marketing Science*, 47:479—498, 2019.
- [34] Saeima. Grozījumi Izglītības likumā — likumi.lv. <https://likumi.lv/ta/id/355811-grozijumi-izglitibas-likuma>, 2024. [Piekļūts 22-05-2025].
- [35] Jon Bingen Sande un Mrinal Ghosh. Endogeneity in survey research. *International Journal of Research in Marketing*, 35(2):185—204, 2018.
- [36] Teodora Sheytanova. The accuracy of the Hausman Test in panel data: A Monte Carlo study, 2015.
- [37] Rahul Singh, Maneesh Sahani un Arthur Gretton. Kernel instrumental variable regression. *Advances in Neural Information Processing Systems*, 32, 2019.

- [38] Xiaogang Su, Xin Yan un Chih-Ling Tsai. Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(3):275—294, 2012.
- [39] Alan O Sykes. An introduction to regression analysis, 1993.
- [40] Xiaofang Sarah Wang, Laura B Perry, Anabela Malpique un Tobias Ide. Factors predicting mathematics achievement in PISA: a systematic review. *Large-Scale Assessments in Education*, 11(1):24, 2023.
- [41] Geoffrey S Watson. Linear least squares regression. *The Annals of Mathematical Statistics*:1679—1699, 1967.
- [42] Sanford Weisberg. *Applied linear regression*, 528. sējums. John Wiley & Sons, 2005.
- [43] Jacob Werblow un Luke Duesbery. The impact of high school size on math achievement and dropout rate. *The High School Journal*, 92(3):14—23, 2009.
- [44] Anna C Wysocki, Katherine M Lawson un Mijke Rhemtulla. Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science*, 5(2):25152459221095823, 2022.
- [45] Liugen Xue un Lixing Zhu. Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika*, 94(4):921—937, 2007.
- [46] Richard York. Control variables and causal inference: A question of balance. *International Journal of Social Research Methodology*, 21(6):675—684, 2018.
- [47] Peixin Zhao un Liugen Xue. Empirical likelihood inferences for semiparametric instrumental variable models. *Journal of Applied Mathematics and Computing*, 43:75—90, 2013.

# PIELIKUMS

## A. SIMULĀCIJU KODS

### A.1 Endogēnu un eksogēnu datu simulācijas

```
#Izmantotās programmu paketes
library(xtable)
library(AER)
library(melt)

#Veido funkciju, kas simulē endogēnus datus
simulate_endogenous_data <- function(n){
  sigma <- matrix(c(40,12,12,20), ncol=2)
  de <- mvtnorm::rmvnorm(n=n, mean=c(0,0), sigma=sigma)
  Z<-rnorm(n,1,2)
  W<-rnorm(n,0,1)
  X <- 3*Z + 0.2*W +de[,1]
  Y <- 2*X + 0.5*W+de[,2]
  df<-data.frame(Y,X,Z,W)
}

#Veido funkciju, kas simulē eksogēnus datus
simulate_exogenous_data<-function(n){
  sigma <- diag(c(40, 20))
  de <- mvtnorm::rmvnorm(n=n, mean=c(0,0), sigma=sigma)
  Z<-rnorm(n,1,2)
  W<-rnorm(n,0,1)
  X <- 3*Z + 0.2*W +de[,1]
  Y <- 2*X + 0.5*W+de[,2]
  df<-data.frame(Y,X,Z,W)
}

#Veido funkciju, kas atkarīga no izlases apjoma,
```

```

# Tā simulētiem datiem pielieto mazāko kvadrātu metodi un
#divfāžu mazāko kvadrātu metodi.
# Veic Hausmena testu endogenitātes pārbaudei.
simulate_once<-function(n){
  #endogēnus vai eksogēnus datus simulē, mainot datu simulācijas funkciju
  df<-simulate_exogenous_data(n)

  #Mazāko kvadrātu metode
  mod1<-lm(Y~X + W, data=df)

  #Divfāžu mazāko kvadrātu metodes pirmā fāze
  mod2_fs<-lm(X~Z + W, data=df)
  X_fitted<-fitted(mod2_fs)

  #Divfāžu mazāko kvadrātu metodes otrā fāze
  mod2<-lm(Y ~ X_fitted + W, data=df)

  #Saglabā pirmās fāzes atlikumus
  df["res_1_stage"]<-residuals(mod2_fs)

  #Hausmena testam regresiju ar pievienotiem pirmās fāzes atlikumiem
  augmented <- lm(Y~X +W+res_1_stage, data=df )
  error_coef_pvalue<-summary(augmented)$coefficients[4,4]

  #Veido rezultējošo matricu
  result<-data.frame(
    beta_ols=mod1$coefficients["X"],
    beta_2sls = mod2$coefficients["X_fitted"],
    Stat = summary(augmented)$coefficients[4,3],
    P_value=error_coef_pvalue
  )
  return(result)
}

```

```

#Veido funkciju simulāciju realizēšanai
#Simulācijas atkārtoti 1000 reizes
results_w_c <- function(n){
  set.seed(421) # For reproducibility
  sim<-do.call(rbind, replicate(1000, simulate_once(c(n)), simplify = FALSE))
  summary_results <- data.frame(
    beta_ols = c(mean(sim$beta_ols)),
    beta_iv = c(mean(sim$beta_2sls)),
    Statistic = c(mean(sim$Stat)),
    P_value = c(mean(sim$P_value))
  )
  return(summary_results)
}

#Simulācijas funkcijas vektorizēšana
vect_results<-Vectorize(results_w_c)
#Izlases apjomu vektors
n_val<-c(100,200,300,400)

#Izsauc simulāciju funkciju un saglabā to mainīgajā
least_squares<-t(vect_results(n_val))
n_val_df<-data.frame(n_val)

#Rezultātus saglabā un izdrukā latex formātā
least_squ_df<-cbind(n_val_df,least_squares)
print(xtable(least_squ_df, type = "latex"))

#Empīriskās ticamības simulāciju kods

#Veido funkciju, kas atkarīga no izlases apjoma,
# Tā simulētiem datiem pielieto empīriskās ticamības metodi
#parastai un instrumentālo mainīgo regresijai
# Veic Hausmena testu endogenitātes pārbaudei
simulate_once_EL<-function(n){

```

```

##Simulē datus
#Endogenitāti un eksogenitāti kontrolē mainot sigma matricas koeficientus
#matrix(c(40,12,12,20), ncol=2) - endogēni dati
#matrix(c(40,0,0,20), ncol=2) - eksogēni dati
sigma <- matrix(c(40,12,12,20), ncol=2)
de <- mvtnorm::rmvnorm(n=n, mean=c(0,0), sigma=sigma)
Z<-rnorm(n,1,2)
W<-rnorm(n,0,1)
X <- 3*Z + 0.2*W +de[,1]
Y <- 2*X + 0.5*W+de[,2]
df<-data.frame(Y,X,Z,W)

#Vienkāršā empīriskās ticamības metode
mod1<-el_lm(Y~X + W, data=df)

#Instrumentālo mainīgo empīriskās ticamības pirmā fāze
D<-cbind(Z,W)
Dt<-t(D)
XDt <- X%%D /nrow(D)
DDt <- Dt%%D /nrow(D)
invDDt <- solve(DDt)
Pi <- XDt%%invDDt
df["X_fit"]<-t(Pi%%Dt)
df["residuals"]<-df["X"]-df["X_fit"]

#Instrumentālo mainīgo empīriskās ticamības otrā fāze
mod2<-el_lm(Y ~ X_fit + W, data=df)

#Hausmena testam veic regresiju ar pievienotiem pirmās fāzes atlikumiem
augmented <- el_lm(Y~X +W+residuals, data=df )
aug_summary<-summary(augmented)
error_coef_pvalue<-summary(augmented)@coefficients[4,3]

#Veido rezultātu tabulu

```

```

result<-data.frame(
  beta_el = mod1@coefficients["X"],
  beta_IV_el = mod2@coefficients["X_fit"],
  Stat=summary(augmented)@coefficients[4,2],
  P_value=error_coef_pvalue
)
return(result)
}
#Funkcija veic empīriskās ticamības novērtētāju un
#Hausmena testa rezultātu simulācijas 1000 reizes.
results_EL <- function(n){
  set.seed(421)
  sim<-do.call(rbind, replicate(1000, simulate_once_EL(c(n)), simplify = FALSE))
  summary_results <- data.frame(
    beta_el = c(mean(sim$beta_el)),
    beta_iv_el = c(mean(sim$beta_IV_el)),
    Statistic = c(mean(sim$Stat)),
    P_value = c(mean(sim$P_value))
  )
  return(summary_results)
}

#Simulācijas funkciju vektorizē
vect_results<-Vectorize(results_EL)
#Izlases apjomi simulācijām
n_val<-c(100,200,300,400)

#Izsauc empīriskās ticamības simulāciju funkciju definētiem izlases apjomiem
emp_likelihood<-(vect_results(n_val))
emp_likelihood_1<-unlist(emp_likelihood)
emp_lik_df<-data.frame(n_val)
emp_df<-cbind(emp_lik_df,t(emp_likelihood))

#Mazāko kvadrātu metodes un empīriskās ticamības metodes rezultātus

```

```
#apvieno un izdrukā latex formātā
full_df<-cbind(least_squ_df,emp_df)
print(xtable(full_df, type = "latex"))
```

## A.2 Jaudas analīze

```
##Jaudas analīzes simulācijas
#Izmantotās programmu paketes
library(xtable)
library(mvtnorm)
library(melt)
#Funkcija simulē endogēnus datus
simulate_endogenous_data <- function(n){
  sigma <- matrix(c(40,12,12,20), ncol=2)
  de <- mvtnorm::rmvnorm(n=n, mean=c(0,0), sigma=sigma)
  Z<-rnorm(n,1,2)
  W<-rnorm(n,0,1)
  X <- 3*Z + 0.2*W +de[,1]
  Y <- 2*X + 0.5*W+de[,2]
  df<-data.frame(Y,X,Z,W)
}
#Funkcija simulē eksogēnus datus
simulate_exogenous_data<-function(n){
  sigma <- diag(c(40, 20))
  de <- mvtnorm::rmvnorm(n=n, mean=c(0,0), sigma=sigma)
  Z<-rnorm(n,1,2)
  W<-rnorm(n,0,1)
  X <- 3*Z + 0.2*W +de[,1]
  Y <- 2*X + 0.5*W+de[,2]
  df<-data.frame(Y,X,Z,W)
}
#Mazāko kvadrātu metode
hausman_rep <- function(n, alpha = 0.05, endog = TRUE) {
  # simulē datus
```

```

df <- if (endog) simulate_endogenous_data(n)
else simulate_exogenous_data(n)
# Veic mazāko kvadrātu metodi
ols      <- lm(Y ~ X + W, data = df)
#Veic pirmās fāzes regresiju
first    <- lm(X ~ Z + W, data = df)
df$rhat  <- resid(first)
#Veic regresiju ar pirmās fāzes atlikumiem
aug      <- lm(Y ~ X + W + rhat, data = df)
## izvelk p-vērtību
p <- summary(aug)$coef["rhat", "Pr(>|t|)"]

## pārbauda nulles hipotēzes noraidījumu
reject   <- as.integer(p < alpha)

## Veido rezultējošo matricu
c(reject = reject,
  pvalue = p,
  beta_ols = coef(ols)["X"],
  beta_2sls = coef(aug)["X"])
}

#Empīriskās ticamības metode
simulate_once_EL<-function(n, endog = TRUE){
  #simulē datus
  df <- if (endog) simulate_endogenous_data(n)
  else simulate_exogenous_data(n)

  #Veic vienkāršo empīriskās ticamības metodi
  mod1<-el_lm(Y~X + W, data=df)

  #Veic instrumentalo mainīgo empīriskās ticamības metodes
  #pirmo fāzi
  D<-as.matrix(df[c("Z", "W")])

```

```

Dt<-t(D)
X<-as.numeric(unlist(df[c("X")]))
XDt <- X%%D /nrow(D)
DDt <- Dt%%D /nrow(D)
invDDt <- solve(DDt)
Pi <- XDt%%invDDt
df["X_fit"]<-t(Pi%%Dt)
df["residuals"]<-df["X"]-df["X_fit"]

#Veic instrumentālo mainīgo regresijas otro fāzi
mod2<-el_lm(Y ~ X_fit + W, data=df)

#Hausmena testa regresija
augmented <- el_lm(Y~X +W+residuals, data=df )
aug_summary<-summary(augmented)
error_coef_pvalue<-summary(augmented)@coefficients[4,3]
p <- summary(augmented)@coefficients[4,3]
#Skata nulles hipotēzes noraidīšanu
reject <- as.integer(p < 0.05)
#Rezultātu matrica
c(reject = reject,
  pvalue = p,
  beta_el = mod1@coefficients["X"],
  beta_IV_el = mod2@coefficients["X_fit"])
}

#Mazāko kvadrātu metodes jaudas simulācijas
hausman_power <- function(Nvec, B = 1000, alpha = 0.05) {
  out <- lapply(Nvec, function(n) {
    rej_alt <- replicate(B, hausman_rep(n, alpha, endog = TRUE)["reject"])
    power <- mean(rej_alt)
    se_pow <- sqrt(power * (1 - power) / B) # Monte-Carlo s.e.
    rej_null <- replicate(B, hausman_rep(n, alpha, endog = FALSE)["reject"])
    size <- mean(rej_null)
  })
}

```

```

    se_size <- sqrt(size * (1 - size) / B)

    data.frame(n, power, size, se_pow, se_size)
  })
do.call(rbind, out)
}

#Empīriskās ticamības metodes jaudas simulācijas
hausman_power_el <- function(Nvec, B = 1000, alpha = 0.05) {
  out <- lapply(Nvec, function(n) {
    rej_alt <- replicate(B, simulate_once_EL(n, endog = TRUE)["reject"])
    power <- mean(rej_alt)
    se_pow <- sqrt(power * (1 - power) / B) # Monte-Carlo s.e.

    rej_null <- replicate(B, simulate_once_EL(n, endog = FALSE)["reject"])
    size <- mean(rej_null)
    se_size <- sqrt(size * (1 - size) / B)

    data.frame(n, power, size, se_pow, se_size)
  })
do.call(rbind, out)
}

set.seed(421)
#Definē izlases apjomu vektoru
n_values <- c(50, 100, 200, 400, 800)

#simulāciju rezultātus saglabā attiecīgajos mainīgajos
results_el <- hausman_power_el(n_values, B = 1000)
results_ls <- hausman_power(n_values, B = 1000)

#veido rezultātu tabulu.
Full_results<-cbind(results_ls, results_el)
print(xtable(Full_results, type = "latex"))

```

## B. PRAKTISKĀS DAĻAS KODS

```
#Izmantotās programmu pakotnes
library(haven)
library(MASS)
library(intsvy)
library(melt)

#Datu ielāde
students<-read_sas("C:/Users/Diana/Documents/University/
4 gads/Bakalaurs/PISA data/CY08MSP_STU_QQQ.SAS7BDAT")
schools<-read_sas("C:/Users/Diana/Documents/University/
4 gads/Bakalaurs/PISA data/CY08MSP_SCH_QQQ.SAS7BDAT")
#Atsevišķā mainīgajā saglabā latvijas studentu atbildes
lv_students = students[students$CNT == 'LVA',]
#Kolonnu pārsaukšana
colnames(lv_students)[colnames(lv_students) == "ST004D01T"] ="gender"
colnames(lv_students)[colnames(lv_students) == "ST273Q01JA"]="Math_discipline_1"
colnames(lv_students)[colnames(lv_students) == "ST273Q02JA"]="Math_discipline_2"
colnames(lv_students)[colnames(lv_students) == "ST273Q03JA"]="Math_discipline_3"
colnames(lv_students)[colnames(lv_students) == "ST273Q04JA"]="Math_discipline_4"
colnames(lv_students)[colnames(lv_students) == "ST273Q05JA"]="Math_discipline_5"
colnames(lv_students)[colnames(lv_students) == "ST273Q06JA"]="Math_discipline_6"
colnames(lv_students)[colnames(lv_students) == "ST273Q07JA"]="Math_discipline_7"
colnames(schools)[colnames(schools) == "SC001Q01TA"]="Location_description"

#Pārbauda vājus instrumentus
#No skolu aptaujas izfiltrē tikai nepieciešamās kolonnas
school_controls <- schools[, c("CNTSCHID", "SC190Q01JA", "SC190Q02JA",
                             "SC190Q05JA", "SC190Q06JA", "SC190Q07JA",
                             "SC190Q08JA", "SC190Q09JA", "SC190Q10JA",
                             "SC190Q11JA", "STRATIO", "SMRATIO",
                             "SC202Q07JA", "SCHSIZE","Location_description",
```

```

"MCLSIZE")]

#apvieno studentu un skolu aptaujas
df <- merge(lv_students, school_controls, by = "CNTSCHID", all.x = TRUE)
#Nedefinē kontroles mainīgos
W<-c("ESCS", "AGE", "GRADE","IMMIG","gender",
      "SCHSIZE","Location_description")
#Nedefinē instrumentālos mainīgos
ins_var<-c("SC190Q01JA", "SC190Q02JA", "SC190Q05JA",
           "SC190Q06JA", "SC190Q07JA", "SC190Q08JA",
           "SC190Q09JA", "SC190Q10JA", "SC190Q11JA",
           "STRATIO", "SMRATIO","SC202Q07JA","MCLSIZE")
#Tukšas tabulas turpmāko rezultātu glabāšanai
Test_weak_all1 <- data.frame()
Test_weak_all2 <- data.frame()
Test_weak_all3 <- data.frame()

#cikls visu mainīgo vājuma pārbaudei skaidrojošajam mainīgajam DISCLIM
for (variable in ins_var){
  ##Instrumentālo mainīgo metode: 2SLS
  #Pirmās fāzes regresija DISCLIM=CELL+W
  LS_fs<-intsvy.reg(y="DISCLIM", c(variable,W),data=df, config=pisa_conf)

  #Pārbauda vājus instrumentus
  LS_FS_reg<-LS_fs$reg
  LS_FS_reg["P value"]<-round(pt(q=abs(LS_FS_reg[,3]), df=4073,
  lower.tail=FALSE),4)
  Test_weak<-LS_FS_reg[2,]
  Test_weak_all1<-rbind(Test_weak_all1,Test_weak)
}

#cikls visu mainīgo vājuma pārbaudei skaidrojošajam mainīgajam Math_discipline_6
for (variable in ins_var){
  ##Instrumentālo mainīgo metode: 2SLS

```

```

#Pirmās fāzes regresija DISCLIM=CELL+W
LS_fs<-intsvy.reg(y="Math_discipline_6", c(variable,W),data=df,
config=pisa_conf)

#Pārbauda vājus instrumentus
LS_FS_reg<-LS_fs$reg
LS_FS_reg["P value"]<-round(pt(q=abs(LS_FS_reg[,3]), df=4073,
lower.tail=FALSE),4)
Test_weak<-LS_FS_reg[2,]
Test_weak_all2<-rbind(Test_weak_all2,Test_weak)
}

#cikls mainīgo vājuma pārbaudei skaidrojošajam mainīgajam Math_discipline_7
for (variable in ins_var){
  ##Instrumentālo mainīgo metode: 2SLS
  #Pirmās fāzes regresija DISCLIM=CELL+W
  LS_fs<-intsvy.reg(y="Math_discipline_7", c(variable,W),data=df,
config=pisa_conf)

  #Pārbauda vājus instrumentus
  LS_FS_reg<-LS_fs$reg
  LS_FS_reg["P value"]<-round(pt(q=abs(LS_FS_reg[,3]), df=4073,
lower.tail=FALSE),4)
  Test_weak<-LS_FS_reg[2,]
  Test_weak_all3<-rbind(Test_weak_all3,Test_weak)
}

#Rezultātus apvieno vienā tabulā un izdrukā latex formātā
Test_weak_all<-cbind(Test_weak_all1, Test_weak_all2,Test_weak_all3)
print(Test_weak_all)
print(xtable(Test_weak_all, type = "latex"))

#####Modelis skaidrojošajam mainīgajam DISCLIM
#Veic pievienojošo mainīgo atlasē algoritmu ar AIC

```

```

Y<-paste0("PV",1:10,"MATH")
W<-c("ESCS", "AGE", "GRADE","IMMIG","gender",
      "STRATIO","SCHSIZE","Location_description")
X<-c("DISCLIM")
no.na.data <- na.omit(df[c(Y,X,W)])
a<-no.na.data[,11]

for (i in 1:10){
  PVIMATH<-no.na.data[,i]
  m <- lm(PVIMATH ~ 1+ Math_discipline_6, data = no.na.data)
  fsAIC<-stepAIC(m, direction="forward",
                scope=list(lower=m,
                           upper=~ Math_discipline_6+ESCS+AGE+GRADE+
                           IMMIG+gender + STRATIO + SCHSIZE +
                           Location_description),trace=0)
  fs_AIC_result<-summary(fsAIC)
  #print(paste0("PV",i,"MATH"))
  # print(formula(fs_AIC_result$terms))
  print((list(attr(terms(fsAIC), "term.labels"))))
}

#Izvēlētās kontroles saglabā mainīgajā W
W<-c("ESCS", "AGE", "GRADE","IMMIG","gender",
      "SCHSIZE","Location_description")

#Mazāko kvadrātu metode
LS<-intsvy.reg.pv(c("DISCLIM",W), pvnames=math, data=df, config=pisa_conf )
LS_res<-LS$reg[c(2),]
LS_res["P value"] = round(pt(q=abs(LS$reg[c(2),3]), df=4119,
lower.tail=FALSE),4)

##Instrumentālo mainīgo metode: 2SLS
#Pirmās fāzes regresija DISCLIM=CELL+W
LS_fs<-intsvy.reg(y="DISCLIM", c("MCLSIZE",W),data=df, config=pisa_conf)

```

```

#Pārbauda vājus instrumentus
LS_FS_reg<-LS_fs$reg
LS_FS_reg["P value"]<-round(pt(q=abs(LS_FS_reg[,3]), df=4073,
lower.tail=FALSE),4)
print("Tests par vājiem instrumentiem.")
print(LS_FS_reg)

all_coef<-LS_fs$reg[,1]
coef<-all_coef[3:length(all_coef)-1]
ZW<-as.matrix(df[c("MCLSIZE",W)])
X_fit<-coef%%t(ZW)+all_coef[1]
X_fit1<-coef%%t(ZW)
df["X_fit"]<-t(coef%%t(ZW)+all_coef[1])

df$residuals <- df$DISCLIM - df$X_fit

#Otrās fāzes regresija: Math=DISCLIM_fit+W
twoSLS<-intsvy.reg.pv(c("X_fit",W), math, data=df, config=pisa_conf )
twoSLS_res<-twoSLS$reg[c(2),]
twoSLS_res["P value"] = round(pt(q=abs(twoSLS_res[,3]), df=4119,
lower.tail=FALSE),4)

#Hausmena tests mazāko kvadrātu metodei
df["residuals"]<-df["DISCLIM"]-df["X_fit"]
X<-c("DISCLIM", W, "residuals")
test<-df[X]
test1<-df[c("DISCLIM", "X_fit", "residuals")]
augmented <- intsvy.reg.pv(X, math, data=df, config=pisa_conf )
LS_hausman<-data.frame(summary(augmented))
LS_hausman["P value"] = round(2*pt(q=abs(augmented$reg[,3]),
df=5372, lower.tail=FALSE),4)
LS_hausman<-LS_hausman[c(nrow(LS_hausman)-1),]

```

```

print(LS_res)
print(twoSLS_res)
colnames(LS_hausman)=c("Estimate", "Std. Error", "t value", "P value")

###Rezultātu matrica mazāko kvadrātu metodēm
LS_results<-rbind(LS_res, twoSLS_res, LS_hausman)
print(xtable(LS_results, type = "latex"))

#Empīriskās ticamības metodes
#Definē datu kopas un mainīgos
df1<-df[c(math,"DISCLIM","MCLSIZE","ESCS", "AGE", "GRADE","IMMIG","gender",
          "SCHSIZE","Location_description","W_FSTUWT")]
no.na.data1 <- na.omit(df1)

Z<-as.numeric(unlist(no.na.data1[c("MCLSIZE")]))
W<-as.matrix(no.na.data1[c("ESCS", "AGE", "GRADE","IMMIG","gender",
                           "SCHSIZE","Location_description")])

#Y<-as.numeric(unlist(no.na.data1[c("PV1MATH")]))
X<-as.numeric(unlist(no.na.data1[c("DISCLIM")]))
Weights<-as.numeric(unlist(no.na.data1[c("W_FSTUWT")]))

#Pirmā fāze nav atkarīga no Y, tādēļ, nav nepieciešams to atkārtot 10 reizes
D<-cbind(Z,W)
Dt<-t(D)
XDt <- X%*%D /nrow(D)
DDt <- Dt%*%D /nrow(D)
invDDt <- solve(DDt)
Pi <- XDt%*%invDDt
no.na.data1["X_fit"]<-t(Pi%*%Dt)
no.na.data1["residuals"]<-no.na.data1["DISCLIM"]-no.na.data1["X_fit"]

#Veido funkciju, kas aprēķina empīriskās ticamības metodes
#un Hausmena tetsu atkarībā no ticamās vērtības math

```

```

EL_func<-function(math){
  #Vienkāršā empīriskās ticamības metode
  mod1<-el_lm(math~DISCLIM + ESCS + AGE + GRADE + IMMIG +
  gender + SCHSIZE + Location_description,
  data=no.na.data1, weights = W_FSTUWT)
  EL_res<-summary(mod1)@coefficients[2,]

  #Instrumentālo mainīgo regresijas otrā fāze
  mod2<-el_lm(math ~ X_fit + ESCS + AGE + GRADE + IMMIG
  + gender + SCHSIZE + Location_description,
  data=no.na.data1, weights = W_FSTUWT)
  IV_EL_res<-summary(mod2)@coefficients[2,]

  #Hausmena testa regresija ar pirmās fāzes atlikumiem
  augmented <- el_lm(math~DISCLIM +ESCS + AGE + GRADE + IMMIG
  + gender + SCHSIZE + Location_description+residuals,
  data=no.na.data1 ,weights = W_FSTUWT )
  aug_summary<-summary(augmented)@coefficients
  error_coef_pvalue<-summary(augmented)@coefficients[nrow(aug_summary),3]
  EL_Hausman<-summary(augmented)@coefficients[nrow(aug_summary),]

  #rezultējošā matrica
  EL_results<-rbind(EL_res, IV_EL_res,EL_Hausman)
  return(EL_results)
}

#empīriskās ticamības funkciju vektorizē
vectorised_EL_func<-Vectorize(EL_func)

#Pielieto empīriskās ticamības funkciju visām ticamajām vērtībām
#PV1MATH, ...,PV10MATH
IV_results<-vectorised_EL_func(no.na.data1[,1:10])
#Pārsauc rindas
rownames(IV_results)<-c("EL estimate", "IV EL est", "Hausman estimate",
"EL Chi sq.,""IV Chi sq.", "HAusman statistic",

```

```

"EL p val.", "IV p val", "Hausman p val.")
IV_results<-data.frame(IV_results)
#Aprēķina vidējo vērtību no rezultātiem
IV_results$Mean <- rowMeans(IV_results, na.rm = TRUE)
print(round(IV_results,4))
#Izdrukā rezultātus latex formātā
print(xtable(t(IV_results), type = "latex"))

#####Modelis skaidrošajam mainīgajam DISCLIM6
#Veic forward selection with AIC

Y<-paste0("PV",1:10,"MATH")
W<-c("ESCS", "AGE", "GRADE","IMMIG","gender",
      "STRATIO","SCHSIZE","Location_description")
X<-c("Math_discipline_6")
no.na.data <- na.omit(df[c(Y,X,W)])
a<-no.na.data[,11]

for (i in 1:10){
  PVIMATH<-no.na.data[,i]
  m <- lm(PVIMATH ~ 1+ Math_discipline_6, data = no.na.data)
  fsAIC<-stepAIC(m, direction="forward",
                scope=list(lower=m,
                            upper=~ Math_discipline_6+ESCS+AGE+GRADE+
                            IMMIG+gender +
                            STRATIO + SCHSIZE +
                            Location_description),trace=0)
  fs_AIC_result<-summary(fsAIC)
  #print(paste0("PV",i,"MATH"))
  # print(formula(fs_AIC_result$terms))
  print((list(attr(terms(fsAIC), "term.labels"))))
}

W<-c("ESCS", "AGE", "GRADE","IMMIG","gender",

```

```

"SCHSIZE", "Location_description")

library(intsvy)
math<-paste0("PV",1:10,"MATH")
W<-c("ESCS", "AGE", "GRADE","IMMIG","gender",
     "SCHSIZE", "Location_description", "STRATIO")
#Mazāko kvadrātu metode
LS<-intsvy.reg.pv(c("Math_discipline_6",W), pvnames=math,
data=df, config=pisa_conf )
LS_res<-LS$reg[c(2,nrow(LS$reg)-1),]
LS_res["P value"] = round(pt(q=abs(LS$reg[c(2,10),3]), df=4119,
lower.tail=FALSE),4)

##Instrumentālo mainīgo metode: 2SLS
#Pirmās fāzes regresija DISCLIM=CELL+W
Z<-c("MCLSIZE", "SC190Q09JA")
LS_fs<-intsvy.reg(y="Math_discipline_6", c(Z,W),data=df, config=pisa_conf)
all_coef<-LS_fs$reg[,1]
coef<-all_coef[3:length(all_coef)-1]
ZW<-as.matrix(df[c(Z,W)])
X_fit<-coef%*%t(ZW)
df["X_fit"]<-t(coef%*%t(ZW))

#Otrās fāzes regresija: Math=DISCLIM_fit+W
twoSLS<-intsvy.reg.pv(c("X_fit",W), math, data=df, config=pisa_conf )
twoSLS_res<-twoSLS$reg[c(2,nrow(twoSLS$reg)-1),]
twoSLS_res["P value"] = round(pt(q=abs(twoSLS_res[,3]), df=4119,
lower.tail=FALSE),4)

#Hausmena tests mazāko kvadrātu metodei
df["residuals"]<-df["Math_discipline_6"]-df["X_fit"]
X<-c("Math_discipline_6", W, "residuals")
augmented <- intsvy.reg.pv(X, math, data=df, config=pisa_conf )
LS_hausman<-data.frame(summary(augmented))

```

```

augmented$reg
LS_hausman["P value"] = round(2*pt(q=abs(augmented$reg[,3]),
df=5372, lower.tail=FALSE),4)
LS_hausman<-LS_hausman[c(nrow(augmented$reg)-1),]

print(LS_res)
print(twoSLS_res)
colnames(LS_hausman)=c("Estimate" , "Std. Error", "t value" , "P value")

###Rezultātu matrica mazāko kvadrātu metodēm
LS_results<-rbind(LS_res, twoSLS_res, LS_hausman)
LS_results<-LS_results[c(1,3,5),c(1,3,4)]
library(xtable)
print(xtable(LS_results, type = "latex"))

library(melt)
math<-paste0("PV",1:10,"MATH")
df1<-df[c(math,"Math_discipline_6","SC190Q02JA","ESCS",
"AGE", "GRADE","IMMIG","gender",
"SCHSIZE","Location_description","STRATIO",
"W_FSTUWT","MCLSIZE","SC190Q09JA")]
no.na.data1 <- na.omit(df1)

Z<-as.matrix((no.na.data1[c("MCLSIZE","SC190Q09JA")]))
W<-as.matrix(no.na.data1[c("ESCS", "AGE", "GRADE","IMMIG","gender",
"SCHSIZE","Location_description","STRATIO")]))

Y<-as.numeric(unlist(no.na.data1[c("PV1MATH")]))
X<-as.numeric(unlist(no.na.data1[c("Math_discipline_6")]))
Weights<-as.numeric(unlist(no.na.data1[c("W_FSTUWT")]))

#Pirmā fāze nav atkarīga no Y, tādēļ, nav nepieciešams to atkārtot 10 reizes
D<-cbind(Z,W)
Dt<-t(D)

```

```

XDt <- X%*%D /nrow(D)
DDt <- Dt%*%D /nrow(D)
invDDt <- solve(DDt)
Pi <- XDt%*%invDDt
no.na.data1["X_fit"]<-t(Pi%*%Dt)
no.na.data1["residuals"]<-no.na.data1["Math_discipline_6"]-no.na.data1["X_fit"]

EL_func<-function(math){
  #Veic vispārīgo empīriskās ticamības modeli
  mod1<-el_lm(math~Math_discipline_6 + ESCS + AGE + GRADE + IMMIG + gender + SCHS
              data=no.na.data1, weights = W_FSTUWT) #simple empirical likelihoo
  mod2_fs<-lm(Math_discipline_6~MCLSIZE+SC190Q09JA + ESCS + AGE +
              GRADE + IMMIG + gender + SCHSIZE
              + Location_description+STRATIO,
              data=no.na.data1, weights = W_FSTUWT)

  EL_res<-summary(mod1)@coefficients[2,]

  mod2<-el_lm(math ~ X_fit + ESCS + AGE + GRADE +
              IMMIG + gender + SCHSIZE +
              Location_description+STRATIO,
              data=no.na.data1, weights = W_FSTUWT)

  IV_EL_res<-summary(mod2)@coefficients[2,]

  #Veic Hausmena testu
  augmented <- el_lm(math~Math_discipline_6 +ESCS + AGE + GRADE + IMMIG + gender
                    data=no.na.data1 ,weights = W_FSTUWT )
  aug_summary<-summary(augmented)@coefficients
  error_coef_pvalue<-summary(augmented)@coefficients[nrow(aug_summary),3]
  EL_Hausman<-summary(augmented)@coefficients[nrow(aug_summary),]

  EL_results<-rbind(EL_res, IV_EL_res,EL_Hausman)
  return(EL_results)
}

```

```

}

vectorised_EL_func<-Vectorize(EL_func)

IV_results<-vectorised_EL_func(no.na.data1[,1:10])
rownames(IV_results)<-c("EL estimate", "IV EL est", "Hausman estimate",
"EL Chi sq.,"IV Chi sq.", "Hausman statistic",
"EL p val.", "IV p val","Hausman p val.")
IV_results<-data.frame(IV_results)
IV_results$Mean <- rowMeans(IV_results, na.rm = TRUE)
print(round(IV_results,4))

IV_EL_results<-matrix(IV_results$Mean,nrow=3, byrow=FALSE)

Full_results<-cbind(LS_results,IV_EL_results)
print(xtable((Full_results), type = "latex"))

### DISCLIM7

Y<-paste0("PV",1:10,"MATH")
W<-c("ESCS", "AGE", "GRADE","IMMIG","gender",
      "STRATIO","SCHSIZE","Location_description")
X<-c("Math_discipline_7")
no.na.data <- na.omit(df[c(Y,X,W)])
a<-no.na.data[,11]

for (i in 1:10){
  PVIMATH<-no.na.data[,i]
  m <- lm(PVIMATH ~ 1+ Math_discipline_7, data = no.na.data)
  fsAIC<-stepAIC(m, direction="forward",
                scope=list(lower=m,
                           upper=~ Math_discipline_7+ESCS+AGE+GRADE
                           +IMMIG+gender +

```

```

        STRATIO + SCHSIZE +
        Location_description),trace=0)

fs_AIC_result<-summary(fsAIC)
#print(paste0("PV",i,"MATH"))
# print(formula(fs_AIC_result$terms))
print((list(attr(terms(fsAIC), "term.labels"))))
}

library(intsvy)
math<-paste0("PV",1:10,"MATH")
skaid_main<-"Math_discipline_7"
W<-c("ESCS", "AGE", "GRADE","IMMIG","gender",
      "SCHSIZE","Location_description")
#Mazāko kvadrātu metode
LS<-intsvy.reg.pv(c(skaid_main,W), pvnames=math, data=df, config=pisa_conf )
LS_reg<-LS$reg
LS_res<-LS$reg[c(2),]
LS_res["P value"] = round(pt(q=abs(LS$reg[c(2),3]), df=2913, lower.tail=FALSE),4)

##Instrumentālo mainīgo metode: 2SLS
Z<-c("MCLSIZE","SC190Q06JA","SC190Q07JA")
#Pirmās fāzes regresija DISCLIM=CELL+W
LS_fs<-intsvy.reg(y=skaid_main, c(Z,W),data=df, config=pisa_conf)
all_coef<-LS_fs$reg[,1]
coef<-all_coef[3:length(all_coef)-1]
ZW<-as.matrix(df[c(Z,W)])
X_fit<-coef%*%t(ZW)
df["X_fit"]<-t(coef%*%t(ZW))

#Otrās fāzes regresija: Math=DISCLIM_fit+W
twoSLS<-intsvy.reg.pv(c("X_fit",W), math, data=df, config=pisa_conf )
twoSLS_res<-twoSLS$reg[c(2),]
twoSLS_res["P value"] = round(pt(q=abs(twoSLS_res[,3]),

```

```

df=2913, lower.tail=FALSE),4)

#Hausmena tests mazāko kvadrātu metodei
df["residuals"]<-df[skaid_main]-df["X_fit"]
X<-c(skaid_main, W, "residuals")
augmented <- intsvy.reg.pv(X, math, data=df, config=pisa_conf )
LS_hausman<-data.frame(summary(augmented))
augmented$reg
LS_hausman["P value"] = round(2*pt(q=abs(augmented$reg[,3]),
df=5372, lower.tail=FALSE),4)
LS_hausman<-LS_hausman[c(nrow(augmented$reg)-1),]

print(LS_res)
print(twoSLS_res)
colnames(LS_hausman)=c("Estimate", "Std. Error", "t value", "P value")

###Rezultātu matrica mazāko kvadrātu metodēm
LS_results<-rbind(LS_res, twoSLS_res, LS_hausman)
LS_results<-LS_results[,c(1,3,4)]
library(xtable)
print(xtable(LS_results, type = "latex"))

###Empīriskā ticamība
library(melt)
math<-paste0("PV",1:10,"MATH")
df1<-df[c(math,skaid_main,Z,"ESCS", "AGE", "GRADE","IMMIG","gender",
"SCHSIZE","Location_description","W_FSTUWT")]
no.na.data1 <- na.omit(df1)

Z<-as.matrix((no.na.data1[Z]))
W<-as.matrix(no.na.data1[c("ESCS", "AGE", "GRADE","IMMIG","gender",
"SCHSIZE","Location_description")])

Y<-as.numeric(unlist(no.na.data1[c("PV1MATH")]))

```

```

X<-as.numeric(unlist(no.na.data1[c(skaid_main)]))
Weights<-as.numeric(unlist(no.na.data1[c("W_FSTUWT")]))

#Pirmā fāze nav atkarīga no Y, tādēļ, nav nepieciešams to atkārtot 10 reizes
D<-cbind(Z,W)
Dt<-t(D)
XDt <- X%%D /nrow(D)
DDt <- Dt%%D /nrow(D)
invDDt <- solve(DDt)
Pi <- XDt%%invDDt
no.na.data1["X_fit"]<-t(Pi%%Dt)
no.na.data1["residuals"]<-no.na.data1[skaid_main]-no.na.data1["X_fit"]

EL_func<-function(math){
  #Veic vispārīgo empīriskās ticamības modeli
  mod1<-el_lm(math~Math_discipline_7 + ESCS + AGE + GRADE +
  IMMIG + gender + SCHSIZE + Location_description,
              data=no.na.data1, weights = W_FSTUWT)
  mod2_fs<-lm(Math_discipline_7~MCLSIZE+ SC190Q06JA+SC190Q07JA +
  ESCS + AGE + GRADE + IMMIG + gender + SCHSIZE + Location_description,
              data=no.na.data1, weights = W_FSTUWT)

  EL_res<-summary(mod1)@coefficients[2,]

  mod2<-el_lm(math ~ X_fit + ESCS + AGE + GRADE + IMMIG +
  gender + SCHSIZE + Location_description,
              data=no.na.data1, weights = W_FSTUWT)

  IV_EL_res<-summary(mod2)@coefficients[2,]

  #Veic Hausmena testu
  augmented <- el_lm(math~Math_discipline_7 +ESCS + AGE + GRADE +
  IMMIG + gender + SCHSIZE + Location_description+residuals,
                    data=no.na.data1 ,weights = W_FSTUWT )

```

```

aug_summary<-summary(augmented)@coefficients
error_coef_pvalue<-summary(augmented)@coefficients[nrow(aug_summary),3]
EL_Hausman<-summary(augmented)@coefficients[nrow(aug_summary),]

EL_results<-rbind(EL_res, IV_EL_res,EL_Hausman)
return(EL_results)
}

vectorised_EL_func<-Vectorize(EL_func)

IV_results<-vectorised_EL_func(no.na.data1[,1:10])
rownames(IV_results)<-c("EL estimate", "IV EL est",
"Hausman estimate","EL Chi sq.,""IV Chi sq.",
"Hausman statistic", "EL p val.", "IV p val","Hausman p val.")
IV_results<-data.frame(IV_results)
IV_results$Mean <- rowMeans(IV_results, na.rm = TRUE)
print(round(IV_results,4))

IV_EL_results<-matrix(IV_results$Mean,nrow=3, byrow=FALSE)
print(xtable((IV_EL_results), type = "latex"))
Full_results<-cbind(LS_results,IV_EL_results)
print(xtable((Full_results), type = "latex"))

```

Bakalaura darbs “Empīriskās ticamības metodes lietojums instrumentālo mainīgo regresijā skolēnu matemātikas zināšanu kauzālajā analīzē” izstrādāts LU Eksakto zinātņu un tehnoloģiju fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors:

\_\_\_\_\_

(paraksts)

(datums)

Rekomendēju darbu aizstāvēšanai:

Vadītājs:

\_\_\_\_\_

(paraksts)

(datums)

Recenzents:

\_\_\_\_\_

(paraksts)

(datums)

Darbs iesniegts Matemātikas nodaļā 2025. gada \_\_\_\_ . jūnijā.

\_\_\_\_\_

(Dekāna pilnvarotā persona: vecākā metodiķe Inita Šneidere)

Darbs aizstāvēts bakalaura gala parbaudījuma komisijas sēdē

2025. gada \_\_\_\_ . jūnijā.