

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
EKONOMETRIJOS BAKALAURO STUDIJŲ PROGRAMA

Aušrinė Abariūtė

**LIETUVOS NEDARBO LYGIO MODELIAVIMAS,
NAUDOJANT PANELINIUS APSKRIČIŲ
DUOMENIS**

**MODELING THE UNEMPLOYMENT RATE IN
LITHUANIA BY USING COUNTIES' PANEL DATA**

Bakalauro baigiamasis darbas

Darbo vadovas prof. Marijus Radavičius

Vilnius
2020

Turinys

1	Įvadas	4
2	Teorinė dalis	6
2.1	Koreliacinė analizė	6
2.1.1	Koreliacijos koeficientas	6
2.1.2	Multikolinearumas	7
2.2	Paneliniai tiesinės regresijos modeliai	7
2.2.1	Pastovaus laisvojo nario modelis	8
2.2.2	Fiksuotų efektų modelis	9
2.2.3	Atsitiktinių efektų modelis	9
2.2.4	Fiksuotų ir atsitiktinių efektų modelių palyginimas	10
2.2.5	Statistiniai testai modelių parinkimui	10
2.3	Automatinės kintamųjų parinkimo procedūros	12
2.3.1	Geriausio poaibio metodas	12
2.3.2	Žingsninė regresija	12
2.4	Modelio vertinimo aspektai	13
3	Praktinė dalis	15
3.1	Duomenų aprašas	15
3.2	Kintamųjų apžvalga	17
3.2.1	Nedarbo raida apskričių grupėse	17
3.2.2	Nedarbui įtaką darančių veiksnių raida apskričių grupėse	18
3.3	Koreliacinė analizė	24
3.4	Ekonometrinė duomenų analizė	27
3.4.1	Tiesinis regresijos modelis	28
3.4.2	Liekanų analizė	33
4	Išvados ir rekomendacijos	35
5	Literatūra	36
A	Priedas	38

Lietuvos nedarbo lygio modeliavimas, naudojant panelinius apskričių duomenis

Santrauka

Bakalauro darbe atliktas Lietuvos nedarbo lygio bei kitų ekonominių ir socialinių rodiklių statistinis tyrimas. Naudoti metiniai 2008-2018 m. laikotarpio apskričių duomenys, kurie buvo surinkti iš Lietuvos Statistikos departamento duomenų bazės. Duomenų apdorojimui ir jų statistinėms išvadoms gauti naudotasi R programa (R 3.5.0 for Windows).

Darbo tikslas – pagal turimus statistinius duomenis atlikti nedarbo lygio priklausomybės nuo kitų ekonominių ir socialinių rodiklių analizę.

Duomenų analizei taikomi šie metodai: koreliacinė analizė, panelinė tiesinė regresija, automatinės kintamųjų parinkimo procedūros.

Atlikus analizę, gauta, kad Lietuvos nedarbo lygio pokyčiams didžiausią įtaką daro BVP vienam gyventojui, bruto darbo užmokesčio ir aukštojo išsilavinimo lygio rodikliai.

Raktiniai žodžiai : nedarbo lygis, Pirsono koreliacijos koeficientas, multikolinearumo patikra, paneliniai tiesinės regresijos modeliai.

Modeling the Unemployment Rate in Lithuania by Using Counties' Panel Data

Abstract

A statistical analysis of the unemployment rate in Lithuania and other economic and social indicators was performed in this bachelor's thesis. The annual 2008-2018 years data of 10 counties of Lithuania is taken from the Lithuanian Department of Statistics database. The R program (R 3.5.0 for Windows) was used to process the data and derive their statistical conclusions.

The main goal of bachelor's thesis is to analyse the dependency of unemployment rate from other economic and social indicators.

Methods applied for data analysis: correlation analysis, panel linear regression, automatic variable selection methods.

The analysis revealed that the changes in unemployment rate in Lithuania is mostly affected by the GDP per capita indicator, gross earnings and the level of higher education.

Key words : unemployment rate, Pearson correlation coefficient, multicollinearity test, panel linear regression models.

1 Įvadas

Darbo rinka visą laiką buvo viena pagrindinių ir svarbiausių ekonominės sistemos rinkų. Užimta visuomenė skatina naujų technologijų, paslaugų ar kitų produktų atsiradimą šalyje, tokiu būdu spartindama ekonominį išsivystymą. Didėjantis darbo lygis kuria aukštesnį gerovės lygį šalyje, auga pragyvenimo lygis ir perkamoji galia. Iš kitos pusės, darbu užimti žmonės jaučiasi stabilesni ir, daugiau ar mažiau, įprasminantys savo gyvenimą. Vadinasi, teigiami darbo rinkos padariniai jaučiami tiek visuomenės makro ir mikro lygiu, tiek atskiro individo ir šeimos lygmenyje.

Lietuva šiuo metu susiduria su rimtomis problemomis darbo organizavimo procese – darbo išteklių šalyje paskirstyti netolygiai, pastebimas darbo jėgos struktūros asimetriškumas skirtinguose šalies regionuose. Tokia situacija šalyje neigiamai koreguoja šalies ekonominio augimo tempą, kadangi dalis profesiją turinčių, bet darbo neradusių, žmonių iškeliauja svetur, o blogai pasiskirstę verslai nesugeba pilnai išnaudoti savo gamybinių pajėgumų. Todėl, norint išvengti neigiamų aukštos nedarbo ir užimtumo rodiklių diferenciacijos pasekmių, šalies regionų padėties darbo rinkoje skirtumų analizė tampa labai svarbi. Proceso identifikavimas ir sąlyginių ryšių nustatymas gali padėti suprasti esamą situaciją ir, galbūt, rasti išeities tašką problemai palengvinti.

Darbo rinka ir joje vykstantys procesai gali būti nagrinėjami įvairiais pjūviais. Kadangi rinka, visų pirma, yra socialinis mechanizmas, kurio epicentre atsiduria žmogiškos būtybės su savo skirtingais ir ne visada racionalių pagrindų grįžtais sprendimais, darbo rinkos studijose yra ypač pasižymėję socialinių – psichologijos, antropologijos, sociologijos, teisės ir politikos – mokslų atstovai. Socialiniai tyrimai ir eksperimentai leidžia geriau pažinti žmogaus elgseną rinkos santykiuose, tačiau nesuteikia galimybės kiekybiškai įvertinti ekonominį reiškinių sąlygojančių veiksnių įtaką. Šiam tikslui turi būti sudaromi ekonometriniai modeliai, kurie, jei adekvatūs, ne tik kiekybiškai įvertina veiksnių poveikį ekonominiam reiškiniui, bet ir garantuoja prognozavimo vientisumą ir suderinamumą.

Tyrimo objektas – 2008-2018 metų laikotarpio Lietuvos apskričių nedarbo lygio, infliacijos, bendro vidaus produkto vienam gyventojui, vidaus migracijos saldo 1000-čiui gyventojų, skurdo lygio, mėnesinio bruto darbo užmokesčio ir aukštojo išsilavinimo lygio rodiklių meti-

niai duomenys.

Darbo tikslas – naudojant panelinius apskričių duomenis sudaryti visos šalies nedarbo lygio pokyčius aprašantį tiesinį regresijos modelį.

Darbo uždaviniai:

1. Grafinės ir statistinės analizės pagalba apžvelgti Lietuvos apskričių nedarbo lygio ir kitų ekonominių ir socialinių veiksnių raidą 2008-2018 m. laikotarpiu;
2. Koreliacinės analizės būdu nustatyti, kurie ekonominiai ir socialiniai veiksniai turi didžiausią įtaką nedarbo lygio pokyčiams;
3. Sudaryti tiesinį regresijos modelį paneliniams apskričių duomenims.

Darbas yra sudarytas iš teorinės ir praktinės dalių. Teorinėje dalyje aprašyti metodai, kurie buvo taikomi praktinėje dalyje: koreliacinė analizė, VIF testas multikolinearumui nustatyti, tiesinės regresijos modeliai paneliniams duomenims, F ir Hausman testai, kintamųjų parinkimo procedūros. Praktinėje darbo dalyje analizuoti turimi statistiniai duomenys – grafinių paveikslų pagalba apžvelgta visų kintamųjų raida 2008-2018 m. laikotarpiu, atlikta koreliacinė analizė ir ekonometrinis modeliavimas. Darbe grafiniai paveikslai braižyti su MS Excel ir R programomis, duomenų apdorojimas ir statistinės išvados gautos su R programa. Kodai su R programa yra pateikti bakalauro darbo priede.

2 Teorinė dalis

2.1 Koreliacinė analizė

Pirminis etapas prieš sudarant tiesinius regresijos modelius yra koreliacinė analizė, kuri nustato, kokio stiprumo tiesinis ryšys sieja kiekybiškai išreikštus veiksnius. Prasminga yra tikrinti ryšius tiek tarp aiškinamojo ir modeliuojamo kintamojo, tiek tarp pačių aiškinamųjų kintamųjų. Pirmuoju atveju stiprus tiesinis ryšys reikš statistiškai reikšmingą kintamųjų priklausomybę (bet ne priežastingumą), antruoju atveju pasireikš vadinamoji multikolinearumo problema.

2.1.1 Koreliacijos koeficientas

Koreliacijos koeficientas – kintamųjų priklausomybės stiprumo matas, nepriklausantis nuo matavimo vienetų. [2] Tiesinio ryšio stiprumui įvertinti populiariausias ir dažniausiai taikomas yra tiesinis Pirsono koreliacijos koeficientas, kuris atsitiktiniams dydžiams X ir Y apskaičiuojamas taikant šią formulę [6]:

$$\rho_{X,Y} = \rho_{Y,X} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y},$$

čia σ_X , σ_Y yra atsitiktinių dydžių X ir Y vidutiniai kvadratiniai nuokrypiai, $\text{cov}(X,Y)$ – X ir Y kovariacija. Koreliacijos koeficientas ρ gali įgyti reikšmes tarp -1 ir 1 , kas pirmuoju atveju reiškia funkcinę tiesinę neigiamą priklausomybę, antruoju atveju – funkcinę tiesinę teigiamą priklausomybę.

Darbe nagrinėjamiems duomenims koreliacijos koeficiento stiprumas bus sprendžiamas laikantis tokio vertinimo [3]:

- $|\rho| < 0,3$ labai silpna koreliacija
- $0,3 \leq |\rho| < 0,5$ silpna koreliacija
- $0,5 \leq |\rho| < 0,7$ vidutinė koreliacija
- $0,7 \leq |\rho| < 0,9$ stipri koreliacija
- $0,9 \leq |\rho| \leq 1$ labai stipri koreliacija

2.1.2 Multikolinearumas

Stipri (bet ne tobula) koreliacija tarp dviejų ar daugiau nepriklausomų kintamųjų yra vadinama multikolinearumu. [12] Esant multikolinearumo situacijai, sunku įvertinti kiekvieno atskiro veiksnio įtaką modeliui kintamajam, kadangi regresijos koeficientai pasižymi labai aukštomis standartinėmis paklaidomis. Multikolinearumui pastebėti, verta atkreipti dėmesį į šiuos problemą signalizuojančius požymius:

- Įverčių stabilumas: nežymūs pokyčiai duomenyse gali stipriai pakeisti modelio koeficientų įverčių reikšmes;
- Nors determinacijos koeficiento R^2 reikšmė aukšta (kas reiškia, kad modelis gerai aprašo duomenis), įverčiai yra mažai reikšmingi, taip pat aukštos jų standartinės paklaidos;
- Koeficientų ženklai yra nelogiški, prieštaraujantys akivaizdžiai logikai.

Jeffrey M. Wooldridge teigimu, neįmanoma tiksliai apibrėžti, kokio stiprumo koreliacija tarp aiškinamųjų kintamųjų bus laikoma „per didelė“. [12] Vis dėlto, Cuthbert Daniel yra pasiūlęs statistinę dispersijos mažėjimo daugiklio (VIF) metodą, kuris padeda nustatyti „per daug“ tarpusavyje koreliuojančius regresorius. Testo statistika skaičiuojama pagal šią formulę [9]:

$$VIF = \frac{1}{1 - R_j^2},$$

čia R_j^2 – regresijos modelio, kuriame X_j – priklausomas kintamasis, $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_k$ – nepriklausomi kintamieji, determinacijos koeficientas. Multikolinearumo problema fiksuojama, kuomet VIF reikšmė viršija 4.

2.2 Paneliniai tiesinės regresijos modeliai

Empiriniams duomenims tirti absoliuti dauguma modelių, atsižvelgiant į duomenų turinį ir struktūrą, yra „statomi“ ant tiesinės regresijos, užrašomos forma:

$$y = \alpha_0 + \beta_1 x_1 + \dots + \beta_k x_k + u,$$

čia y – priklausomas (modeliuojamas) kintamasis, $x_1, x_2, \dots, x_k$ – nepriklausomi (aiškinamieji) kintamieji (kitais dar – regresoriai), α_0 – laisvasis narys, $\beta_1, \beta_2, \dots, \beta_k$ – parametrai prie nepriklausomų kintamųjų, u – atsitiktinis dydis. Labiau sofistikuoti ekonometriniai modeliai kyla susiduriant su problema, kuomet tiesinės regresijos negalima pritaikyti turimiems duomenims.

Panelinių (blokuotų) duomenų įvertinimas yra dažnai apibūdinamas kaip efektyvus analitinis metodas, nagrinėjant ekonometrinius duomenis. Kombinuotą panelinių duomenų matricos rinkinį sudaro du pjūviai: pjūvis pagal laiką t ir pjūvis pagal individus (firmas, regionus, šalis ar pan.), žymimus indeksu i . Paneliniams duomenims aprašyti yra sudaromas tiesinis regresijos modelis, kuris turi tokią baigtinę išraišką [5]:

$$y_{it} = \alpha_{it} + \sum_{k=1}^K \beta_{kit} x_{kit} + u_{it},$$

čia i yra individualaus pjūvio indeksas ($i = 1, \dots, N$), k žymi nepriklausomų kintamųjų x_{kit} indeksą ($k = 1, \dots, K$), t žymi laiko momentą ($t = 1, \dots, T$), u_{it} – normaliai pasiskirsčiusios paklaidos su nuliniu vidurkiu. Šiame regresijos modelyje tiek posvyrio, tiek laisvojo nario koeficientai tarp individų i ir laiko momentų t yra skirtingi. Pritaikyti šį modelį empiriniams duomenims yra nelengva užduotis, todėl pagal savo struktūrą paneliniai regresijos modeliai toliau specifikuojami į tris grupes: 1) pastovaus laisvojo nario modelis, 2) fiksuotų efektų modelis ir 3) atsitiktinių efektų modelis. [1]

2.2.1 Pastovaus laisvojo nario modelis

Populiariausia praktika yra laikyti, kad visi panelės elementai (individai) yra homogeniški, t.y. $\alpha_{it} = \alpha \ \forall i, t$ ir $\beta_{it} = \beta \ \forall i, t$. Toks modelis, užrašomas lygtimi

$$y_{it} = \alpha + \beta^T x_{it} + u_{it},$$

vadinamas pastovaus laisvojo nario tiesiniu regresijos modeliu.

2.2.2 Fiksuotų efektų modelis

Fiksuotų efektų panelinis regresijos modelis gali būti trijų tipų: 1) fiksuotų individualių efektų, 2) fiksuotų laiko efektų ir 3) kombinuotas fiksuotų individualių ir laiko efektų. [11] Aptarsime tik pirmus du tipus.

Taikant fiksuotų individualių efektų metodą, kiekvienam panelės elementui pažymėti yra įvedami fiktyvūs kintamieji, tokiu būdu atsižvelgiant į neįtrauktą į modelį (t.y. nestebėtų) kintamųjų poveikį, kuris yra būdingas atskiriems individams. Fiksuotų individualių efektų modelis užrašomas šia lygtimi:

$$y_{it} = \alpha_i + \beta^T x_{it} + u_{it},$$

čia laisvasis narys α kiekvienam individui i skiriasi, bet yra pastovus laike. Fiksuotų laiko efektų paneliniame regresijos modelyje atvirkščiai: laisvasis narys α yra pastovus individams, tačiau laiko momentams t – skiriasi. Pastarasis modelis yra užrašomas tokia lygtimi:

$$y_{it} = \alpha_t + \beta^T x_{it} + u_{it}.$$

Paneliniams duomenims taikant fiksuotų efektų metodą svarbu žinoti, kad tam tikrais atvejais yra galimas labai didelis kiekis fiktyvių laisvųjų narių įtraukimas į modelį. [1] Vertinant daug parametrų, sumažėja ir laisvės laipsnių skaičius, o tai gali tapti problema turint mažą individų imtį.

2.2.3 Atsitiktinių efektų modelis

Atsitiktinių efektų panelinis regresijos modelis, lygiai kaip ir fiksuotų efektų atveju, yra išskiriamas į atsitiktinių individualių, atsitiktinių laiko ir kombinuotą atsitiktinių individualių ir laiko efektų modelius. Pirmuoju atveju panelės elementai nėra homogeniški, tačiau jų skirtingumas, apibrėžiamas per efektus, yra ne pastovaus, o atsitiktinio pobūdžio. Tai reiškia, kad kiekvieno individo i laisvasis narys α traktuojamas kaip atsitiktinis parametras.

Atsitiktinių individualių efektų paneliniame regresijos modelyje

$$y_{it} = \alpha_i + \beta^T x_{it} + u_{it}$$

laisvasis narys α_i užrašomos lygtimi $\alpha_i = \alpha + v_i$, čia v_i yra atsitiktinis kintamasis, kurio vidurkis lygus 0. Pastaroji lygtis gali būti perrašoma į sekančią išraišką:

$$\begin{aligned} y_{it} &= (\alpha + v_i) + \beta x_{it} + u_{it} \\ &= \alpha + \beta x_{it} + (u_{it} + v_i) \end{aligned}$$

Atsitiktinių laiko efektų metodu sudarytame paneliniame regresijos modelyje laisvasis narys α yra pastovus visiems individams, bet atsitiktinis laike.

2.2.4 Fiksuotų ir atsitiktinių efektų modelių palyginimas

Esminis skirtumas tarp fiksuotų ir atsitiktinių efektų modelių yra tas, kad fiksuotų, tegul individualių, efektų modelyje panelės elementai skiriasi per laisvąjį narį, o atsitiktinių efektų modelyje – per paklaidą. Kalbant apie pasirinkimą tarp šių modelių, Dimitrios Asteriou ir Stephen G. pataria: kuomet paneliniai duomenys yra subalansuoti (visi panelės elementai stebimi tuo pačiu fiksuotu laiko intervalu), tikėtina, kad fiksuotų efektų modelis veiks geriau. Kitais atvejais, kai turimas ribotas individų skaičius, tinkamesnis bus atsitiktinių efektų modelis. [1] Kita vertus, sudarant atsitiktinių efektų modelį, galima gauti paslinktus ir nesuderintus įverčius, jei atsitiktiniai efektai koreliuoja su aiškinamaisiais modelio kintamaisiais, o tai sukuria griežtus apribojimus dirbant su paneliniais duomenimis. Todėl renkantis teisingą modelį svarbu yra palyginti abiem metodais gautų įverčių skirtumus.

2.2.5 Statistiniai testai modelių parinkimui

Siekiant nustatyti, kuris iš metodų – pastovaus laisvojo nario, fiksuotų ar atsitiktinių efektų – yra geriausias modeliuojant turimus panelinius duomenis, skirtinguose modelių sudarymo etapuose yra taikomi įvairūs statistiniai testai. Populiariausi jų – F ir Hausman testas.

F testas. Statistinis F testas yra skirtas patikrinti fiksuotų efektų statistinį reikšmingumą paneliniuose duomenyse. Tikrinama tokia hipotezė [1]:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_N$$

$$H_1: \alpha_s \neq \alpha_j \text{ bent vienam } s, j \in \{1, \dots, N\}$$

Neatmetus nulinės hipotezės daroma išvada, kad fiksuoti efektai paneliniuose duomenyse yra statistiškai nereikšmingi, t.y. visi panelės elementai homogeniški. Taigi, duomenis šiuo atveju gerai aprašo pastovaus laisvojo nario modelis. Priešingu atveju, kuomet nulinė hipotezė yra atmetama, prasminga yra taikyti fiksuotų efektų metodą.

Svarbu pažymėti – F testas neužtikrina, kad efektai, kurie skiriami į individualius arba laiko, bus būtinai fiksuoti – jie gali būti ir atsitiktiniai. Todėl, jei efektai nustatomi statistiškai reikšmingais, sekantis žingsnis yra palyginti fiksuotų ir atsitiktinių efektų metodais sudarytus modelius. Šiam tikslui atlikti yra taikomas Hausman testas.

Hausman testas. Jerry A. Hausman (1978) pasiūlytas statistinis testas tikrina, kuris iš fiksuotų ir atsitiktinių efektų modelių yra tinkamesnis nagrinėjamiems paneliniams duomenims. Hipotezė formuluojama taip:

$$H_0: \beta^{FE} \text{ ir } \beta^{AE} \text{ nesiskiria,}$$

$$H_1: \beta^{FE} \text{ ir } \beta^{AE} \text{ skiriasi,}$$

čia β^{FE} yra fiksuotų efektų metodu apskaičiuoti įverčiai, β^{AE} – atsitiktinių efektų metodu apskaičiuoti įverčiai. Testo statistika turi tokią išraišką [1]:

$$H = (\hat{\beta}^{FE} - \hat{\beta}^{AE})'[Var(\hat{\beta}^{FE}) - Var(\hat{\beta}^{AE})]^{-1}(\hat{\beta}^{FE} - \hat{\beta}^{AE}) \sim \chi^2(k)$$

Jei kriterijaus statistika yra didelė, tuomet skirtumas tarp fiksuotų ir atsitiktinių efektų metodais apskaičiuotų įverčių laikomas statistiškai reikšmingu (H_0 atmetama). Ši situacija yra suvokiama kaip atsitiktinių efektų metodo neveiksnumas, kuris sąlygoja paslinktus ir nesuderintus įverčius. Todėl atmetus nulinę hipotezę, paneliniams duomenims yra taikytinas fiksuotų efektų metodas, priešingu atveju tinkamesnis atsitiktinių efektų metodas.

2.3 Automatinės kintamųjų parinkimo procedūros

Sudarius regresijos modelį, dažnu atveju susiduriama su situacija, kuomet dalis aiškinamųjų kintamųjų statistiškai reikšmingai nepaaiškina modeliuojamo kintamojo. Kadangi mažiausių kvadratų metodas (MKM) tik įvertina modelio koeficientus, bet nenurodo, kuriuos kintamuosius vertėtų pašalinti iš modelio, naudinga yra taikyti automatinės kintamųjų parinkimo procedūras. Kintamųjų skaičiaus sumažinimas, nors ne visada, tačiau gali pagerinti modelio tikslumą bei jo interpretavimą.

Geriausio modelio paieškai darbe bus taikomos dvi populiarios procedūros: pirma, geriausio poaibio (angl. best subset) parinkimo metodas ir, antra, žingsninė regresija (angl. stepwise regression). Minėtosios procedūros automatiškai vertina modelius, sudarytus iš įvairių regresorių kombinacijų (poaibių), ir tarp jų išrenka geriausią. Geriausiu modeliu yra laikomas tas, kuris minimizuoja pasirinktą modelio parinkimo kriterijų (AIC, BIC ir kt.).

2.3.1 Geriausio poaibio metodas

Geriausio poaibio metodas įvertina atskirus modelius su visomis įmanomomis regresorių kombinacijoms. Vadinasi, jei aiškinamųjų kintamųjų skaičius yra k , geriausio poaibio metodas patikrins 2^k kombinacijas. [4]

2.3.2 Žingsninė regresija

Žingsninė procedūra pagal atlikimo algoritmą yra skiriama į dvi modifikacijas – kintamųjų įrašymo (angl. Forward stepwise) ir kintamųjų išbraukimo (angl. Backward stepwise). [4] Pirmoji ypatinga tuo, kad į tuščią (regresorių neturintį) modelį po vieną įtraukiami regresoriai, suteikiantys geriausią „papildomą pagerinimą“ modeliui. Žingsninė kintamųjų išbraukimo procedūra, atvirkščiai, pradedama nuo pilno (turinčio visus regresorius) modelio, toliau po vieną eliminuojant mažiausiai reikšmingus. Tiek kintamųjų įrašymo, tiek kintamųjų išbraukimo žingsninių procedūrų metu iš viso yra sudaromi $1 + k(k + 1)/2$ modeliai.

Pasak Bruce E. Hansen, pasirinkimas tarp geriausio poaibio ir žingsninės procedūrų yra grindžiamas norimu į modelį įtraukti regresorių skaičiumi. [4] Jei regresorių skaičius yra didelis, geriau veiks žingsninė procedūra, kadangi ji, lyginant su geriausio poaibio metodu,

padės išvengti didelių skaičiavimų.

2.4 Modelio vertinimo aspektai

Duomenims pritaikius regresijos modelį, svarbu patikrinti šias jo specifikas – įvertintų parametrų reikšmingumą, įvertinto modelio tikslumą ir patį modelio reikšmingumą. Toliau trumpai aptarsime, kaip tai yra daroma.

Parametrų reikšmingumas. Regresijos koeficientų reikšmingumui nustatyti gali būti taikomi daugybė statistinių testų. Vienas populiariausių – Stjudento kriterijus. Reikšmingumo lygmuo visame bakalauro darbe standartiškai bus laikomas $\alpha = 0.05$, tad jei Stjudento testo p -reikšmė bus mažesnė už α , parametro įvertis bus laikomas statistiškai reikšmingu, ir atvirkščiai.

Įvertinto modelio tikslumas. Regresijos modelio tikslumo vertinimui yra naudojamas determinacijos koeficientas R^2 , kuris parodo, kokią dalį nagrinėjamo rodiklio sklaidos lemia regresijos lygties nepriklausomi kintamieji. [7] Determinacijos koeficientas skaičiuojamas pagal šią formulę:

$$R^2 = 1 - \frac{\sum_i^N (\hat{y}_i - \bar{y})^2}{\sum_i^N (y_i - \bar{y})^2},$$

čia $\hat{y}_i$ – kintamojo y_i įvertis, $\bar{y}$ – kintamojo y_i vidurkis, $0 \leq R^2 \leq 1$. Kuo koeficiento reikšmė yra arčiau 1, tuo modelis geriau aprašo duomenis. $R^2 = 1$ reiškia, kad visi stebėjimai yra išsidėstę regresijos tiesėje, t.y. tarp priklausomo ir nepriklausomų kintamųjų egzistuoja tobulas tiesinis ryšys.

Koreguotas determinacijos koeficientas $\hat{R}^2$, kuris yra nepaslinktas determinacijos koeficiento R^2 įvertinys, skaičiuojamas naudojant šią formulę [7]:

$$\hat{R}^2 = R^2 - \frac{k}{n - k - 1} \cdot (1 - R^2),$$

čia n – stebėjimų skaičius, k – nepriklausomų kintamųjų skaičius. Koreguotas determinacijos koeficientas atsižvelgia į imties didumą ir nepriklausomų kintamųjų skaičių, todėl yra tinkamesnis turint mažai stebėjimų ir daug regresorių.

Įvertinto modelio reikšmingumas. Sudaryto regresijos modelio reikšmingumo vertinimui tikrinama tokia hipotezė:

H_0 : visi regresijos lygties koeficientai lygūs nuliui,

H_1 : bent vienas regresijos lygties koeficientas nelygus nuliui.

Jeigu nulinė hipotezė yra teisinga, reiškia, kad generalinės visumos visi dauginės regresijos lygties koeficientai lygūs nuliui, t.y. nėra priklausomo ir nepriklausomų kintamųjų realaus ryšio. [7] Atmetus nulinę hipotezę daroma išvada, kad bent vienas veiksnys paaiškina modeliuojamą kintamąjį, taigi, regresijos lygtis yra statistiškai reikšminga.

Nulinei hipotezei patikrinti yra naudojamas F kriterijus. F santykio formulė gali būti išreikšta naudojant determinacijos koeficientą R^2 [7]:

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)},$$

čia n – stebėjimų skaičius, k – nepriklausomų kintamųjų skaičius. Vadinasi, F kriterijus iš tiesų tikrina determinacijos koeficiento statistinį reikšmingumą. Jei šis pasirodo reikšmingas, modelis laikomas adekvačiu.

3 Praktinė dalis

3.1 Duomenų aprašas

2008 metais prasidėjusi pasaulio ekonomikos krizė smarkiai paveikė Lietuvos ūkį. Pokyčiai, kurie nuosmukio metu įvyko ekonominiame sektoriuje, vienareikšmiškai didino darbo išteklių šalyje trūkumą, o tai koregavo žmonių socialinį gyvenimą, ir gerovę šalyje apskritai. Vadinas, nedarbo, kaip ekonominio ir socialinio reiškinių, svyravimai šalies raidoje negali būti nagrinėjami vien ekonominiu aspektu – privaloma atsižvelgti ir į socialinių rodiklių pokyčius. Šiam tikslui pasiekti, panelinė analizė buvo atlikta naudojant 2008-2018 metų laikotarpio ekonominių ir socialinių rodiklių metinius duomenis, kurie buvo surinkti iš Lietuvos Statistikos departamento laisvai prieinamos duomenų bazės. [13]

Dėl ekonominių ir socialinių rodiklių gausos, nedarbo lygio analizei pasirinkti tik šeši svarbiausi: metinė infliacija, bendrasis vidaus produktas vienam gyventojui, neto vidaus migracija, asmenys, gyvenantys skurdo rizikoje ar socialinėje atskirtyje (toliau sutrumpintai vadinsime „skurdo lygis“), mėnesinis bruto darbo užmokestis ir gyventojų aukštasis išsilavinimas. Pastarieji rodikliai, kartu su Statistikos departamento naudojamais matavimo vienetais ir darbo autorės sukurtais pažymėjimais, surašyti žemiau pateiktoje lentelėje.

1 lentelė: Ekonominiai ir socialiniai rodikliai, jų matavimo vienetai ir pažymėjimai

Rodiklis	Matavimo vienetai	Pažymėjimai
nedarbo lygis	proc.	<i>unem</i>
metinė infliacija	proc.	<i>inf</i>
BVP vienam gyventojui	EUR	<i>bvp</i>
neto vidaus migracija	asmenys	<i>migr</i>
skurdo lygis	proc.	<i>povrate</i>
mėnesinis bruto darbo užmokestis	EUR	<i>wage</i>
gyventojų aukštasis išsilavinimas	tūkst.	<i>educ</i>

Šaltinis: sudaryta darbo autorės.

Visi 1 lentelėje surašyti rodikliai, išskyrus metinę infliaciją, yra pamatuoti apskritims individualiai. Kadangi infliacijos lygis jau yra kainų lygio santykinis pokytis (matuojamas

kainų indekso padidėjimu per metus), šiam rodikliui nebuvo taikytos jokios duomenų transformacijos. Likę darbe naudojami ekonominiai ir socialiniai rodikliai dėl matavimo vienetų nevienodumo buvo transformuoti šiais būdais:

- Apskričių rodikliai, kurie matuojami procentiškai, būtent, nedarbo lygis, skurdo lygis ir aukštojo išsilavinimo lygis (apskaičiuotas kaip aukštą išsilavinimą turinčių ir darbo jėgos santykis), buvo transformuoti, panaudojant pirmos eilės skirtumus;
- Pinigines išraiškas turintys apskričių rodikliai – bendrasis vidaus produktas vienam gyventojui ir mėnesinis bruto darbo užmokestis – buvo logaritmuoti ir vieną kartą diferencijuoti;
- Neto vidaus migracijos rodiklis buvo paverstas į vidaus migracijos saldo (skirtumas tarp atvykusiųjų ir išvykusiųjų) 1000-čiui gyventojų rodiklį, kuris buvo logaritmuotas ir vieną kartą diferencijuotas.

Atlikus reikalingas duomenų transformacijas, dėl skirtumų panaudojimo buvo prarasta po vieną visų transformuotų rodiklių stebėjimą kiekvienai apskričiai. Dėl to tiriamasis laikotarpis susitraukė iki 2009-2018 metų, o kiekvienam rodikliui liko po 100 stebėjimų (10 apskričių, 10 laiko momentų). Infliacijos duomenys, kadangi jie nebuvo transformuoti, atitinkamai buvo surinkti tam pačiam 2009-2018 m. laikotarpiui.

Aptariant Lietuvos apskričių žymėjimą, kuriuo buvo naudotasi regresinėje analizėje, apskritys pagal Statistikos departamentą numeruojamos tokia tvarka: 1) Vilniaus, 2) Alytaus, 3) Kauno, 4) Klaipėdos, 5) Marijampolės, 6) Panevėžio, 7) Šiaulių, 8) Tauragės, 9) Telšių, 10) Utenos.

Kintamųjų apžvalgai buvo naudoti Statistikos departamento matavimo vienetais skaičiuoti rodiklių 2008-2018 metų duomenys, tik vietoje neto vidaus migracijos rodiklio stebint vidaus migracijos saldo 1000-čiui gyventojų rodiklį ir vietoje gyventojų aukštojo išsilavinimo stebint aukštojo išsilavinimo lygį apskrityse. **Regresinei** duomenų analizei naudoti transformuoti rodiklių duomenys, apimantys 2009-2018 m. laikotarpį.

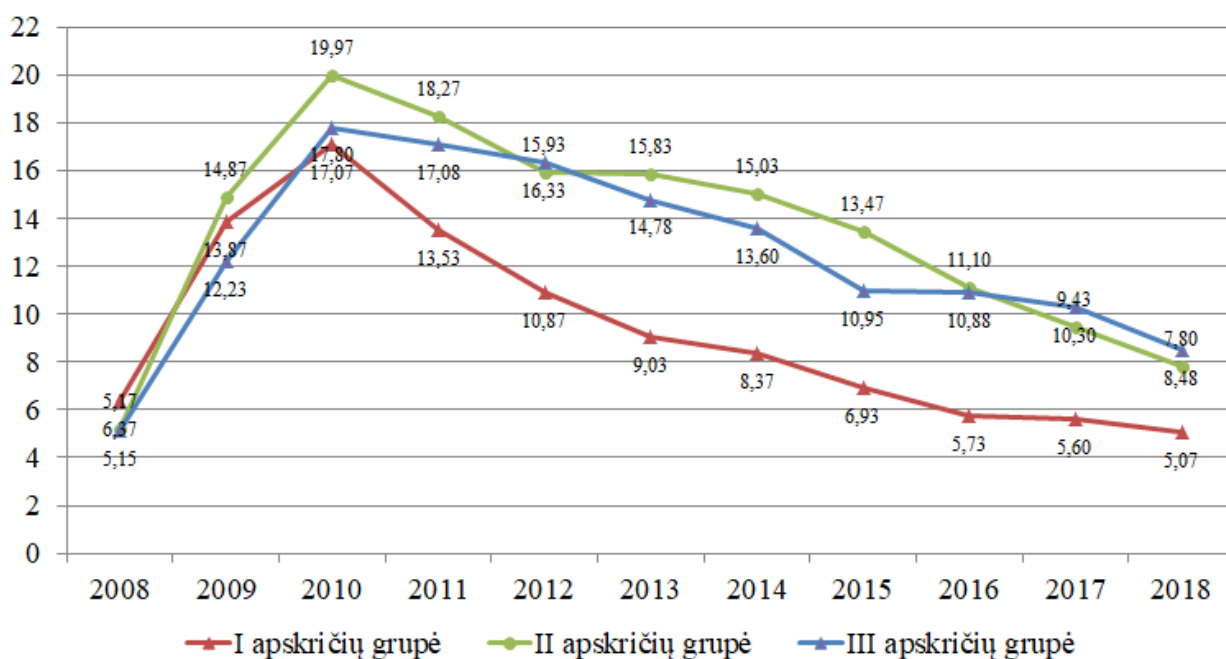
3.2 Kintamųjų apžvalga

Šioje darbo dalyje, pasitelkiant aprašomosios statistikos duomenų sisteminimo ir grafinio vaizdavimo metodus, bus apžvelgta nedarbo ir jį apibūdinančių veiksnių raida tiriamu 2008-2018 m. laikotarpiu. Tam, kad analizė būtų informatyvesnė, A. Pocius ir L. Neverauskienė (LSTC) Lietuvos apskritis siūlo skirstyti į atskiras grupes pagal užimtųjų skaičiaus rodiklį. [8] Pirmąją regionų grupę sudaro didžiausios šalies apskritys (Vilniaus, Kauno, Klaipėdos), antrąją – vidutinės (Alytaus, Panevėžio, Šiaulių), trečiąją – likusios mažosios apskritys (Marijampolės, Tauragės, Telšių ir Utenos).

3.2.1 Nedarbo raida apskričių grupėse

2008 metais pradėjus smukti realiai ekonomikai, dalis žmonių neteko pastovaus darbo. Dėl šios priežasties nedarbo rodikliai smarkiai šovė į viršų, o didžiausią reikšmę nedarbas pasiekė 2010 m., kai jis Lietuvos Respublikos mastu sudarė 17,8 proc. Sekantis grafinis paveikslas vaizduoja nedarbo lygio kaitą apskričių grupėse 2008-2018 m. laikotarpiu.

1 pav.: Nedarbo lygis apskričių grupėse 2008-2018 m. (proc.)



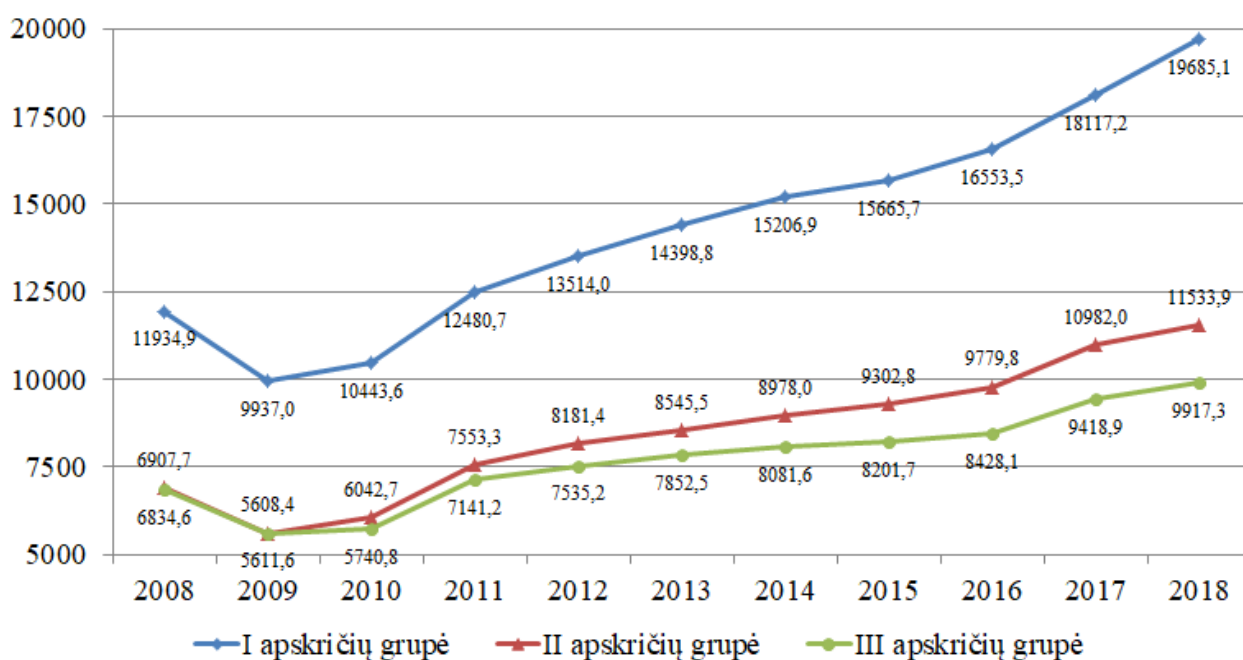
Šaltinis: autorės skaičiavimai remiantis Lietuvos Statistikos departamento duomenimis.

Kaip matome iš grafinio paveikslo, krizės laikotarpio pradžioje Lietuvos apskritys pasižymėjo labai panašiais nedarbo lygio rodikliais. Tačiau jau nuo 2011 metų atsirado ryški didžiųjų ir mažesniųjų regionų diferenciacija. Didžiausias atotrūkis buvo pasiektas 2013 metais, kuomet I-ąją ir II-ąją grupes skyrė 6,8 proc. nedarbas, I-ąją ir III-ąją – 5,74 proc. Vis dėlto, nedarbo lygis visose apskrityse turi tendenciją mažėti.

3.2.2 Nedarbui įtaką darančių veiksnių raida apskričių grupėse

BVP vienam gyventojui. Empiriškai pagrįstas A. M. Okun (1962) dėsnis teigia, kad galioja atvirkštinis ryšys tarp nedarbo lygio ir bendrojo vidaus produkto: nedarbui didėjant, bendrasis vidaus produktas mažėja. Kaip matysime sekančiame BVP vienam gyventojui rodiklio raidos grafiniame paveiksle, panaši tendencija stebėtu 2008-2018 m. laikotarpiu pasireiškė ir Lietuvos apskritims.

2 pav.: Bendrasis vidaus produktas vienam gyventojui apskričių grupėse 2008-2018 m. (EUR)



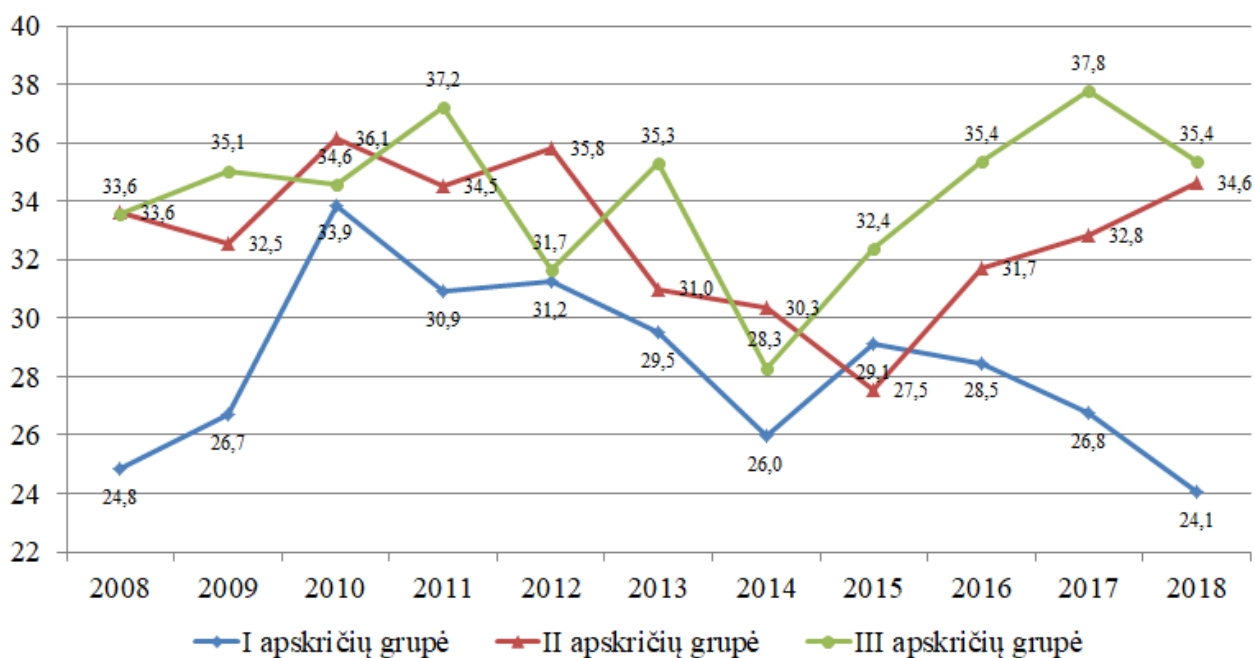
Šaltinis: autorės skaičiavimai remiantis Lietuvos Statistikos departamento duomenimis.

Iš grafiko matyti, kad 2008-2018 m. laikotarpiu Lietuvoje BVP vienam gyventojui rodiklis, išskyrus 2009-uosius, sulig kiekvienais metais vis didėjo. Diferenciacija tarp didžiųjų

ir mažesniųjų apskričių yra akivaizdi: 2018 metų duomenimis I-osios grupės BVP vienam gyventojui rodiklis buvo 1,7 kartus didesnis už II-osios grupės rodiklį ir net 1,98 didesnis už III-iosios grupės rodiklį. II-osios ir III-iosios grupių BVP vienam gyventojui rodikliai iki krizės pradžios buvo labai panašūs, tačiau bėgant metams atskirtis ryškėjo.

Skurdo lygis. Lietuvos Statistikos departamento duomenys rodo, kad įsibėgėjus krizei, kuomet dalis žmonių prarado darbus, šoktelėjo ir skurdo rodiklis – asmenų, gyvenančių skurdo rizikoje ar socialinėje atskirtyje (trumpiau, „skurdo lygis“), lyginant su 2008 m., 2010 m. Lietuvos Respublikos mastu padaugėjo 5,7 proc. Ekonominė krizė paveikė visas šalies apskritis – atotrūkis tarp apskričių skurdo rodiklių buvo labai nežymus. Kita vertus, kaip matysime sekančiame grafiniame paveiksle, didžiosios šalies apskritys, lyginant su mažesnėmis, atsigavo žymiai greičiau. 2015-2018 m. laikotarpiu stebėjome smarkiai augančius II-osios ir III-iosios apskričių grupių skurdo rodiklius, kai tuo tarpu didesnės apskritys pasižymėjo geresnėmis gyvenimo sąlygomis.

3 pav.: Skurdo lygis apskričių grupėse 2008-2018 m. (proc.)



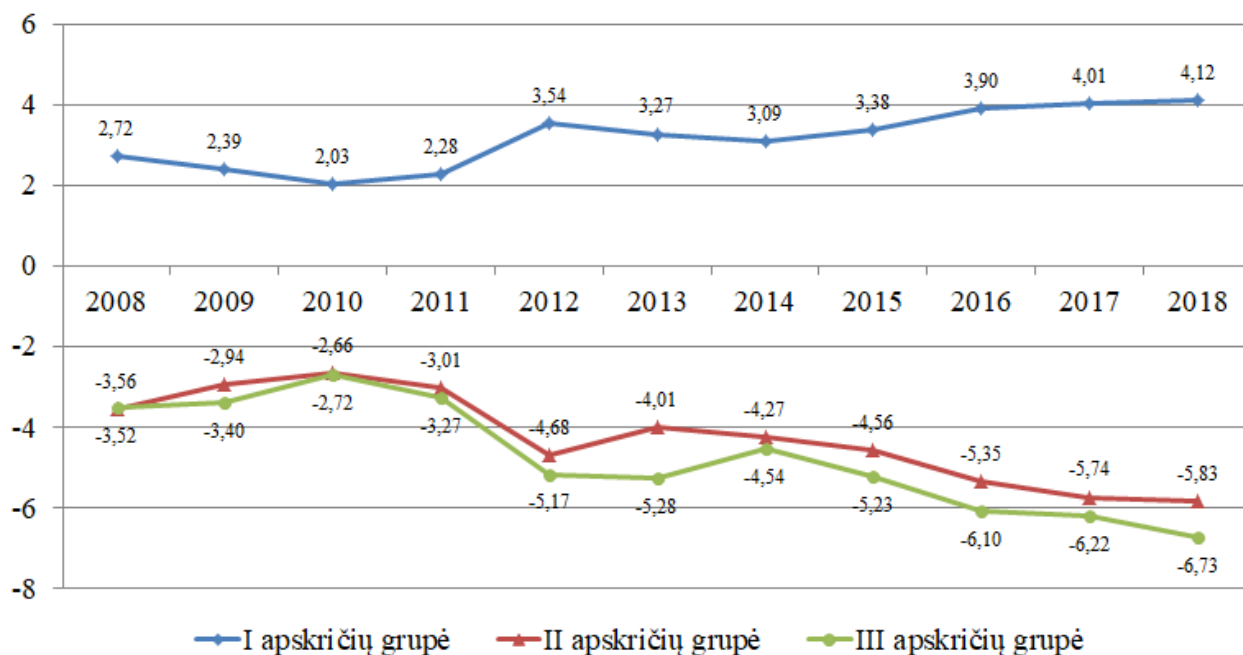
Šaltinis: autorės skaičiavimai remiantis Lietuvos Statistikos departamento duomenimis.

Prieš tiriant skurdo ir nedarbo lygio priklausomumą, svarbu pažymėti, kad nors čia galioja priežasties – pasekmės dėsnis (praradę darbus, žmonės labiau skursta), skurdo padėtis

šalyje ypatingai priklauso ir nuo valstybės aparato funkcionavimo – kuo geriau veikia skurdo mažinimo politika, tuo galima pasiekti geresnių rezultatų šalies socialinės gerovės vystymo procese. Todėl tikėtina, kad tiesinis ryšys tarp skurdo ir nedarbo lygio nebus reikšmingas.

Vidaus migracijos saldo. Stebėtu 2008-2018 m. laikotarpiu vidaus migracijos saldo 1000-čiui gyventojų rodiklis buvo teigiamas tik didžiosiose Lietuvos apskrityse, kitose jis kiekvienais metais buvo neigiamas. Vadinasi, dėl teigiamos vidinės migracijos didieji šalies regionai išlošė, galėdami sparčiau vystyti darbo paklausos ir pasiūlos santykius. Mažosios apskritys, tuo tarpu, susidūrė su darbo jėgos mažėjimo problema. Sekantis grafikas iliustruoja 2008-2018 metų vidaus migracijos saldo 1000-čiui gyventojų raidą išskirtose apskričių grupėse.

4 pav.: Vidaus migracijos saldo 1000-čiui gyventojų apskričių grupėse 2008-2018 m. (asmenys)



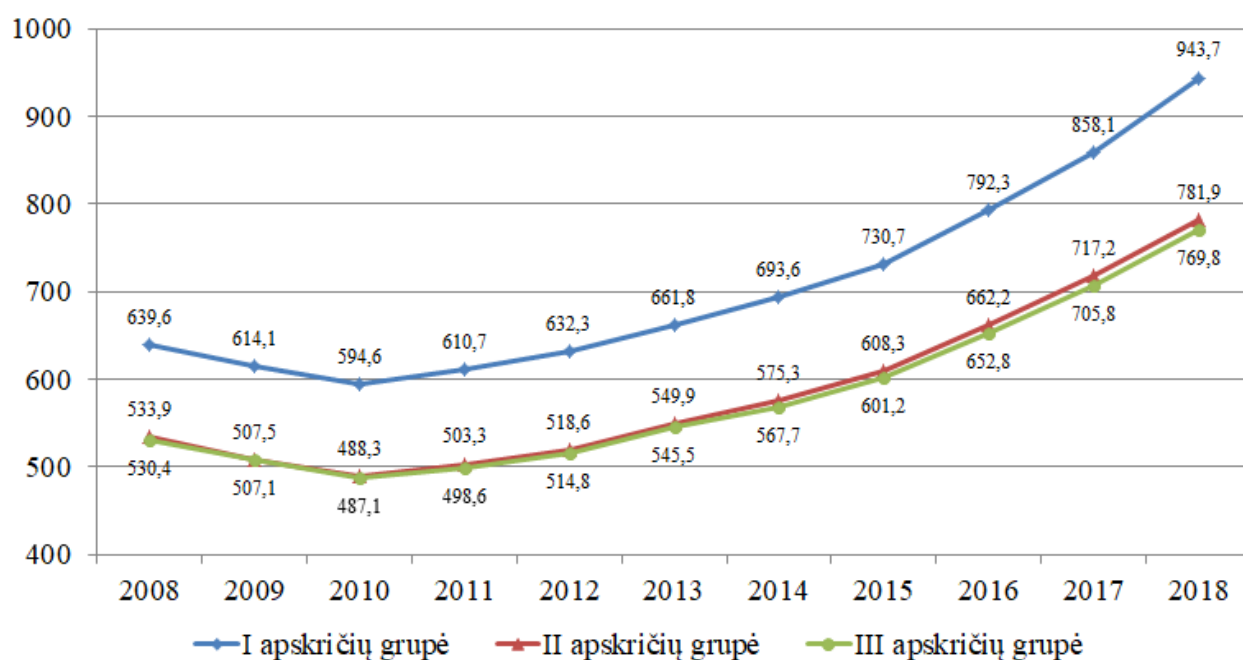
Šaltinis: autorės skaičiavimai remiantis Lietuvos Statistikos departamento duomenimis.

Pažvelgus į grafiką, matome beveik asimetrinį vaizdą, lyginant I-ąją apskričių grupę su II-ąja ir III-ąja grupėmis. Kuomet vidaus migracijos saldo rodiklis didžiosiose apskrityse augo, mažesnėse apskrityse jis atitinkamai mažėjo. Tad akivaizdu, kad gyventojai iš mažesniųjų apskričių migravo į didesnes, o nuo 2014 m. ši tendencija dar labiau suintensyvėjo.

Aptariant nedarbo ir migracijos procesų sąveiką, galima manyti, kad teritorijose, iš kurių išvykstama, nedarbo lygis turėtų mažėti (visgi, pagrindinė migravimo priežastis yra nedarbas). Tačiau dėl mažėjančios darbo pasiūlos teritorijose, į kurias atvykstama, galima ir priešinga situacija. Dėl to nedarbo ir migracijos sąryšis nėra vienareikšmis ir turi būti analizuojamas detalai. Kita vertus, vidinės (bet ne tarptautinės) migracijos mastai Lietuvos atveju nėra tokie dideli, kad koreguotų darbo pasiūlos ir paklausos procesus. Be to, kaip ir skurdo padėtis, vidaus migracijos procesai didele dalimi priklauso nuo valdžios veiksmų.

Mėnesinis bruto darbo užmokestis. Nagrinėjant bruto darbo užmokesčio kitimą 2008-2018 m. laikotarpiu, pastebėtina, kad nuo krizės įsibėgėjimo pradžios iki sunkiausių metų vadinamųjų 2010-ųjų, vidutinis darbo užmokestis krito. Vėliau šis rodiklis tik augo, o sparčiausias kilimas fiksuotas 2016-2018 metų laikotarpiu (Lietuvos Respublikos mastu tai buvo 8-9 proc. metinis augimas). Pažvelkime, kaip atrodė mėnesinio bruto darbo užmokesčio kitimas analizuojamu laikotarpiu apskričių grupėse.

5 pav.: Mėnesinis bruto darbo užmokestis apskričių grupėse 2008-2018 m. (Eur.)

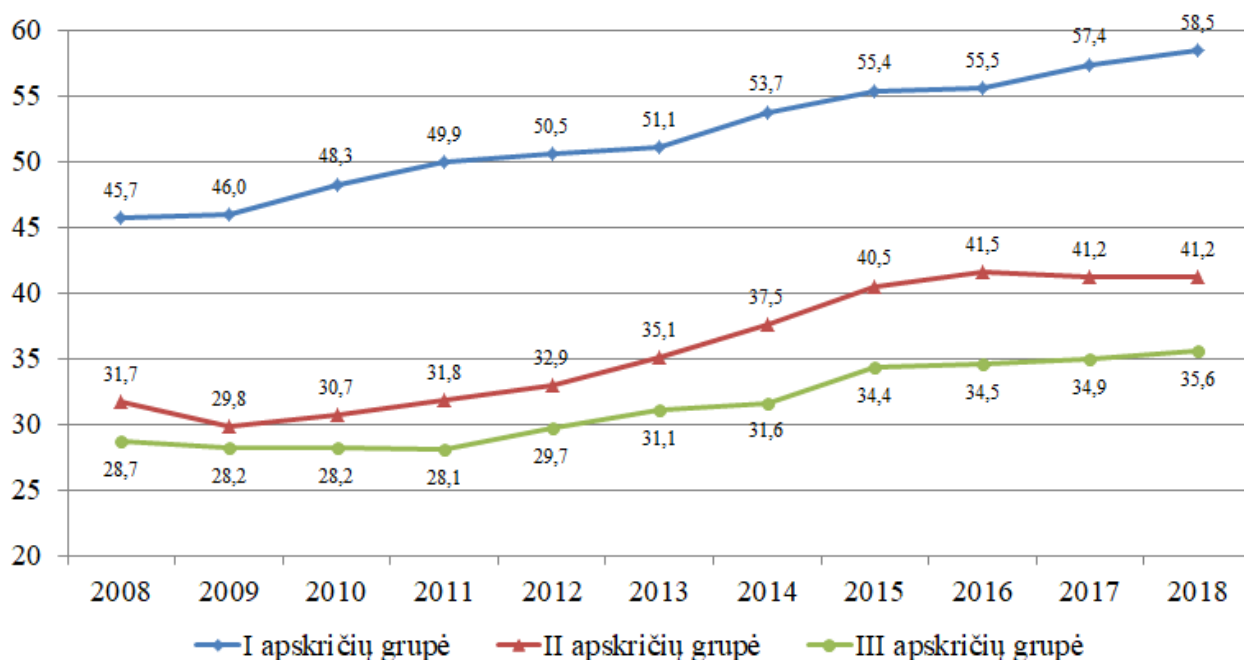


Šaltinis: autorės skaičiavimai remiantis Lietuvos Statistikos departamento duomenimis.

Iš grafiko matyti, kad yra ryški I-osios apskričių grupės diferenciacija su II-ąja ir III-iaja. Šis atotrūkis tarp grupių turi tendenciją augti – jei 2008 metais I-ąją ir II-ąją grupes vidutiniškai skyrė 105,8 Eur darbo užmokesčio skirtumas, o I-ąją ir III-iąją – 109,2 Eur, tai 2018 metais šis skirtumas buvo atitinkamai 161,8 ir 173,9 Eur. II-oji ir III-ioji grupės 2008-2018 m. laikotarpiu pasižymėjo labai panašiu darbo užmokesčio dydžiu, tačiau paskutiniaisiais metais skirtumas dar grupių nežymiai padidėjo.

Aukštojo išsilavinimo lygis. Sekančiame grafiniame paveiksle pažvelkime, kaip kito aukštojo išsilavinimo lygis apskričių grupėse 2008-2018 m. laikotarpiu.

6 pav.: Aukštojo išsilavinimo lygis apskričių grupėse 2008-2018 m. (proc.)



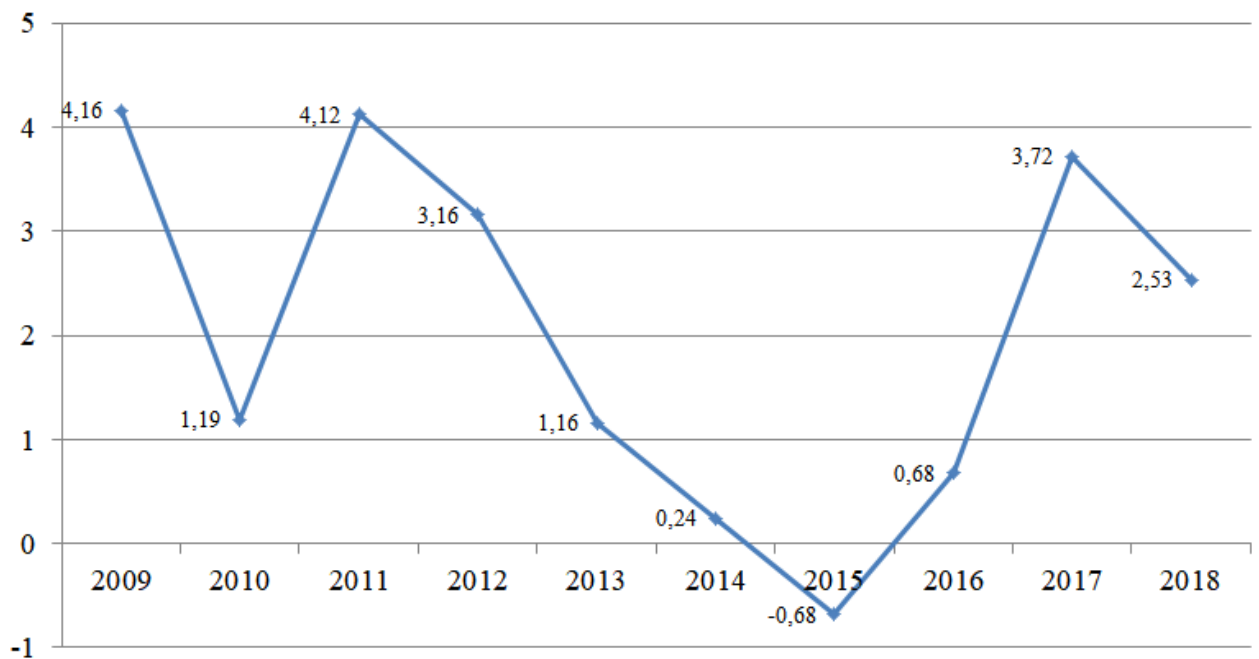
Šaltinis: autorės skaičiavimai remiantis Lietuvos Statistikos departamento duomenimis.

Aukštojo išsilavinimo padėtis šalyje tiriamu laikotarpiu sulig kiekvienais metais vis gerėjo: per dešimt metų nuo 2008-ųjų aukštą išsilavinimą turinčių žmonių Lietuvos Respublikos mastu padaugėjo net 10 proc. Natūralu, kad didžiosios apskritys visu stebėtu laikotarpiu pasižymėjo geresniu išsilavinimo lygiu, kadangi didesnės apskritys pasižymi aukštesne švietimo paslaugų kokybe. O tai tampa viena iš priežasčių, kodėl nemaža dalis gyventojų iš mažesniųjų apskričių migruoja į didžiuosius regionus.

Išsilavinimo ir nedarbo lygio rodiklių ryšys teorine prasme yra gana akivaizdus – geresnį išsilavinimą įgiję žmonės turi didesnes galimybes susirasti norimą darbą. Todėl didėjant išsilavinimo lygiui, nedarbas turi tendenciją mažėti.

Infliacijos lygis. Infliacijos ir nedarbo sąveiką teoriškai galima paaikškinti Filipo kreive – kuo mažesnis infliacijos lygis, tuo didesnis nedarbo lygis, ir atvirkščiai. [10] Lietuvos infliacijos lygio rodiklis, kaip minėta duomenų apraše, yra skaičiuojamas bendrai šalies mastu. Pažvelkime, kaip atrodė infliacijos pokyčiai 2009-2018 m. laikotarpiu.

7 pav.: Infliacijos lygis Lietuvos Respublikoje 2009-2018 m. (proc.)



Šaltinis: sudaryta darbo autorės remiantis Lietuvos Statistikos departamento duomenimis.

Iš grafiko matyti, kad analizuojamu laikotarpiu, išskyrus 2015-uosius metus, bendras kainų lygis Lietuvoje kilo. 2015 m. šalyje pasireiškė priešingas reiškinys inflacijai – defliacija, kuri buvo sąlygota Lietuvoje įvesto euro.

3.3 Koreliacinė analizė

Atliekant koreliacinę analizę, buvo siekta dviejų tikslų: 1) nustatyti, kokio stiprumo koreliacinis ryšys sieja priklausomą kintamąjį – nedarbo lygį – su nepriklausomais kintamaisiais – ekonominiais ir socialiniais rodikliais; 2) ištirti nepriklausomų kintamųjų tarpusavio koreliacinius ryšius.

Suskaičiuotos porinės Pirsono koreliacijos parodė, kad statistiškai reikšmingas ryšys sieja nedarbo lygį su šiais rodikliais: infliacijos lygis (koreliacijos reikšmė $\rho = 0.24$), BVP vienam gyventojui ($\rho = -0.67$), vidaus migracijos saldo ($\rho = -0.25$), mėnesinis bruto darbo užmokestis ($\rho = -0.75$) ir aukštojo išsilavinimo lygis ($\rho = -0.30$). Sekančioje lentelėje pateikta Pirsono metodu skaičiuotos porinės koreliacinės analizės rezultatų suvestinė.

2 lentelė: Pirsono porinės koreliacinės analizės rezultatų suvestinė

	ρ	p -reikšmė	Lauktas poveikis	Gautas poveikis	Reikšmingumas
Infliacijos lygis	0.24	0.0145	Neigiamas	Teigiamas	Reikšminga
BVP vienam gyv.	-0.67	0.0000	Neigiamas	Neigiamas	Reikšminga
Vidaus migracijos saldo	-0.25	0.0106	Neigiamas	Neigiamas	Reikšminga
Skurdo lygis	0.12	0.2477	Teigiamas	Teigiamas	Nereikšminga
Bruto darbo užmokestis	-0.75	0.0000	Neigiamas	Neigiamas	Reikšminga
Aukštojo išsilav. lygis	-0.30	0.0021	Neigiamas	Neigiamas	Reikšminga

Šaltinis: sudaryta darbo autorės.

Pirsono porines koreliacijas galima interpretuoti tokiu būdu:

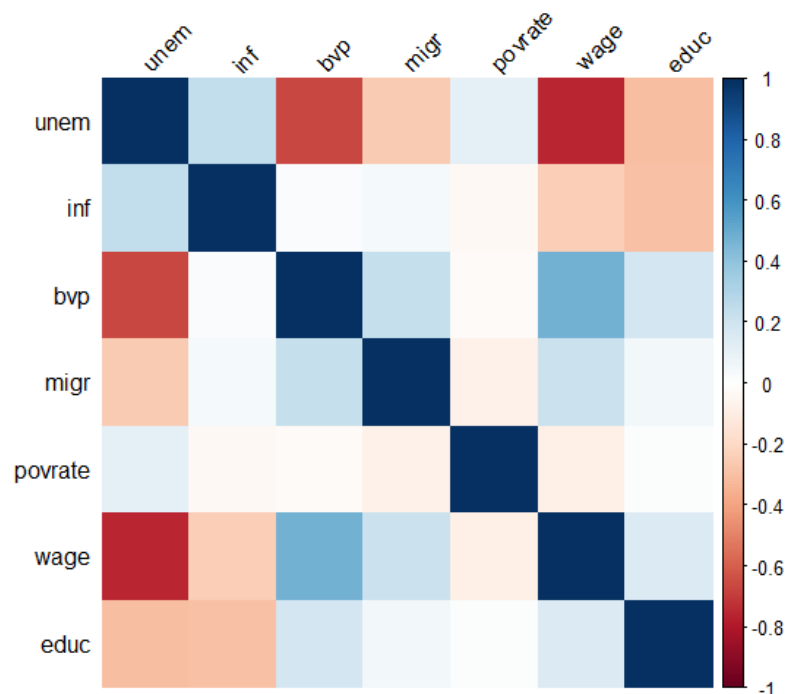
- Logikai prieštaraujantis teigiamas koreliacinis ryšys nustatytas tarp nedarbo ir infliacijos lygio rodiklių. Teoriškai, didėjant infliacijai, nedarbas turėtų mažėti, o infliacijai traukiantis – nedarbas turėtų augti.
- BVP vienam gyventojui ir mėnesinio bruto darbo užmokesčio porinės koreliacijos su nedarbo lygiu pasirodė reikšmingiausios. Neigiamos koreliacinės priklausomybės reiškia, kad kuo didesnis BVP vienam gyventojui rodiklis ir kuo aukštesnis bruto darbo užmokestis, tuo nedarbo lygis mažesnis.
- Nevienareikšmis ryšys gali sieti vidaus migracijos saldo ir nedarbo lygio rodiklius. Kaip matome iš Pirsono porinės koreliacinės analizės rezultatų, Lietuvos atveju vidaus migracijos saldo teigiamai veikia nedarbo rodiklį: augant migracijos saldo (atvykstant

darbo jėgai), nedarbo lygis mažėja. Ši neigiama koreliacinė priklausomybė yra labai silpna, tačiau statistiškai reikšminga.

- Skurdo ir nedarbo lygio koreliacijos ryšys labai silpnas ir statistiškai nereikšmingas. Kaip ir buvo tikėtasi, ryšys tarp kintamųjų yra teigiamas – didėjant skurdo lygiui, nedarbo lygis taip pat auga.
- Silpno stiprumo neigiamas koreliacinis ryšys sieja nedarbą ir aukštojo išsilavinimo lygį. Tai reiškia, kad didėjant aukštąjį mokslą baigusių žmonių procentinei daliai, nedarbo lygis mažėja.

Pereinant prie nepriklausomų kintamųjų tarpusavio koreliacinių ryšių vertinimo, toliau buvo apskaičiuota bendra visų kintamųjų Pirsono koreliacinė matrica, kuriai iliustruoti nubraižytas sekantis grafinis paveikslas. Toks grafinis vaizdavimas leidžia vizualiai pastebėti, kokio stiprumo koreliaciniai ryšiai sieja kintamuosius. Čia dešinė skalė žymi koreliacinio ryšio stiprumą: kuo dviejų kintamųjų susikirtimo laukelyje spalva yra raudonesnė, tuo stipresnis neigiamas ryšys, ir atvirkščiai – kuo spalva mėlynėsnė, tuo stipresnis teigiamas ryšys.

8 pav.: Pirsono koreliacijos koeficientų grafikas



Šaltinis: sudaryta darbo autorės.

Sekančioje lentelėje pateiktos Pirsono koreliacijos koeficientų reikšmės. Čia lentelės pirmoje eilutėje ir pirmame stulpelyje surašyti kintamųjų žymėjimai, kurių paaiškinimai yra pateikti darbo apraše.

3 lentelė: Pirsono koreliacijos koeficientų reikšmės

	<i>unem</i>	<i>inf</i>	<i>bvp</i>	<i>migr</i>	<i>povrate</i>	<i>wage</i>	<i>educ</i>
<i>unem</i>	1.000	0.244	-0.668	-0.255	0.117	-0.754	-0.304
<i>inf</i>	0.244	1.000	0.023	0.042	-0.032	-0.244	-0.296
<i>bvp</i>	-0.668	0.023	1.000	0.238	-0.029	0.477	0.184
<i>migr</i>	-0.255	0.042	0.238	1.000	-0.070	0.219	0.058
<i>povrate</i>	0.117	-0.032	-0.029	-0.070	1.000	-0.084	0.019
<i>wage</i>	-0.754	-0.244	0.477	0.219	-0.084	1.000	0.155
<i>educ</i>	-0.304	-0.296	0.184	0.058	0.019	0.155	1.000

Šaltinis: sudaryta darbo autorės.

Koreliacinių priklausomybių statistiniam reikšmingumui nustatyti, sekančioje lentelėje pateiktos Pirsono koreliacijos koeficientų p -reikšmės.

4 lentelė: Pirsono koreliacijos koeficientų p -reikšmės

	<i>unem</i>	<i>inf</i>	<i>bvp</i>	<i>migr</i>	<i>povrate</i>	<i>wage</i>	<i>educ</i>
<i>unem</i>	0.000	0.014	0.000	0.011	0.248	0.000	0.002
<i>inf</i>	0.014	0.000	0.821	0.680	0.753	0.014	0.003
<i>bvp</i>	0.000	0.821	0.000	0.017	0.775	0.000	0.066
<i>migr</i>	0.011	0.680	0.017	0.000	0.487	0.029	0.563
<i>povrate</i>	0.248	0.753	0.775	0.487	0.000	0.406	0.849
<i>wage</i>	0.000	0.014	0.000	0.029	0.406	0.000	0.123
<i>educ</i>	0.002	0.003	0.066	0.563	0.849	0.123	0.000

Šaltinis: sudaryta darbo autorės.

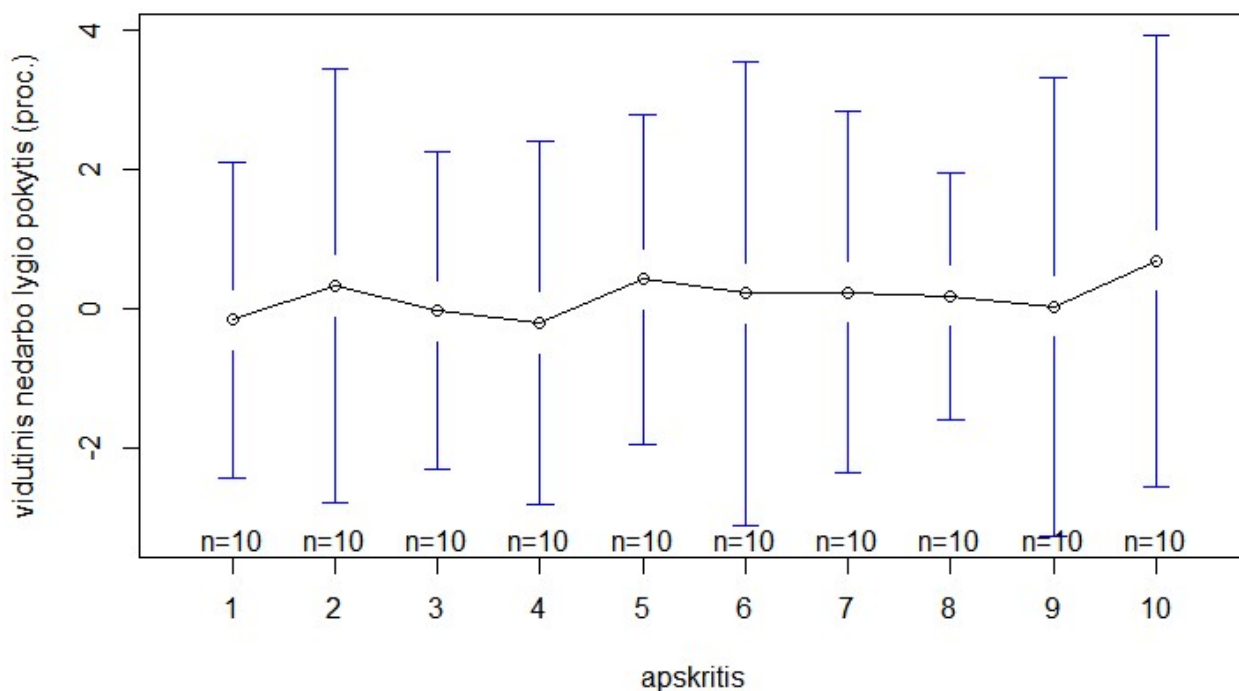
Iš 4 lentelės matyti, kad egzistuoja statistiškai reikšmingi koreliaciniai sąryšiai tarp nepriklausomų kintamųjų. Sudarant regresijos modelį atsižvelgti į nepriklausomų kintamųjų tarpusavio koreliuotumą yra būtina, kadangi tai signalizuoja, kad galimai susiduriama su multikolinearumo problema.

3.4 Ekonometrinė duomenų analizė

Koreliacinės analizės pagalba buvo apskaičiuotos tik porinės kintamųjų koreliacijos. Siekiant ištirti bendrą nepriklausomų kintamųjų poveikį priklausomam kintamajam, šiame skyriuje bus sudaromi regresijos modeliai, kurie remiasi visų kintamųjų kovariacine, taigi ir koreliacine matrica. Statistinės išvados, gautos remiantis regresine analize, yra ženkliai patikimesnės, negu išvados, kurios remiasi tik porinėmis koreliacijomis.

Regresinei analizei pradėti pirmiausiai buvo patikrinta modeliuojamo kintamojo – nedarbo lygio – heterogeniškumo savybė atskirai tarp apskričių ir tarp laiko momentų. Žemiau pateiktame grafiniame paveiksle išbrėžtos kiekvienos apskrities vidutinės nedarbo lygio pokyčių reikšmės 2009-2018 m. laikotarpiu.

9 pav.: Heterogeniškumas tarp apskričių



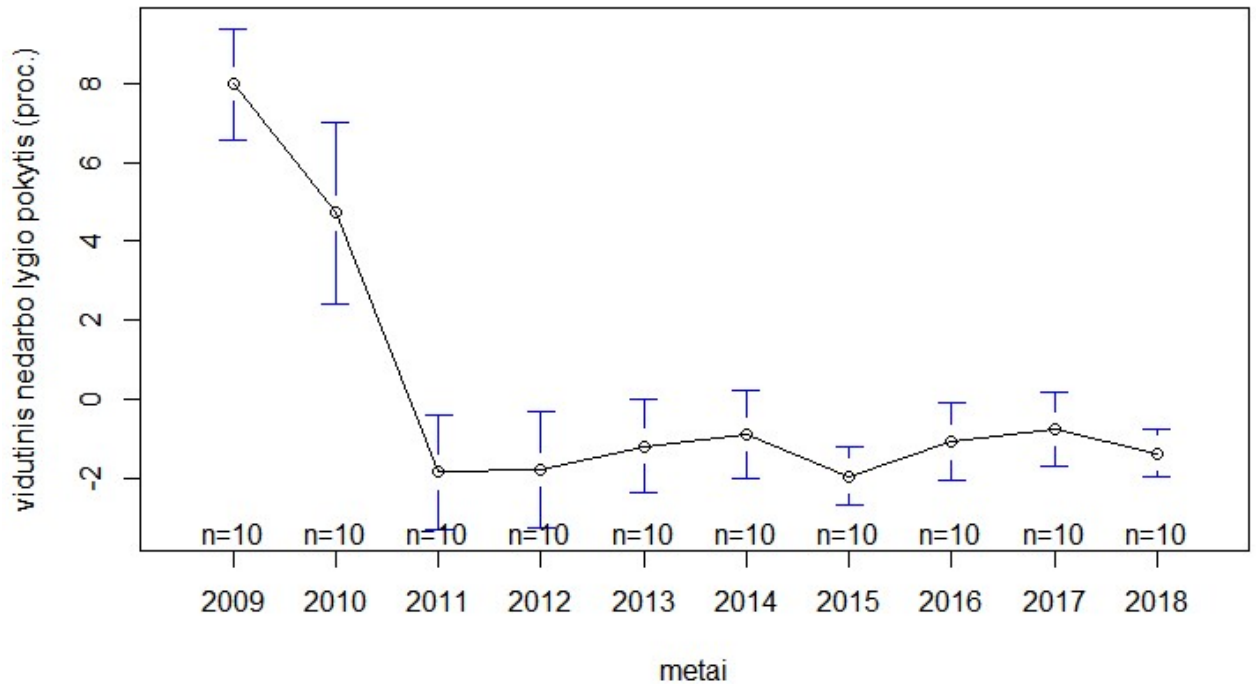
Šaltinis: sudaryta darbo autorės.

Iš grafiko matyti, kad Lietuvos apskritys pasižymi pakankamai panašiomis vidutinio nedarbo lygio pokyčių reikšmėmis. Kiek aukštesni vidutiniai nedarbo lygio pokyčiai stebėti Alytaus (apskrities nr. 2), Marijampolės (nr. 5) ir Utenos (nr. 10) apskrityse. Mažesni vidutiniai nedarbo pokyčiai fiksuoti didžiosiose Lietuvos apskrityse – Vilniaus, Kauno ir

Klaipėdos (atitinkamai, nr. 1, nr. 3 ir nr. 4).

Toliau pažvelkime, kaip heterogeniškumo savybė pasireiškia tarp stebėtų laiko momentų $t = 2009, \dots, 2018$. Sekantis grafikas iliustruoja vidutinės nedarbo lygio pokyčių reikšmes visoms apskritims pamečiui.

10 pav.: Heterogeniškumas tarp laiko momentų



Šaltinis: sudaryta darbo autorės.

Grafike aiškiai matyti ekonominio nuosmukio įtaka nedarbo lygio svyravimams: Lietuvos Respublikos mastu vidutinis nedarbo lygis 2009 m. augo net 7,99 proc., 2010 m. – 4,72 proc. Šios dvi išsiskiriančios reikšmės sąlygoja duomenų nestacionarumą ir, kaip matysime tolimesnėje analizėje, sukelia problemų parenkant analizuojamus duomenis aprašantį modelį.

3.4.1 Tiesinis regresijos modelis

Regresinio modeliavimo procese siekta priklausomą kintamąjį – nedarbo lygį – paaiškinti nepriklausomais kintamaisiais – metine infliacija, BVP vienam gyventojui, vidaus migracijos saldo 1000-čiui gyventojų rodikliu, skurdo lygiu, mėnesiniu bruto darbo užmokesčiu ir aukštojo išsilavinimo lygiu. Regresinei analizei buvo naudoti visi 100 stebėjimų (10 apskričių, 10

metų), sujungti į vieną didelę duomenų matricą. Žemiau pateikta transformuotų duomenų santraukos lentelė, čia *apskr* žymi apskrities numerį, *year* – metus.

5 lentelė: Transformuotų duomenų santrauka

<i>apskr</i>	<i>year</i>	<i>unem</i>	<i>inf</i>	<i>bvp</i>	<i>migr</i>	<i>povrate</i>	<i>wage</i>	<i>educ</i>
1	2009	8.0	4.16	-8.51	-2.30	1.7	-2.15	0.64
1	2010	2.0	1.19	1.93	-14.16	4.0	-1.01	2.61
1	2011	-1.9	4.12	6.22	9.83	1.8	1.12	0.92
...	...	...	...	...	...	...	...	...
10	2017	1.2	3.72	2.97	-7.21	-0.8	3.18	-1.14
10	2018	-2.6	2.53	1.33	-2.89	4.4	3.39	3.28

Šaltinis: sudaryta darbo autorės.

Pastovaus laisvojo nario modelis. Sudarinėjant pastovaus laisvojo nario regresijos modelį Lietuvos apskričių duomenims, laikytasi prielaidos, kad visos apskritys yra homogeniškos, t.y. pasižymi statistiškai reikšmingai nesiskiriančiais rodikliais. Sudarytasis modelis ir mažiausių kvadratų metodu (MKM) apskaičiuotų koeficientų standartinės paklaidos pateiktos sekančioje lygtyje.

$$\begin{aligned}
 \Delta unem = & 2.21 + 0.21inf - 0.33\Delta \ln(bvp) - 0.01\Delta \ln(migr) + 0.05\Delta povrate - \\
 & (0.43) \quad (0.13) \quad (0.05) \quad (0.02) \quad (0.04) \\
 & - 0.94\Delta \ln(wage) - 0.23\Delta educ \\
 & (0.12) \quad (0.11)
 \end{aligned}$$

Šio modelio determinacijos koeficientas $R^2 = 0.72493$ rodo, kad nepriklausomų kintamųjų reikšmės suformuoja maždaug 72,5 proc. priklausomojo kintamojo – nedarbo lygio – dispersijos. Koreguoto determinacijos koeficiento $\hat{R}^2$ reikšmė yra 0.70718. Modelio F statistikos reikšmingumą nusakanti p -reikšmė $< 2.22e-16$, kas reiškia, kad sudarytas pastovaus laisvojo nario modelis yra adekvatus aprašomiems duomenims. Statistiškai reikšmingi modelio parametrai yra šie: laisvasis narys (statistinis reikšmingumas $p = 0,0013$), BVP vienam gy-

ventojui ($p = 0,000025$), bruto darbo užmokestis ($p = 0,00000001$) ir aukštojo išsilavinimo lygis ($p = 0.0364$).

Sekančiame žingsnyje atliktas VIF testas nustatė, kad multikolinearumo problemos duomenyse nėra. VIF testo rezultatai kiekvienam nepriklausomam kintamajam surašyti žemiau pateiktoje lentelėje.

6 lentelė: VIF testo rezultatai

	<i>inf</i>	<i>bvp</i>	<i>migr</i>	<i>povrate</i>	<i>wage</i>	<i>educ</i>
VIF testo reikšmė	1.209	1.402	1.087	1.014	1.454	1.143

Šaltinis: sudaryta darbo autorės.

Iš lentelės matyti, kad VIF testo reikšmės kiekvienam kintamajam yra mažesnės už leistiną normą 4. Vadinas, galima daryti išvadą, kad statistiškai reikšmingos koreliacijos tarp nepriklausomų kintamųjų nenustatyta ir į regresijos modelį galima įtraukti visus turimus regresorius. Ši išvada galioja ir tolimesniems sudarytiems regresijos modeliams.

Kadangi pastovaus laisvojo nario regresijos modelis paneliniams duomenims gali būti suvokiamas kaip atskiras klasikinės tiesinės regresijos atvejis, šiam modeliui galima taikyti automatines kintamųjų parinkimo procedūras. Sudarytam pastovaus laisvojo nario modeliui buvo taikytos automatinės geriausio poaibio ir žingsninė procedūros. Abi procedūros geriausiu modeliu parinko iš trijų statistiškai reikšmingiausių kintamųjų kombinacijos. Naujasis modelis užrašomas šia forma:

$$\Delta unem = 2.76 - 0.32\Delta \ln(bvp) - 1.02\Delta \ln(wage) - 0.28\Delta educ,$$

(0.27) (0.05) (0.12) (0.10)

determinacijos koeficientas $R^2 = 0.7123$, koreguotas determinacijos koeficientas $\hat{R}^2 = 0.7033$. Šio modelio koeficientų reikšmės ir ženklai gali būti interpretuojami tokiu būdu:

- BVP vienam gyventojui, mėnesinio bruto darbo užmokesčio ir aukštojo išsilavinimo lygio rodiklių pokyčiams esant nuliniais, nedarbo lygis vidutiniškai padidėja 2,76 proc.;
- BVP vienam gyventojui padidėjus 1 proc., nedarbo lygis sumažėja 0,32 proc.;

- Mėnesiniam bruto darbo užmokesčiui padidėjus 1 proc., nedarbo lygis sumažėja 1,02 proc.;
- Aukštojo išsilavinimo lygiui padidėjus 1 proc., nedarbo lygis sumažėja 0,28 proc.

Fiksuotų efektų modelis. Skirtingai nei pastovaus laisvojo nario, fiksuotų efektų panelinis regresijos modelis atsižvelgia į skirtumus tarp panelės elementų. Buvo sudaryti dviejų tipų – fiksuotų individualių ir fiksuotų laiko efektų paneliniai regresijos modeliai. Gauta, kad fiksuotų individualių efektų regresijos modelio įverčiai mažai skiriasi nuo pastovaus laisvojo nario modelio įverčių, pastarasis modelis užrašomas šia regresijos lygtimi:

$$\begin{aligned} \Delta unem = \alpha_i + 0.21inf - 0.34\Delta \ln(bvp) - 0.01\Delta \ln(migr) + 0.05\Delta povrate - \\ (0.14) \quad (0.06) \quad (0.02) \quad (0.04) \\ - 0.95\Delta \ln(wage) - 0.22\Delta educ, \\ (0.13) \quad (0.12) \end{aligned}$$

determinacijos koeficientas $R^2 = 0.72965$, koreguotas determinacijos koeficientas $\hat{R}^2 = 0.68138$. Fiksuoti **individualūs efektai** α_i kiekvienai apskričiai $i = 1, \dots, 10$ pateikti sekančioje lentelėje.

7 lentelė: Fiksuoti individualūs efektai

	1	2	3	4	5	6	7	8	9	10
įvertis	1.93	2.15	2.35	1.98	2.47	2.35	2.56	2.64	1.78	1.93

Šaltinis: sudaryta darbo autorės.

Visi šie apskričių individualūs efektai yra statistiškai reikšmingi, t.y. jų p -reikšmės mažesnės už reikšmingumo lygmenį $\alpha = 0.05$. Visgi, kaip jau pastebėjome iš heterogeniškumo tarp Lietuvos apskričių (9 pav.) ir tarp laiko momentų (10 pav.) grafinių paveikslų, būtent laiko, o ne apskričių, efektai yra ryškesni. Taigi, sekančiame žingsnyje buvo įvertintas

fiksuotų laiko efektų modelis, kuris yra užrašomas šia forma:

$$\begin{aligned} \Delta unem = & \alpha_t - 0.30\Delta \ln(bvp) + 0.02\Delta \ln(migr) + 0.003\Delta povrate - 0.58\Delta \ln(wage) - \\ & (0.16) \qquad (0.02) \qquad (0.04) \qquad (0.53) \\ & - 0.15\Delta educ, \\ & (0.10) \end{aligned}$$

determinacijos koeficientas $R^2 = 0.14218$, koreguotas determinacijos koeficientas $\hat{R}^2 = 0.000886$. Matyti, kad fiksuotų laiko efektų metodas neįtraukė į modelį infliacijos lygio kintamojo. Taip yra todėl, kad infliacijos rodiklis yra matuojamas visos šalies mastu, taigi, tarp apskričių jis nesiskiria. Lyginant su jau sudarytais pastovaus laisvojo nario ir fiksuotų individualių efektų modeliu, fiksuotų laiko efektų modelis neturi nei vieno statistiškai reikšmingo kintamojo, jo ypač žemas determinacijos koeficientas. Kita vertus, modelio koeficientų ženklai, išskyrus prie kintamojo „vidaus migracijos saldo“, yra tokie patys kaip ir prieš tai sudarytų modelių. F statistika su p -reikšme 0.02105 priima alternatyvią hipotezę, kuri teigia, kad sudarytasis modelis yra statistiškai reikšmingas (t.y. adekvatus).

Fiksuoti **laiko efektai** α_t kiekvienam laiko momentui $t = 2009, \dots, 2018$ pateikti sekančioje lentelėje.

8 lentelė: Fiksuoti laiko efektai										
	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
įvertis	4.38	4.59	1.63	-0.21	0.95	1.06	0.09	1.51	2.63	1.68

Šaltinis: sudaryta darbo autorės.

Iš aukščiau sudarytos lentelės matyti, kad fiksuoti laiko efektai tarpusavyje stipriai skiriasi. Statistiškai reikšmingi yra tik 2009 ir 2010 metų efektai.

F testas. Pastovaus laisvojo nario ir fiksuotų efektų panelinės regresijos modeliams palyginti buvo taikytas statistinis F testas. Testas buvo atliktas du kartus – pastovaus laisvojo nario metodu sudarytą modelį lyginant, pirma, su įvertintu fiksuotų individualių efektų modeliu, antra, su įvertintu fiksuotų laiko efektų modeliu. Testo statistika su p -reikšme = 0.9921

priėmė nulinę hipotezę, kad fiksuoti individualūs efektai yra statistiškai nereikšmingi, ir su p -reikšme = 0.0002 atmetė statistiškai nereikšmingų laiko efektų pasireiškimą duomenyse.

Vadinasi, iš iki šiol sudarytų modelių, kol kas geriausiai pasirodė fiksuotų laiko efektų panelinis regresijos modelis. Sekantis žingsnis yra patikrinti, ar duomenyse pasireiškiantys laiko efektai iš tiesų yra fiksuoti, ar verčiau yra atsitiktinio pobūdžio.

Atsitiktinių laiko efektų modelis. Šiuo metodu sudarytame modelyje laiko momentai $t = 2009, \dots, 2018$ traktuojami kaip atsitiktiniai parametrai. Gautas toks modelis:

$$\begin{aligned} \Delta unem = & 2.11 + 0.22inf - 0.33\Delta \ln(bvp) - 0.003\Delta \ln(migr) + 0.02\Delta povrate - \\ & (0.60) \quad (0.19) \quad (0.08) \quad (0.02) \quad (0.04) \\ & - 0.95\Delta \ln(wage) - 0.18\Delta educ, \\ & (0.17) \quad (0.10) \end{aligned}$$

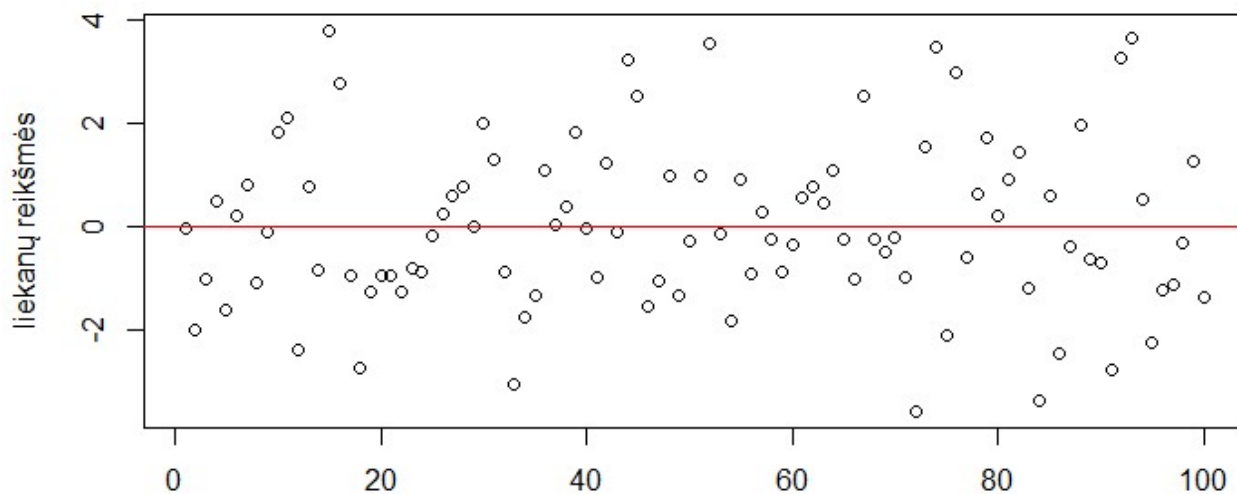
determinacijos koeficientas $R^2 = 0.55257$, koreguotas determinacijos koeficientas $\hat{R}^2 = 0.5237$. Sekantis žingsnis yra palyginti fiksuotų ir atsitiktinių laiko efektų metodais sudarytus panelinius regresijos modelius.

Hausman testas. Hausman testo p -reikšmė lygi $2.256e - 08$, kas reiškia, kad skirtumas tarp fiksuotų ir atsitiktinių laiko efektų metodais apskaičiuotų įverčių yra statistiškai reikšmingas. Tai indikuoja, kad atsitiktinių efektų modelio įverčiai yra paslinkti ir nesuderinti. Dėl šios priežasties turimiems duomenims aprašyti tinkamesnis fiksuotų laiko efektų metodas.

3.4.2 Liekanų analizė

Liekanų analizės metu buvo siekta patikrinti, ar Lietuvos apskričių duomenims sudaryto fiksuotų laiko efektų panelinio regresijos modelio liekanos yra stacionarios. Stacionarumo savybė reiškia, kad liekanų vidurkis statistiškai reikšmingai nesiskiria nuo nulio, o jų dispersija yra nekintanti laike (pastovi). Šiai savybei ištirti, sudarytas sekantis grafinis paveikslas, kuris vaizduoja modelio liekanų išsibarstymą. Čia raudona linija žymi liekanų vidurkį, kuris yra artimas nuliui.

11 pav.: Fiksuotų laiko efektų panelinio regresijos modelio liekanų sklaidos grafikas



Šaltinis: sudaryta darbo autorės.

Panašu, kad aukščiau pavaizduotos liekanos yra stacionarios – jos yra tolygiai išsibarsčiusios aplink savo vidurkį, dispersija neatrodo kintanti laike. Diagnostiniam patikrinimui buvo atliktas su Ljung-Box statistiniu testu, kuris su p -reikšme = 0.4121 neatmetė nulinės hipotezės. Tai reiškia, kad duomenims pritaikyto fiksuotų laiko efektų panelinio regresijos modelio liekanos atitinka joms keliamą stacionarumo reikalavimą.

4 Išvados ir rekomendacijos

Šiame bakalauro darbe, naudojant 2008-2018 m. laikotarpio panelinius apskričių metinius duomenis, buvo analizuota nedarbo lygio tiesinė priklausomybė nuo kitų ekonominių ir socialinių rodiklių. Pasiiekti tokie rezultatai:

- Kintamųjų apžvalgos metu, apskritis suskirstius į tris grupes pagal užimtųjų skaičiaus rodiklį, pastebėta ryški ekonominė ir socialinė grupių diferenciacija – didžiosios apskritys, lyginant su vidutinėmis ir mažosiomis, ekonominį sunkmetį išgyveno pastebimai lengviau ir stabiliau;
- Koreliacinės analizės pagalba apskaičiuotos Pirsono porinės kintamųjų koreliacijos parodė, kad didžiausią įtaką nedarbui darantys veiksniai yra BVP vienam gyventojui, mėnesinis bruto darbo užmokestis ir aukštojo išsilavinimo lygis;
- Problema, kurią pavyko identifikuoti regresinio modeliavimo dalyje – tai krizinio nuosmukio išbalansuoti nestacionarūs pokyčiai duomenyse. Nestacionarumas, kuris pasireiškė per laiko efektus, kėlė sunkumų sudarant duomenis aprašantį regresijos modelį.
- Nors atrodė, kad neparametrinį trendą visoms apskritims eliminuojantis fiksuotų laiko efektų panelinis regresijos modelis padės nustatyti tikruosius priklausomybės ryšius tarp kintamųjų, visi įvertinto modelio parametrai buvo statistiškai nereikšmingi. Taip pat, žemas determinacijos koeficientas leido suprasti, kad sudaryto modelio kintamieji nedarbo lygio pokyčius aprašo labai menkai. Nepaisant to, fiksuotų laiko efektų panelinis regresijos modelis yra adekvatus stebėtiems duomenims, jo liekanos stacionarios.

Tolimesniam darbo vystymui apskričių grupavimą tikslingiau būtų atlikti klasterinės analizės metodu, remiantis kiekvienos iš apskričių regresijos modelių kokybine informacija. Tokiu būdu būtų įvertintas apskričių panašumas pagal individualius jų regresijos modelių rezultatus. Regresinei analizei vertėtų pabandyti taikyti sudėtingesnes ekonometrinių modelių specifikacijas, pavyzdžiui, krizės sąlygotus rodiklių pokyčius laikant struktūriniu lūžiu duomenyse. Taip pat, patikimesniems rezultatams gauti tikslinga būtų analizuoti ilgesnio laikotarpio duomenis.

5 Literatūra

- [1] D. Asteriou, S. G. Hall. *Applied econometrics*, Palgrave Macmillan, 2007 m., p. 344-349
- [2] V. Čekanavičius, G. Murauskas. *Statistika ir jos taikymai, I dalis*, Vilnius: Leidykla TEV, 2000 m., p. 98
- [3] V. Čekanavičius, G. Murauskas. *Taikomoji regresinė analizė socialiniuose tyrimuose*, Vilniaus universitetas, 2014 m., p. 22
- [4] B. E. Hansen. *Econometrics*, University of Wisconsin, 2020 m., p. 784-785
- [5] C. Hsiao. *Analysis of panel data*, Cambridge University Press, 1986 m., p. 9
- [6] S. A. Martišius, V. Kėdaitis. *Statistika. I dalis. Statistinės analizės teorija ir metodai*, Vilniaus universitetas, 2013 m., p. 246
- [7] S. A. Martišius, V. Kėdaitis. *Statistika. II dalis. Statistinės išvados ir sprendimai*, Vilniaus universitetas, 2011 m., p. 311, 326
- [8] A. Pocius, L. Okunevičiūtė Neverauskienė. Lietuvos darbo rinkos regioninių rodiklių diferenciacija demografinių pokyčių kontekste. Iš *Darbo rinkos pokyčiai: problemos ir galimybės. Lietuvos socialinė raida Nr. 6*, Lietuvos socialinių tyrimų centras, 2017 m., p. 35
- [9] J. O. Rawlings, S. G. Pantula, D. A. Dickey. *Applied regression analysis: a research tool*, Springer, 1998 m., p. 372
- [10] E. Smith. *Modern labor economics: theory and public policy*, HarperCollins College Publishers, 1994 m., p. 602-604
- [11] J. H. Stock, M. W. Watson. *Introduction to econometrics*, Pearson Education, 2015 m., p. 356-364
- [12] J. M. Wooldridge. *Introductory econometrics: a modern approach*, Cengage Learning, 2013 m., p. 95, 97

[13] Lietuvos Statistikos departamentas (prieiga per internetą 2020.02.29:
<https://www.stat.gov.lt/>)

A Priedas

```
library(stats)
library(plm)
library(car)
library(leaps)
library(gplots)

# DUOMENŲ NUSKAITYMAS
duomenys_apskritys <- read.csv2("apskriciu_duomenys.csv")

# HETEROGENIŠKUMAS
plotmeans(unem ~ factor(apskr), xlab = "apskritis",
          ylab = "vidutinis nedarbo lygio pokytis (proc.)",
          data = duomenys_apskritys)
plotmeans(unem ~ year, xlab = "metai",
          ylab = "vidutinis nedarbo lygio pokytis (proc.)",
          data = duomenys_apskritys)

#####
#                KORELIACINĖ ANALIZĖ                #
#####
koreliacijos <- cor(duomenys_apskritys[,c(-1,-2)])
cor(duomenys_apskritys[,c(-1,-2)], method = "pearson")

# KORELIACIJŲ VAIZDINIS GRAFIKAS
source("http://www.sthda.com/upload/rquery_cormat.r")
require("corrplot")
corrplot(koreliacijos, order = "original",
        tl.col = "black", tl.srt = 45,
        method = "color")
```

```
#####
#      TIESINIS REGRESIJOS MODELIS PANELINIAMS DUOMENIMS      #
#####

paneliniai_apskritys <- pdata.frame(duomenys_apskritys,
                                     index = c("apskr", "year"),
                                     drop.index = TRUE, row.names = TRUE)

# PASTOVAUS LAISVOJO NARIO MODELIS
pooled <- plm(unem ~ inf + bvp + migr + povrate + wage + educ,
              data = paneliniai_apskritys, model = "pooling")
round(vif(pooled), 3)    # VIF testas

# STEPWISE PROCEDŪRA
selectedMod <- step(lm(unem ~ inf + bvp + migr + povrate + wage + educ,
                       data = duomenys_apskritys))

all_vifs <- car::vif(selectedMod)
signif_all <- names(all_vifs)

while(any(all_vifs > 4)){
  var_with_max_vif <- names(which(all_vifs == max(all_vifs)))
  signif_all <- signif_all[!(signif_all) %in% var_with_max_vif]
  myForm <- as.formula(paste("unem ~", paste(signif_all, collapse = " + "),
                              sep = ""))
  selectedMod <- lm(myForm, data = duomenys_apskritys)
  all_vifs <- car::vif(selectedMod)
}

all_vars <- names(selectedMod[[1]])[-1]
```



```

summ <- summary(selectedMod)
pvals <- summ[[4]][, 4]
not_significant <- character()
not_significant <- names(which(pvals > 0.05))
not_significant <- not_significant[!not_significant %in% "(Intercept)"]

while(length(not_significant) > 0){
  all_vars <- all_vars[!all_vars %in% not_significant[1]]
  myForm <- as.formula(paste("unem ~ ", paste (all_vars, collapse = " + "),
                             sep = ""))
  selectedMod <- lm(myForm, data = duomenys_apskritys)

  summ <- summary(selectedMod)
  pvals <- summ[[4]][, 4]
  not_significant <- character()
  not_significant <- names(which(pvals > 0.05))
}
summary(selectedMod)

# GERIAUSIO POAIBIO PROCEDŪRA
models <- regsubsets(unem ~ inf + bvp + migr + povrate + wage + educ,
                     data = paneliniai_apskritys, nvmax = 5)
res.sum <- summary(models)
data.frame(Adj.R2 = which.max(res.sum$adjr2),
           CP = which.min(res.sum$cp),
           BIC = which.min(res.sum$bic))
# ==> Pagal BIC kriterijų siūloma 3 regresorių kombinacija
summary(models)
# ==> Reikšmingi kintamieji: bvp, wage, educ

```

```

# FIKSUOTŲ INDIVIDUALIŲ EFEKTŲ MODELIS
fixed.individual <- plm(unem ~ inf + bvp + migr + povrate + wage + educ,
                        data = paneliniai_apskritys, effect = "individual",
                        model = "within")

# Fiksuoti individualūs efektai:
summary(fixef(fixed.individual, type = "level"))

# FIKSUOTŲ LAIKO EFEKTŲ MODELIS
fixed.time <- plm(unem ~ inf + bvp + migr + povrate + wage + educ,
                  data = paneliniai_apskritys, effect = "time",
                  model = "within")

# Fiksuoti laiko efektai:
summary(fixef(fixed.time, type = "level"))

# F TESTAS
pFtest(fixed.individual, pooled)    # p-value = 0.9921
pFtest(fixed.time, pooled)         # p-value = 0.0001934

# ATSITIKTINIŲ LAIKO EFEKTŲ MODELIS
random.time <- plm(unem ~ inf + bvp + migr + povrate + wage + educ,
                   data = paneliniai_apskritys, effect = "time",
                   model = "random")

# HAUSMAN TESTAS
phtest(fixed.time, random.time)    # p-value = 2.256e-08

```

```
#####
#                               #
#                               #
#####

res.fixed.time <- fixed.time$residuals
plot(res.fixed.time, xlab = NA, ylab = "liekanų reikšmės")
abline(h = mean(res.fixed.time), col = "red")

# Ljung-Box stacionarumo testas:
Box.test(res.fixed.time)          # p-value = 0.4121
```