



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas

Pridėtinės vertės mokesčio (PVM) mokėtojų tipinės elgsenos analizė

Baigiamasis magistro studijų projektas

Agnė Miliušienė
Projekto autorė

Doc. dr. Audrius Kabašinskas

Vadovas

Prof. dr. Borisas Seminogovas

Vadovas

Kaunas, 2020



Kauno technologijos universitetas
Matematikos ir gamtos mokslų fakultetas

Pridėtinės vertės mokesčio (PVM) mokėtojų tipinės elgsenos analizė

Baigiamasis magistro studijų projektas
Didžiųjų verslo duomenų analitika (6213AX001)

Agnė Miliušienė

Projekto autorė

Doc. dr. Audrius Kabašinskas
Vadovas

Prof. dr. Borisas Seminogovas
Vadovas

Dr. Kęstutis Lukšys
Recenzentas

Doc. dr. Kristina Kundelienė
Recenzentė

Kaunas, 2020



Kauno technologijos universitetas

Matematikos ir gamtos mokslų fakultetas

Agnė Miliušienė

Pridėtinės vertės mokesčio (PVM) mokėtojų tipinės elgsenos analizė

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Agnės Miliušienės, baigiamasis projektas tema „Pridėtinės vertės mokesčio (PVM) mokėtojų tipinės elgsenos analizė“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

Miliušienė Agnė. Pridėtinės vertės mokesčio (PVM) mokėtojų tipinės elgsenos analizė. Magistro baigiamasis projektas / vadovai doc. dr. Audrius Kabašinskas ir prof. dr. Borisas Seminogovas; Kauno technologijos universitetas, matematikos ir gamtos mokslų fakultetas. Studijų kryptis ir sritis (studijų krypčių grupė): Taikomoji matematika (Matematikos mokslai).

Reikšminiai žodžiai: Pridėtinės vertės mokestis (PVM), mokestinis sukčiavimas, tipinė elgsena, saviorganizuojantys neuroniniai tinklai (SOM), hierarchinis klasterizavimas, sprendimų medis.

Kaunas, 2020. 64 p.

Santrauka

Pridėtinės vertės mokestis yra daugiausiai pajamų į šalies biudžetą surenkantis mokestis. Tačiau PVM mokestinis sukčiavimas Lietuvoje yra vienas didžiausių Europos Sąjungoje ir reikšmingai nemažėjantis. Todėl šiame darbe atliktas tyrimas yra orientuotas į sukčiavimo mažinimą, didinant tikimybę efektyviai surasti sukčiaujančias įmones. Tyrimas išanalizuoja pridėtinės vertės mokesčio mokėtojų tipinę elgesį pažeidžiamiausiuose sukčiauti ekonominėse veiklose. Darbe apželgiami pridėtinės vertės mokesčio mokėjimo, skaičiavimo, sukčiavimo būdų susijusių su šiuo mokesčiu ir atskirų ekonominių sektorių veiklos teoriniai aspektai. Be to, išanalizuojami panašumų identifikavimui skirti duomenų analitikos metodai. Ištiriami statybų, žemės ūkio ir elektronikos pardavimų veiklų PVM ataskaitų duomenys išskiriant mokėtojų panašumus ir rizikingas sukčiauti įmones. Analizuojami įmonių duomenys, kurie yra pateikiami pridėtinės vertės mokesčio deklaracijose nuo 2011 metų sausio mėnesio iki 2016 metų gruodžio mėnesio, pridedant papildomą išorinį kintamąjį - darbuotojų skaičių. Naudojamos absoliučios kintamųjų reikšmės ir apskaičiuoti išvestiniai rodikliai. Tyrimo struktūra susideda iš duomenų gavybos, paruošimo, apžvalgos, išskirčių radimo, klasterių nustatymo, tipinių ir netipinių pridėtinės vertės mokėtojų elgsenos nustatymo, tyrimo validavimo, kategorinių kintamųjų pridėjimo ir rezultatų aptarimo. Tyrime klasterizavimui naudojami saviorganizuojantys neuroniniai tinklai ir hierarchinis klasterizavimas, taisyklių sukūrimui ir klasterių validavimui naudojamas sprendimų medis. Atliktas tyrimas parodo, kad visuose tirtose veiklose yra tokia įmonių dauguma, kuri, pagal analizuojamus kintamuosius, turi bendrą elgseną. Visų trijų veiklų tipinių pridėtinės vertės mokesčio mokėtojų kintamųjų medianos visais ketvirčiais nenukrypsta nuo vidutinių, lyginant su kitais klasteriais. Ketvirtiniu sezoniškumu daugiausiai išsiskiria žemės ūkio prekių pardavimo veiklos tipinės įmonės. Saviorganizuojančių neuroninių tinklų (SOM) ir hierarchinio klasterizavimo rezultatai stipriai atitinka sprendimų medžio taisyklėmis grįstas suskirstymas. Pagal netipinių pridėtinės vertės mokėtojų klasterių rodiklius, išryškėja rizikingiausios sukčiauti įmonės. Pridėjus kategorinius kintamuosius, suskirstymas apskritimis neprideda papildomos naudingos informacijos. Tačiau suskirstius pagal dydį, pastebėta, kad didelės ir vidutinės įmonės visuose tirtose veiklose nepatenka arba mažai patenka prie tipinių PVM mokėtojų klasterio.

Agnė Miliušienė. Typical Behaviour Analysis of Value-added Tax Payers. Master's Final Degree Project / supervisor assoc. prof. dr. Audrius Kabašinskas ir prof. dr. Borisas Seminogovas; Department of Applied Mathematics, Kaunas University of Technology.

Study field and area (study field group): Applied Mathematics.

Keywords: Value-added Tax (VAT), tax evasion, typical behaviour, self-organising map (SOM), hierarchical clustering, decision tree.

Kaunas, 2020. 64.

Summary

Value-added tax brings the biggest revenue of all the taxes. Nevertheless, value-added tax evasion is one of the biggest in Lithuania of all the European Union. And it is not substantially decreasing. Thus, the research of this work is oriented towards decreasing tax evasion, by increasing the probability to effectively find evaders. The research analyses value-added tax payers' typical behavior in the most vulnerable to evasion economy fields. Value-added tax payment, counting, evasion methods related to this tax and separate economic fields' behavior theoretical aspects is reviewed in the work. Also, data analysis methods, which are developed for the similarity analysis, are reviewed. Construction, agriculture and electronics sales fields' data of value-added tax declarations is analysed by identifying payers' similarities and companies that are suspected to evade tax. Companies' data is from tax declarations between 2011 January and 2016 December, together with added employees number as external variable. Derivative indexes are also used in the analysis. The structure of the analysis is: gathering the data, data preparation, review of the data, outliers detection, clustering, typical and atypical value-added tax payers' behaviour identification, validation, introduction of categorical variables and introduction of the results. The analysis uses self-organizing map and hierarchical clustering for the separation of groups and decision tree for the validation of results and creation of separation rules. The research identified that the majority of companies exist, which has the common behaviour according to the variables used. All variables' medians of all three economical fields' value-added tax payers is around the middle in all the quarters compared to all clusters. Agriculture products sales field typical tax payers stand out as the most quarterly seasonal. Self-organizing map and hierarchical clustering result is strongly similar to the decision tree based rules separation. The companies, which are suspected to evade the tax, is identified according to the atypical value-added tax payers' clusters' indicators. The information of the location of business does not give any additional useful insights in this research. Although, the information of the size of the companies shows that the big and average sized companies is not included in the typical payers' clusters in all three economical fields.

Turinys

Paveikslų sąrašas	7
Lentelių sąrašas	8
Įvadas.....	9
1. Literatūros apžvalga	10
1.1. Pridėtinės vertės mokestis ir jo skaičiavimas	10
1.1.1. Pagrindiniai PVM privalumai ir trūkumai.....	10
1.1.2. PVM skaičiavimas.....	12
1.2. Netipinė elgsena: sukčiavimas pridėtinės vertės mokesčio srityje.....	13
1.2.1. Pagrindiniai žinomi PVM sukčiavimo būdai	13
1.2.2. Priemonės taikomos PVM sukčiavimo mažinimui	16
1.3. Lietuvos pridėtinės vertės mokesčio mokėtojų sektorių apžvalga	19
1.4. Literatūros apžvalgos išvados.....	23
2. Duomenų tyrimo metodai ir tyrimo objektas	25
2.1. Aprašomosios analizės metodai	26
2.1.1. Klasterizavimo metodai.....	26
2.1.2. Saviorganizuojantis neuroninis tinklas (SOM)	27
2.1.3. Sprendimų medis	29
2.2. Tyrimo objektas.....	30
2.2.1. Imties ir kintamųjų parinkimas.....	30
2.2.2. Išskirtys duomenyse	33
2.3. Naudota įranga.....	34
3. Tyrimų rezultatai ir jų aptarimas.....	35
3.1. Žemės ūkio prekių pardavimų veiklos tyrimo rezultatai	36
3.2. Su statybų sektoriumi susijusių pardavimų veiklos tyrimo rezultatai	45
3.3. Elektronikos prekių pardavimų veiklos tyrimo rezultatai	53
Išvados	63
Literatūros sąrašas	64
Priedai.....	69
1 priedas. Žemės ūkio pardavimų veiklos kintamųjų stačiakampės diagramos, suskirstytos klasteriais ir ketvirčiais.	69
2 priedas. Su statybų sektoriumi susijusių pardavimų veiklos tiriamų kintamųjų suvestinė suskirstyta ketvirčiais.	73
3 priedas. Elektronikos prekių pardavimų veiklos kintamųjų stačiakampės diagramos, suskirstytos klasteriais ir ketvirčiais.	77
4 priedas. Žemės ūkio pardavimų veiklos netipinių įmonių sąrašas, jų klasteriai ir priskyrimas prie išskirčių pirminio išskirčių tyrimo metu. Paryškintos įmonės yra priskirtos prie rizikingų sukčiauti.....	84
5 priedas. Su statybų sektoriumi susijusių pardavimų veiklos netipinių įmonių sąrašas, jų klasteriai ir priskyrimas prie išskirčių pirminio išskirčių tyrimo metu. Paryškintos įmonės yra priskirtos prie rizikingų sukčiauti.	86
6 priedas. Elektronikos prekių pardavimų veiklos netipinių įmonių sąrašas, jų klasteriai ir priskyrimas prie išskirčių pirminio išskirčių tyrimo metu. Paryškintos įmonės yra priskirtos prie rizikingų sukčiauti.	95

Paveikslų sąrašas

1 pav. Pagrindiniai PVM privalumai ir trūkumai.....	10
2 pav. Karuselės schemas pavyzdys (sudaryta remiantis M. Keen'o ir S. Smith'o [22] ir Europos audito rūmų [23] pavyzdžiu)	16
3 pav. Statybų sektoriaus bendroji pridėtinė vertė Lietuvoje, indeksas: 2010 m. = 100 (Eurostat [28])	22
4 pav. Didmeninės ir mažmeninės prekybos, transporto, apgyvendinimo ir maitinimo paslaugų sektorių bendroji pridėtinė vertė Lietuvoje, indeksas: 2010 m. = 100 (Eurostat [28])	22
5 pav. Žemės ūkio sektoriaus bendroji pridėtinė vertė Lietuvoje, indeksas: 2010 m. = 100 (Eurostat [28])	23
6 pav. Dviejų lygių klasterizavimas naudojant saviorganizuojančius neuroninius tinklus (J. Vesanto ir E. Alhoniemi [47])	28
7 pav. SOM rezultato kiekių ir atstumų grafikai.....	36
8 pav. SOM rezultato kintamųjų sąlyginiai dydžiai suskirstyti ketvirčiais.....	37
9 pav. Teorinių metodų balsai už optimalų klasterių skaičių.....	38
10 pav. SOM rezultato pasiskirstymas klasteriais ir ketvirčiais	39
11 pav. Didžiausio klasterio centruotų ir bendro mastelio kintamųjų medianų reikšmės, suskirstytos ketvirčiais	40
12 pav. Sprendimų medžio rezultatas	42
13 pav. SOM rezultato stebėjimų pasiskirstymas pagal apskritį, skaitinius kirntamuosius ir įmonės dydį.....	43
14 pav. Apskrities ir įmonės dydžio kintamųjų centruotos ir bendro mastelio medianų reikšmės, suskirstytos klasteriais	44
15 pav. SOM rezultato kiekių ir atstumų grafikai.....	45
16 pav. SOM rezultato kintamųjų sąlyginiai dydžiai suskirstyti ketvirčiais.....	46
17 pav. Teorinių metodų balsai už optimalų klasterių skaičių.....	47
18 pav. SOM rezultato pasiskirstymas klasteriais ir ketvirčiais	48
19 pav. Didžiausio klasterio centruotų ir bendro mastelio kintamųjų medianų reikšmės, suskirstytos ketvirčiais	49
20 pav. Sprendimų medžio rezultatas	51
21 pav. SOM rezultato stebėjimų pasiskirstymas pagal apskritį, skaitinius kirntamuosius ir įmonės dydį.....	52
22 pav. Didžiausio klasterio apskrities ir įmonės dydžio kintamųjų centruotos ir bendro mastelio medianų reikšmės	53
23 pav. SOM rezultato kiekių ir atstumų grafikai.....	53
24 pav. SOM rezultato kintamųjų sąlyginiai dydžiai suskirstyti ketvirčiais.....	54
25 pav. Teorinių metodų balsai už optimalų klasterių skaičių.....	55
26 pav. SOM rezultato pasiskirstymas klasteriais ir ketvirčiais	56
27 pav. Didžiausio klasterio centruotų ir bendro mastelio kintamųjų medianų reikšmės, suskirstytos ketvirčiais	57
28 pav. Sprendimų medžio rezultatas	60
29 pav. SOM rezultato stebėjimų pasiskirstymas pagal apskritį, skaitinius kirntamuosius ir įmonės dydį.....	61
30 pav. Didžiausio klasterio apskrities ir įmonės dydžio kintamųjų centruotos ir bendro mastelio medianų reikšmės	62

Lentelių sąrašas

1 lentelė. Literatūroje pateikiami rizikingų įmonių atrankos kriterijai.	17
2 lentelė. Lietuvos PVM įstatyme numatytų lengvatų vertinimas (Lietuvos Bankas [2]).....	19
3 lentelė. PVM tarifai kai kuriose Europos Sąjungos šalyse, % [27]	22
4 lentelė. Mokesčių administratorių taikomi duomenų tyrybos metodai mokestinio sukčiavimo aptikimui (P. C. Gonzalez‘as, J. D. Velasquez‘as [45])	25

Įvadas

Pridėtinės vertės mokestis 2019 m. sudarė 45,2% visų Lietuvos mokesčių biudžeto, administruojamo valstybinės mokesčių inspekcijos [1]. Skaičiuojama, kad 2015 metais 26,5% šių mokesčių nebuvo sumokama į valstybės biudžetą dėl sukčiavimo. Šis atotrūkis yra vienas didžiausių Europos Sąjungoje ir reikšmingai nemažėjantis [2]. Todėl yra būtina efektyviau mažinti sukčiavimo atvejus. Šiam tikslui reikalinga tiksliau identifikuoti rizikingas sukčiauti įmones.

Kadangi mokesčių mokėtojai pateikdami duomenis sukuria didžiulius informacijos srautus [3], jų analizė yra būtina [9], kad efektyviau panaudoti sukčiavimams atrasti skiriamus išteklius. Šiame darbe remiantis prielaida, kad didžioji dalis įmonių pateikia teisingus mokesčius duomenis, pagal PVM ataskaitų duomenis, yra išskiriami tipiniai mokėtojai, o kiekviena įmonė, kuri išsiskiria iš tipinio mokėtojo elgsenos laikoma kaip potencialiai vengianti mokėti mokesčius [4].

Darbo tikslas – atlikti pridėtinės vertės mokesčio mokėtojų (įmonių) elgesio analizę, kuri padės identifikuoti netipinę mokėtojų elgseną ekonominėje veikloje.

Problema – kokie yra įmonių PVM ataskaitų panašumai skirtinguose ekonominiuose sektoriuose?

Darbo uždaviniai:

1. apžvelgti pridėtinės vertės mokesčio mokėjimo, skaičiavimo, sukčiavimo būdų susijusių su šiuo mokesčiu ir atskirų ekonominių sektorių veiklos teorinius aspektus;
2. išanalizuoti panašumų identifikavimui skirtus duomenų analitikos metodus, kuriuos naudoja kitų valstybių mokesčių administratoriai;
3. ištirti statybų, žemės ūkio ir elektronikos pardavimų veiklą PVM ataskaitų duomenis išskiriant tipinius ir netipinius mokėtojus.

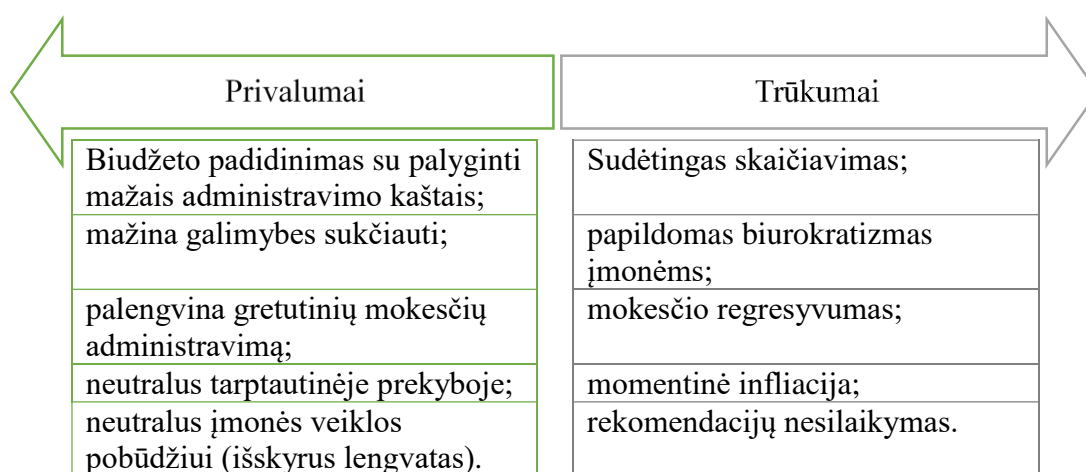
Tyrimui atlikti taikoma duomenų tyrybos aprašomoji analizė. Pirmiausiai, apžvelgiami kintamieji ir padaromos pirminės išvalgos naudojantis saviorganizuojančius neuroninius tinklus. Po to parenkami klasterizavimo parametrai ir atliekamas apmokymas be mokytojo. Galiausiai, patikrinamas klasterių atsiskyrimo stiprumas pagal sprendimo medžio suskirstymo taisykles.

1. Literatūros apžvalga

1.1. Pridėtinės vertės mokestis ir jo skaičiavimas

Lietuvoje pridėtinės vertės mokestį administruoja valstybinė mokesčių inspekcija (VMI) [5]. Pagal VMI [6], šio mokesčio mokėtojai yra „Lietuvos bei užsienio fiziniai ir juridiniai asmenys, kurie Lietuvoje vykdo bet kokio pobūdžio ekonominę veiklą, taip pat juridinio asmens statuso neturintis Lietuvos Respublikoje įsteigtas kolektyvinio investavimo subjektas, kurio veiklos forma yra investicinis fondas.“ Įmonė privalo tapti PVM mokėtoju jei per pastaruosius 12 mėnesių jos gaunamos pajamos viršijo 45000 € [20]. Tačiau PVM yra vartojimo mokestis, todėl šio mokesčio našta tenka ne mokėtojui, o vartotojui [7].

1.1.1. Pagrindiniai PVM privalumai ir trūkumai



1 pav. Pagrindiniai PVM privalumai ir trūkumai

Nuo 1960 m. PVM tapo vienu iš pagrindinių mokesčių daugiau nei 140 valstybių [13]. Taip yra todėl, kad pridėtinės vertės mokestis gali padidinti valstybės biudžetą su palyginti mažais administravimo kaštais. Pavyzdžiui, kadangi kiekvieno pardavėjo pardavimas yra kažkieno kito pirkimas, o įmonės, norėdamos susigrąžinti sumokėtą PVM, pateikia tiek pirkimo, tiek pardavimo sąskaitas. Dėl šios priežasties mokesčių administratorius gali pertikrinti ar pateikiama informacija yra tikra, tai yra, ar pirkimai sutampa su tiekėjo pardavimais [10, 11]. Taip pat PVM surinkimui nėra reikalinga informacija apie rinkos pakilimus, nuolaidas, sektoriaus susitarimus ir specialius pardavimų bei gamybinių mokesčių reguliavimus [10].

Dėl savo skaičiavimo sudėtingumo tiek mokesčių mokėtojams, tiek jų administratoriams, PVM įvedimas yra sudėtingas trečiosiose šalyse, kuriose net ir paprasčiausias įmonės įrašų generavimas yra žemos kokybės. Tai gali lemti ir neteisingą sąskaitų pateikimą, kuriuo įvykdomi nepastebėti sukčiavimai. Tačiau svarbiausia ne pats pridėtinės vertės mokesčių surinkimo ir apmokėjimo sudėtingumas, tačiau ar jis yra sudėtingesnis ir brangesnis už kitų mokesčių surinkimą ir apmokėjimą. L.P. Ebrill'as, M. Keen'as, J.P. Bodin'as ir V.P. Summers'as [17] ataskaitoje teigia, kad pagal ištirtas besivystančias šalis PVM reikalauja mažiau kaštų nei pajamų mokestis. Jie teigia, kad dėl šio atradimo nėra pagrindo teigti, kad PVM yra brangesnis ar sudėtingesnis už vartojimo mokesčius, kuriuos jis pakeičia.

Net jei pridėtinės vertės mokestis sukuria nemažai papildomo biurokratizmo įmonėms [16], pagal D. Pomeranz'o [9] atliktą tyrimą, PVM sukurtas informacijos srautas teigiamai veikia atbaidant įmones nuo mokestinio sukčiavimo. Tyrimas parodė, kad įmonės, kurios turi pateikti daug informacijos dėl pridėtinės vertės mokesčio, beveik visiškai nebereaguoja į audito tikimybės įmonėje padidėjimą. Taip pat, jei įmonėse padidėja audito tikimybė ir jos turi pateikti informaciją, susijusią su pridėtinės vertės mokesčiu, padidėja ir jų tiekėjų skaidrumas, kadangi gauta informacija yra ir apie jų pardavimus.

Pagal A.A. Tait'ą [10], šis mokestis netgi palengvina kitų mokesčių administravimą. Duomenys gaunami apie įmonių pirkimus ir pardavimus pateikia naudingos informacijos apie pelnus, kurie yra svarbūs skaičiuojant pajamų mokestį. Taip pagerinamas tiesioginių mokesčių administravimas. Tačiau tas pats autorius kontraargumentuoja, kad pajamų mokesčio ir pridėtinės vertės mokesčio administravimo ypatumai stipriai skiriasi, todėl juos administruoti kartu gali būti ir neefektyvu. Pavyzdžiui, PVM auditas yra paprastesnis ir reikalauja mažiau finansinių žinių nei pajamų mokesčio, todėl PVM audituoti galima būtų žymiai dažniau, jei jis būtų atskirtas nuo pajamų mokesčio administravimo. PVM mokesčių administratoriai yra labiau susikoncentravę į einamuosius mokestinius metus, o pajamų mokesčio – į praėjusius. Kita vertus, administruojant šiuos mokesčius kartu galima lanksčiau panaudoti žmogiškuosius išteklius. Galima sugeneruoti tikslingesnių idėjų abiejų mokesčių problemoms spręsti. Mažiau panaudojama išteklių toms pačioms užduotims atlikti. Mažesni mokesčių mokėtojų kaštai pateikiant reikiamą informaciją mokesčių administratoriui.

PVM taip pat užtikrina neutralumą tarptautinėje prekyboje, kadangi neapmokestina eksporto, o importas yra apmokestinamas lygiai taip pat kaip vietiniai produktai [8]. Kartu šis mokestis skatina eksportą, kadangi eksportuojami produktai yra neapmokestinami. Tačiau pagal Hines ir Desai [12] empirinį tyrimą, dėl valiutų kurso skirtumo, šis privalumas išnyksta ir netgi pastebėta tendencija, kad valstybėse, kuriose įvestas PVM, stebimas eksporto sumažėjimas.

PVM sąskaita yra tokia pati produktui, gaminamam korporaciniame ir ne korporaciniame sektoriuose, kapitalo pagrindu ar žmogiškųjų išteklių pagrindu dirbančiose įmonėse, taip pat tokia pati pagamintam integruotose ar specializuotose įmonėse [8, 14]. Išskyrus tam tikras lengvatas (apie jas bus plačiau kalbama 1.3 skyrelyje), galima teigti, kad pridėtinės vertės mokestis, kitaip nei pajamų mokestis, neišskiria formų ar metodų, kuriais vykdoma įmonės veikla.

Pagal M.L. Weidenbaum'ą ir E.S. Christian'ą [14], vienas iš pridėtinės vertės mokesčio trūkumų yra tas, kad, kaip ir bet koks vartojimu grįstas mokestis, jis yra regresyvus. Mokestinė našta yra didžiausia tiems, kas uždirba mažiausiai. Šis trūkumas yra švelninamas nustatant mokestines lengvatas būtiniausioms prekėms. Tačiau šis metodas suteikia galimybių sukčiaujančioms įmonėms vengti mokesčių [15]. Be to, skirtingi tarifai įgalina interesų grupes varžytis dėl palankių nuolaidų ir lengvatų [2]. Lietuvos bankas [2] kaip variantą šiai problemai spręsti siūlo naudoti vieną PVM tarifą, bet kartu efektyviai pritaikyti tiesioginių mokesčių ir socialinių išmokų sistemas.

Kadangi PVM yra įtraukiamas į pirkimo kainą, jis pakelia kiekvieno produkto kainą ir todėl sukelia infliaciją. Tačiau šis efektas yra vienkartinis, atsirandantis tik mokesčio įvedimo ar padidinimo momentu. Be to, infliacija gali būti sušvelninama tinkamomis monetarinėmis priemonėmis, bet kartais jos lemia neigiamą efektą gamybos ir įdarbinimo apimtims [14].

Be viso to, L.P. Ebrill'as, M. Keen'as, J.P. Bodin'as ir V.P. Summers'as [17] empiriniame tyrime pastebėjo, kad didžiojoje dalyje valstybių yra tam tikrų nukrypimų nuo rekomenduojamo PVM

dizaino. Tokių kaip platesnis nei rekomenduojamas nulinio tarifo naudojimas, rekomenduojamų daugkartinių laiku įvestų kreditų gamybinėms prekėms trūkumas, ir per dažnai įvedamo skirtingo tarifo PVM. Ir turbūt esminis nukrypimas nuo rekomendacijų yra žemesnės PVM mokėtoju tapimo ribos nustatymas. Šie nukrypimai gali pakenkti mokesčio efektyvumui ilgoju periodu.

1.1.2. PVM skaičiavimas

Norint tiksliau pažvelgti į PVM mokėtojo elgesį, reikalinga išnagrinėti šio mokesčio sudedamąsias dalis. Daugumoje valstybių pridėtinės vertės mokestis skaičiuojamas netiesioginiu atimties metodu, tačiau egzistuoja dar ir netiesioginis sudėties ir tiesioginiai atimties ir sudėties metodai, kurie gali apmokestinti tik įmonėje pridėtą vertę, o ne visą pardavimą [10, 18]. Tai reiškia, kad skaičiavimas gali būti atliakamas atimant arba sudedant (1) [10].

$$\text{Pridėtinė vertė} = \text{pajamos} - \text{sąnaudos žaliavoms} = \text{darbo užmokestis} + \text{pelnas.} \quad (1)$$

Į formulę įtraukus tarifą (t), skaičiavimas suskirstomas į keturis metodus, kuriais gaunami identiški rezultatai (2, 3, 4, 5) [10].

$$t (\text{darbo užmokestis} + \text{pelnas}): \text{tiesioginis sudėties metodas;} \quad (2)$$

$$t (\text{darbo užmokestis}) + t (\text{pelnas}): \text{netiesioginis sudėties metodas;} \quad (3)$$

$$t (\text{pajamos} - \text{sąnaudos žaliavoms}): \text{tiesioginis atimties metodas;} \quad (4)$$

$$t (\text{pajamos}) - t (\text{sąnaudos žaliavoms}): \text{netiesioginis atimties metodas.} \quad (5)$$

Nors dauguma mokesčių būna apmokestinami tiesiogiai – pirmiausiai apskaičiuojant mokesčio bazę, o tada pritaikant tarifą, skaičiuojant PVM, netiesioginio atimties metodo (5) populiarumą lėmė keli jo privalumai. Pirmiausia, šis metodas apmokestina kiekvieną sandorį. Tai padaro sąskaitą faktūrą svarbiausiu dokumentu sekant tiek sandorį, tiek mokestinę prievolę [10]. Dėl to šis metodas dažnai dar vadinamas sąskaitos faktūros metodu.

Sąskaitos faktūros metodas sukuria galimybes tinkamam auditui. Pagal A.A. Taitą [10], šalių, kurios naudoja tiesioginį atimties metodą patirtis rodo, kad nepaliekant sąskaitų faktūrų kaip sandorių įrodymų, trūksta informacijos ar neatimamos sąnaudos, už kurias nemokėta PVM. Taip pat kyla skaičiavimo problemos, kai sąnaudos viršija pajamas. L.P. Ebrill'as, M. Keen'as, J.P. Bodin'as ir V.P. Summers'as [17] teigia, kad dėl šių priežasčių kitoks skaičiavimo būdas mažiau apsaugo nuo sukčiavimo.

Atimties metodai yra pranašesni už sudėties metodus dėl sunkumų, kylančių įmonėse bandant nustatyti pelną atskiroms prekių grupėms [10]. Taip pat pasidaro neįmanoma taikyti skirtingų PVM tarifų dėl darbo užmokesčio apmokestinimo išskaidymo. Todėl sudėties metodai gali būti taikomi tik jei PVM yra vieno tarifo. Jei tarifai išskaidomi keliais lygiais, tinkamas metodas tampa tik netiesioginis atimties [10, 18].

Galiausiai, nors atrodytų, kad paprasčiausias skaičiavimo metodas yra pirmiausia apskaičiuoti pardavimų ir pirkimų skirtumą ir tada taikyti tarifą, praktikoje yra kitaip. Įmonės tokiu būdu negali skaičiuoti mėnesinės apskaitos dėl to, kad atsargų kiekis nuolatos kinta nebūtinai mėnesiniu terminu, tai priklauso nuo produkcijos tipo, sezoniškumo ar galimų tiekimo sutrikimų. Šį skaičiavimo procesą palengvina netiesioginis metodas, kuriuo galima paprastai apskaičiuoti netgi dienių PVM [10, 18].

Netiesioginiu atimties metodu skaičiuojamas PVM gali būti apmokestinamas įprastiniu ir atvirkštiniu būdu. Įprastinio apmokestinimo atveju įmonė A parduoda prekes įmonei B, kuri patobulintas prekes parduoda klientui C. Įmonės A ir B yra PVM mokėtojos. Įmonė A pateikė sąskaitą įmonei B, kur nurodyta prekės kaina be PVM ir priskaičiuota PVM suma. Įmonė B sumoka ir sumą be PVM, ir PVM. Įmonė A yra įpareigota valstybei sumokėti tą PVM sumą. Tas pats vyksta ir B, ir C veikėjų atžvilgiu, tik šiuo atveju B įmonė sumoka valstybei PVM sumą, kurią gavo iš kliento C, o PVM sumą, kurią sumokėjo tiekėjui A susigrąžina [2].

Atvirkštinis apmokestinimas yra taikomas atskiroms produktų ar paslaugų grupėms. Šio apmokestinimo principas yra tas, kad tiekėjas neišrašo PVM sumos pardavėjui, o pardavėjas tampa atsakingas už šio sandorio PVM sumokėjimą valstybei. Praktikoje, įmonė, kuri yra PVM mokėtojas, paskelbia ir išskaičiuoja PVM tuo pačiu metu be realaus mokesčio sumokėjimo, todėl nėra susigrąžinimo [19]. Tokiu būdu yra trukdoma vykdyti „karuselės“ sukčiavimo schemą. Kadangi moka ir susigrąžina PVM sumą ta pati įmonė, nebelieka „pradingstančios“ įmonės problemos [2].

Tačiau pagal Lietuvos banką [2], atvirkštinis apmokestinimas yra tik laikina priemonė stabdant sukčiavimą. Ji Europos Sąjungoje yra numatyta naudoti tik iki 2022 metų. Kadangi šis mechanizmas yra taikomas tik konkrečiuose sektoriuose, tai gali sukelti sukčiavimą kituose sektoriuose. Be to, išplėstas atvirkštinio apmokestinimo taikymas gali PVM paversti pardavimo mokesčiu.

1.2. Netipinė elgsena: sukčiavimas pridėtinės vertės mokesčio srityje

Mokestinis sukčiavimas šio darbo prielaidoje laikomas netipine PVM mokėjojo elgsena. Galutinis darbo tikslas yra sudaryti galimybes atskirti netipinius mokėtojus, kurie yra galimai sukčiaujantys. Todėl šiame poskyryje giliau nagrinėjami sukčiavimo būdai susiję su PVM ir kas daroma iki šiol, siekiant to išvengti.

Nors PVM ir turi save tikrinantį mechanizmą sąskaitų pavidalu, šio mokesčio sukčiavimas sukelia didžiausius finansinius nuostolius [2]. A.A. Tait'as [10] įvardina godumą ir nepasitenkinimą kaštais, susijusiais su PVM mokesčio mokėjimu, kaip pagrindines priežastis sukčiauti.

Yra daug būdų sukčiauti PVM srityje, kuriais, pagal VMI [21], siekiama:

- neteisėtai susigrąžinti PVM;
- pasisavinti pirkėjo sumokėtą PVM, o ne sumokėti jį į biudžetą;
- išvengti prievolių sumokėti PVM į valstybės biudžetą;
- sumažinti PVM prievoles.

Norint suprasti pagrindines rizikos sritis, toliau bus pateikti ir išskirstyti pagrindiniai žinomi sukčiavimo metodai.

1.2.1. Pagrindiniai žinomi PVM sukčiavimo būdai

Neprireistravimas PVM mokėtoju. Pagal M. Keen'ą ir S. Smith'ą [22], dažniausiai sukčiauja tokios įmonės, kurios yra palyginus mažos ir yra arti pajamų ribos, kai prisireistravimas yra privalomas. Neprireistruodamos jos sutaupo tiek kaštus, reikalingus mokesčio mokėjimui, tiek PVM mokesťį, kurį turėtų sumokėti (nors jos moka PVM už savo perkamas žaliavas, jų pardavimai lieka neapmokestinti). Prie šios grupės taip pat priskiriamos įmonės, kurios yra iš viso nežinomos mokesčių inspekcijai, jos nemoka ne tik PVM, tačiau ir pajamų mokesčio. Situaciją daro prastesne

tai, kad, pagal apklausą, atliktą Didžiojoje Britanijoje, net 70 procentų respondentų nemato nieko blogo nusipirkti prekes iš tokių įmonių grynaisias, jei jiems nereikės mokėti PVM mokesčio [10].

Sumažinta pardavimų suma. Pardavėjas gali suklastoti sąskaitas, parduoti kelias prekes ar paslaugas visai be sąskaitų ir pateikti tik dalį savo pardavimų. Šiam būdui, kaip prevencinė priemonė, veikia sąskaitų sutapimo tikrinimas, ar visas PVM susigrąžinimui skirtas sąskaitas pateikė ir žaliavų pardavėjai. Sukčiaujant šiuo būdu daugiausiai pasipelno įmonės, kurios pardavinėja galutiniam vartotojui, o jų žaliavų kaštai nėra dideli (pvz. kirpėjai, renginių organizatoriai, statybininkai dirbantys fiziniams asmenims ir kiti). Jų pridėtinės vertės suma yra didelė, o susigrąžinamas PVM mažas [10, 22].

Neteisingai suklasifikuotos prekės. Ši problema egzistuoja, kai įmonė pardavinėja taikomų skirtingų tarifų prekes. Tokios įmonės gali sumažinti mokesčius priskirdamos didesnio tarifo pardavimus prie mažesniojo tarifo pardavimų [22]. Šioje kategorijoje svarbu paminėti tai, kad dėl tarifų skirtumų sudėtingėja PVM apskaita, todėl įmonėse gali atsirasti netyčinės klaidos sumažinančios PVM [10].

Pačios įmonės pagamintų ir suvartotų prekių ar paslaugų praleidimas. Prekės ar paslaugos, kurias pagamina įmonė ir sunaudoja patys darbuotojai ir, kurios yra apmokestinamos, gali būti nepateiktos kaip mokestinės. Ši schema dažniausiai būdinga žemės ūkio sektoriui. Tačiau tokią riziką sumažina gan didelė pajamų riba, kurią pasiekusios įmonės turi registruotis PVM mokėtojomis. Tokiu atveju, didelė dalis mažų verslų, kurie naudojami savo produkcija ir taip nėra PVM mokėtojai [22].

Apgaulingas susigrąžinimas. Pagal A.A. Tait'ą [10], 44 % Olandijoje aptiktų PVM sukčiavimų buvo susiję su apgaulingu susigrąžinimu, pavyzdžiui, suklastotų pirkimo padidavimo ar nevykusio pirkimo sąskaitų pateikimu. Tačiau šis apgavimo metodas turi limitą. Nesunku pastebėti, jei įmonė kurį laiką perka daugiau žaliavų nei parduoda arba dirba su labai nedideliu pardavimų plusu. Todėl tokiu būdu dažniau sukčiauja naujai atsidariusios įmonės, kurių laikinos didelės sąnaudos gali būti pateisinamos investavimu [22].

Apgaulingam susigrąžinimui gali būti būdinga tokia schema, kai perkama iš smulkaus neregistruoto tiekėjo, o inspekcijai pateikiama sąskaita su melagingu PVM numeriu ir taip susigrąžinamas nesumokėtas PVM. Dažniausiai sumos nėra didelės, nes didesni tiekėjai automatiškai turi būti užregistruojami PVM mokėtojais [10].

Taip pat nelegalus PVM susigrąžinimas gali įvykti tuo atveju, jei įmonė parduoda nulinio tarifo ir pilnai apmokestinamas prekes. Tuomet yra galimybė susigrąžinti PVM už daugiau žaliavų, kuriomis buvo gaminamos neapmokestinamos prekės [10, 22].

Kitas pakankamai dažnas apgaulingo susigrąžinimo pavyzdys, kai PVM susigrąžinamas už prekes, kurių mokestis neturėtų būti gražintinas. A.A. Tait'as [10] kaip pavyzdį pateikia automobilio pirkimą. Jei jis naudojamas vien verslo reikalams, už automobilį susigrąžintas PVM yra teisėtas. Tačiau realybėje dažnai verslo automobilis būna naudojamas ir darbuotojo asmeniniais reikalais. Šios sukčiavimo sumos paprastai nėra didelės, tačiau sukuria nelygybę tarp PVM mokėtojų ir nemokėtojų.

Melagingas eksportas. Toks sukčiavimas remiasi tuo, kad eksportas yra PVM neapmokestinamas. Todėl atsiranda galimybė pateikti suklastotus dokumentus apie prekių eksportą, tačiau jas parduoti vietinėje šešėlinėje rinkoje. Kadangi eksportuojančios įmonės dažnai susigrąžina PVM daugiau nei

sumoka, toks sandėris gali likti nepastebėtas. Tačiau eksportą lydi kiti dokumentai, tokie kaip muitai ar transportavimas, todėl yra sudėtinga padirbti visus reikalingus dokumentus. Sunkiau patikrinti informacinių technologijų paslaugų eksportą [10].

Nepateikti pirkimai ir pardavimai grynaisiais. Paprasta sukčiavimo schema, kai perkama žaliava be sąskaitos ir parduodama prekė be sąskaitos. Dažniausiai pasitaiko trumpose tiekimo grandinėse. Tai nebūtų rimta problema, nes dauguma PVM mokėtojų nori registruoti pirkimus, kad susigrąžintų mokesť. Tačiau jei pajamų mokesťis yra didelis, tokiu būdu būna išvengiama ne tik PVM, bet ir šio mokesčio [10].

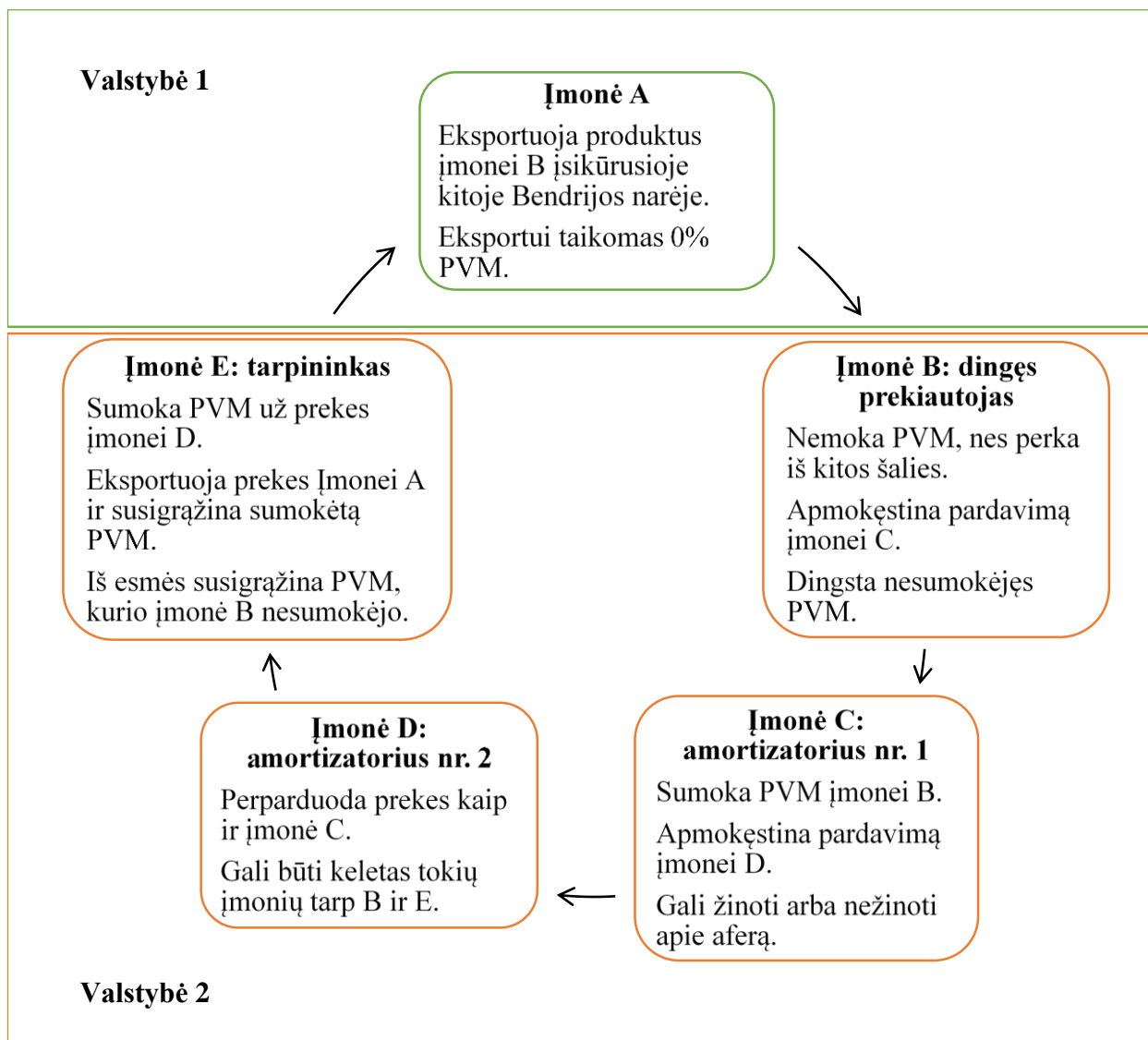
Tyčinės apskaitos klaidos. Padarius apskaitos klaidą, nesvarbu ar ji tyčinė, ar ne, paprastai baudos dydis yra tik toks, koks būna už pavėluotą mokesčių sumokėjimą. Tačiau sukčiai atranda galimybių pasipelninti ir iš tokios situacijos. A.A. Tait'as [10] pateikia pavyzdį, kaip dvi seserinės įmonės gali viena kitai pardavinėti prekes. Nusipirkusi prekes įmonė susigrąžina sumokėtą PVM. Tačiau pardavusi prekes įmonė gali išrašyti kreditinę sąskaitą ir laukti kelis mėnesius neparodydama, kad PVM buvo gautas. Kol viskas išaiškėja kaip klaida ir PVM gražinamas, įmonė-pirkėja gali naudoti susigrąžintą PVM savo tikslams be jokio nuostolio, tai yra, ji gauna paskolą be palūkanų.

Mainų susitarimai. Tai sukčiavimo forma, kai tiekėjas ir pirkėjas susitaria sąskaitoje pakeisti realų pardavimo objektą į palankesnę mokesčiui susigrąžinti. Tai yra, jei parduodama tokia prekė, už kurią, pagal galiojančius vietas įstatymus, būtų negalima susigrąžinti PVM, oficialiai sąskaitoje ji pavadinama žaliavoms skirta preke. Tokius susitarimus gan sudėtinga sudaryti, išlaikyti ateityje ir slėpti [10].

Mokesčiai surinkti, bet nesumokėti valstybei. Šis būdas galimas arba sumažintos pardavimų sumos atveju, arba suklastojus įmonės bankrotą prieš apmokant mokesčius. Šis sukčiavimo būdas populiarus mažesnėse statybų įmonėse. Belgijoje tokie sukčiavimai buvo sumažinti padarant užsakovą atsakingu už PVM sumokėjimą, jei jo pasamdyta įmonė yra neregistruota [10]. Naujausias tokių sukčiavimo būdų stabdymo mechanizmas yra 1.1.2. skyrelyje aptartas atvirkštinio apmokestinimo metodas.

Dingęs prekiautojas. Sukuriama fiktyvi trumpalaikė įmonė, kuri gali suklastoti netikrus eksporto dokumentus ir atsiimti PVM už žaliavas, kurios niekada nebuvo gautos arba parduoti prekes vietinėje rinkoje, o dokumentus sutvarkyti, lyg tai būtų eksportas.

Tai viena didžiausių problemų PVM sukčiavimo srityje su kuria susiduria Europos sąjunga [23]. Šis principas yra pagrindinė „karuselės“ sukčiavimo schemas sudedamoji dalis [22]. Tokioje scheme dalyvauja kitoje valstybėje (Bendrijos narėje) esantis eksportuotojas, dingęs prekiautojas, amortizuojančios įmonės (gali ir nežinoti, kad dalyvauja tokioje scheme) ir tarpininkas (žr. 2 pav.). Visi schemas veikėjai, išskyrus dingusį prekeivį, atrodo įprastai prekiaujantys, todėl juos ypač sunku sugauti. Pardavimui dažniausiai naudojamos didelės vertės prekės, pavyzdžiui elektronikos prekės ar taurieji metalai [23].



2 pav. Karuselės schemos pavyzdys (sudaryta remiantis M. Keen'o ir S. Smith'o [22] ir Europos audito rūmų [23] pavyzdžiu)

Yra daug žinomų būdų kaip sukčiaujama PVM srityje. Tačiau vis atsiranda ir naujų dar iki tol nežinomų sukčiavimo atvejų. Todėl ir sukčiavimo atvejų radimo būdai privalo tobulėti, pasitelkti naujausius metodus.

1.2.2. Priemonės taikomos PVM sukčiavimo mažinimui

Didelė dalis PVM sukčiavimo būdų yra būdingi tik šiam mokesčiui dėl jo sandaros specifikos ir skirtingų tarifų. Todėl ir priemonių sukčiavimo stabdymui ieškoma išskirtinių, pritaikytų galimybių ir paskatų mažinimui savo tikslais naudotis pridėtinės vertės apmokestinimo mechanizmu.

Primityvios sukčiavimo schemos Lietuvoje nuo 2016 metų eliminuojamos naudojantis i.SAF duomenų rinkmena. Tai yra, elektroninė sąskaitų faktūrų pateikimo ir tvarkymo sistema, bendros i.MAS sistemos dalis [80]. Tokia sistema siekiama sekti ir analizuoti įmonių pirkimo ir pardavimo ryšius su kitomis įmonėmis. Tokiu būdu atrandant, ar nėra padirbinėjamų pardavimo sąskaitų.

Didžiausius nuostolius padaro karuselės schemos dalyviai, be to, jie yra ir sunkiai surandami. Europos komisija šiai schemai stabdyti yra nutarusi naudoti laikiną atvirkštinio apmokestinimo (žr. 1.1.2.

skyrelyje) priemonę [2]. Pagal M. Keen'ą ir S. Smith'ą [22], kitos priemonės galėtų būti tokios kaip atvirkštinis sulaikymas – principas ganėtinai panašus į atvirkštinio apmokestinimo, tik šiuo atveju pardavėjas lieka atsakingas už pardavimo PVM, kol pirkėjas sulaiko PVM sumos sumokėjimą. Naudoti PVM banko sąskaitas – pardavėjai turi atsidaryti atskirą banko sąskaitą, į kurią pervestu PVM sumą, atsiradus įtartinais veiksiais galima būtų tikrinti tiekimo grandinėje esančias sąskaitas ir išaiškinti karuselės dalyvius. Kitas pasiūlymas yra sustiprinti muitų patikrinimus ir nepraleisti prekių, kol neišaiškinama ar PVM apmokėtas. Taip pat siūlomas metodas yra eksportus apmokestinti šalies viduje, o surinktas sumas pervesti kitoms šalims kur buvo importuota. Visi pateikti metodai apsunkina administravimą ir taip sudaro kitas galimybes sukčiauti.

PVM, kaip ir kitų mokesčių, sukčiavimo mažinimui svarbiausia yra tinkamo dydžio baudos ir pakankamai didelė tikimybė, kad įmonė bus patikrinta mokesčių administratoriaus [10, 22, 24]. Pagal L.P. Ebrill'ą, M. Keen'ą, J.P. Bodin'ą ir V.P. Summers'ą [17], pagrindiniai audito sistemų trūkumai yra:

- efektyvios rizikingų įmonių atrankos sistemos nebuvimas;
- aiškių auditavimo tikslų ir laiko sąnaudų naudojimo taisyklių nebuvimas;
- primityvios technikos, pavyzdžiui, tokios kaip visų įmonių auditavimas;
- stiprios už auditą atsakingos institucijos nebuvimas, pavyzdžiui, skirtingų mokesčių administravimo padalinimas skirtingoms institucijoms.

Kiekvienais metais turėtų būti patikrinama 25-30 procentų PVM mokėtojų, kad audito programa, pagal mokesčių administravimo ekspertus, būtų efektyvi [17]. Tačiau auditui atrinkus 25-30 % PVM mokėtojų atsitiktine tvarka negaunamas efektyvus rezultatas. Naudojantis Pareto principu [29] plačiąja prasme ir atrinkus 25-30 % rizikingiausių įmonių būtų pasiektas tinkamas audito kaštų panaudojimo ir sukčiavimo atvejų išsiaiškinimo balansas. Pagal H. W. Cecil [41], kai JAV mokesčių administratorius pradėjo naudoti neatsitiktinį audituojamų įmonių parinkimą, įmonių „be pakitimų“ sumažėjo nuo 40 % iki 15 % visų atliekamų auditų.

1 lentelė. Literatūroje pateikiami rizikingų įmonių atrankos kriterijai.

Audito atrankos kriterijai	Šaltinis
Palyginami pirkimų ir pardavimų koeficientai, tos pačios įmonės skirtingų mokamų mokesčių informacija bei informacija iš prieš tai buvusių auditų. Taip pat galima panaudoti informaciją iš muitinių apie tiekimus ir išvežamus produktus ar susijusių įmonių pirkimus ir pardavimus.	L.P. Ebrill'as, M. Keen'as, J.P. Bodin'as ir V.P. Summers'as [17]
Atrenkamos įmonės, kuriose aptinkami didesni išskaitymų svyravimai, tiekia paslaugas, praleidžia terminus ir registracijos adresus yra kitoje valstijoje.	J. Alm'as, C. Blackwell'as, M. McKee [30]
Svarbus įmonės gyvavimo laikotarpis, susijusių įmonių skaičius ir ar savininkas iš kitos valstijos (šie kriterijai didina tikimybę sukčiauti). Kuo didesni įmonės pardavimai, tuo didesni galimo sukčiavimo mastai. Didesni svyravimai tarp pateiktų mėnesinių/ketvirtinių pardavimų siejami su mažesne tikimybe sukčiauti. Jei yra daugiau neapmokestinamų pardavimų tikėtina, kad sukčiavimo mastai tokiose įmonėse bus mažesni.	M. N. Murray [31]

Sektoriaus gairės pirkimų-pardavimų santykiui ir efektyvaus PVM rodiklis (santykis tarp PVM sumos ir pardavimų).	R. G. Manasan'as [32]
Pateiktos informacijos matematinis tikslumas, ar informacija suderinama su kitų įmonių, neįprastai dideli išskaičiavimai ar paradimai.	W. H. Wiersema [33]
Apyvartos apimtis arba bendrojo pelno rodiklis.	D. Ho, P. Lau [34]
Apyvartos augimas, esamos apyvartos rodiklis, turto ir apyvartos santykis ir apimtis.	P. Kumar'as, V. Nagadevara [35]
Įmonės tipas, veiklos vieta, pardavimo mokesčių mokėjimui pateikta informacija, kitų mokesčių mokėjimui pateikta informacija, darbuotojų skaičius, atlyginimų dydis, prieš tai buvusių auditų rezultatai.	D. Micci-Barreca, S. Ramachandran'as [36]
Įmonės profilis: ar naujai užregistruotas, ar veikla vykdoma didelių mokesčių srityje, ar yra kitų verslų registruotų tuo pačiu adresu, ar yra kitų įmonių su tuo pačiu telefonu. Pelno atitikimas: ar buvęs neparodytas pelno trūkumas, pelno deklaravimo vėlavimas, kiek pelno yra neapmokestinama. Apskaitos duomenys ir rodikliai: mokesčių santykis su apyvarta, bendrasis pelnas, neapmokestinamų pardavimų santykis su apyvarta, atsargų santykis su apyvarta, pardavimų santykis su pirkimais, gražintinos sumos, nesumokėti mokesčiai (skola), mokesčių patikslinimų dydis. Pelno svyravimai per mokestinius periodus: apyvartos augimas, mokesčių sumos augimas, apyvartos svyravimas, pakitimai pardavimuose (vietiniai, eksportas), pakitimai pirkimuose (vietiniai, importas), pakitimai tarp perkamų ar parduodamų produktų apmokestinamumo (neapmokestinami, lengvatinis tarifas, pilnas tarifas). To pačio sektoriaus įmonių palyginimas: mokesčių santykis su apyvarta, bendrasis pelnas, apmokestinamų pardavimų santykis su apyvarta, atsargų kiekio santykis su apyvarta, apyvartos augimas, mokesčių sumos augimas.	M. Gupta, V. Nagadevara [37]
Z-balas (kapitalas/bendras turtas*1,2 + pelnas/bendras turtas*1,4 + pelnas prieš pajamų mokesť/bendras turtas*3,3 + nuosavo kapitalo rinkos vertė/bendra apskaitinė vertė + įsipareigojimai*0,6 + pardavimai/bendras turtas*0,999), didelis įsiskolinimas padidina tikimybę, kad įmonė sukčiaus (bandro įsiskolinimo ir bendro turto santykis, įsiskolinimo ir nuosavybės santykis), neįprasta nuolatinio augimo reikšmė (pavyzdžiui, pardavimų ir augimo santykis), finansinės ataskaitos reikšmės (gautinos sumos, atsargos, bendroji marža), kiti finansiniai rodikliai (grynojo pelno ir bendro turto santykis, apyvartinio kapitalo ir bendro turto santykis, grynojo pelno ir pardavimų santykis, trumpalaikio turto ir trumpalaikių įsipareigojimų santykis). Dažni įmonės vadovo ir finansų direktoriaus pokyčiai gali parodyti, kad įmonėje egzistuoja mokesčių slėpimas. Be to, tikimybę sukčiauti gali parodyti ir prieš tai auditavusių auditorių kvalifikacija.	P. Ravisankar'as [38]
Išnuomoto nekilnojamo turto analizė, ar logine prasme sutampa mokesčių deklaracijose pateikta informacija su galima veikla visose išnuomotose vietose. Paslėpta ekonomika, jei atrandama neaiški su šešėline veikla susijusi informacija apie įmonės savininką. Keli PVM mokėtojo numeriai susieti su tuo pačiu kontaktiniu telefono numeriu.	M. S. Khwaja, R. Awasthi, J. Loeprick [39]

Tikrinti visų įmonės vadovų, akcininkų ir jų šeimos narių ryšius su kitomis įmonėmis. Įvertinti individualių asmenų, kurie yra įmonės vadovai ar savininkai asmeninį rizikingumą.	
Mokesčių naštos rodiklis, mokesčių tarifas, akcijų kainos, kritinio likvidumo koeficientas, turto grynojo pelno marža, pardavimų kainos rodiklis, pardavimų finansavimo mokesčio norma.	X. Liu, D. Pan'as, S. Chen'as [40]

Pagal informaciją patektą 1 lentelėje, galima teigti, kad rizikingų įmonių atrinkimui dauguma tyrimų siūlo naudoti įmonės vidinius duomenis, o modelio iš vidinių duomenų papildymui galima prijungti ir išorinius duomenis.

Net ir naudojant dalį šių priemonių, yra pripažįstama, kad kova su sukčiaujančiomis įmonėmis Europos sąjungoje nėra efektyvi [17], o pagrindinės rekomendacijos yra susijusios su tinkamesniu informacijos apsikeitimu tarp valstybių [23].

Pagal Lietuvos banko [2] apžvalgą, aktyvus elektroninių priemonių naudojimas sumažintų sukčiaujančių įmonių skaičių ar mastą. Tai yra, jei kiekvienas sandoris patektų į duomenų bazę ir būtų žinomas valstybinei mokesčių inspekcijai, atrinkti galimus mokesčius būtų paprasčiau, o tikimybė būti sugautam didesnė. Kitas Lietuvos Banko [2] siūlymas yra didelės mokesčių slėpimo rizikos asmenų sąrašas. Jei asmuo yra vadovavęs įmonei, kuri žinoma kaip sukčiavusi įmonė, kita jo vadovaujama įmonė automatiškai patektų į rizikingų sąrašą.

1.3. Lietuvos pridėtinės vertės mokesčio mokėtojų sektorių apžvalga

Pridėtinės vertės mokestis dėl savo struktūros reikalauja papildomų sąlygų, kad būtų efektyvus vartojimo mokestis ir kad būtų sudaroma kuo mažiau galimybių sukčiauti. Todėl skirtingi ekonomikos sektoriai ir net skirtingos prekės ar paslaugos apmokestinamos skirtingais tarifais ir sąlygomis. Šiame poskyryje aptariama kam ir kokios sąlygos yra taikomos, kodėl kai kurie sektoriai turi būti daugiau analizuojami ir kokios tipinės elgsenos galima tikėtis iš tokio sektoriaus.

2 lentelė. Lietuvos PVM įstatyme numatytų lengvatų vertinimas (Lietuvos Bankas [2])

Lengvata	Teorinis pagrindimas	Kylančios problemos	PVM pajamų netekimai dėl lengvatų (mln. Eur ir proc., palyginti su BVP)
Vaistai ir medicinos pagalbos priemonės, neįgalųjų techninės pagalbos priemonės.	Perskirstymas, mažas pajamas gaunančių asmenų rėmimas.	Naudinga tiek mažas, tiek didelės pajamas gaunantiems namų ūkiams, tikslinė išmokų (pašalpų) sistema užtikrintų geresnį perskirstymą.	97,7 (0,22 %)
Centrinis šildymas ir karštas vanduo.	Perskirstymas, mažas pajamas gaunančių asmenų rėmimas.	Lengvata teikia naudą tiek mažas, tiek didelės pajamas gaunantiems namų ūkiams, nemažas valstybės pajamų praradimas.	47,0 (0,11 %)
Keleivių vežimo nustatytais reguliaraus susisiekimo maršrutais, keleivių bagažo vežimo paslaugos.	Viešojo transporto sistemos patrauklumas ir prieinamumas mažas pajamas gaunantiems asmenims.	Lengvata teikia naudą tiek mažas, tiek didelės pajamas gaunantiems namų ūkiams, nemažas valstybės pajamų praradimas.	31,0 (0,07 %)

Viešbučiai ir specialaus apgyvendinimo paslaugos.	Padidėja žemos kvalifikacijos darbuotojų paklausa, didinamas sektoriaus konkurencingumas.	Ekonominis poveikis darbo rinkai laikinas arba nereikšmingas; vartotojų kainų analizė rodo, kad sektoriaus konkurencingumas nebūtinai padidėja.	14,2 (0,03 %)
Knygos ir periodiniai leidiniai.	Prekių ir paslaugų paklausos ir pasiūlos skatinimas, įvertinant ilgalaikę naudą.	Daugelis šalių taiko šią lengvatą.	12,0 (0,03 %)

2 lentelėje parodyta 9 ir 5 procentų lengvatinio PVM tarifo prekės ir paslaugos bei jų poveikis. Daugelis šių lengvatų Lietuvoje pritaikytos dėl pridėtinės vertės mokesčio regresinio pobūdžio. Kitais atvejais skatinamas didesnis prekės ar paslaugos vartojimas. Tačiau matoma, kad šios priemonės ne visada suteikia norimą rezultatą ir yra itin brangiai kainuojančios valstybei.

Kitos išskirtinių PVM sąlygų grupės yra neapmokestinamos ir 0 % PVM tarifo prekės ir paslaugos. Pardavėjas, kuris parduoda PVM neapmokestinamas prekes ar paslaugas, turi sąlygas lygiai tokias pačias kaip ir įmonės, kurios nėra registruotos kaip PVM mokėtojai. Tai yra, pardavėjas negauna PVM iš pirkėjo ir neturi jo sumokėti valstybei, tačiau moka PVM sumą už pirktas žaliavas ir negali šio mokesčio atsiimti. Įmonės, parduodančios prekes su nuliniu tarifu PVM, apmokeština savo pirkėjus nuliniu tarifu (iš esmės PVM suma šioms prekėms visada yra lygi 0), tačiau žaliavoms taikomas įprastinis PVM tarifas ir sumokėta suma būna susigrąžinama [10]. Nulinio tarifo mokesčio, kaip ir bet kurio kito lengvatinio tarifo, poveikį junta tik galutinės tiekimo grandinės dalyvis. O neapmokestinamo PVM poveikis priklauso nuo to, kurioje tiekimo grandinės vietoje jis atsiranda [25].

Prekės ir paslaugos, kurios PVM neapmokestinamos, pagal Lietuvos Respublikos Seimo [5] 20-33 straipsnius:

- su sveikatos priežiūra susijusios prekės ir paslaugos;
- socialinės paslaugos ir susijusios prekės;
- švietimo ir mokymo paslaugos;
- kultūros ir sporto paslaugos;
- politinių partijų, profesinių sąjungų, religinių bendruomenių ir pelno nesiekiančių juridinių asmenų veikla;
- pašto paslaugos;
- radijas ir televizija;
- nepriklausomų grupių veikla;
- draudimo paslaugos;
- finansinės paslaugos;
- specialūs ženklai;
- azartiniai lošimai ir loterijos;
- nekilnojamųjų pagal prigimtį daiktų nuoma;
- nekilnojamųjų pagal prigimtį daiktų pardavimas ar kitoks perdavimas.

Prekės ir paslaugos, kurioms taikomas 0% PVM, pagal Lietuvos Respublikos Seimo [5] 41-51 straipsnius:

- prekių išgabėnimas iš Europos Sąjungos teritorijos;
- keleivių išgabėnamos prekės;

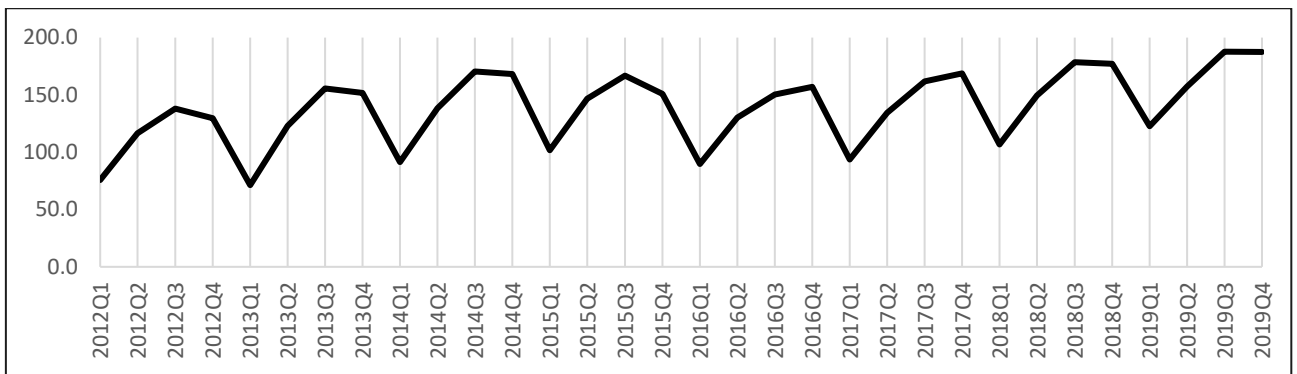
- laivų ir orlaivių tiekimas, nuoma, kt. sandoriai, susiję su laivais ir orlaiviais;
- laivų ir orlaivių atsargų tiekimas;
- vežimo paslaugos ir susiję sandoriai;
- draudimo ir finansinėms paslaugoms, susijusios su prekių eksportu;
- diplomatinėms atstovybėms, konsulinėms įstaigoms, Europos Sąjungos institucijoms, jų įsteigtoms įstaigoms ir tarptautinėms organizacijoms ar jų atstovybėms, taip pat šių atstovybių ir konsulinių įstaigų personalui ir jo šeimos nariams skirtos prekės ir paslaugos;
- aukso tiekimas Europos centrinių bankų sistemai ir Europos centriniam bankui;
- prekių tiekimas į kitą valstybę narę;
- prekių tiekimas paramos gavėjams, kai prekės yra išgabenamos iš Europos Sąjungos;
- kilnojamųjų daiktų aptarnavimas ir perdirbimas, kai tokie daiktai laikinai įvežti į Europos Sąjungą;
- atstovavimas tam tikruose aukščiau nurodytuose sandoriuose;
- tam tikrais su tarptautine prekyba susijusiais atvejais taikomos specialios taisyklės.

Nulinio tarifo PVM ar neapmokestinimas dažnai taikomas dėl tų pačių priežasčių kaip ir lengvatinis tarifas – mokesčio regresyvumo mažinimas ar tiesiog produktas ar paslauga yra pagirtina visuomenėje ir atneša valstybei ilgalaikės naudos. Be viso to, nulinis tarifas ar neapmokestinimas dar naudojamas ir prekėms ar paslaugoms, kurias tiesiog sudėtinga apmokestinti ir administravimo kaštai būtų potencialiai didesni nei gaunama nauda [10]. Neapmokestinamų prekių ar paslaugų įmonė gali iš viso nerinkti sąskaitų susijusių su pridėtinu vertės mokesčiu [25].

2019 metais Lietuvoje buvo išplėstas atvirkštinio apmokestinimo PVM mechanizmo taikymas. Pagal Lietuvos Respublikos Vyriausybę [26], šis apmokestinimo būdas naujai pritaikytas sritims, kuriose nustatyta daugiausiai sukčiavimo atvejų. Šiuo metu atvirkštinis apmokestinimas Lietuvoje taikomas [5]:

- nuosavybės teisės objektams;
- perduodamam pastato esminiam pagerinimui;
- statybos darbams;
- pardavimams, kai pardavėjui pradėta bankroto procedūra;
- juodųjų ir spalvotųjų metalų atliekoms ir laužui;
- tam tikrai medienai;
- mobiliesiems telefonams (nuo 2019 m.);
- planšetėms ir nešiojamiems kompiuteriams (nuo 2019 m.);
- standiesiems diskams (nuo 2019 m.);

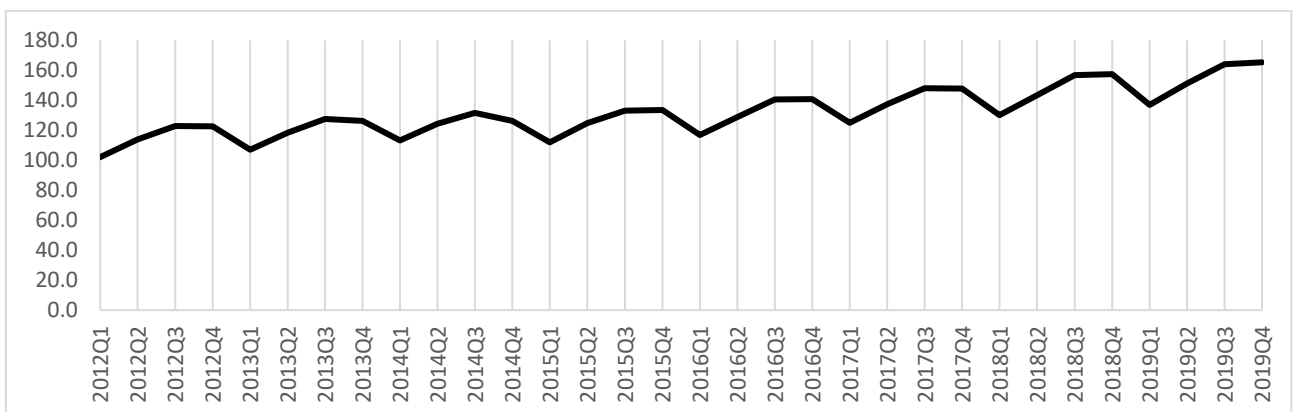
Statybos pramonėje taikomas atvirkštinis apmokestinimas parodo, kad šis sektorius yra rizikingas nemokėti PVM. Pagal prieš tai aptartas paskatas sukčiauti, galima daryti išvadą, kad šis sektorius labiau rizikingas dėl didelės pridėtinės vertės paskutiniame statybų paslaugos tiekimo etape. Statybų sektoriuje būdingiausias „sumažintos pardavimų sumos“ (žr. 1.2.1. skyrelyje) sukčiavimas.



3 pav. Statybų sektoriaus bendroji pridėtinė vertė Lietuvoje, indeksas: 2010 m. = 100 (Eurostat [28])

Trečiame paveiksle matoma, kad statybų sektoriaus pridėtinei vertei būdingas sezoniškumas. Pardavimai sumažėja pirmąjį ketvirtį. To priežastis gali būti statybų atviraime lauke stabdymas dėl nepalankių oro sąlygų. Bendra tendencija nuo 2012 metų yra didėjanti.

Kita, naujai suformuota rizikinga sritis, kurioje imtasi sukčiavimo mažinimo priemonių tik 2019 metais yra elektronikos prietaisai tokie kaip mobilieji telefonai, planšetės, nešiojami kompiuteriai ir standieji diskai. Šių prietaisų vertė yra itin didelė, todėl jie tapo vienais iš pagrindinių „karuselės“ (žr. 1.2.1. skyrelyje) sukčiavimo schemas įrankių.



4 pav. Didmeninės ir mažmeninės prekybos, transporto, apgyvendinimo ir maitinimo paslaugų sektorių bendroji pridėtinė vertė Lietuvoje, indeksas: 2010 m. = 100 (Eurostat [28])

Analizuojant elektronikos prietaisų PVM ataskaitas galima tikėtis mažesnio sezoniškumo nei statybų ar žemės ūkio sektoriuje (4 pav.). Todėl sukčiavimo schemas šiame sektoriuje yra dar pavojingesnės, kadangi galimai nėra labia aiškios įmonių elgsenos. Bendrosios pridėtinės vertės tendencija nuo 2012 metų didmeninės ir mažmeninės prekybos, transporto, apgyvendinimo ir maitinimo paslaugų sektoriuose yra didėjanti.

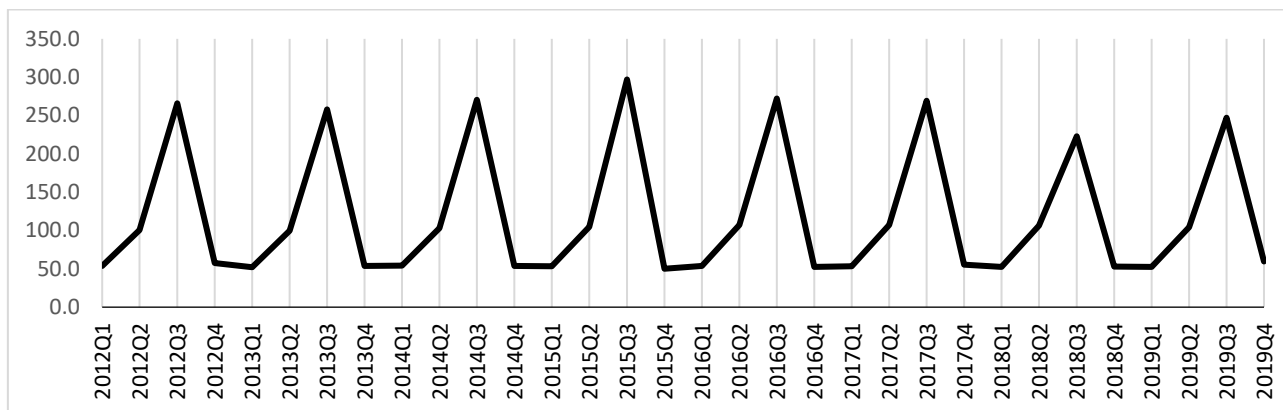
Daug diskusijų moksliniuose straipsniuose sukelia ir PVM tarifų taikymas žemės ūkio sektoriuje [10, 17, 27]. Dažniausiai vienaip ar kitaip šiam sektoriui yra suteikiamos tam tikros PVM lengvatos. Paprastai tai būna žaliavų lengvatinis tarifas, pavyzdžiui kuro. Pagal L. P. Ebrill'o, M. Keen'o, J. P. Bodin'o ir V. P. Summers'o [17] tyrimą, šalys, kuriose yra didesnė žemės ūkio dalis, įsivedusios PVM gauna palyginti mažesnes surenkamų mokesčių pajamas. To priežastis gali būti lengvatiniai tarifai ar agrarinių valstybių prastesnis administracijos išvystymas.

3 lentelė. PVM tarifai kai kuriose Europos Sąjungos šalyse, % [27]

Šalys	Lietuva	Lenkija	Latvija	Airija	J. Karalystė	Ispanija	Prancūzija	Nyderlandai	Belgija	Vokietija	Čekija	Austrija	Slovakija	Danija	Švedija	Estija	Suomija
Standartinis tarifas	21	22	21	21,5	15	16	19,6	19	21	19	19	20	19	25	25	18	22
Mėsa	21	3/7	21	0	0	4	5,5	6	6	7	9	10	19	25	12	18	17
Daržovės ir vaisiai	21	3	21	0	0	4	5,5	6	6	7	9	10	19	25	12	18	17
Pasikeitimai nuo 2009-01-01	+3		+3								+4						
Pasikeitimai žaliai mėšai, žuviai	+16		+3								+4						

Šešėlinę ekonomiką Lietuvos žemės ūkyje iš dalies paveikė 2009 metais panaikinti lengvatiniai tarifai žaliai mėšai, žuviai, daržovėms ir vaisiams. Pagal A. Gapšį ir O. Eičaitę [27], dėl šių tarifų pokyčio susidarė kainų skirtumas tarp kaimyninių valstybių. Atsirado paskata ir galimybės perparduoti atvežtinius produktus šešėlinėje ekonomikoje nemokant mokesčių. Todėl žemės ūkyje susidarė didžiausia mokesstinio sukčiavimo grėsmė mėsos ir vaisių ir daržovių sektoriuose.

Žemės ūkis išsiskiria mažesniu apskaitos sekimu. Be to, šiame sektoriuje yra didelis laiko skirtumas tarp žaliavų pirkimo ir tarp pardavimų, kas apsunkina PVM apskaičiavimą ir mokėjimą. Todėl žemės ūkis turi aukštus administravimo ir mokesčių mokėjimo kaštus [17]. Tą patį galima pamatyti ir 5 paveiksle, pardavimai daugiausiai vyksta trečiąjį ketvirtį, todėl ir bendra pridėtinė vertė yra ryškiai sezoninė. Bendra tendencija nuo 2012 metų išlikusi pastovi.



5 pav. Žemės ūkio sektoriaus bendroji pridėtinė vertė Lietuvoje, indeksas: 2010 m. = 100 (Eurostat [28])

1.4. Literatūros apžvalgos išvados

Iš literatūros apžvalgos matoma, kad pridėtinės vertės yra pakankamai efektyvus ir svarbus mokeskis šalies biudžetui. Tačiau, dėl savo struktūros sudėtingumo ir trūkumų kompensavimo įvairiomis lengvatomis, PVM sudaro galimybes sudėtingam mokesčiniam sukčiavimui. Norint šios struktūros dar labiau neapsunkinti ir neužkrauti didesnių mokėjimo kaštų mokesčio mokėtojams, būtina efektyviau identifikuoti rizikingas sukčiauti įmones.

Tikimybė sugauti sukčiaujančią įmonę didėja, kai efektyviai išskiriamos įmonės, kurioms patikrinimų nereikia. Be to, literatūros analizė parodė, kad šiuo metu Europos Sąjungoje yra skatinami tarpvalstybiniai mainai ir visų sandorių elektronizavimas, todėl surenkama daug informacijos elektroninių duomenų pavidalu. Todėl atsižvelgus į šiuos siūlymus ir turimos duomenų informacijos svarbą, šio darbo tyrime bus analizuojami pateiktų PVM ataskaitų duomenys ir ieškoma bendrų nerizikingų mokėtojų elgsenos bruožų, kartu išskitiant rizikingas sukčiauti įmones.

Pagal literatūros analizę, duomenų tyrimui atrenkamos aktualiausios ekonomikos veiklos PVM sukčiavimui tirti. Pirmiausia, tai yra statybų sektorius, kuris pasižymi didele pridėtine verte galutiniame tiekimo grandinės etape, todėl sukuria didesnes paskatas slėpti tikrąją pridėtinę vertę. Taip pat rizikingas sektorius yra elektronikos, kadangi juo dažniausiai naudojasi daug nuostolių padarančios „karuselės“ schemos dalyviai. Trečiasis parinktas tyrimui sektorius yra žemės ūkis. Šis sektorius prarado dalį rinkos dėl padidinto PVM tarifo ir kaimyninių šalių mažesnių tarifų žemės ūkio produktams, kas sukūrė tiek paskatas tiek galimybes sukčiauti.

2. Duomenų tyrimo metodai ir tyrimo objektas

Šis tyrimas skirtas panaudoti duomenų tyrybą kaip įrankį PVM sukčiavimui mažinti. Duomenų tyryba tai įdomių, netikėtų ar vertingų sandarų atradimas dideliuose duomenų rinkiniuose [42]. Šie atradimai įmanomi, nes naudojamos įvairių disciplinų apjungimu, tokių kaip dirbtinis intelektas, informacinės technologijos, mašininis mokymasis, duomenų rinkinių valdymas, duomenų vizualizavimas, matematiniai algoritmai ir statistika [43].

Pagal J. Han'ą, J. Pei ir M. Kamber'ą [44], duomenų tyryba klasifikuojama į dvi pagrindines atšakas: aprašomoji analitika ir prognozavimo analitika. Aprašomajai analitikai būdinga šablonų paieška neturint išankstinės informacijos apie duomenyse esančias struktūras. Tokiai analizei priskiriama dažnumo įvertinimas, klasterinė analizė ir segmentavimas, priklausomybių analizė [42]. Prognozavimo analitika, naudodama žinomą informaciją, nustato nežinomus kintamuosius. Tam tikslui naudojama klasifikacija (kategoriniams kintamiesiems prognozuoti) arba regresija (kiekybiniais kintamiesiems prognozuoti) [42]. Mokesčių administratoriai įvairiose šalyse šešėlinės ekonomikos mažinimui naudoja tiek aprašomąją, tiek prognozavimo analitiką (žr. 4 lentelę).

4 lentelė. Mokesčių administratorių taikomi duomenų tyrybos metodai mokesčio sukčiavimo aptikimui (P. C. Gonzalez'as, J. D. Velasquez'as [45])

Taikomas metodas	JAV	Kanada	Australija	JK	Bulgarija	Brazilija	Peru	Čilė
Neuroniniai tinklai	Taip	Taip	-	Taip	Taip	-	Taip	Taip
Sprendimų medžiai	Taip	Taip	Taip	-	-	-	Taip	Taip
Logistinė regresija	Taip	-	Taip	Taip	Taip	-	-	-
SOM	-	-	Taip	-	-	-	-	Taip
K-vidurkių	-	-	Taip	-	-	-	-	Taip
Atramos vektorių mašina	Taip	-	Taip	-	-	-	-	-
Vizualizavimas	Taip	-	-	-	-	Taip	-	-
Bajeso tinklai	-	-	Taip	-	-	-	-	-
K-artimiausių kaimynų	-	-	Taip	-	-	-	-	-
Asociacijų taisyklės	-	-	-	-	-	-	Taip	-
Neryškios taisyklės	-	-	-	-	-	-	Taip	-
Markovo grandinės	-	-	-	-	-	Taip	-	-
Laiko eilutės	-	Taip	-	-	-	-	-	-
Regresija	-	-	-	Taip	-	-	-	-
Simuliacija	Taip	-	-	-	-	-	-	-

Prognozavimo analitikos modeliai slepiančioms mokesčius įmonėms atrinkti dažniausiai būna kuriami jau ant audituotų įmonių [46], kad būtų aiškūs sukčiavimo atvejai. Tačiau audituojamos įmonės jau būna parinktos pagal išankstinį, dažniausiai ekspertų patirtimi grįstą, algoritmą [46]. Šio

darbo tyrimo tikslas išskirti taisykles, pagal kurias įmonės turėtų būti atrinktos auditavimui, todėl tinkamesni yra aprašomosios analitikos metodai, kurie gali būti taikomi visoms įmonėms, nesvarbu ar jos prieš tai buvo audituotos ar ne.

2.1. Aprašomosios analizės metodai

Pagal M. S. Khwaja, R. Awasthi ir J. Loeprick'ą [39], tipinės elgsenos analizei ir kartu rizikingų įmonių išskyrimo tikslui yra tinkamiausi neuroninių tinklų, sprendimų medžių ir klasterizavimo metodai.

2.1.1. Klasterizavimo metodai

Du pagrindiniai duomenų klasterizavimo būdai yra padalijimo ir hierarchinis [47]. Duomenų rinkinys padalijimo būdu skaidomas į nurodytą grupių kiekį. Stebėjimų priklausymas kuriai nors grupei paprastai nustatomas minimizuojant tam tikrą kriterijų ar paklaidą. Stebėjimai esantys tame pačiame klasteryje turi būti arti vienas kito (panašūs), o skirtingų klasterių stebėjimai kuo toliau nuo vienas kito (labai skirtingi). Kiekvienas stebėjimas pagrindinėje algoritmo versijoje yra priklausantis tik vienam klasteriui. Klasteriai būna sferinės formos [44]. Dažniausiai dalinimui naudojamas k-vidurkių metodas, o padalijimo algoritmas paprastai turi tokius žingsnius [47]:

1. Klasterių skaičiaus nustatymas.
2. Inicijuojami klasterių centrai.
3. Apskaičiuojamas duomenų padalinimas.
4. Perskaičiuojami klasterių centrai.
5. Jei padalinimas nesikeičia, sustojama. Kitu atveju kartojamas algoritmas nuo 3 žingsnio.

Pagal H. Xiong'ą, J. Wu, J. Chen'ą [48], k-vidurkių algoritmas bando sukurti sąlyginai tolygų stebėjimų kiekio paskirstymą klasteriuose. Todėl šis metodas nėra tinkamas iškreipto pasiskirstymo duomenims.

Hierarchinis metodas sukuria duomenų stebėjimų hierarchinę dekompoziciją. Tokia dekompozicija dažniausiai būna atvaizduojama dendograma. Hierarchinis metodas skirstomas į aglomeratyvius ir skaidančius algoritmus. Aglomeratyvaus būdo principas yra pradėti dalijimą nuo kiekvieno stebėjimo kaip atskiro klasterio, o skaidančio – pradėti dalinimą apjungus visus stebėjimus į vieną klasterį [44]. Aglomeratyvus algoritmas dažniausiai turi tokius žingsnius [47]:

1. Inicijuojamas kiekvieno vektoriaus priskyrimas atskiram klasteriui.
2. Apskaičiuojamas atstumas tarp kiekvieno klasterio.
3. Apjungiami du klasteriai, kurie yra arčiausiai vienas kito.
4. Kartojamas algoritmas nuo žingsnio numeris 2, kol lieka tik vienas klasteris.

Pagal J. Han'ą, J. Pei ir M. Kamber'ą [44], hierarchiniai klasterizavimo algoritmai yra nelankstūs, nes negali pakeisti ankstesniame žingsnyje jau sudaryto klasterio. Be to, skirtingi atstumo tarp klasterių nurodymai tam pačiam duomenų rinkiniui dažnai nulemia skirtingus klasterizavimo rezultatus. Atstumas tarp klasterių turi būti nurodomas prieš algoritmo inicijavimą.

Šiame darbe tyrimui atlikti pasirinktas aglomeratyvus hierarchinis klasterizavimas kaip tinkamiausias tikslui pasiekti. Naudojami Ward'o [49] ir grupės vidurkio [56] apjungimo metodai, kadangi abu šie metodai, pagal B. S. Everitt'ą [57], daugumoje atvejų atlieka klasterizavimą itin gerai [57]. Ward'o metodo esmė ta, kad klasteriai apjungiami minimizuojant paklaidų sumos kvadratą [50, psl. 30]:

$$D(c_1, c_2) = \delta^2(c_1, c_2) = (|c_1| \times |c_2|) / (|c_1| + |c_2|) \times \|c_1 - c_2\|^2 \quad (6)$$

Tai funkcija, kuri matuoja bendros kvadratų sumos pokytį apjungus c_1 ir c_2 stebėjimus.

Ward'o yra vienintelis metodas tarp aglomeratyvių, kuris suranda grupes su mažiausiomis vidinėmis dispersijomis kiekvieno dvinario apjungimo metu [51]. Pagal G. W. Milligan'o [52] ir G. W. Milligan'o, M. C. Cooper [53] tyrimus, šitas metodas puikiai suranda sferinius klasterius, tačiau prastai atlieką suskirstymą, kai stebėjimų kiekis klasteriuose nevienodas ar duomenyse egzistuoja išskirtys. Kadangi tyrimo duomenys būtent tokie (žr. 2.2. poskyrį) pasirinkta atlikti klasterizavimą ir su grupės vidurkio metodu (rezultatas pagerintas nuo 85,8% klasterių perskirstymo sprendimo medžiu atitikimo iki 90,63%). Todėl rezultatai pateikiami klasterizuojant grupės vidurkio metodu.

Grupės vidurkio metodas yra pripažintas kaip tinkamiausias hierarchinis algoritmas duomenims turintiems išskirčių [52, 53]. Tai metodas, kuris apjungia klasterius su minimaliu vidutiniu atstumu tarp objektų porų [54]. Tokiu būdu metodas bando suformuoti klasterių hierarchiją taip, kad visuose padalinimo lygiuose kiekvienas klasterio stebėjimas turėtų didesnę vidutinę panašumą su likusiais klasterio stebėjimais nei su kitų klasterių [55]. Kai klasteriai yra apjungti, suformuojant didesnę klasterį, atstumai tarp to naujo klasterio ir kitų klasterių yra apskaičiuojami pritaikant vienodą svorį kiekvienam stebėjimui, nepriklausomai nuo klasterių struktūros [56, psl. 52]:

$$D(X_i, X_j) = \frac{1}{|X_i||X_j|} \sum_{x \in X_i} \sum_{y \in X_j} d(x, y). \quad (7)$$

Tyrime naudojamas euklido atstumo matas skaitinėms reikšmėms ir tanimoto atstumo matas kategoriniams kintamiesiems. Euklido atstumas matuoja tiesią liniją tarp stebėjimų. Šis atstumas tarp objektų $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ ir $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ yra aprašomas [44, psl. 72]:

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}. \quad (8)$$

Euklido atstumo matas nėra tinkamiausias, kai reikalinga sugrupuoti kategorinius kintamuosius [58]. Todėl šiai užduočiai naudojamas Tanimoto atstumo matas. Šis matas yra cosinuso panašumo funkcijos variacija [44, psl. 78]:

$$\text{sim}(x, y) = \frac{x \times y}{x \times x + y \times y - x \times y}, \quad (9)$$

čia yra kelių bendrų x ir y kintamųjų ir kintamųjų priklausančių x arba y santykis.

Klasterių skaičiui nustatyti tyrime naudojami įvairūs (kl, ch, hartigan, ccc, scott, marriot, troww, tracew, friedman, rubin, cindex, db, duda, pseudot2, beale, ratkowsky, ball, ptbiserial, mcclain, gamma, gplus, tau, dunn, sindex, sdbw, silhouette, wss) metodai. Kiekvieno iš jų rekomenduojamas klasterių skaičius yra laikomas kaip vienas balsas. Renkantis tinkamą klasterių skaičių atsižvelgiama į metodų balsavimą ir į gaunamą duomenų padalinimo santykį.

2.1.2. Saviorganizuojantis neuroninis tinklas (SOM)

Saviorganizuojantis neuroninis tinklas (SOM), pristatytas T. Kohonen'o [59], yra vienas iš apmokymo be mokytojo metodų. Jo tikslas nustatyti statistines savybes be išskirtinio atsako iš mokytojo [60], tai yra metodas nereikalaujantis iš anksto pateikto galutinio atsakymo. Saviorganizuojantis neuroninis tinklas iš visų dirbtinių neuroninių tinklų algoritmų išsiskiria efektyviu įvesčių bruožų ir jų apibendrinimų pateikimu erdvėje [61].

Pagal Kohonen'ą [61], paprasčiausioje versijoje SOM algoritme tik viena ląstelė ar vietinė grupė ląstelių vienu metu duoda aktyvų atsaką įvestims. Atsako vietos tampa surikiuotos į prasmingą koordinacių sistemą skirtingiems įvesties bruožams tinkle. Erdvinė lokacija ar ląstelės koordinatės tinkle atitinka tam tikrą įvesties signalų modelio sritį. Kiekviena ląstelė ar ląstelių grupė yra lyg atskiras tos pačios įvesties išanalizuotojas. Todėl buvimas ar nebuvimas vietinio aktyvaus signalo interpretuoja įvestą informaciją. Saviorganizuojančio neuroninio tinklo algoritmą galima būtų aprašyti taip [62]:

1. Iš visų įvesties vektorių atsitiktinai parenkamas \vec{x} vektorius.
2. Nustatomas neuronas $c = i$, kur $\|\vec{x} - \vec{w}_i^{(t)}\|$ yra minimalus su visais i .

Čia $\vec{w}_i^{(t)}$ svorio vektorius neurone i su iteracijų skaičiumi t .

3. Kiekvienam neuronui i :

$$\vec{w}_i^{(t+1)} = \vec{w}_i^{(t)} + \alpha^{(t)} (\vec{x} - \vec{w}_i^{(t)}), \text{ jei } i \in N_c(t)$$

$$= \vec{w}_i^{(t)}, \text{ jei } i \notin N_c(t)$$

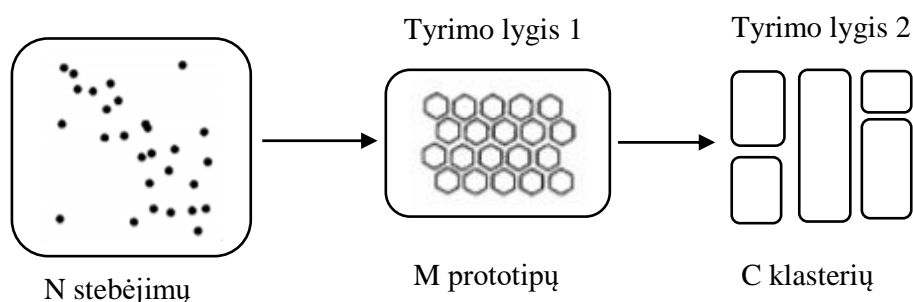
čia $\alpha^{(t)}$ yra maža dalis, naudojama konvergavimo kontrolei;

$N_c(t)$ neurono c kaiminystė;

4. Padidinti t ir grįži prie 1 žingsnio, kol sustabdymo sąlyga yra pasiekta.

Dėl savo savybių saviorganizuojantys neuroniniai tinklai gali būti priskiriami tiek prie klasterinės analizės metodų grupės, tiek prie dimensijos mažinimo metodų, tokių kaip pagrindinių komponentų analizė. SOM skirsto duomenis su tam tikra klasterių santykių tvarka, todėl šį metodą galima naudoti kaip įrankį skiriamajai analizei [62].

Šio darbo tyrimas atliekamas dviejų lygių sistema. Tai yra metodas, kai klasterizuojami ne pradiniai duomenys, o pirmiausiai atliekama saviorganizuojančių neuroninių tinklų analizė ir tuomet neuronai būna klasterizuojami [47]. 6 paveiksle galima matyti, kad, pirmiausiai naudojantis SOM algoritmu iš visų stebėjimų sugeneruojamas didelis kiekis prototipų. Kitu žingsniu šie prototipai būna apjungiami į klasterius. Kiekvienas originalių duomenų vektorius priklauso tam pačiam klasteriui kaip ir jo artimiausias prototipas.



6 pav. Dviejų lygių klasterizavimas naudojant saviorganizuojančius neuroninius tinklus (J. Vesanto ir E. Alhoniemi [47])

Dviejų lygių klasterizavimo vienas iš ryškiausių privalumų yra mažesni skaičiavimo kaštai. Net ir su pakankamai mažu stebėjimų kiekiu, daug klasterizavimo algoritmų (ypač hierarchiniai) pasidaro

sunkiai apdorojami. Todėl patogiau klasterizuoti jau sukurtus prototipus, o ne tiesiogiai duomenis [47].

Kitas privalumas yra sumažinamas duomenyse esantis triukšmas. Prototipai atspindi lokalius duomenų vidurkius, todėl yra mažiau jautrūs atsitiktinėms pradinį duomenų variacijoms [47]. Pagal P. Mangiameli, S. K. Chen'o ir D. West'o [63] tyrimą, toks metodas kaip SOM smarkiai pagerino hierarchinių metodų rezultatus klasterizuojant nenormalaus pasiskirstymo duomenis. Išskirtys SOM duomenų padalinime yra ne problema, kadangi yra mažai išskirčių taškų ir todėl jų įtaka vektoriaus skaičiavimo rezultatams yra nedidelė.

Be to, saviorganizuojantys neuroniniai tinklai kartu su klasterizavimu sudaro interpretuojamą dviejų dimensijų klasterių atvaizdą. Kiekvienas daug matmenų turintis duomenų vektorius yra nubraižomas kaip mažos dimensijos diskrečioji reikšmė, todėl lyginant šias reikšmes iš esmės galima palyginti originalius atstumus [64].

Dviejų lygių klasterizavimas yra pranašesnis už klasikinius klasterizavimo metodus tuo, kad naudoja pritaikomą atstumo matą. Pavyzdžiui, k-vidurkių ar isodata metodai klasterius, kurie yra per dideli dalina pusiau, o tuos, kurie yra per maži apjungia kartu iki norimo klasterių dydžio. Praktikoje, yra sudėtinga nustatyti tinkamą klasterių dydį ir paprastai dydis nebūna vienodas. Dviejų lygių klasterizavimas matuoja atstumą taip, kad būtų įtraukti visi klasterio taškai [64].

2.1.3. Sprendimų medis

Sprendimų medžio tikslas yra nustatyti kuo panašesnes stebėjimų grupes pagal nurodytus kintamuosius. Pavyzdžiui, klasifikacijos ir regresijos medžio (CART) technika, kuri yra viena iš dažniausiai naudojamų, grupes skirsto minimizuojant Gini indeksą, kuris matuoja grupės statistinę dispersiją [39].

Medžio sudarymo procesas prasideda nuo kintamojo, kuris, pagal indeksą, geriausiai padalina duomenis į dvi grupes, radimas. Kai duomenys padalinami, tas pats procesas taikomas kiekvienam pogrupiui. Tai kartojama iki kol sustabdymo sąlyga yra pasiekama [65]. Tokio proceso algoritmas gali būti aprašomas taip [44]:

1. Sukuriamas susikirtimo taškas N.
2. Jei duomenyse (D) stebėjimai yra tos pačios klasės (C), tai N pažymimas kaip lapas C klasės.
3. Jei kintamųjų sąrašas yra tuščias, tai N pažymimas kaip lapas pavadintas klase, kurios yra duomenyse daugiausiai.
4. Pritaikoma procedūra, nustatanti padalinimo kriterijų, kuris tinkamiausiai padalina duomenis į individualias klases.
5. Taškui N nurodomas tinkamiausias dalinimo kriterijus.
6. Jei padalinimo kintamasis susideda iš diskrečių skaičių ir pasikartojantys keliai leidžiami (nėra apribojama tik dvinariu medžiu), tai galimi kintamieji \leftarrow galimi kintamieji – padalinantis kintamasis (panaikinamas padalinantis kintamasis).
7. Kiekvienam padalinimo kriterijaus rezultatui j (padalina stebėjimus ir išaugina šakas kiekvienam padalinimui), tarkim D_j yra grupė duomenų stebėjimų, kur D patenkinamai padalintas rezultatu j, jei D_j yra tuščias, tai prie N prijungiamas lapas pavadintas daugumos klase. Kitaip, prijungti susikirtimo tašką, sukurtą algoritmo su D_j ir kintamųjų sąrašu, prie taško N.
8. Sukurti N.

Pagal M. Khanbabaie ir M. Alborzi [71], pagrindiniai sprendimų medžio trūkumai yra godumas medžio augimo procese ir vietinis optimizavimas kiekviename žingsnyje atskiriant šakas, be to, yra kitų problemų su nestabilumu ir šališkumu parenkant padalinimo kriterijų. Taip pat, tendencija sukonstruoti per didelius medžius ir permokėti apmokymo imtį ir perdidelio apibendrinimo problema.

Vienas iš esminių klausimų sudarinėjant sprendimo medžius, kokia yra sustabdymo sąlyga. Jei medis suformuojamas iki paskutinių lapų, gali atsirasti anksčiau minėta permokymo problema. Tai yra, sprendimų medis sudarys idealius ar beveik idealius rezultatus apmokymo imtyje, tačiau pritaikius tokį modelį kitiems stebėjimams, tinkamų rezultatų jis nepateiks. Tokiai problemai spręsti dažniausiai taikomas medžio apkarpymas ar genėjimas kryžminio validavimo principu [65].

Sprendimų medžiai dažniausiai naudojami klasifikavimo arba regresijos uždaviniams, kai reikia nustatyti vieno kintamojo įverčius. Tačiau, kadangi sprendimų medis natūraliai padalina duomenis, galima sakyti, kad medis yra klasterių hierarchija [66]. Panašias išvalgas pateikia ir Langley [69], kad kiekvienas sprendimų medžio taškas atitinka klasterio konceptą ir pats medis atitinka tam tikrą hierarchiją. Tokias hierarchijas rezultate sudaro ne tik sprendimų medžiai, bet ir kai kurie klasterizavimo algoritmai, pavyzdžiui COBWEB [70]. D. Yook'o [68] tyrimas parodė, kad sprendimų medis, adaptuotas klasterizuoti be mokytojo, susitvarko tiek su duomenų nepakankamumu, tiek su norimo rezultato radimu.

Pagal J. Basak'ą ir R. Krishnapuram'ą [67], sprendimų medžio vienas iš privalumų yra aiški interpretacija. Tai yra, sprendimai gali būti parodomi kaip taisyklių rinkinys. Atsišakojimo sprendimas kiekviename taške yra nustatomas pagal tam tikrą kintamąjį ir pasirinkimas to kintamojo yra pagrįstas tam tikru dalinimo kriterijumi. Kiekvienas lapo taškas medyje nurodo klasę ir yra interpretuojamas pagal kelią nuo kamieno iki pat to lapo.

Kitaip nei sprendimo medyje, hierarchinio klasterizavimo algoritmai kiekviename dendogramos lygyje atspindi tik klasterio skiriamąją gebą ar detalumą ir, norint interpretuoti hierarchiją, klasteriai kiekviename taške turi būti analizuojami atskirai [67]. Todėl, sprendimo medžio taisyklės papildė šio darbo tyrimą apie tipinio PVM mokėtojo (didžiausio klasterio) savybes.

Be to, remiantis pateikta teorija, darbo tyrime sprendimų medis naudojamas klasterių atskyrimo su saviorganizuojančiais neuroniniais tinklais ir hierarchiniu klasterizavimu vertinimui. Tyrime taikomas klasifikavimo sprendimų medis, kur prognozuojamas kintamasis yra SOM ir hierarchiniu klasterizavimu suformuoti klasteriai. Naudojamas Gini indekas kaip padalinimo kriterijus.

2.2. Tyrimo objektas

2.2.1. Imties ir kintamųjų parinkimas

Tyrime analizuojami PVM mokėtojų pateikti deklaracijų duomenys. PVM deklaraciją sudaro 36 laukeliai. 1-10 laukeliuose pateikiama informacija apie mokėtoją, mokesčinį laikotarpį, PVM deklaracijos rūšį. 11-14 laukeliuose nurodoma įvairaus apmokestinamumo patiektų prekių ir suteiktų paslaugų apmokestinama vertė. 15 laukelyje - turto apmokestinamoji vertė, 16 laukelyje - apmokestinamoji vertė sandorių, kuriems taikoma speciali apmokestinimo schema. 17-19 laukeliuose - patiektų eksportui prekių ir suteiktų paslaugų su 0 % PVM, išskirstant į ES ir iš ES išvežtas prekes, apmokestinamoji vertė. 20-24 laukeliuose - kitų pardavimo sandorių su ES įsikūrusiomis įmonėmis apmokestinamoji vertė. 25-27 - laukeliuose importo PVM. 28 laukelyje - proporcinis PVM atskaitos

procentas. 29-35 laukeliuose – apskaičiuotos PVM sumos, išskirstytos pagal tarifus, susiejimą su užsieniu, turtą ir atskaitą. 36 laukelyje pateikiama bendra mokėtina arba grąžintina PVM suma [81].

Tiriami kintamieji iš PVM deklaracijos parenkami pagal literatūros analizėje pateiktus kitų tyrimų pavyzdžius ir rekomendacijas. Pirmiausiai, analizuojami skaitiniai kintamieji: bendri pardavimai („Sum of Pard“), bendri pirkimai („Sum of Pirk“), pardavimai Lietuvoje („Sum of PardLT“), pirkimai Lietuvoje („Sum of PirkLT“), PVM suma („Sum of d36_PVM_SUMA“).

Be to, iš deklaracijoje pateiktų duomenų išvesti papildomi rodikliai. Apskaičiuotas bendrų pirkimų ir bendrų pardavimų santykis („Pirk/Pard“). Apskaičiuotas pirkimų Lietuvoje ir pardavimų Lietuvoje santykis („PirkLT/PardLT“). Naudojami PVM sumos svyravimai („Var of d36_PVM_SUMA“) ir atskaitos svyravimai („Var of d35_PVM_ATSKAITA“). Apskaičiuotas efektyvus PVM tarifas, tai yra, apskaičiuotas PVM sumos ir prekių tiekimo ir paslaugų teikimo sandorių santykis („R1“).

Literatūros apžvalgoje pastebėta, kad rekomenduojama tirti ne tik deklaracijoje pateiktus duomenis, bet ir išorinius kintamuosius. Todėl tyrime deklaracijų duomenys papildomi įmonės darbuotojų skaičiumi. Apskaičiuota bendrų pardavimų vienam darbuotojui suma („Pard/DARBSK“).

Duomenų apimtis yra nuo 2011 metų sausio mėnesio iki 2016 metų gruodžio mėnesio. Tyrimas orientuotas į daugumos įmonių panašių PVM mokėjimo tendencijų atradimą. Todėl svarbu atskirai tirti skirtingas ekonomines veiklas. Literatūroje išryškėja sritys, į kurias reikia atkreipti didesnę mokesčių administratoriaus dėmesį, nes jose matomi daugiau ar didesnių mastų sukčiavimai. Dėl darbo apimties pasirinkta tirti tris ekonomines veiklas.

Pirmoji veikla yra pardavimai susiję su žemės ūkio veikla. Duomenyse tokie stebėjimai priklauso prie kategorijų: „Žemės ūkio mašinų, įrangos ir reikmenų didmeninė prekyba“, „Žemės ūkio žaliavų, gyvų gyvulių, tekstilės žaliavų ir pusegaminų pardavimo agentų veikla“ ir „Žemės ūkio žaliavų ir gyvų gyvulių didmeninė prekyba“. Tokių stebėjimų yra 331.

Antroji tiriama veikla siejama su statybų sektoriumi. Duomenyse tokie stebėjimai priklauso prie kategorijų: „Medienos, statybinių medžiagų ir sanitarinių įrenginių didmeninė prekyba“, „Statybinio miško ir statybinių medžiagų pardavimo agentų veikla“, „Kasybos, statybos ir statybos inžinerijos mašinų didmeninė prekyba“. Tokių stebėjimų yra 1776.

Trečioji veikla yra susijusi su tomis elektronikos prekėmis, kurios dažniausiai naudojamos karuselės sukčiavimo schemeje. Stebėjimai duomenyse priskiriami prie kategorijų: „Elektroninės ir telekomunikacinės įrangos ir jos dalių didmeninė prekyba“, „Informacijos ir ryšių technologijų (IRT) įrangos didmeninė prekyba“, „Kompiuterių, jų išorinės ir programinės įrangos didmeninė prekyba“. Tokių stebėjimų yra 577.

Visi įmonių kintamieji yra agreguoti ketvirčiais, kad patikrinti sezoninius PVM mokėtojų panašumus ir skirtumus. Be to, kadangi duomenų rinkinyje didelės dalies įmonių pateiktų mėnesinių duomenų datos nesutampa, kiekvienos įmonės sumos dalintos iš ketvirtyje turimų duomenų mėnesių kiekio. Todėl visų kintamųjų pateiktos ir naudojamos reikšmės yra mėnesinės. Parenkamos tik tos įmonės, kurių duomenys buvo visuose ketvirčiuose.

Iš žemės ūkio prekių pardavimų veiklos duomenų suvestinės matoma, kad per mėnesį vidutiniškai (dėl išskirčių stebimi mediana ir kvantiliai) daugiausiai pardavimų ir pirkimų tiek Lietuvoje tiek iš viso būna antrą ketvirtį. Antrą ketvirtį žemės ūkio žaliavų didmeninėms parduoda daugiausiai, nes

vyksta sėjimai ir pasiruošimas derliaus nuėmimui, tai sutampa su teorija (žr. 1.3 poskyryje), kad žemės ūkis sukuria didžiausią pridėtinę vertę trečiąjį ketvirtį (dėl didelių pirkimų antrą ketvirtį ir derliaus nuėmimo sezoniškumo Lietuvoje). Todėl ir pardavimai vienam darbuotojui vidutiniškai didžiausi antrą ketvirtį.

Taip pat, išsiskiria žemės ūkio didmenininkų pirkimų ir pardavimų (Lietuvoje ir bendras) santykio mediana, kuri pirmą ketvirtį yra artimiausia vienetui. Tai reiškia, kad pirmąjį ketvirtį pirkimai sąlyginai padidėja, o pardavimai sąlyginai sumažėja vienas kito atžvilgiu. Šią informaciją papildė ir pirmame ketvirtyje vidutiniškai mažiausias efektyvus PVM tarifas. Todėl ir PVM suma vidutiniškai pirmąjį ketvirtį mažiausia. Pirkimų Lietuvoje ir pardavimų Lietuvoje santykis vidutiniškai mažesnis už bendrų pirkimų ir bendrų pardavimų santykį. Tai reiškia, kad įmonės sąlyginai mažiau perka Lietuvoje ir daugiau parduoda Lietuvoje.

Statybų sektoriaus prekių pardavimo suvestinėje pastebima, kad vidutiniškai (pagal medianą) mažiausiai pirkimų ir pardavimų (bendrų, Lietuvoje ir vienam darbuotojui) būna pirmąjį ketvirtį. Teorinėje dalyje matoma, kad kiekvienais metais pirmąjį ketvirtį pastebimas bendrosios pridėtinės vertės statybų sektoriuje sumažėjimas. Tai galima paaiškinti sezoniškumu Lietuvoje, kai šaltaisiais mėnesiais dalis lauko statybos darbų negalimi.

Iš visų tiriamų veiklų, su statybų sektoriumi susijusių prekių pardavimų efektyvus PVM tarifas (R1) vidutiniškai (pagal medianą) yra mažiausias. Palyginus kitus rodiklius, pastebima, kad šios veiklos pirkimų ir pardavimų Lietuvoje santykis vidutiniškai yra didžiausias. Taip pat nežymiai vidutiniškai didesnis yra bendrų pirkimų ir pardavimų santykis. Tai reiškia, kad R1 šioje veikloje yra vidutiniškai mažesnis dėl mažesnio skirtumo tarp pirkimų ir pardavimų. Todėl ir PVM sumos mediana visais ketvirčiais yra mažesnė.

Elektronikos prekių pardavimo veiklos suvestinė parodo, kad pardavimai ir pirkimai (Lietuvoje ir bendri) yra vidutiniškai (pagal medianą) mažiausi pirmą ketvirtį, o ketvirtą ketvirtį yra vidutiniškai didžiausi. Pirmo ketvirčio susitraukimas yra sutampantis su bendros pardavimų pridėtinės vertės tendencija (žr. 1.3 poskyris).

Vidutiniai elektronikos didmeninių pardavimų įmonių PVM atskaitos svyravimai yra mažiausi antrą ketvirtį. O tų pačių įmonių vidutiniai PVM sumos svyravimai antrą ketvirtį yra beveik didžiausi (pastebima, kad didesni tik ketvirtą ketvirtį). Tai reiškia, kad antrą ketvirtį įmonių pirkimų suma kiekvienais metais yra itin pastovi, o pardavimų suma besikeičianti.

Visų duomenų kintamųjų mediana nuo vidurkio pastebimai skiriasi. Tokia tendencija pastebima visuose ketvirčiuose. Todėl tikėtina, kad duomenyse egzistuoja išskirtys, o kintamųjų skirstiniai nėra normalūs.

Metodai taikomi skaitiniams kintamiesiems, kurie turi vienodą mastelį. Duomenys yra centruojami – iš kiekvieno stebėjimo reikšmės atimamas kintamojo vidurkis. Centruoti duomenys yra padalinami iš jų standartinio nuokrypio [72].

Suskirsčius klasteriais skaitinius kintamuosius, pridedami kategoriniai kintamieji, su tikslu tikrinti ar skiriasi stebėjimų panašumai pagal įmonės dydį ir vietovę. Vietovė duomenyse pažymėta kintamuoju „Apskritis“, o įmonės dydis nustatomas pagal darbuotojų skaičių ir žymimas „Darb_sk._gr“. Mikro kategorijai priskiriamos įmonės turinčios nuo 1 iki 9 darbuotojų imtinai, mažų kategorijai – nuo 10

iki 49 darbuotojų imtinai, vidutinių kategorijai – nuo 50 iki 249 darbuotojų imtinai, o didelių kategorijai – virš 250 darbuotojų.

2.2.2. Išskirtys duomenyse

Pagal J. Han'ą, J. Pei ir M. Kamber'ą [44], išskirtys gali būti globalios, kontekstinės ir kolektyvinės. Globalios išskirtys tai tokios kintamųjų reikšmės, kurios reikšmingai nukrypsta nuo visų kitų. Mokesčių sukčiavimo atveju, globalios išskirties pavyzdys galėtų būti pervedimai atlikti ne pagal įstatymus.

Kontekstinės išskirtys yra tokios kintamųjų reikšmės, kurių nukrypimas nuo visumos priklauso nuo bendro duomenų konteksto. Pavyzdžiui, įmonės, kuri užsiima didmenine prekyba, pardavimų suma gali būti išskirtis, jei visi kiti duomenys yra mažmeninio pardavimo įmonių, bet nebus išskirtis, jei visos kitos pardavimų sumos yra didmeninės prekybos.

Kolektyvinės išskirtys, tai keli stebėjimai, kurių reikšmės bendrai nukrypę nuo visų kitų. Šie stebėjimai atskirai po vieną gali ir neišsiskirti, tačiau visi, bendrai kaip grupė, tampa išskirtimi. Mokestiniame sukčiavime tai galėtų būti kelios tos pačios veiklos ir tokiu pačiu metodu sukčiaujančios įmonės. Tokios įmonės dažnai neoficialiai susijusios ir bendru vadovu.

Duomenyse su daug kintamųjų vienas iš metodų rasti išskirtis yra Mahalanobio atstumas [82]. Tokio metodo principas yra matuoti atstumą nuo kintamųjų centroidų daugiamatėje aplinkoje. Mahalanobio atstumas suvienodina kintamųjų mastelį ir centruoja duomenis. Tokiu būdu gautos išskirtys yra pažymėtos 4, 5 ir 6 prieduose.

Žemės ūkio pardavimų veikloje aptikta 53 išskirtys, 13 iš jų aptinkamos visuose klasteriuose, tiriant kiekvieną klasterį atskirai. Klasterizavimo tyrimo metu trys išskirtys, priskirtos tipinių PVM mokėtojų grupei, o netipinių įmonių išskirta 22, kurios nebuvo Mahalanobio atstumu parinktos kaip išskirtys.

Su statybų sektoriumi susijusių pardavimų veiklos įmonėse aptiktos 39 išskirtys, 20 iš jų aptinkamos visuose klasteriuose, tiriant kiekvieną klasterį atskirai. Klasterizavimo tyrimo metu šešios išskirtys, priskirtos tipinių PVM mokėtojų grupei, o netipinių įmonių išskirta 215, kurios nebuvo Mahalanobio atstumu parinktos kaip išskirtys.

Elektronikos prekių pardavimų veiklos įmonėse aptikta 40 išskirčių, 11 iš jų aptinkamos visuose klasteriuose, tiriant kiekvieną klasterį atskirai. Klasterizavimo tyrimo metu keturios išskirtys, priskirtos tipinių PVM mokėtojų grupei, o netipinių įmonių išskirta 99, kurios nebuvo Mahalanobio atstumu parinktos kaip išskirtys.

Išskirtis gali aptikti klasterizavimas. Reikšmės ar grupės reikšmių, kurios atsiranda toli nuo klasterio, yra laikomos išskirtimis. Tokio išskirčių aptikimo pagrindinė prielaida yra ta, kad normalūs stebėjimai išlaiko tam tikrą struktūrą daug dažniau nei išskirtys [44]. Šio darbo tyrimas taip pat remiasi tokia prielaida. Todėl duomenyse paliekamos galimos išskirtys, o rezultate aptinkama normalių stebėjimų struktūra, kuri būdinga nesukčiaujantiems PVM mokėtojams. Išskirti netipiniai mokėtojai palyginami su prieš tai aprašytais išskirtimis.

2.3. Naudota įranga

Tyrime naudojama programa R [73] dėl jos pakankamų galimybių ir prieinamumo. R yra statistinio skaičiavimo ir grafinio atvaizdavimo programa. Ji suteikia programavimo kalbą, geros kokybės grafinius sprendimus, sąsajas su kitomis kalbomis ir derinimo galimybes [74].

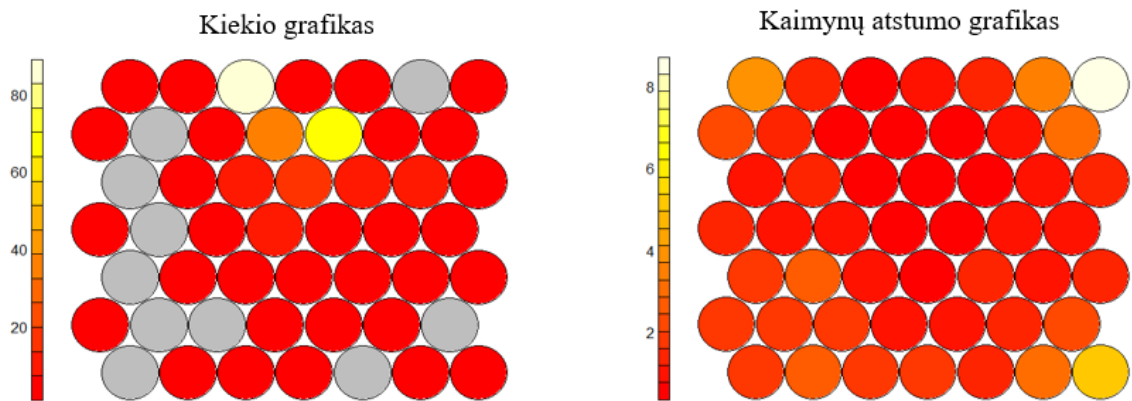
Tai atviro kodo programa kurioje papildymus gali kurti visi. Tyrime naudotas saviorganizuojančių neuroninių tinklų algoritmas pagal R. Wehrens ir J. Kruisselbrink'ą [75]. Su šiuo paketu galima pateikti kelis duomenų sluoksnius, kurie būtų apdoroti apskaičiuojant bendrus atstumus, pagal kuriuos būtų sudėliojamas tinklas. Išskirtims, pagal Mahalanobio atstumą, surasti naudotas chemometrics algoritmas [83]. Dėl R kodo paprastumo naudojami paketai pagal H. Wickham'ą ir kitus [76] ir pagal S. M. Bache ir H. Wickham'ą [77]. Naudota spalvų paletė pagal E. Neuwirth ir R. Brewer [78]. Duomenims įkelti naudotas H. Wickham'o ir kitų [79] sukurtas paketas.

3. Tyrimų rezultatai ir jų aptarimas

Atlikto tyrimo struktūra yra:

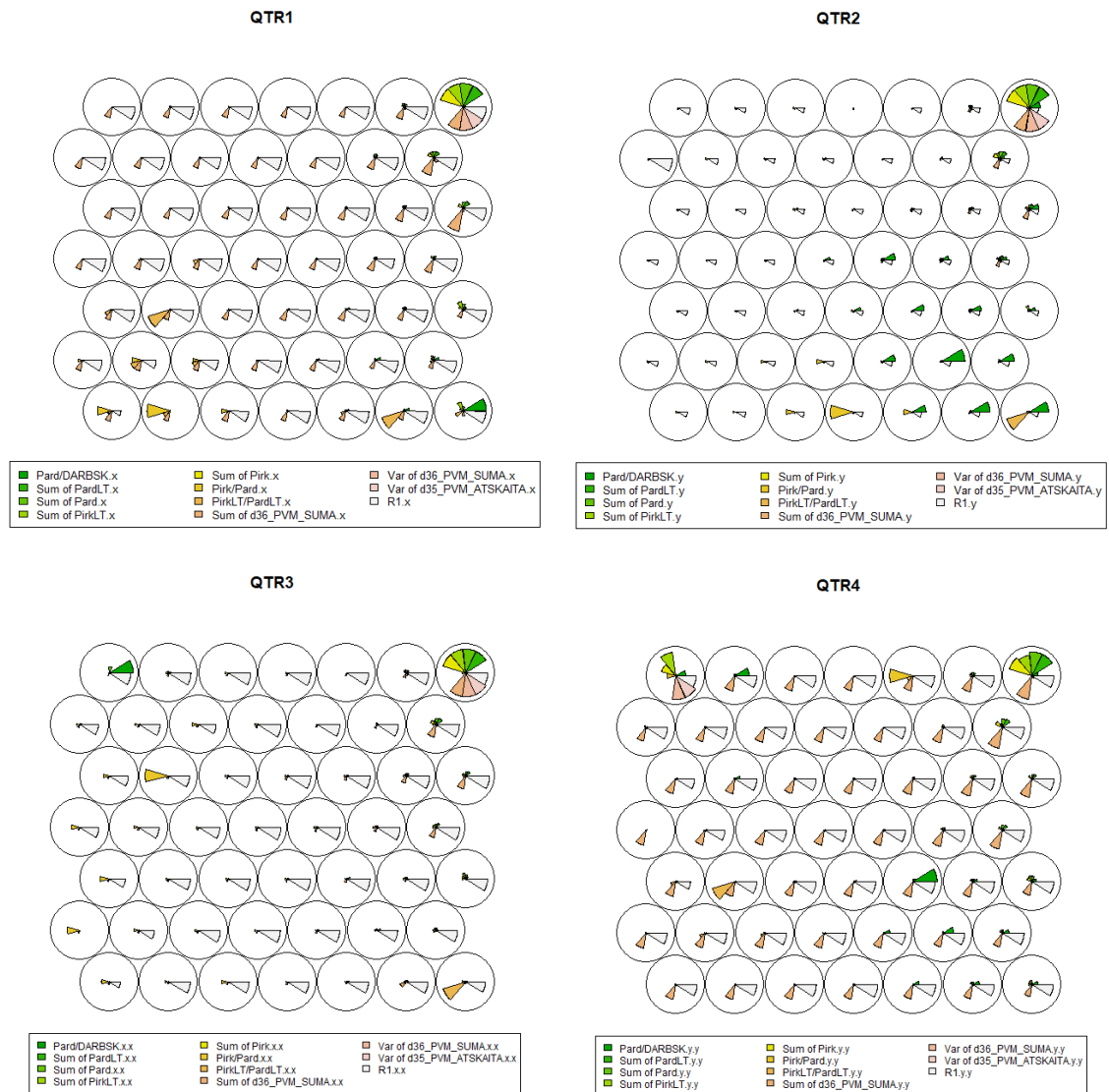
1. Duomenų gavyba iš Valstybinės mokesčių inspekcijos (pagal 2.2.1. skyrelį).
2. Duomenų paruošimas (pagal 2.2.1. skyrelį). Duomenų išvalymas ir agregavimas, kad atspindėtų reali situacija, nepriklausoma nuo trūkstamų reikšmių. Papildomų kintamųjų pridėjimas ir apskaičiavimas.
3. Duomenų apžvalga (aprašyta 2.2.1. skyrelyje). Susipažįstama su duomenų struktūra, kintamųjų reikšmių suvestine, bendra situacija ekonominėje veikloje. Sudaroma pirminė nuomonė, kaip kintamieji susiję, atskiriamas priežastingumas nuo atsitiktinumo.
4. Išskirčių radimas (pagal 2.2.2. skyrelį). Sudaromas pirminis netipinių PVM mokėtojų sąrašas.
5. Klasterių nustatymas (pagal 2.1.1. ir 2.1.2. skyrelius). Įmonės suskirstomos į tam tikrą grupių skaičių, kuris parenkamas pagal klasterių skaičiaus nustatymo metodų rekomendacijas ir klasteryje priskiriamų įmonių skaičiaus proporcijas.
6. Tipinių ir netipinių PVM mokėtojų elgsenos nustatymas:
 - Didžiausio(-ių) klasterio(-ių) įmonės priskiriamos prie tipinių PVM mokėtojų. Priklausomai nuo duomenų pasiskirstymo, tipiniams mokėtojams priskiriama 70-80% visų įmonių, todėl tai gali būti vienas arba keli klasteriai. Išskiriamos tipiniams mokėtojams būdingos kintamųjų reikšmės, kurios lyginamos su bendromis ekonominės veiklos reikšmėmis;
 - įmonės esančios kituose klasteriuose priskiriamos prie netipinių PVM mokėtojų (potencialiai sukčiaujančių įmonių). Išskiriamos netipiniams mokėtojams būdingos kintamųjų reikšmės, kurios palyginamos su bendromis ekonominės veiklos reikšmėmis. Išsiskiriančios reikšmės analizuojamos bendrame kontekste ir sprendžiama, ar tokia situacija gali susidaryti dėl kitos priežasties nei sukčiavimas. Plačiau analizuojamas efektyvus PVM tarifas: grafiškai stebima įmonių elgsena pagal klasterius ir ketvirčius. Neigiamai išsiskiriančios įmonės priskiriamos prie rizikingų. Visi netipiniai mokėtojai ir rizikingos įmonės pateiktos 4, 5 ir 6 prieduose.
7. Tipinės ir netipinės elgsenos įmonių atrankos validavimas:
 - sprendimų medžio taisyklės (pagal 2.1.3. skyrelį). Apžvelgiamos sprendimų medžio sudarytos didžiausios šakos taisyklės, kurios apibūdina tipinę PVM mokėtojo elgseną. Palyginamas medžio stebėjimų suskirstymas su klasterizavimo suskirstymu, jei sutampa daugiau nei 85% rezultatų, laikoma, kad klasteriai yra pastovūs;
 - netipinės elgsenos įmonės lyginamos su rastomis išskirtimis. Palyginami 6 punkto antros dalies rezultatai su 4 punkto rezultatais.
8. Rezultatų išsidėstymas pagal apskrities ir įmonės dydžio kintamuosius. Tikrinamas priežastinis klasteryje išsiskiriančių skaitinių kintamųjų ryšys su klasteriams būdingomis kategorinių kintamųjų reikšmėmis.

3.1. Žemės ūkio prekių pardavimų veiklos tyrimo rezultatai



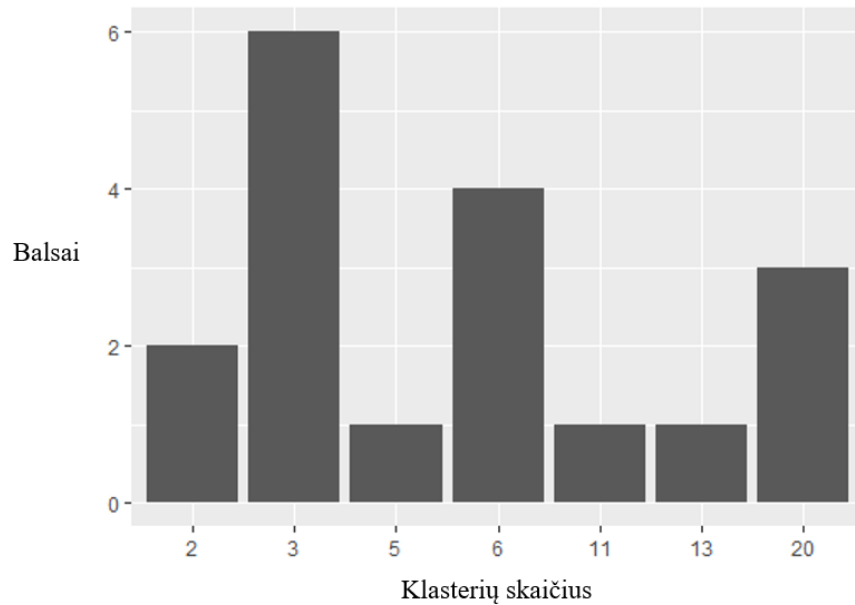
7 pav. SOM rezultato kiekių ir atstumų grafikai

Atliktas saviorganizuojančių neuroninių tinklų algoritmas suskirstė stebėjimus, pagal jų panašumą, taip, kad jie būtų atvaizduojami dvimačiame 49 pozicijų žemėlapyje. 7 paveiksle matoma, kad daugumoje pozicijų stebėjimų kiekis yra panašus (mažiau nei 20), trys pozicijos išsiskiria dideliu (virš 40) priskirtų stebėjimų kiekiu, o 9 pozicijos liko tuščios. Visos pozicijos, išskyrus dvi, yra sąlyginai arti viena kitos.



8 pav. SOM rezultato kintamųjų sąlyginiai dydžiai suskirstyti ketvirčiais

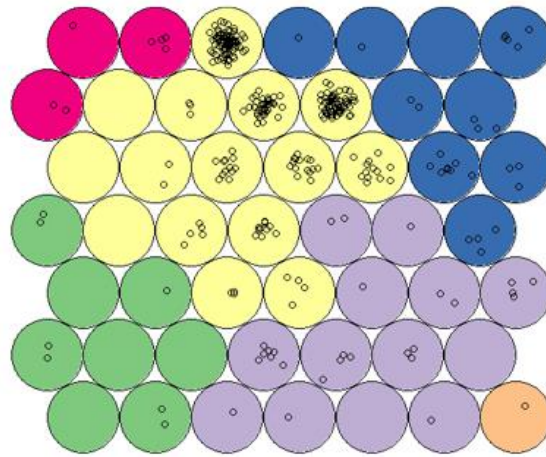
Saviorganizuojančių neuroninių tinklų algoritmo rezultate pastebimi bendri stebėjimų skirtumai skirtingais ketvirčiais. Antrame ketvirtyje pastebimas daugumos įmonių efektyvaus PVM tarifo (R1) rodiklio sąlyginis sumažėjimas. Pagal absoliučius rodiklius, matoma, kad R1 didžiausią reikšmę pasiekia būtent antrame ketvirtyje. Todėl sąlyginis dydis ir parodomas mažesnis. Taip pat pastebima, kad antro ketvirčio metu dalyje pozicijų padidėja sąlyginiai pardavimai vienam darbuotojui. Kitais ketvirčiais dauguma tų pačių pozicijų įmonių neišsiskiria šiuo rodikliu.



9 pav. Teorinių metodų balsai už optimalų klasterių skaičių

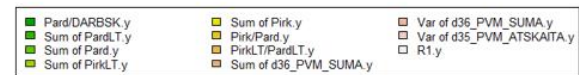
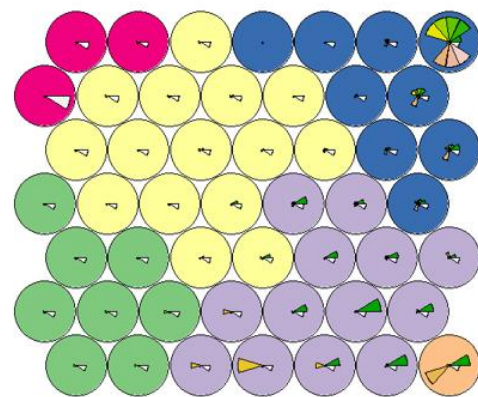
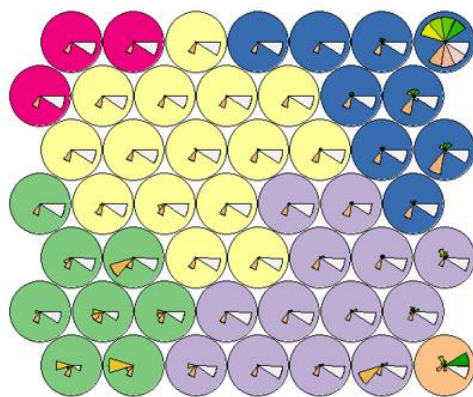
SOM rezultato objektus skirstant į klasterius daugiausiai teorinių metodų parodė, kad optimalus klasterių skaičius yra 3 (žr. 9 paveikslą). Tačiau, suskirsčius duomenis į tris klasterius, didžiausiam klasteriui priskiriami 83,38 % visų stebėjimų. Pagal teorijos apžvalgą, audituojamos turėtų būti apie 25-30 % rizikingų įmonių, o ir sukčiaujančios įmonės tikėtina bus nebūtinai prie labiausiai išsiskiriančių. Todėl, pagal prie didžiausio klasterio priskirtų stebėjimų proporcijas ir teorinių modelių rekomendacijas, klasterių analizei parenkami 6 klasteriai.

Klasteriai ir stebėjimų kiekis



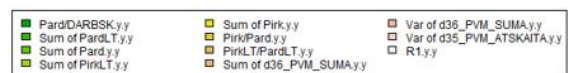
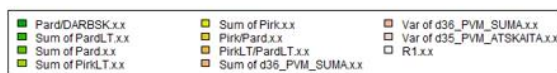
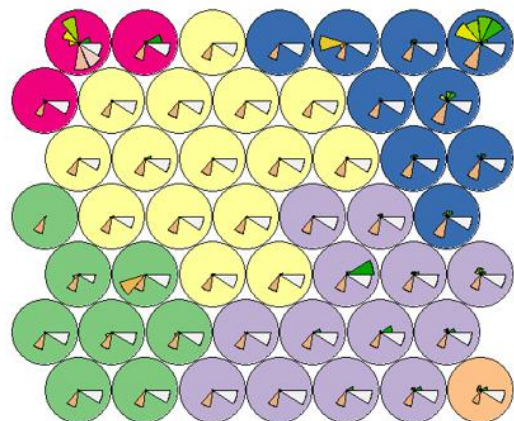
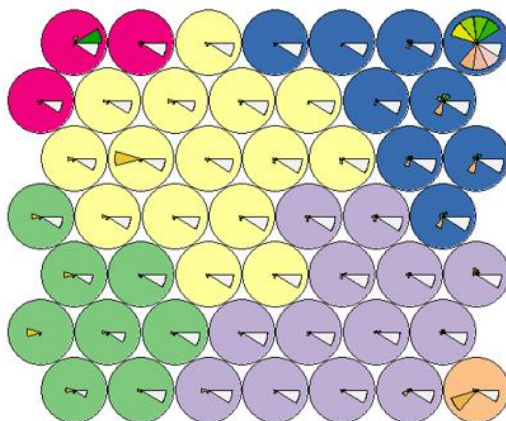
QTR1

QTR2



QTR3

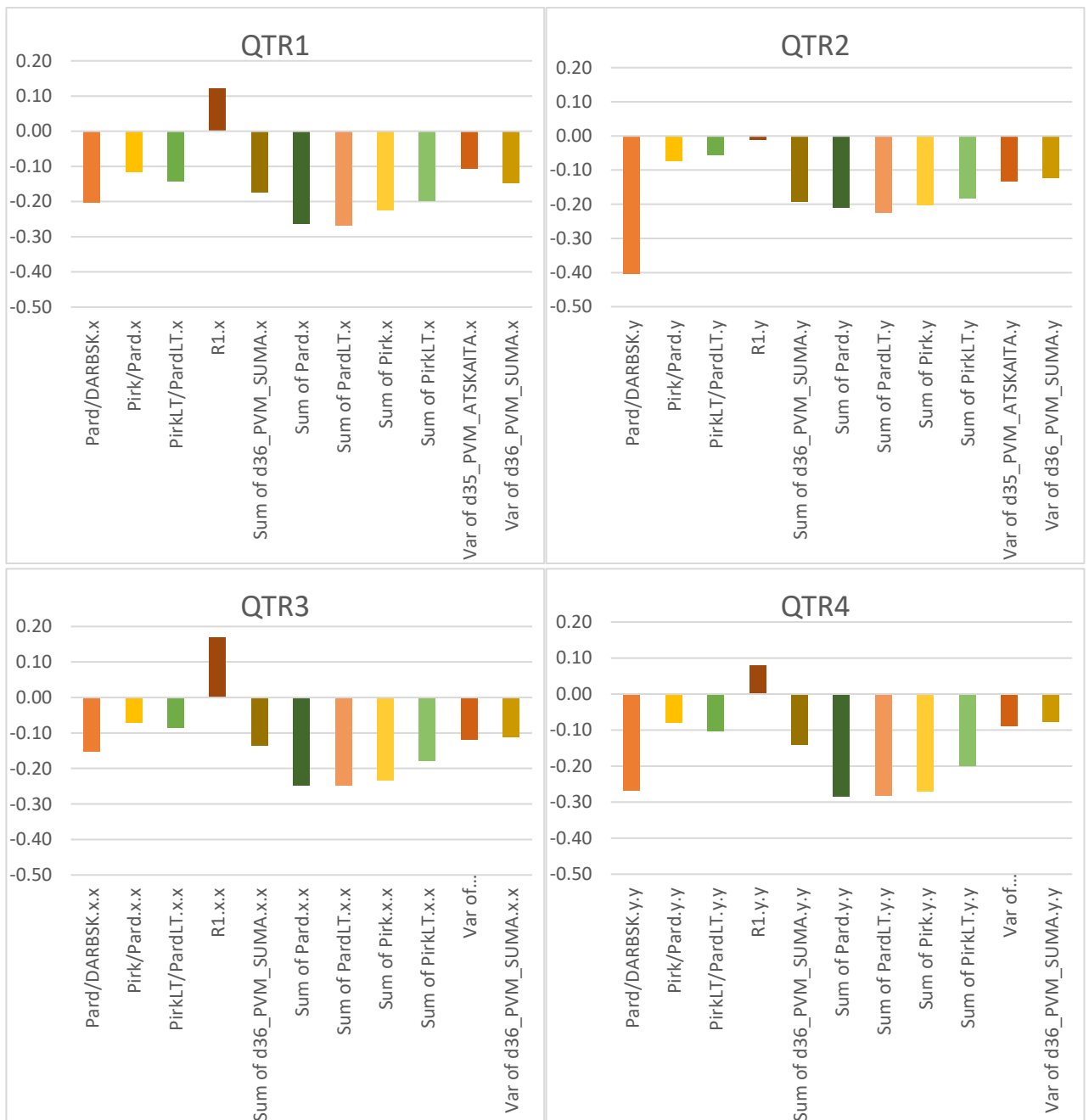
QTR4



10 pav. SOM rezultato pasiskirstymas klasteriais ir ketvirčiais

6 klasterių modelis didžiausiam klasteriui, kuriam priklausančios įmonės ir yra šio darbo vadinamos tipinėmis PVM mokėtojomis, priskiria 262 stebėjimus (79,15 % visų stebėjimų). 10 paveiksle didžiausias klasteris pažymėtas geltona spalva. Matoma, kad pardavimo ir pirkimų sumos (Lietuvoje

ir bendros) šiame klasteryje yra sąlyginai mažos. Pardavimai vienam darbuotojui didžiausiame klasteryje taip pat yra nedideli visuose ketvirčiuose palyginus su kitų klasterių įmonėmis. Tarp prieš tai pastebėtų dalies SOM tinklo pozicijų, išsiskiriančių sąlyginu pardavimų dydžiu vienam darbuotojui antrame ketvirtyje, tiriamos veiklos tipinių PVM mokėtojų įmonių nėra.



11 pav. Didžiausio klasterio centruotų ir bendro mastelio kintamųjų medianų reikšmės, suskirstytos ketvirčiais

Daugelio kintamųjų centruotų medianų reikšmės ketvirtame klasteryje yra mažesnės už 0 (žr. 11 paveikslą). Tai yra, tiriamos veiklos tipinių PVM mokėtojų įmonių kintamųjų medianos reikšmės yra mažesnės už vidutines. Tokia tendencija pastebima visuose ketvirčiuose visų rodiklių, išskyrus efektyvų PVM tarifą (R1). Jis pirmame, trečiame ir ketvirtame ketvirtyje yra didesnis už visų klasterių vidutines reikšmes. R1 rodiklio vidutinį sumažėjimą antrame ketvirtyje galima paaiškinti pastebėjus 10 paveiksle vienos pozicijos sąlyginį R1 padidėjimą. Antrame ketvirtyje, taip pat stebima tendencija,

kad pardavimai tenkantys vienam darbuotojui didžiausiame klasteryje tampa dar mažesni už visų klasterių vidutinius lyginant su kitais ketvirčiais. Iš tokios tendencijos galima teigti, kad antrame ketvirtyje tipiniai PVM mokėtojai išsiskiria mažesniu pardavimų vienam darbuotojui padidėjimu nei kitos įmonės.

Netipiniais PVM mokėtojais laikomos visos įmonės nepriklausančios didžiausiam klasteriui. Pagal 1 priedą, matoma, kad pirmojo klasterio visų kintamųjų mediana yra pastovi ir panaši į kitų klasterių medianas visuose ketvirčiuose. Tačiau pirkimų ir pardavimų bendro ir Lietuvoje santykių matoma sklaida į didesnę pusę. Taip pat stebima gerokai didesnių santykių išskirtis. Tokia tendencija ypač ryški pirmame ketvirtyje. Galima daryti prielaidą, kad skirtumas tarp pirkimų ir pardavimų yra sąlyginai didelis, tačiau absoliučios sumos nėra didelės, kadangi jos neišsiskiria iš tipiškų klasterių medianos. Kadangi efektyvus PVM tarifas taip pat parodo išskirtis neigiamoje rodiklio pusėje, o sklaida į šią pusę yra didesnė, įmonės esančios 1 klasteryje ir turinčios mažesnę nei vidutinis R1 rodiklį priskiriamos prie rizikingų.

Antrąjo klasterio įmonės išsiskiria tik sąlyginai didesne pardavimų suma vienam darbuotojui antrame ir ketvirtame ketvirčiuose. Šis klasteris turi pastebimas efektyvaus PVM tarifo neigiamas išskirtis pirmame ir trečiame ketvirčiuose. Tokia situacija gali susidaryti, kad įmonės pardavimai yra itin sezoniški, šiuo atveju antrame ir ketvirtame ketvirtyje, o pirkimai ir atsargų papildymai pastovūs visus metus. Tokiu atveju, kai pardavimai pirmą ir trečią ketvirčius sumažėja, o pirkimai lieka pastovūs, sumažėja ir efektyvus PVM tarifas. Tokia tendencija pirmame ir antrame ketvirčiuose sutampa su teorija, tačiau ketvirtame ketvirtyje pardavimų padidėjimas tokioje veikloje yra netikėtas. Todėl 2 klasterio R1 rodiklio neigiamos išskirtys pažymimos kaip rizikingos.

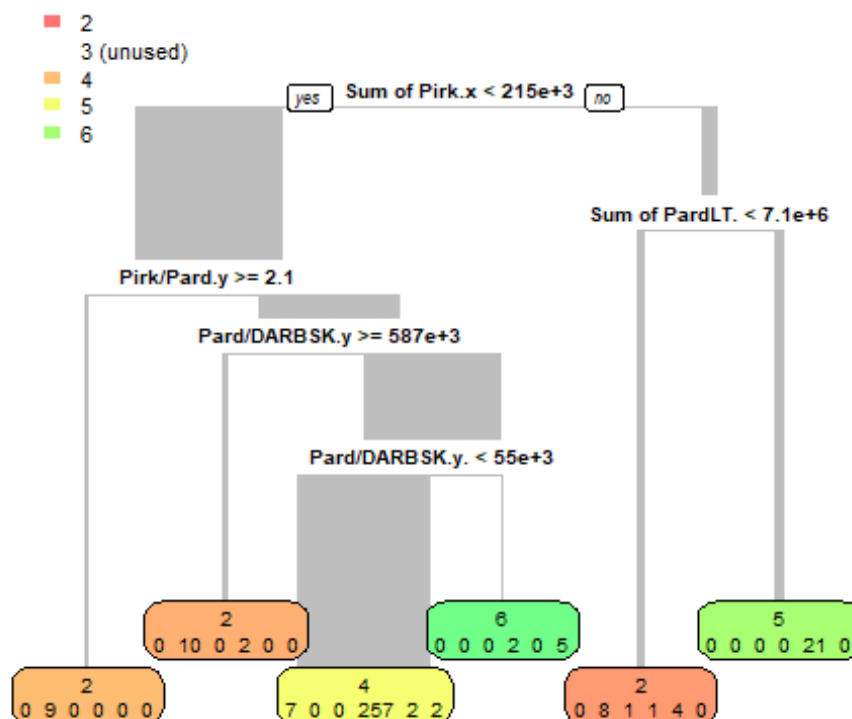
Trečiame klasteryje priskirta tik viena įmonė. Jos efektyvus PVM tarifas yra mažesnis nei kitų klasterių mediana visuose ketvirčiuose, o trečiame ketvirtyje dar ryškiau neigiama prasme išsiskiria iš kitų klasterių. Pagal pirkimų ir pardavimų santykį ir PVM sumą, matoma, kad pirkimai tiek Lietuvoje tiek bendri sąlyginai didesni visuose ketvirčiuose. Nors pirmame ketvirtyje išryškėja didesnė pardavimų suma vienam darbuotojui, tą patį ketvirtį ir pirkimų Lietuvoje suma didesnė nei kitų klasterių, todėl PVM suma pastebimai mažesnė. Be to, ir PVM sumos svyravimai per stebimus metus yra didesni nei kitų klasterių. Toks pirkimų ir pardavimų santykis ir PVM sumos nepastovumas yra netipinis šioje veikloje, dėl to 3 klasterio įmonė priskiriama prie rizikingų.

Ketvirtas klasteris yra didžiausias, todėl jame esantys stebėjimas priskiriami prie tipinių PVM mokėtojų. Tačiau, pagal R1 rodiklį, šitas klasteris turi neigiamų išskirčių, jos priskiriamos prie rizikingų įmonių.

Penktas klasteris išsiskiria pardavimų ir pirkimų sumų didesne sklaida ir daugiau išskirčių į teigiamą pusę. Kadangi pardavimai vienam darbuotojui taip stipriai neišsiskiria, tikėtina, kad šiame klasteryje yra priskirtos didesnės įmonės veikiančios didesniu mastu. 5 klasteryje efektyvus PVM tarifas mažesnis antrame ketvirtyje vienoje išskirtyje, o trečiame ketvirtyje dvejose. Tik tokios įmonės iš šio klasterio priskiriamos prie rizikingų.

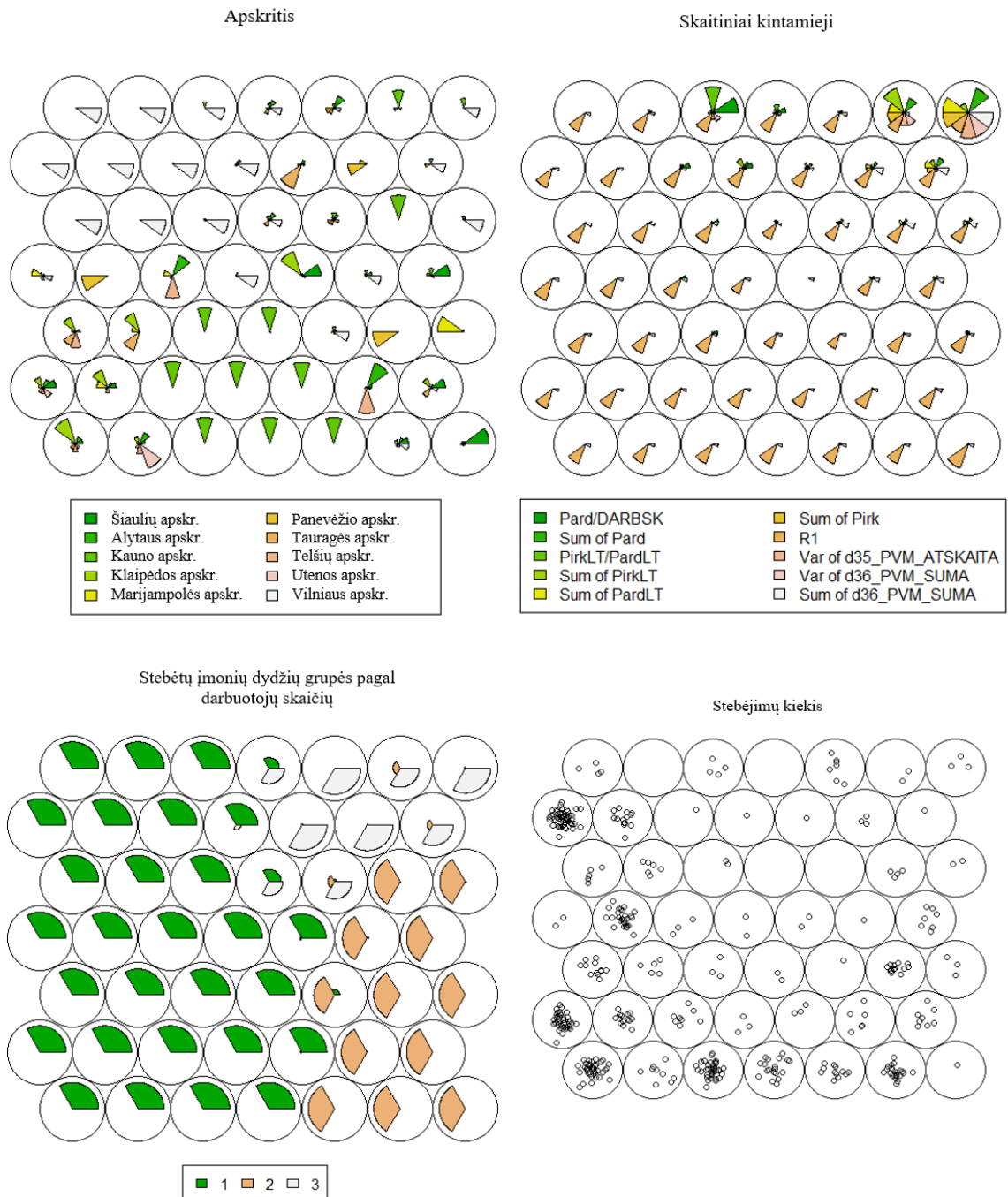
Šeštas klasteris išsiskiria tik didesne pardavimų suma vienam darbuotojui ketvirtame ketvirtyje ir viena įmone – išskirtimi. Ši įmonė ypač išsiskiria ketvirtame ketvirtyje didesniais pirkimais Lietuvoje ir bendrais. To pasekmė yra mažesnė bendra PVM suma. Efektyvus PVM tarifas taip pat yra mažesnis. Tačiau pagal tai, kad PVM atskaitos ir PVM sumos svyravimai irgi stipriai išsiskiria, galima daryti prielaidą, kad tokia įmonė didelę sumą investavo vienu metu. Antrame ketvirtyje ta pati

įmonė pastebimai išsiskiria pagal efektyvų PVM tarifą, tačiau kitaip nei paprastai, jos R1 rodiklis yra didesnis nei visų kitų įmonių. Įmonė turinti didelį efektyvų PVM rodiklį paprastai nesukeltų įtarimų dėl sukčiavimo, tačiau ji išsiskiria ir iš bendros veiklos konteksto ir iš klasterio, o R1 rodiklis labai svyruoja kiekvieną ketvirtį, todėl ši įmonė priskiriama prie rizikingų.



12 pav. Sprendimų medžio rezultatas

Sprendimų medžio algoritmas ir klasterizavimas suskirsto vienodai 93,66 % visų stebėjimų. Iš sprendimo medžio taisyklių (žr. 12 paveikslą) galima matyti, kad tipiniai PVM mokėtojai yra tie, kurių bendra pirkimų suma pirmame ketvirtyje yra mažesnė už 214608,8 €, bendrų pirkimų ir pardavimų santykis antrame ketvirtyje yra mažesnis už 2,05, pardavimų suma vienam darbuotojui antrame ketvirtyje yra mažesnė už 586983 €, o pardavimų suma vienam darbuotojui ketvirtame ketvirtyje yra mažesnė už 54967 €. Pagal tokias taisykles, teisingai atrenkamos 257 iš 262 įmonių ir priskiriamos 11 įmonių, kurios, pagal klasterizavimo algoritmą, pateko į kitus klasterius.



13 pav. SOM rezultato stebėjimų pasiskirstymas pagal apskritį, skaitinius kintamuosius ir įmonės dydį

SOM algoritmu suskirstyti duomenys pagal kategorinius kintamuosius (apskritį ir įmonės dydį) parodo (žr. 13 paveikslą), kad įmonės, kurios išsiskiria didesniu pirkimų Lietuvoje ir pardavimų Lietuvoje santykiu, yra įsikūrusios daugiausiai Vilniaus apskrityje. Taip pat, tokios įmonės išsiskiria didesniais pardavimais vienam darbuotojui. Tiek Kauno, tiek ir Vilniaus apskrityse yra įmonių, kurios išsiskiria didesniu pardavimų ir pirkimų kiekiu. Taip pat didesniu PVM atskaitos ir PVM sumos svyravimu. Vienoje iš aptariamų tinklo pozicijų pastebima sąlyginai maža PVM suma, tačiau tai galima pateisinti dideliais pirkimais Lietuvoje.

Tiriamoje veikloje didelių įmonių nėra, o daugiausiai yra mikro įmonių. Didesni pirkimai ir pardavimai yra būdingi vidutinio dydžio įmonėms. Mikro ir mažos įmonės, pagal tiriamus kintamuosius, yra panašios. Tačiau pastebimos keturios įmonės, kurios priskiriamos mikro dydžiui,

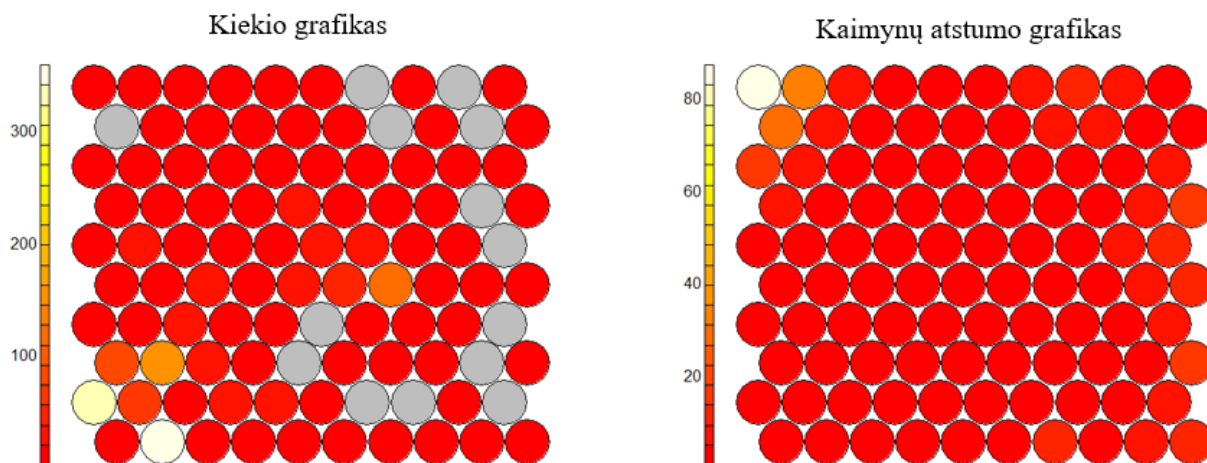
o jų pardavimai vienam darbuotojui ir pirkimų Lietuvoje santykis su pardavimais Lietuvoje yra palyginus dideli. Tokia situacija gali būti, kai darbuotojų skaičius mažas (matoma, kad įmonė yra mikro pagal darbuotojų skaičių), o pardavimai ne Lietuvoje dideli ir pirkimai Lietuvoje sąlyginai dideli arba pardavimai Lietuvoje sąlyginai maži.



14 pav. Apskritis ir įmonės dydžio kintamųjų centruotos ir bendro mastelio medianų reikšmės, suskirstytos klasteriais

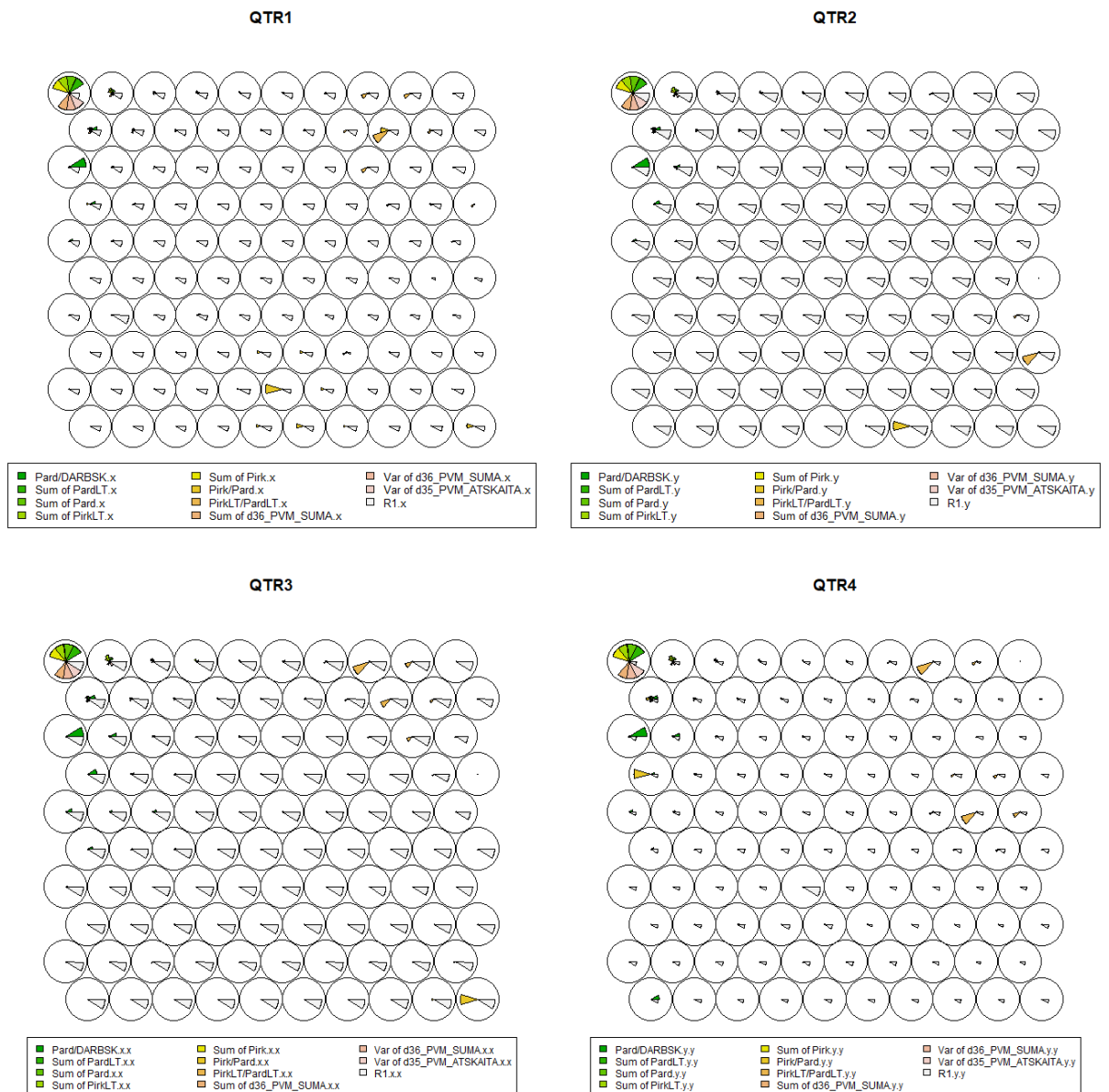
14 paveiksle matoma, kad didžiausiame klasteryje (nr. 4) yra visų apskričių įmonių (daugiausiai Vilniaus ir Kauno apskrities). Tiriamos veiklos tipinio PVM mokėtojo klasteryje tik nedidelį procentą sudaro vidutinės įmonės, o daugiausiai yra mikro įmonių. Todėl papildančiam tyrimui rekomenduotina atskirai analizuoti mikro ir mažo dydžio grupės įmones.

3.2. Su statybų sektoriumi susijusių pardavimų veiklos tyrimo rezultatai



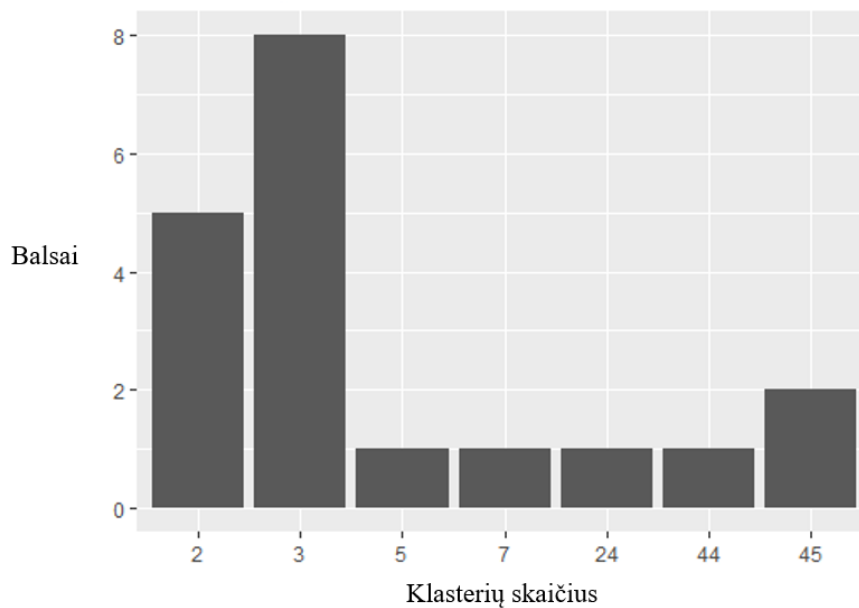
15 pav. SOM rezultato kiekių ir atstumų grafikai

Atliktas saviorganizuojančių neuroninių tinklų algoritmas suskirstė stebėjimus, pagal jų panašumą, taip, kad jie būtų atvaizduojami dvimačiame 100 pozicijų žemėlapyje. 15 paveiksle matoma, kad daugumoje pozicijų stebėjimų kiekis yra panašus (mažiau už 100), dvi pozicijos išsiskiria dideliu (virš 300) priskirtų stebėjimų kiekiu, o 14 pozicijų liko tuščios. Visos pozicijos, išskyrus tris, yra sąlyginai arti viena kitos.



16 pav. SOM rezultato kintamųjų sąlyginiai dydžiai suskirstyti ketvirčiais

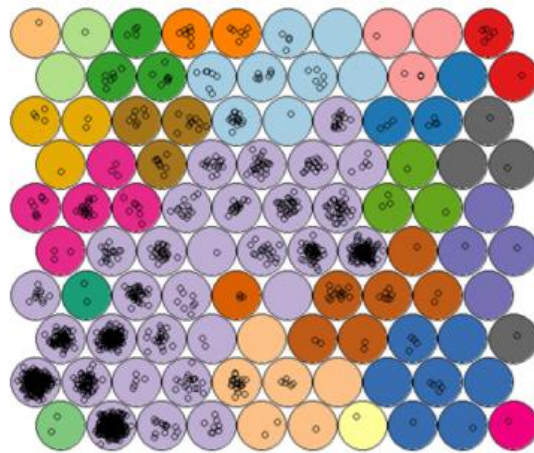
Saviorganizuojančių neuroninių tinklų algoritmo rezultate pastebimi bendri stebėjimų skirtumai skirtingais ketvirčiais. Antrame ir trečiame ketvirčiuose matoma, kad daugiausiai pozicijų sąlyginai didžiausias yra R1 rodiklis. Pirmame ir ketvirtame ketvirčiuose yra daugiau įmonių išsiskiriančių didesniu R1 rodikliu. Visuose ketvirčiuose yra matoma pozicija, kurioje priskirtos įmonės išsiskiria beveik visais didesniais rodikliais, išskyrus anksčiau paminėtą R1 rodiklį, taip pat pirkimų ir pardavimų (Lietuvos ir bendrų) santykį ir pardavimus vienam darbuotojui.



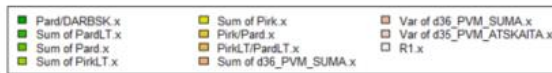
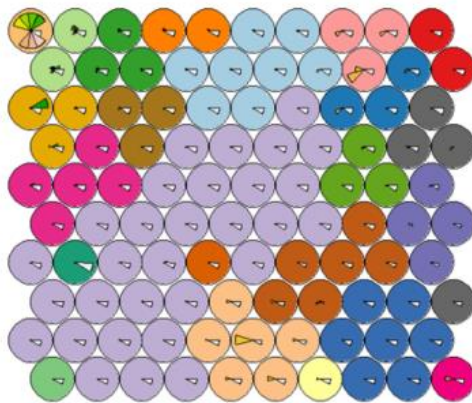
17 pav. Teorinių metodų balsai už optimalų klasterių skaičių

SOM rezultato objektus skirstant į klasterius daugiausiai teorinių metodų parodė, kad optimalus klasterių skaičius yra 3 (žr. 17 paveikslą). Tačiau, suskirsčius duomenis į tris klasterius, didžiausiam klasteriui priskiriami 95,98% visų stebėjimų. Pagal teorijos apžvalgą, audituojamos turėtų būti apie 25-30% rizikingų įmonių, o ir sukčiaujančios įmonės tikėtina bus nebūtinai prie labiausiai išsiskiriančių. Todėl, pagal prie didžiausio klasterio priskirtų stebėjimų proporcijas ir teorinių modelių rekomendacijas, klasterių analizei parenkami 24 klasteriai. Su tokia parinktimi didžiausio klasterio proporcija vis dar yra didelė (85,79%). Tačiau dar didinant klasterių skaičių ši proporcija mažėja nedaug. Todėl atsižvelgus į interpretavimo ir skaičiavimo sudėtingumą bei abejotiną naudą paliekamas 24 klasterių kiekis.

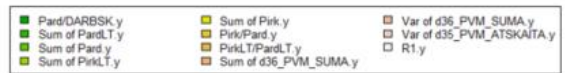
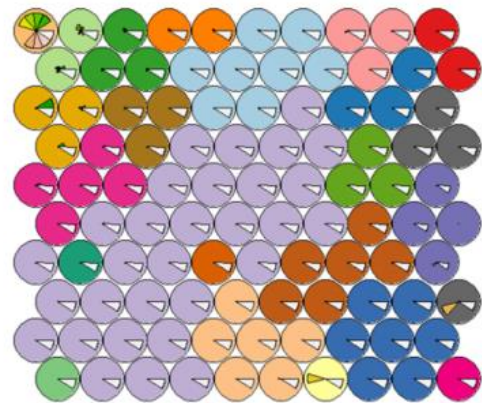
Klasteriai ir stebėjimų kiekis



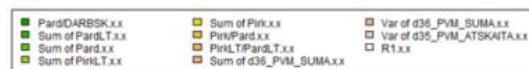
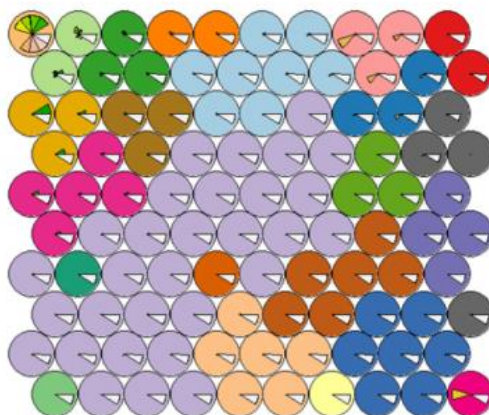
QTR1



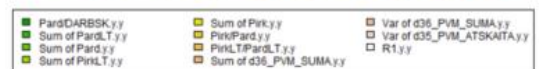
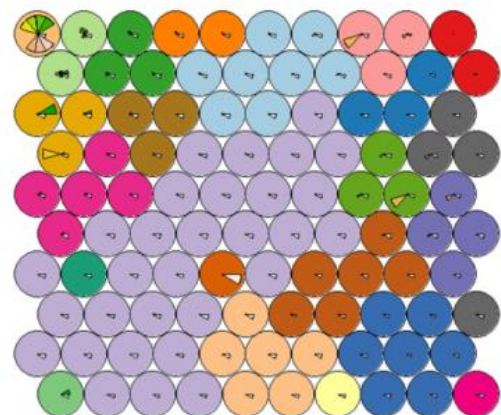
QTR2



QTR3

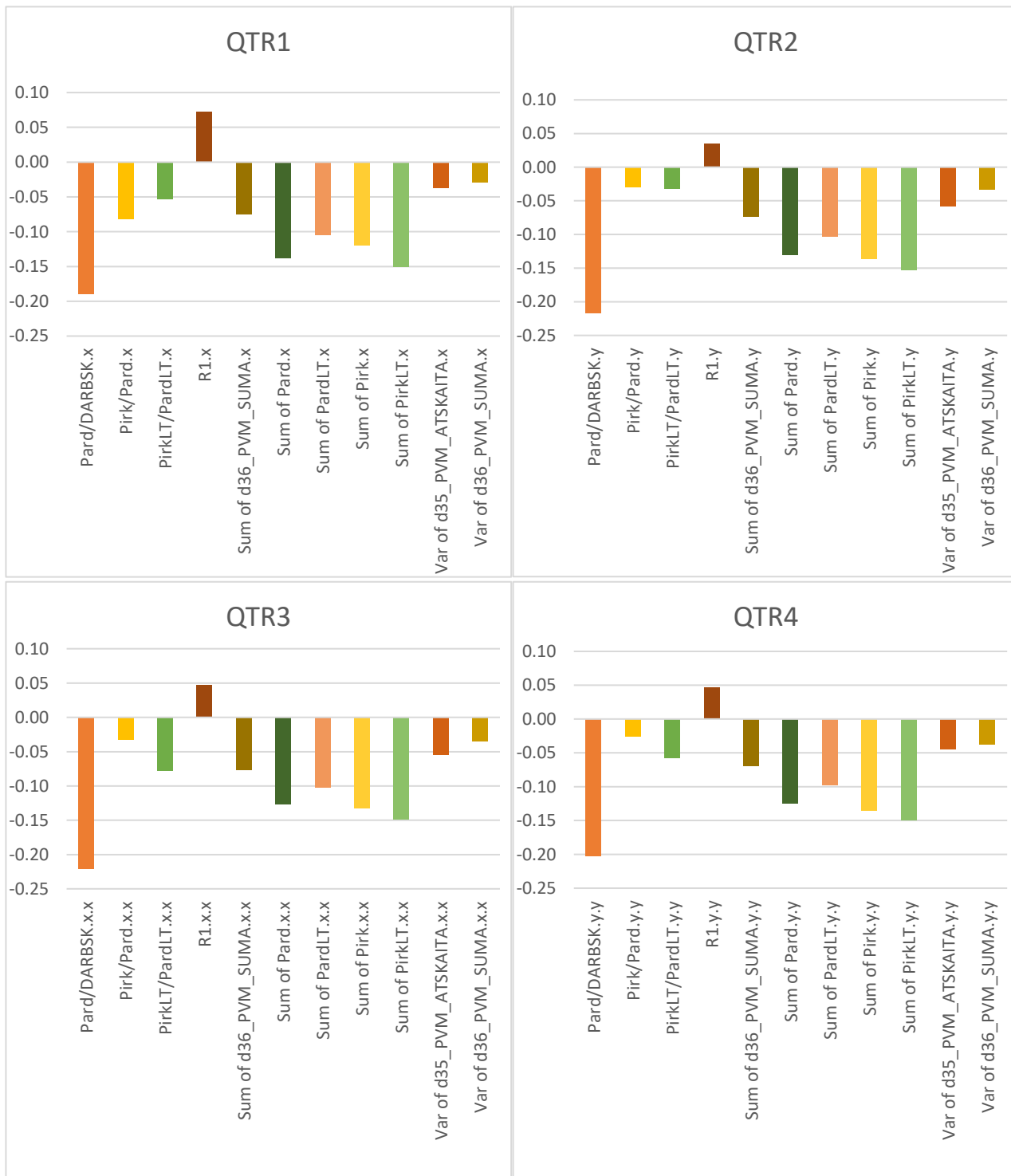


QTR4



18 pav. SOM rezultato pasiskirstymas klasteriais ir ketvirčiais

24 klasterių modelis didžiausiam klasteriui, kuriam priklausančios įmonės ir yra šio darbo laikomos kaip tipinės PVM mokėtojos, priskiria 1515 stebėjimus. 18 paveiksle matoma, kad pardavimo ir pirkimų sumos (Lietuvoje, bendros ir vienam darbuotojui) šiame klasteryje yra sąlyginai mažos. R1 rodikliu išsiskiriančių pozicijų, kurios buvo pastebėtos pirmame ir ketvirtame ketvirčiuose, didžiausiame klasteryje nėra.

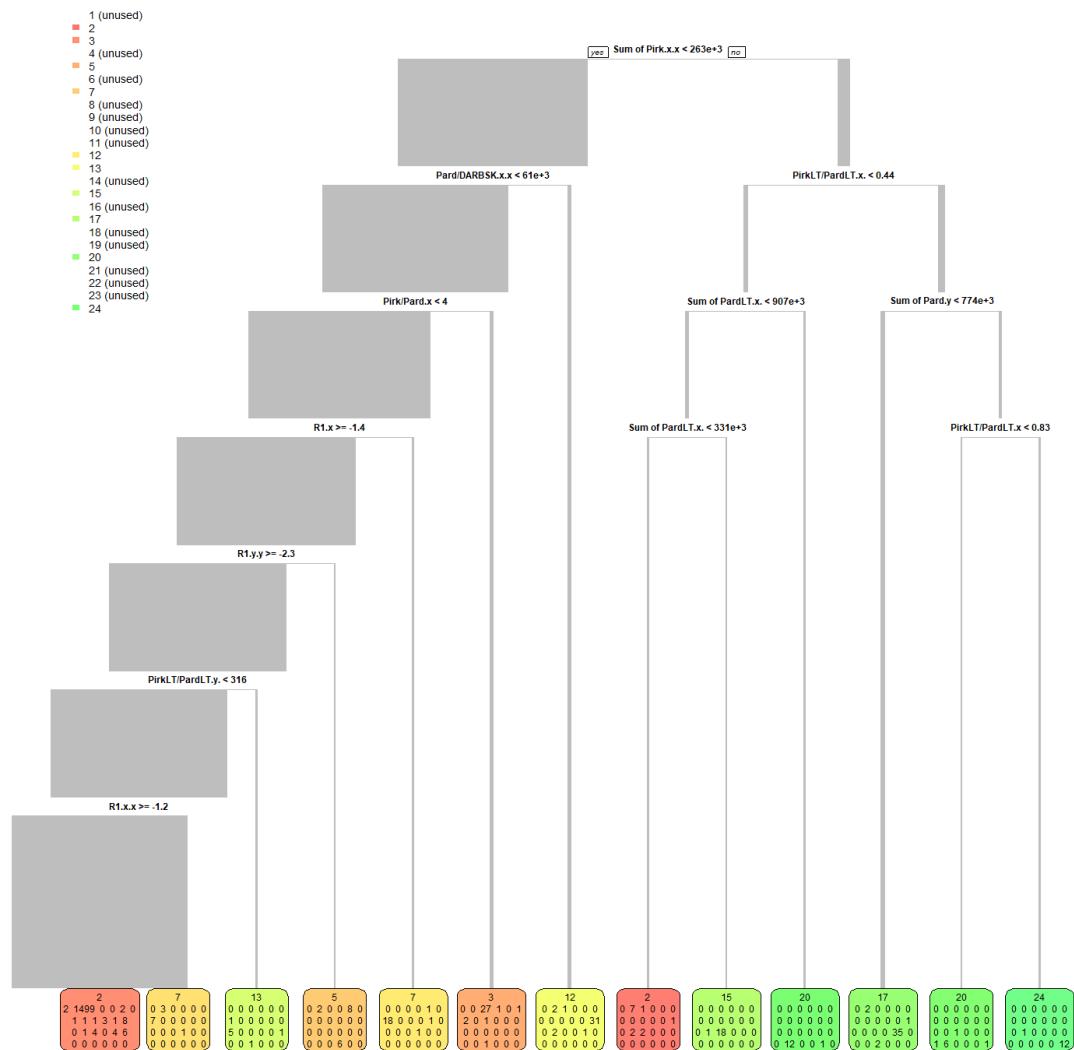


19 pav. Didžiausio klasterio centruotų ir bendro mastelio kintamųjų medianų reikšmės, suskirstytos ketvirčiais

Beveik visų kintamųjų centruotų medianų reikšmės didžiausiame klasteryje yra mažesnės už 0 (žr. 19 paveikslą). Tai yra, tiriamos veiklos tipinių PVM mokėtojų įmonių kintamųjų medianos reikšmės yra šiek tiek mažesnės už vidutines. Tokia tendencija pastebima visuose ketvirčiuose visų rodiklių, išskyrus efektyvų PVM tarifą. Jis visuose ketvirčiuose yra didesnis už visų klasterių vidutines reikšmes. Rodiklių vidutinės reikšmės skirtinguose ketvirčiuose skiriasi nežymiai.

Netipiniais PVM mokėtojais laikomos visos įmonės nepriklausančios didžiausiam klasteriui. Pagal 2 priedą matoma, kad efektyvus PVM tarifas yra mažesnis 5, 7, 11, 16, 22 klasteriuose. Be to, mažesnio efektyvaus PVM tarifo išskirtys pastebimos didžiausiame klasteryje (2), 3, 17 ir 21 klasteriuose. Šio rodiklio išskirtys pažymimos kaip rizikingos įmonės.

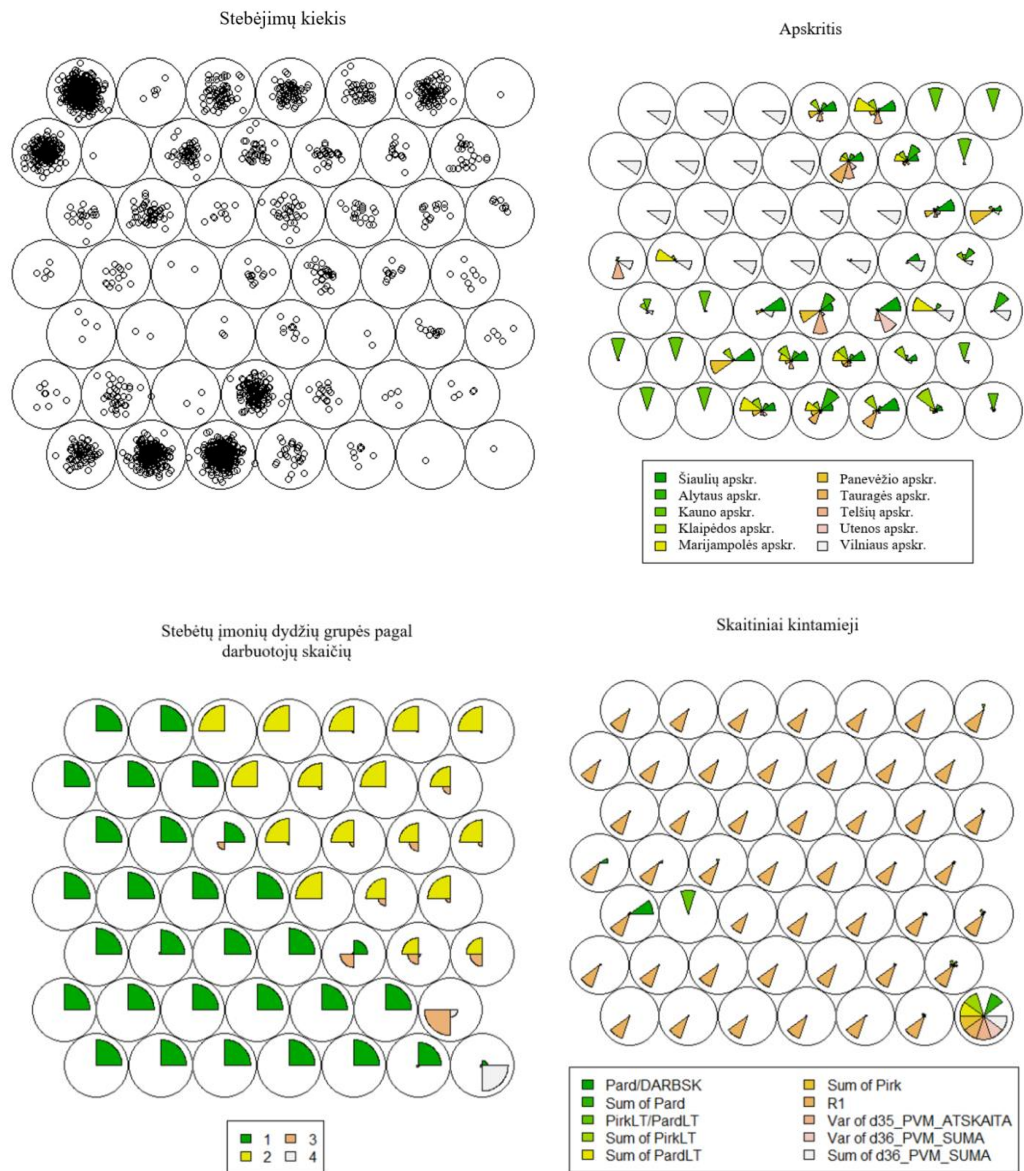
5 klasteris žemesniu R1 išsiskiria ketvirtame ketvirtyje, kiti rodikliai nesiskiria nuo daugumos. 7 klasteryje žemesnis efektyvus PVM tarifas yra tik pirmame ketvirtyje. Pirmame ketvirtyje taip pat išsiskiria žemesniu R1 rodikliu 11 klasteris. Be to, šis klasteris antrame ketvirtyje yra vienintelis, kurio visos R1 reikšmės yra mažesnės nei kitų klasterių įmonių. Trečiame ketvirtyje taip stipriai, pagal R1, išsiskiriantis vienintelis klasteris yra 16. Šio klasterio įmonės taip pat turi vidutiniškai mažesnę efektyvų PVM tarifą nei daugumos klasterių pirmame ketvirtyje. Ketvirtame ketvirtyje mažiausias R1 reikšmes turi 22 klasteris, tačiau kituose ketvirčiuose rodiklis šiame klasteryje neišsiskiria. Todėl 5 ir 7 klasterio įmonės, kurios turi mažesnę nei klasterio vidutinę efektyvų PVM tarifą yra priskiriamos prie rizikingų, o 11, 16 ir 22 klasterio visos įmonės pažymimos rizikingomis.



20 pav. Sprendimų medžio rezultatas

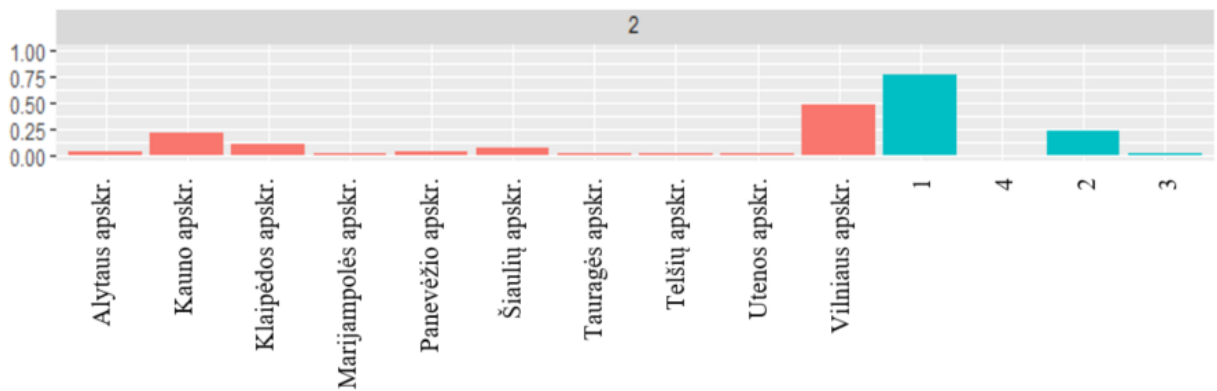
Sprendimų medžio algoritmas ir klasterizavimas suskirsto vienodai 95,41% visų stebėjimų. Iš sprendimo medžio taisyklių (žr. 20 paveikslą) galima matyti, kad sprendimų medis išskiria tipinius PVM mokėtojus tokius, kurių pirkimų suma trečiame ketvirtyje yra mažesnė už 262950,1 €, pardavimų suma vienam darbuotojui trečiame ketvirtyje yra mažesnė už 61101 €, bendrų pirkimų ir pardavimų santykis pirmame ketvirtyje yra mažesnis už 3,96, R1 rodiklis pirmame ketvirtyje didesnis už -1,41, o ketvirto ketvirčio R1 rodiklis yra didesnis už -2,25, pirkimų ir pardavimų Lietuvoje santykis antrame ketvirtyje yra mažesnis už 316, R1 rodiklis trečiame ketvirtyje didesnis už -1,17. Pagal tokias taisykles, teisingai atrenkamos 1499 iš 1766 įmonių ir priskiriamos 34 įmonės, kurios, pagal klasterizavimo algoritmą, pateko į kitus klasterius.

Galima pastebėti, kad statybos sektoriaus prekių pardavimo veiklos tipiniai PVM mokėtojai, pagal sprendimų medį, turi daugiau taisyklių nei su žemės ūkiu susijusi veikla ar elektronikos prekių pardavimo veikla. Tai reiškia, kad su statybų sektoriumi susijusių pardavimų veikloje yra mažiau ryškūs įmonių išsiskyrimai, pagal kitamuosius, nei kitose tiriamose veiklose.



21 pav. SOM rezultato stebėjimų pasiskirstymas pagal apskritį, skaitinius kintamuosius ir įmonės dydį

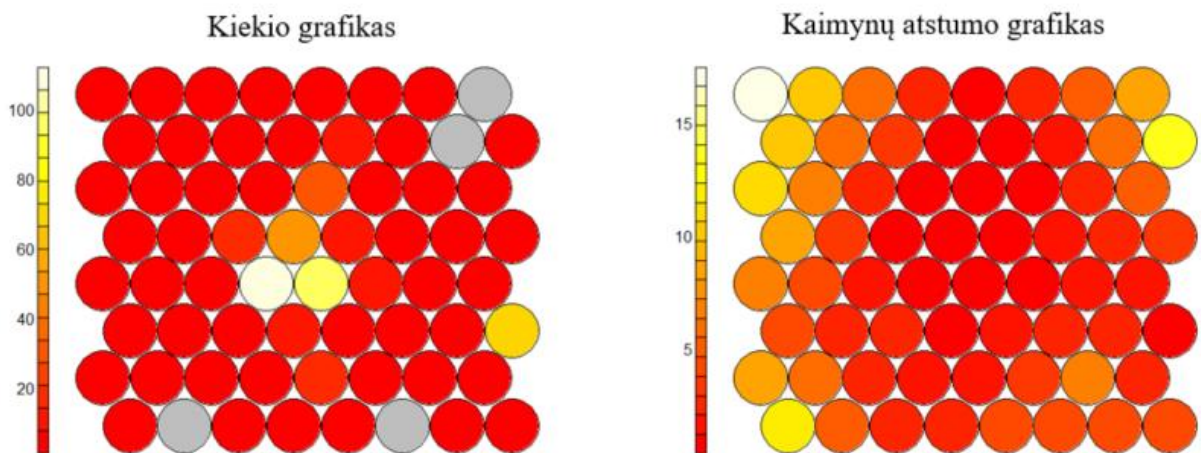
SOM algoritmu suskirstyti duomenys, pagal kategorinius kintamuosius (apskritį ir įmonės dydį), parodo (žr. 21 paveikslą), kad įmonės, kurios, pagal įvairius rodiklius, pastebimai išsiskira daugiausiai yra įsikūrusios Kaune. Įmonė išsiskirianti visais rodikliais, išskyrus pardavimus vienam darbuotojui, pirkimų ir pardavimų Lietuvoje santykį ir R1 rodiklį yra priskiriama prie didelių įmonių grupės. Įmonės išsiskiriančios didesnia pardavimų suma vienam darbuotojui ir įmonės, kurių didesnis pirkimų ir pardavimų Lietuvoje santykis, o efektyvus PVM tarifas (R1) mažesnis priskiriamos prie mikro įmonių dydžio grupės. Mikro, mažos ir vidutinės įmonės, be prieš tai paminėtų išimčių, pagal tiriamus kintamuosius, yra panašios.



22 pav. Didžiausio klasterio apskrities ir įmonės dydžio kintamųjų centruotos ir bendro mastelio medianų reikšmės

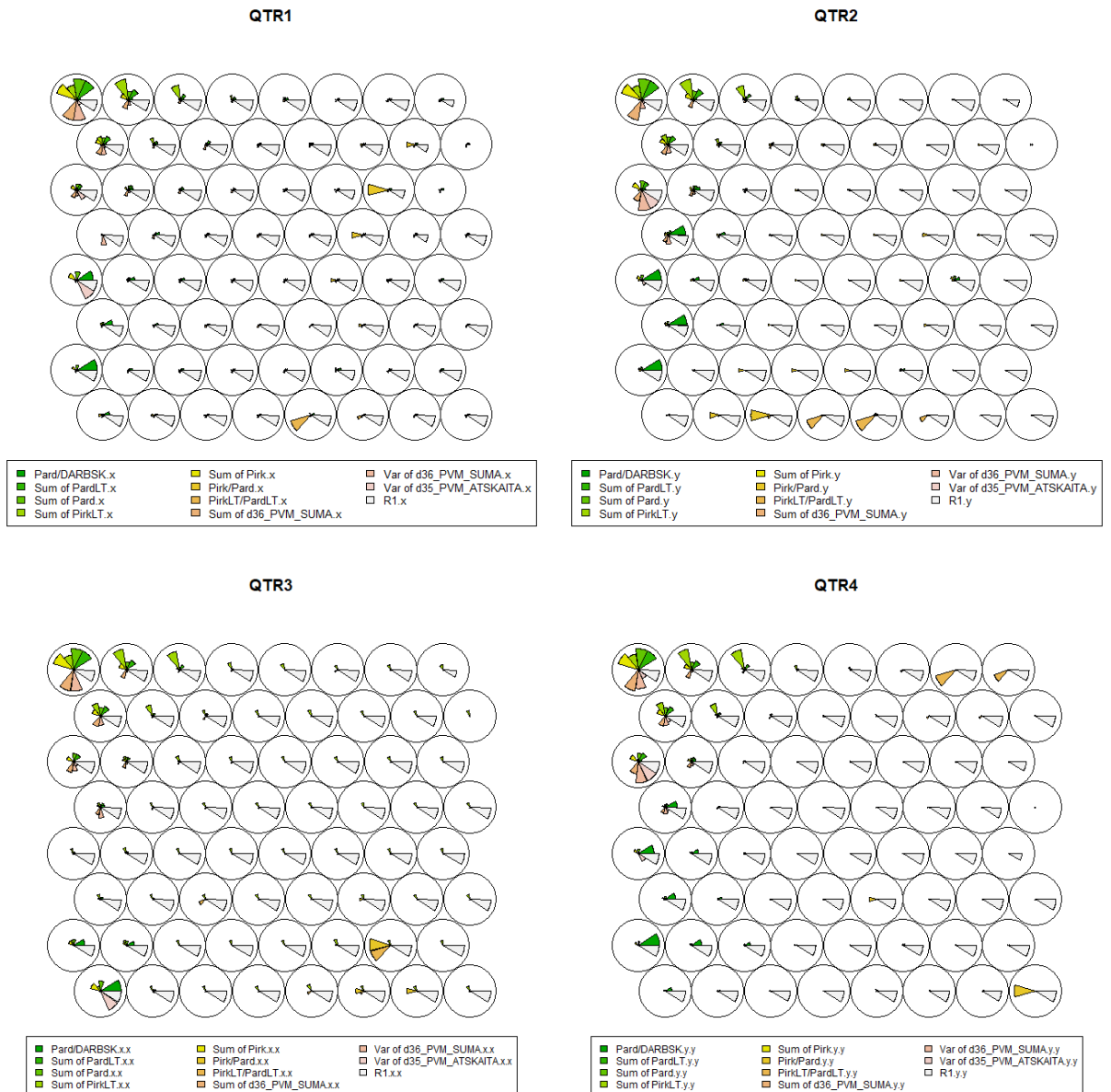
22 paveiksle matoma, kad didžiausiame klasteryje (nr. 2) yra visų apskričių įmonių (daugiausiai Vilniaus ir Kauno apskrities). Tiriamos veiklos tipinio PVM mokytojo klasteryje tik nedidelį procentą sudaro vidutinės įmonės, o daugiausiai yra mikro įmonių. Didelių įmonių nepriskirta. Todėl papildančiam tyrimui rekomenduotina atskirai analizuoti mikro ir mažo dydžio grupės įmones.

3.3. Elektronikos prekių pardavimų veiklos tyrimo rezultatai



23 pav. SOM rezultato kiekių ir atstumų grafikai

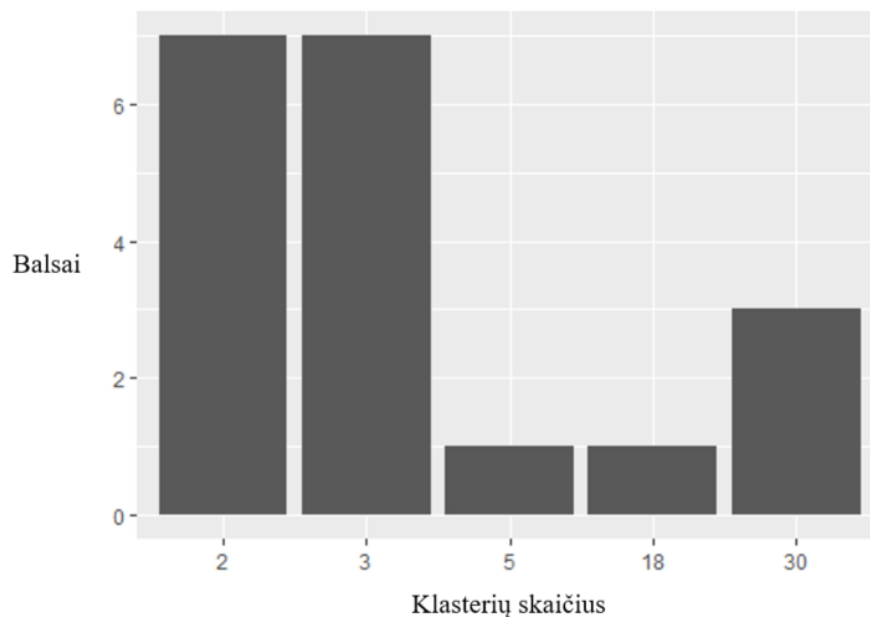
Atliktas saviorganizuojančių neuroninių tinklų algoritmas suskirstė stebėjimus, pagal jų panašumą, taip, kad jie būtų atvaizduojami dvimačiame 64 pozicijų žemėlapyje. 23 paveiksle matoma, kad daugumoje pozicijų stebėjimų kiekis yra panašus (mažiau už 20), keturios pozicijos išsiskiria dideliu (virš 60) priskirtų stebėjimų kiekiu, o 4 pozicijos liko tuščios. Pastinklo kraštuose pastebimas atstumo tarp pozicijų padidėjimas.



24 pav. SOM rezultato kintamųjų sąlyginiai dydžiai suskirstyti ketvirčiais

Saviorganizuojančių neuroninių tinklų algoritmo rezultate pastebima, kaip ir kitose tirtose veiklose, daugumoje pozicijų sąlyginai didžiausias yra efektyvus PVM tarifas (R1). Tik elektronikos prekių pardavimo veikloje šis rodiklis nėra palygintinai svyruojantis skirtinguose ketvirčiuose.

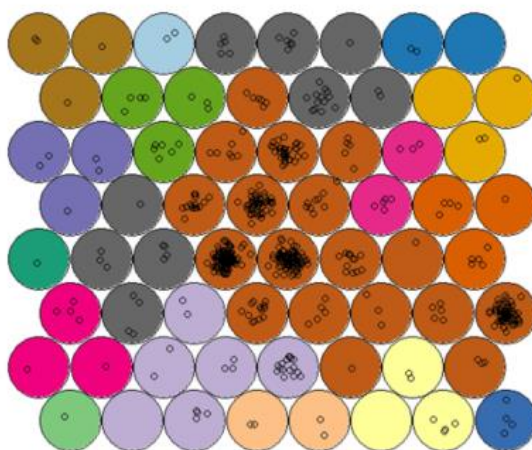
Pastebimos grupės pozicijų, kurios skirtingais ketvirčiais išsiskiria pagal kurį nors kintamąjį ar jų grupę. Tokie išskiriantys kintamieji yra pardavimų suma vienam darbuotojui, pirkimų ir pardavimų (bendrų ir Lietuvoje) santykis bei PVM sumos ir PVM atskaitos įmonėje per tiriamus metus svyravimai.



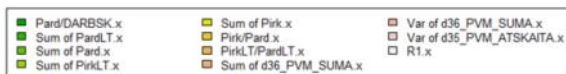
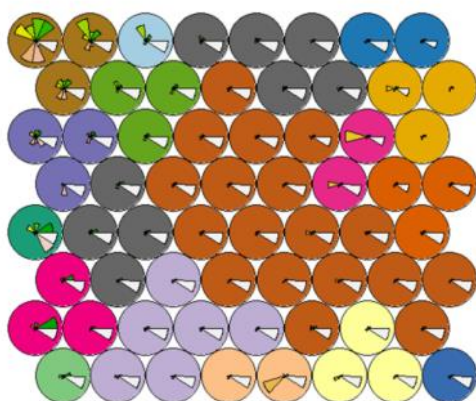
25 pav. Teorinių metodų balsai už optimalų klasterių skaičių

SOM rezultato objektus skirstant į klasterius daugiausiai teorinių metodų parodė, kad optimalus klasterių skaičius yra 2 arba 3 (žr. 25 paveikslą). Tačiau, suskirsčius duomenis į tris klasterius, didžiausiam klasteriui priskiriami 88,39% visų stebėjimų. Pagal teorijos apžvalgą, audituojamos turėtų būti apie 25-30% rizikingų įmonių, o ir sukčiaujančios įmonės tikėtina bus nebūtinai prie labiausiai išsiskiriančių. Todėl, pagal prie didžiausio klasterio priskirtų stebėjimų proporcijas ir teorinių modelių rekomendacijas, klasterių analizei parenkami 18 klasterių.

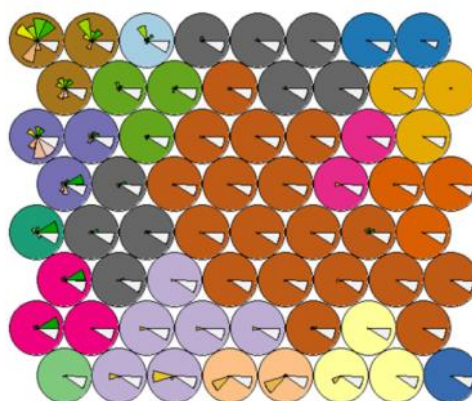
Klasteriai ir stebėjimų kiekis



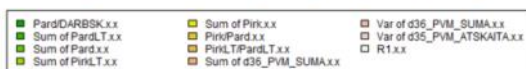
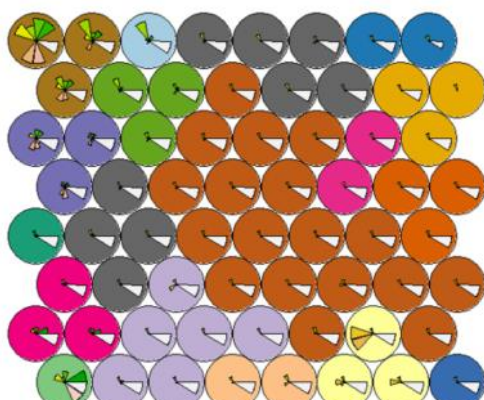
QTR1



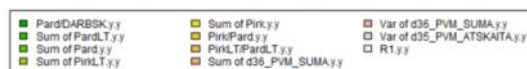
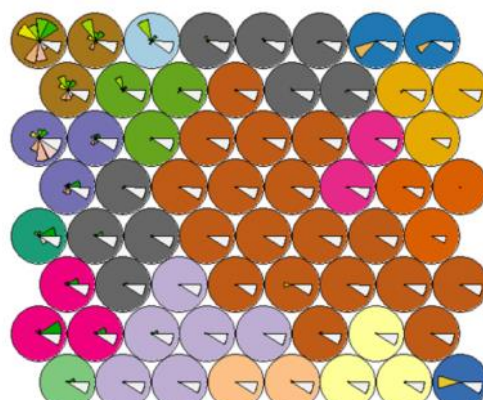
QTR2



QTR3

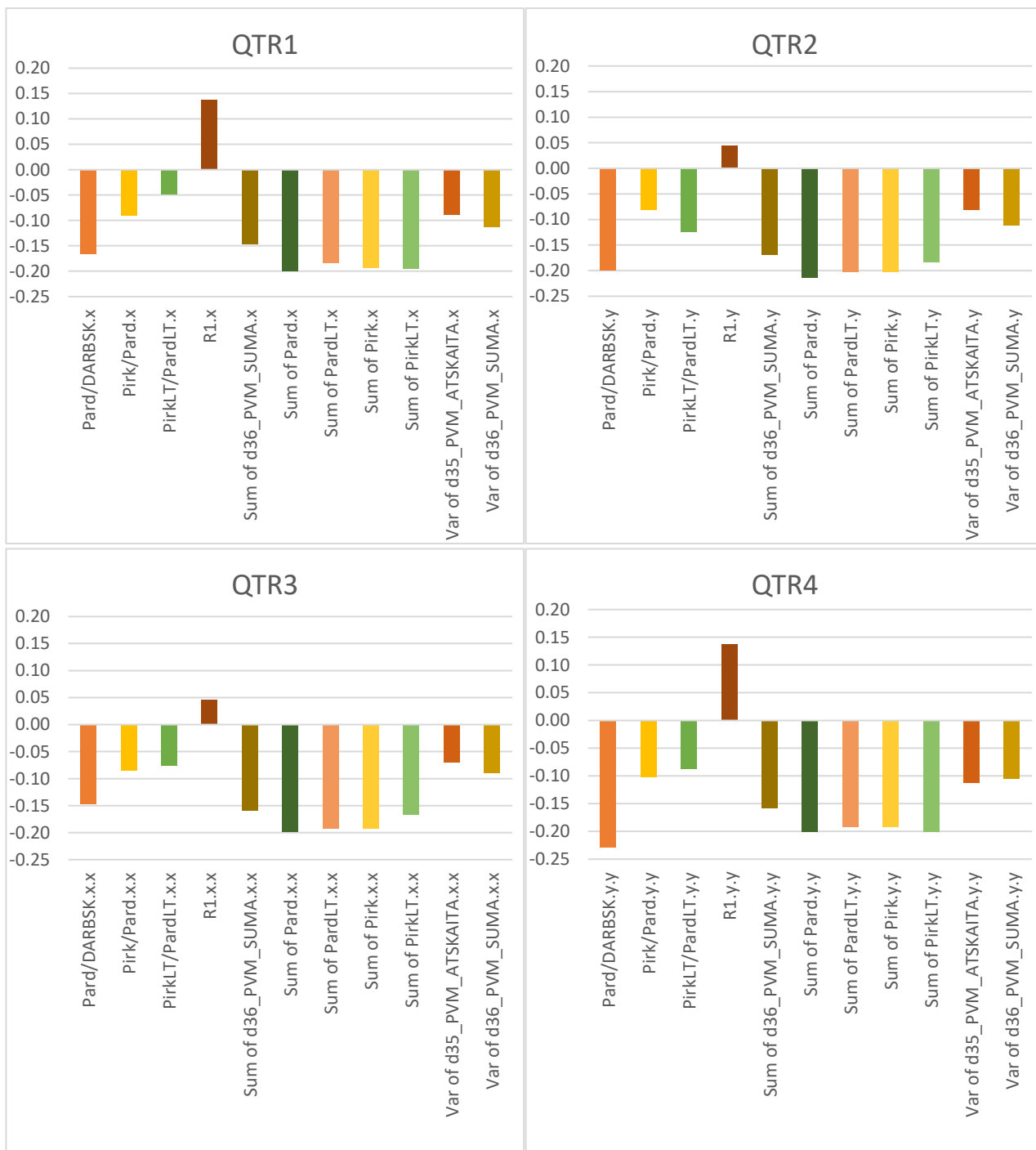


QTR4



26 pav. SOM rezultato pasiskirstymas klasteriais ir ketvirčiais

18 klasterių modelis didžiausiam klasteriui, kuriam priklausančios įmonės ir yra šio darbo laikomos kaip tipinės PVM mokėtojos, priskiria 442 stebėjimus (76,6% visų stebėjimų). 26 paveiksle matoma, kad pardavimo ir pirkimų sumos (Lietuvoje ir bendros) šiame klasteryje yra sąlyginai mažos. Pardavimai vienam darbuotojui didžiausiam klasteryje taip pat yra nedideli visuose ketvirčiuose palyginus su kitų klasterių įmonėmis. Prieš tai pastebėtų pozicijų su tam tikrų kintamųjų išsiskyrimu kai kuriuose ketvirčiuose, didžiausiam klasteryje nėra. Pastebimų kintamųjų reikšmių sąlyginių pokyčių skirtingais metų ketvirčiais didžiausiam klasteryje nėra.



27 pav. Didžiausio klasterio centruotų ir bendro mastelio kintamųjų medianų reikšmės, suskirstytos ketvirčiais

Kaip ir kitose tirtose veiklose, beveik visų kintamųjų centruotų medianų reikšmės didžiausiame klasteryje yra mažesnės už 0 (žr. 27 paveikslą). Tai yra, tiriamos veiklos tipinių PVM mokėtojų įmonių kintamųjų medianos reikšmės yra šiek tiek mažesnės už vidutines. Tokia tendencija pastebima visuose ketvirčiuose visų rodiklių, išskyrus efektyvų PVM tarifą. Jis visuose ketvirčiuose yra didesnis už visų klasterių vidutines reikšmes. Be to, R1 rodiklio mediana salyginai didesnė pirmą ir ketvirtą ketvirtį. Pardavimų sumos vienam darbuotojui mediana ketvirtame ketvirtyje yra mažesnė nei kituose ketvirčiuose, lyginant su kitų klasterių to paties kintamojo medianomis. Pirkimų ir pardavimų Lietuvoje santykio mediana santykinai sumažėja antrąjį ketvirtį. Kitų kintamųjų medianų ketvirtinis pokytis nėra toks ryškus.

Netipiniais PVM mokėtojais laikomos visos įmonės nepriklausančios didžiausiam klasteriui. Pagal 3 priedą, matoma, kad pardavimų vienam darbuotojui didžiausios sumos priskiriamos 9 klasteriui. 9 klasteris dar išsiskiria didesne bendra pardavimų suma ir bendra pirkimų suma. PVM atskaita šio klasterio įmonėse yra daugiau kintanti skirtingais metais. Tokia tendencija pastebima visus, išskyrus trečiąjį, ketvirtį. Šio klasterio įmonės, tikėtina, turi nedaug darbuotojų tačiau daug perka ir tai paaiškina didelius pardavimus vienam darbuotojui. O didelė PVM atskaitos kaita galėtų būti paaiškinama dideliais pirkimais atsargoms papildyti, tačiau kiekvienais metais atsargos pildomos skirtingu metu. Kadangi efektyvus PVM tarifas ir bendra PVM suma nėra išsiskirianti iš kitų klasterių, 9 klasteris neįtraukiamas tarp rizikingiausių.

Tų pačių kintamųjų ir tos pačios tendencijos tik mažesnis išskirtinumas nei 9 klasterio yra pastebimas 6 klasteryje. 6 klasteris pasižymi didesne sklaida išskirtiniuose kintamuosiuose. Be to, šis klasteris yra pastovesnis visuose keturiuose ketvirčiuose nei 9 klasterio reikšmės.

Bendrų pirkimų ir pardavimų santykio reikšmė išsiskiria 12 klasteris. Pirmame ketvirtyje pastebima viena klasterio išskirtis, kurios pirkimų ir pardavimų santykis yra didesnis. Kiti kintamieji ryškiai neišsiskiria iš bendro įmonių konteksto, todėl galima manyti, kad 12 klasteriui priskirtos įmonės perka ir parduoda nedidelėmis sumomis. Pastebima efektyvaus PVM tarifo didesnė sklaida pirmame ketvirtyje. Todėl šiam klasteriui priskiriamos įmonės, kurios turi mažesnę nei vidutinis R1 rodiklį pirmame ketvirtyje, priskiriamos prie rizikingų.

Pirkimų ir pardavimų Lietuvoje santykiu išsiskiria 3 ir 18 klasteriai. 3 klasteryje pirmame ketvirtyje pastebima didesnio santykio išskirtis. Antrame ketvirtyje šio klasterio pirkimų ir pardavimų Lietuvoje santykio mediana yra didesnė nei kitų klasterių. Tokį efektą sukelia pastebimas klasterio pirkimų sumos Lietuvoje medianos padidėjimas. Pirkimų padidėjimą įmonėse tam tikrą ketvirtį kiekvienais metais gali lemti numatyti sezoniniai atsargų užpirkimai. Kadangi efektyvaus PVM tarifo mediana neišsiskiria iš kitų klasterių nei viename ketvirtyje, 3 klasterio įmonės nepriskiriamos prie rizikingų.

18 klasterio pirkimų ir pardavimų Lietuvoje santykio mediana pirmame ketvirtyje yra sąlyginai nedaug didesnė nei kitų klasterių, o ketvirtame ketvirtyje pastebimai didesnė. Kiti kintamųjų nei medianos reikšmė nei stebėjimų sklaida neišsiskiria nuo daugelio kitų klasterių. Dėl tos pačios sezoninių atsargų užpirkimo prielaidos ir efektyvaus PVM tarifo neišsiskyrimo, 18 klasterio įmonės nepriskiriamos rizikingoms.

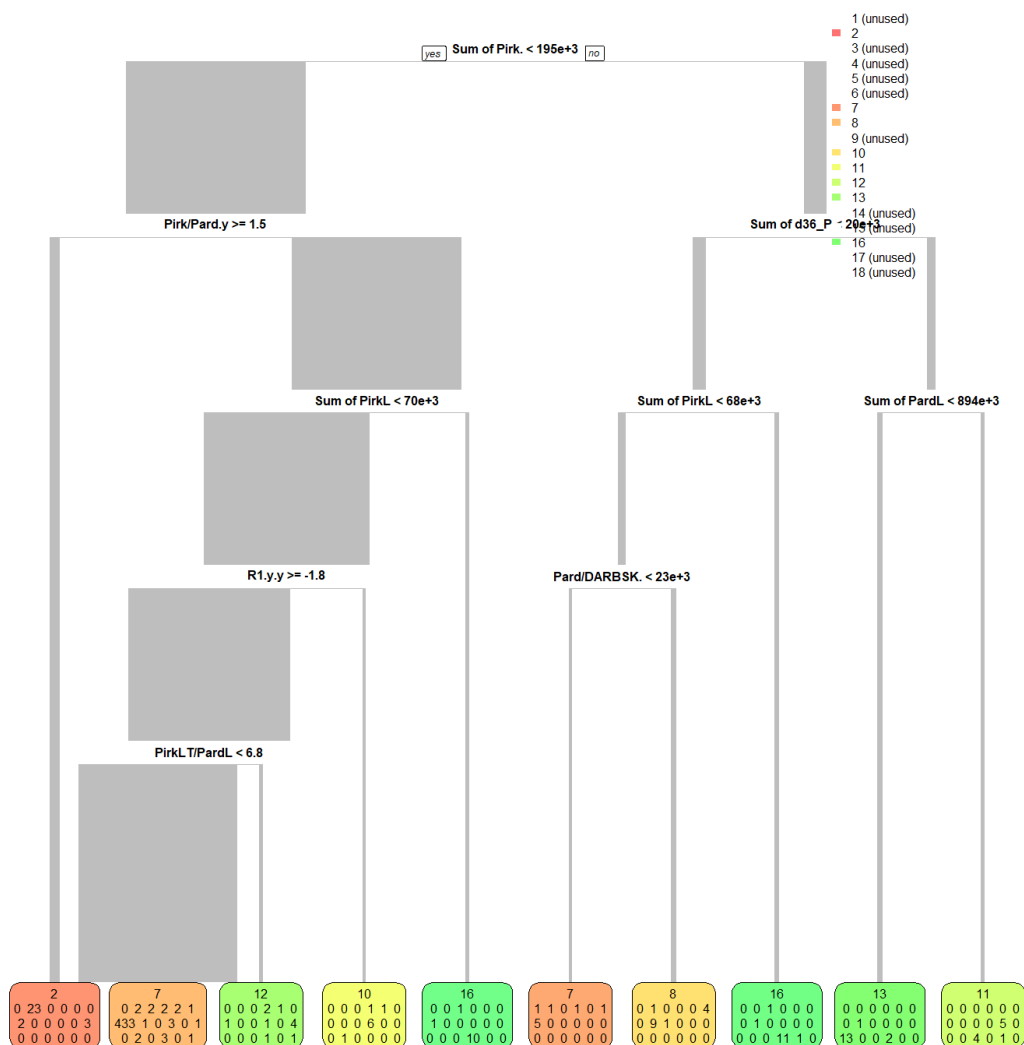
Didesne bendra PVM suma išsiskiria 11 ir 15 klasteriai. Abu klasteriai pasižymi didesniais pardavimais Lietuvoje ir bendrais, didesniais bendrais pirkimais ir didesnia PVM sumos kaita kiekvienais metais tame pačiame ketvirtyje. Tokios tendencijos išlieka visuose ketvirčiuose. 11 klasteris išsiskiria antrame ir ketvirtame ketvirčiuose didesniais pardavimais vienam darbuotojui ir

visuose ketvirčiuose mažesniais pirkimais Lietuvoje. Pagal tokius rodiklius, galima daryti prielaidą, kad 11 klasteryje yra nedidelės, tačiau daug parduodančios įmonės, o 15 klasteryje yra didelės ir daug parduodančios įmonės. Kadangi efektyvus PVM tarifas abiejų klasterių yra neišsiskiriantis iš kitų, 11 ir 15 klasterių įmonės nėra priskiriamos prie rizikingų.

Panašių rodiklių tendenciją kaip ir 15 klasteris turi ir 17 klasteris, tačiau šio klasterio medianos visų minėtų kintamųjų daugelyje ketvirčių yra mažesnės nei 15 klasterio. Tadėl tikėtina, kad šiame klasteryje yra panašios, tačiau kuriančios mažesnę pridėtinę vertę nei 15 klasterio įmonės. 17 klasterio efektyvus PVM tarifas neišsiskiria, todėl įmonės nepriskiriamos prie rizikingų.

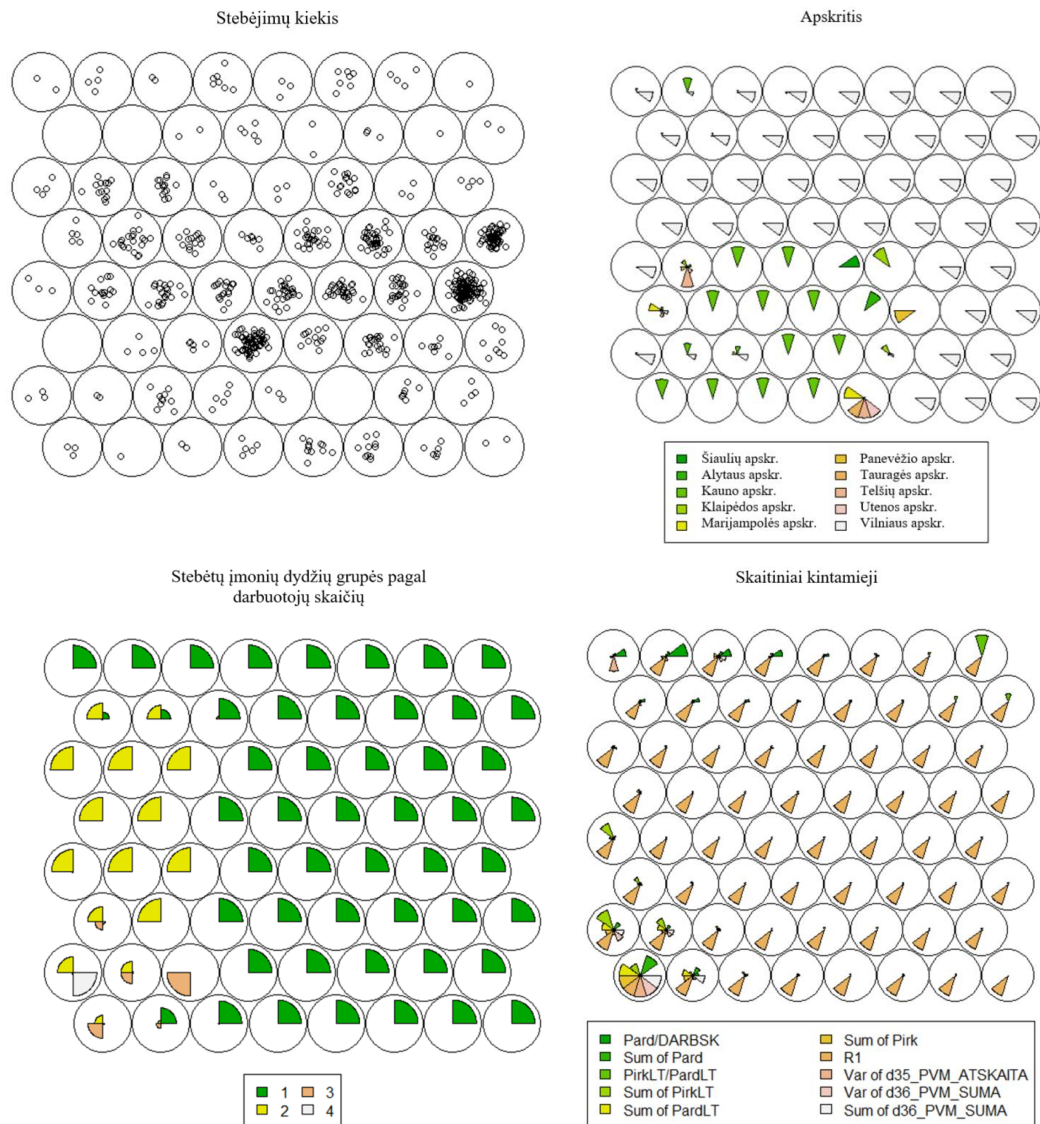
Efektyvus PVM tarifas išsiskiria, be jau minėto 12 klasterio, 10 ir 14 klasteriuose. Taip pat matomos kelios R1 rodiklio išskirtys 2, 4 ir didžiausiame (7) klasteryje. 10 ir 14 klasterių įmonių kiti rodikliai skiriasi nedaug nuo kitų klasterių. Tikėtina, kad šios įmonės vykdo veiklas nedidelėmis sumomis. Kadangi R1 rodiklis visų įmonių esančių šiuose klasteriuose yra mažesnis nei kitų klasterių medianos, visos 10 ir 14 klasterių įmonės priskiriamos prie rizikingų.

2, 4 ir 7 klasterių kiti rodikliai, išskyrus išskirtis, neišsiskiria nuo bendrų tendencijų. Todėl visos šios efektyvaus PVM tarifo išskirtys pažymimos kaip rizikingos.



28 pav. Sprendimų medžio rezultatas

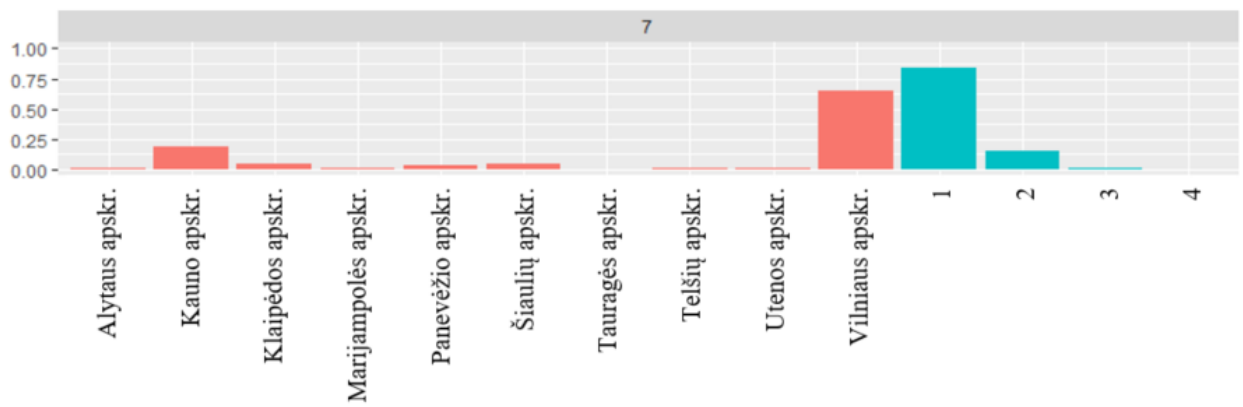
Sprendimų medžio algoritmas ir klasterizavimas suskirsto vienodai 89,95% visų stebėjimų. Iš sprendimo medžio taisyklių (žr. 28 paveikslą) galima matyti, kad, pagal sprendimų medžio atskyrimą, tipiniai PVM mokėtojai yra tie, kurių bendra pirkimų suma ketvirtame ketvirtyje yra mažesnė už 195084,4 €, bendrų pirkimų ir pardavimų santykis antrame ketvirtyje yra mažesnis už 1,5, pirkimų Lietuvoje suma trečiame ketvirtyje yra mažesnė už 70344,68 €, R1 ketvirtame ketvirtyje yra didesnis už -1,75 bei pirkimų ir pardavimų Lietuvoje santykis pirmame ketvirtyje mažesnis už 6,78. Pagal tokias taisykles, teisingai atrenkamos 433 iš 442 įmonių ir priskiriamos 20 įmonių, kurios, pagal klasterizavimo algoritmą, pateko į kitus klasterius.



29 pav. SOM rezultato stebėjimų pasiskirstymas pagal apskritį, skaitinius kintamuosius ir įmonės dydį

SOM algoritmu suskirstyti duomenys, pagal kategorinius kintamuosius (apskritį ir įmonės dydį), parodo (žr. 29 paveikslą), kad dauguma tiriamų įmonių yra įsikūrusios Vilniuje ir Kaune. Daugiausiai yra mikro dydžio grupės įmonių.

Kitaip nei kitose tirtose veiklose, elektronikos prekių pardavimų veiklos įmonės, kurių pardavimų suma (Lietuvoje ir bendra) yra didžiausia priklauso mažų ir vidutinių įmonių dydžio kategorijoms. O daugiausiai pirkimų sumos Lietuvoje priklauso ne tik didelioms įmonėms, bet ir mažoms. Pirkimų ir pardavimų Lietuvoje santykis didžiausias ir pardavimų suma vienam darbuotojui yra didžiausia mikro dydžio įmonėse.



30 pav. Didžiausio klasterio apskrities ir įmonės dydžio kintamųjų centruotos ir bendro mastelio medianų reikšmės

30 paveiksle matoma, kad didžiausiame klasteryje (nr. 7) yra daugumos apskričių įmonių (daugiausiai Vilniaus ir Kauno apskrities), nėra tik Tauragės apskrities. Tiriamos veiklos tipinio PVM mokėtojo klasteryje daugiausiai yra mikro įmonių. Didžiausiame klasteryje tik nedidelį procentą sudaro vidutinės įmonės, o didelių įmonių nėra. Todėl, kaip ir kitose tirtose veiklose, papildančiam tyrimui rekomenduotina atskirai analizuoti mikro ir mažo dydžio grupės įmones.

Išvados

1. Pridėtinės vertės mokestis, dėl sudėtingos skaičiavimo ir mokėjimo struktūros, sudaro galimybes mokestiniam sukčiavimui. Norint stabdyti sukčiavimą, mokesčių administratoriui rekomenduojama plačiau naudotis surenkamais duomenimis apie įmones. Vienos iš rizikingiausių sukčiauti įmonės užsiima statybų, žemės ūkio ar elektronikos pardavimų veiklomis.
2. Tikslui pasiekti tinkamiausi tyrimo metodai yra saviorganizuojantys neuroniniai tinklai kartu su hierarchiniu klasterizavimu ir sprendimų medis.
3. Tiriant kiekvieną ekonominę veiklą atskirai, didžiosios įmonių dalies rodikliai yra panašūs, todėl aiškiai atsiskiria veiklos tipinės grupės elgsena. Kiekvienos tirtos veiklos tipinė įmonė savo rodikliais daug neišsiskiria nuo visos veiklos vidutinių rodiklių. Vidutinė sąžiningai dirbanti įmonė turi visų rodiklių reikšmes mažesnes už veiklos vidutines, išskyrus efektyvų pridėtinės vertės mokesčio tarifą, kuris yra didesnis nei veiklos vidutinis. Įmonės, kurios neatitinka šių kriterijų yra laikomos netipinėmis ir galimai sukčiaujančios.

Žemės ūkio prekių pardavimo veikloje tipiniai pridėtinės vertės mokesčio mokėtojai išsiskiria ketvirtiniu sezoniškumu. Iš šios veiklos sąžiningai dirbančių įmonių galima tikėtis pardavimų sumos ir pirkimų sumos kitimo skirtingais ketvirčiais. Dėl pardavimų ir pirkimų netolygaus svyravimo gali atsirasti ir efektyvaus pridėtinės vertės mokesčio tarifo kitimas skirtingais ketvirčiais.

Su statybų sektoriumi susijusių pardavimų veiklos tipiniai pridėtinės vertės mokesčio mokėtojai iš bendrų veiklos įmonių ketvirtiniu sezoniškumu neišsiskiria. Šios veiklos tipinių įmonių išskirtinumas yra mažiau ryškūs nei kitų tirtų veiklų tipinių įmonių. Dėl to, su statybomis susijusių pardavimų veikloje sudėtingiau pastebėti sukčiaujančias įmones.

Elektronikos prekių pardavimo veikloje tipiniai pridėtinės vertės mokesčio mokėtojai pasižymi pardavimų sumos vienam darbuotojui ir pirkimų ir pardavimų santykio kitimu skirtingais ketvirčiais. Todėl galima tikėtis ir to, kad sąžiningai dirbanti tipinė įmonė turi efektyvaus pridėtinės vertės mokesčio tarifo svyravimus skirtingais ketvirčiais.

4. Rekomenduojama papildančiuose tyrimuose tirti mikro ir mažas įmonių grupes atskirai nuo vidutinių ir didelių. Be to, rekomenduojama duomenis papildyti kiekvienos įmonės visų jos mokamų mokesčių deklaracijų informacija, įtraukti papildomus išorinius kintamuosius, tokius kaip asmeninė informacija apie įmonės savininką, jo artimuosius ir susijusias įmones.

Literatūros sąrašas

1. VALSTYBINĖ MOKESČIŲ INSPEKCIJA. VMI administruojamų nacionalinio biudžeto pajamų surinkimo apžvalga. 2019.
2. BANKAS, Lietuvos. Lietuvos ekonomikos apžvalga. 2018, birželis.
3. BANARESCU, Adrian. Detecting and preventing fraud with data analytics. *Procedia economics and finance*, 2015, 32: 1827-1836.
4. BAESSENS, Bart; VAN VLASSELAER, Veronique; VERBEKE, Wouter. Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection. *John Wiley & Sons*, 2015.
5. SEIMAS, Lietuvos Respublikos. Lietuvos Respublikos pridėtinės vertės mokesčio įstatymas. Nr. IX-751. *Valstybės žinios*, 2002, 35-1271.
6. VALSTYBINĖ MOKESČIŲ INSPEKCIJA, prieiga per: <https://www.vmi.lt/cms/pridetines-vertes-mokestis4>
7. ČERKA, Paulius; GRIGIENĖ, Jurgita. Piktnaudžiavimas mokesčių teisėje, kurio pagrindu prarandama teisė į PVM atskaitą. *Teisės apžvalga= Law review [elektroninis išteklius]*. Kaunas: Vytauto Didžiojo universitetas, 2017, nr. 2 (16), 2017.
8. CNOSSEN, Sijbren. Global trends and issues in value added taxation. *International Tax and Public Finance*, 1998, 5.3: 399-428.
9. POMERANZ, Dina. No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 2015, 105.8: 2539-69.
10. TAIT, Mr Alan A. Value added tax: International practice and problems. *International Monetary Fund*, 1988.
11. KEEN, Michael; MINTZ, Jack. The optimal threshold for a value-added tax. *Journal of Public Economics*, 2004, 88.3-4: 559-576.
12. HINES Jr, J., & DESAI, M. A. Value-added taxes and international trades: the evidence. *Unpublished Manuscript*, 2005.
13. BIRD, Richard, et al. The VAT in developing and transitional countries. *Cambridge Books*, 2007.
14. WEIDENBAUM, Murray L.; CHRISTIAN JR, Ernest S. The Allure of Value-Added Taxes: Examining the Pros and Cons. 1989.
15. PECHMAN, J. A consumption tax is not desirable for the united states. *Walker. In Walker, C. & Bloomfield, M., editors, The Consumption Tax. Ballinger Publishing Company*, 1987.
16. ROTHBARD, Murray N. The Consumption Tax: A Critique. *The Review of Austrian Economics*, 1994, 7.2: 75-90.
17. EBRILL LP, KEEN M, BODIN JP, SUMMERS VP. The modern VAT. *International Monetary Fund*, 2001.
18. PAJUODIENĖ, Gina Marija. Mokesčių sistemos formavimas išsivysčiusiose šalyse. Vilnius: Lietuvos informacijos institutas, 1993, 29-33.
19. KOGELS, Han. VAT fraud with emission allowances trading. 2010.
20. VALSTYBINĖ MOKESČIŲ INSPEKCIJA PRIE LIETUVOS RESPUBLIKOS FINANSŲ MINISTERIJOS. Dėl įregistravimo į Pridėtinės vertės mokesčio mokėtojų registrą / išregistravimo iš Pridėtinės vertės mokesčių mokėtojų registro taisyklių. *Valstybės žinios*, 2002, Nr. 66-2744

21. VALSTYBINĖ MOKESČIŲ INSPEKCIJA, prieiga per:
<https://www.vmi.lt/cms/web/guest/pvm-sukciavimas>
22. SMITH, Stephen; KEEN, Mr Michael. VAT fraud and evasion: what do we know, and what can be done?. *International Monetary Fund*, 2007.
23. EUROPOS AUDITO RŪMAI. Kova Su Sukčiavimu PVM Bendrijos Viduje: Reikia Imtis Daugiau Veiksmų : (Pagal SESV 287 Straipsnio 4 Dalies Antrą Pastraipą). *Liuksemburgas: Europos Sąjungos leidinių biuras*, 2016. ISBN 9789287238108.
24. ALLINGHAM, Michael G.; SANDMO, Agnar. Income tax evasion: A theoretical analysis. *Journal of public economics*, 1972, 1.3-4: 323-338.
25. JENSEN, Thomas C.; WANHILL, Stephen. Tourism's taxing times: value added tax in Europe and Denmark. *Tourism Management*, 2002, 23.1: 67-79.
26. VYRIAUSYBĖ, Lietuvos Respublikos. Dėl Lietuvos stabilumo 2019 metų programos. Nr. 385. *TAR*, 2019, Nr. 6902.
27. GAPŠYS, Albertas; EIČAITĖ, Ovidija. Šešėlinės ekonomikos žemės ūkio ir maisto produktų rinkoje įtaka šalies biudžetui. *Vadybos mokslas ir studijos-kaimo verslų ir jų infrastruktūros plėtrai*, 2010, 3: 15-24.
28. Eurostat, <https://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do>
29. SANDERS, Robert. The Pareto principle: its use and abuse. *Journal of Services Marketing*, 1987.
30. ALM, James; BLACKWELL, C.; MCKEE, Michael. Audit Selection and Firm Compliance with a Broad-based Sales Tax. *National Tax Journal*, 2004, 57, (2), 1, 209-27.
31. MURRAY, Matthew N. Sales tax compliance and audit selection. *National Tax Journal*, 1995, 515-530.
32. MANASAN, Rosario G. *Estimating Industry Benchmarks for the Value Added Tax*. PIDS Discussion Paper Series, 2002.
33. WIERSEMA, William H. Will the IRS Audit you. *Electrical Apparatus*, 1997, 50: 38-40.
34. HO, Daniel; LAU, Peter. Tax Audits in Hong Kong. *Int'l Tax J.*, 1999, 25: 61.
35. KUMAR, Anuj; NAGADEVARA, Vishnuprasad. Development of hybrid classification methodology for mining skewed data sets-a case study of indian customs data. In: *IEEE International Conference on Computer Systems and Applications*, 2006. IEEE, 2006. p. 584-591.
36. MICCI-BARRECA, Daniele; RAMACHANDRAN, Satheesh. Improving tax administration with data mining. *White paper. Elite Analytics LLC*, 2004.
37. GUPTA, Manish; NAGADEVARA, Vishnuprasad. Audit selection strategy for improving tax compliance: application of data mining techniques. In: *Foundations of Risk-Based Audits. Proceedings of the eleventh International Conference on e-Governance, Hyderabad, India, December. 2007*. p. 28-30.
38. RAVISANKAR, Pediredla, et al. Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 2011, 50.2: 491-500.
39. KHWAJA, Muwer Sultan; AWASTHI, Rajul; LOEPRICK, Jan (ed.). *Risk-Based Tax Audits: Approaches and Country Experiences*. The World Bank, 2011.
40. LIU, Xiaoqing; PAN, Ding; CHEN, Shihong. Application of hierarchical clustering in tax inspection case-selecting. In: *2010 International Conference on Computational Intelligence and Software Engineering*. IEEE, 2010. p. 1-4.

41. CECIL, H. Wayne. Assuring individual taxpayer compliance: Audit rates, selection methods, and electronic auditing. *The CPA Journal*, 1998, 68.12: 60.
42. HAND, David J. Principles of data mining. *Drug safety*, 2007, 30.7: 621-622.
43. LIAO, Shu-hsien. Knowledge management technologies and applications—literature review from 1995 to 2002. *Expert systems with applications*, 2003, 25.2: 155-164.
44. HAN, Jiawei; PEI, Jian; KAMBER, Micheline. Data mining: concepts and techniques. *Elsevier*, 2011.
45. GONZALEZ, Pamela Castellón; VELASQUEZ, Juan D. Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Systems with Applications*, 2013, 40.5: 1427-1436.
46. BONCHI, Francesco, et al. A classification-based methodology for planning audit strategies in fraud detection. In: *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. 1999. p. 175-184
47. VESANTO, Juha; ALHONIEMI, Esa. Clustering of the self-organizing map. *IEEE Transactions on neural networks*, 2000, 11.3: 586-600.
48. XIONG, Hui; WU, Junjie; CHEN, Jian. K-means clustering versus validation measures: a data-distribution perspective. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2008, 39.2: 318-331.
49. WARD JR, Joe H. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 1963, 58.301: 236-244.
50. KAUFMANN, Leonard; ROUSSEEUW, Peter J. Finding groups in data: an introduction to cluster analysis. *New York: John Wiley*, 1990.
51. MURTAGH, Fionn; LEGENDRE, Pierre. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. *Journal of classification*, 2014, 31.3: 274-295.
52. MILLIGAN, Glenn W. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *psychometrika*, 1980, 45.3: 325-342.
53. MILLIGAN, Glenn W.; COOPER, Martha C. Methodology review: Clustering methods. *Applied psychological measurement*, 1987, 11.4: 329-354.
54. SZEKELY, Gabor J.; RIZZO, Maria L. Hierarchical Clustering via Joint Between-Within Distances: Extending Ward's Minimum Variance Method. *Journal of classification*, 2005, 22.2.
55. MILLIGAN, Glenn W.; SOKOL, Lisa M. A two-stage clustering algorithm with robust recovery characteristics. *Educational and psychological measurement*, 1980, 40.3: 755-759.
56. FERNÁNDEZ, Alberto; GÓMEZ, Sergio. Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms. *Journal of Classification*, 2008, 25.1: 43-65.
57. EVERITT, Brian S. Unresolved problems in cluster analysis. *Biometrics*, 1979, 169-181.
58. MODHA, Dharmendra S.; SPANGLER, W. Scott. Feature weighting in k-means clustering. *Machine learning*, 2003, 52.3: 217-237.
59. KOHONEN, T. Self-organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 1982, 43, 59-69.
60. OBERMAYER, Klaus, et al. (ed.). Self-organizing map formation: foundations of neural computation. *Mit Press*, 2001.
61. KOHONEN, Teuvo. The self-organizing map. *Proceedings of the IEEE*, 1990, 78.9: 1464-1480.

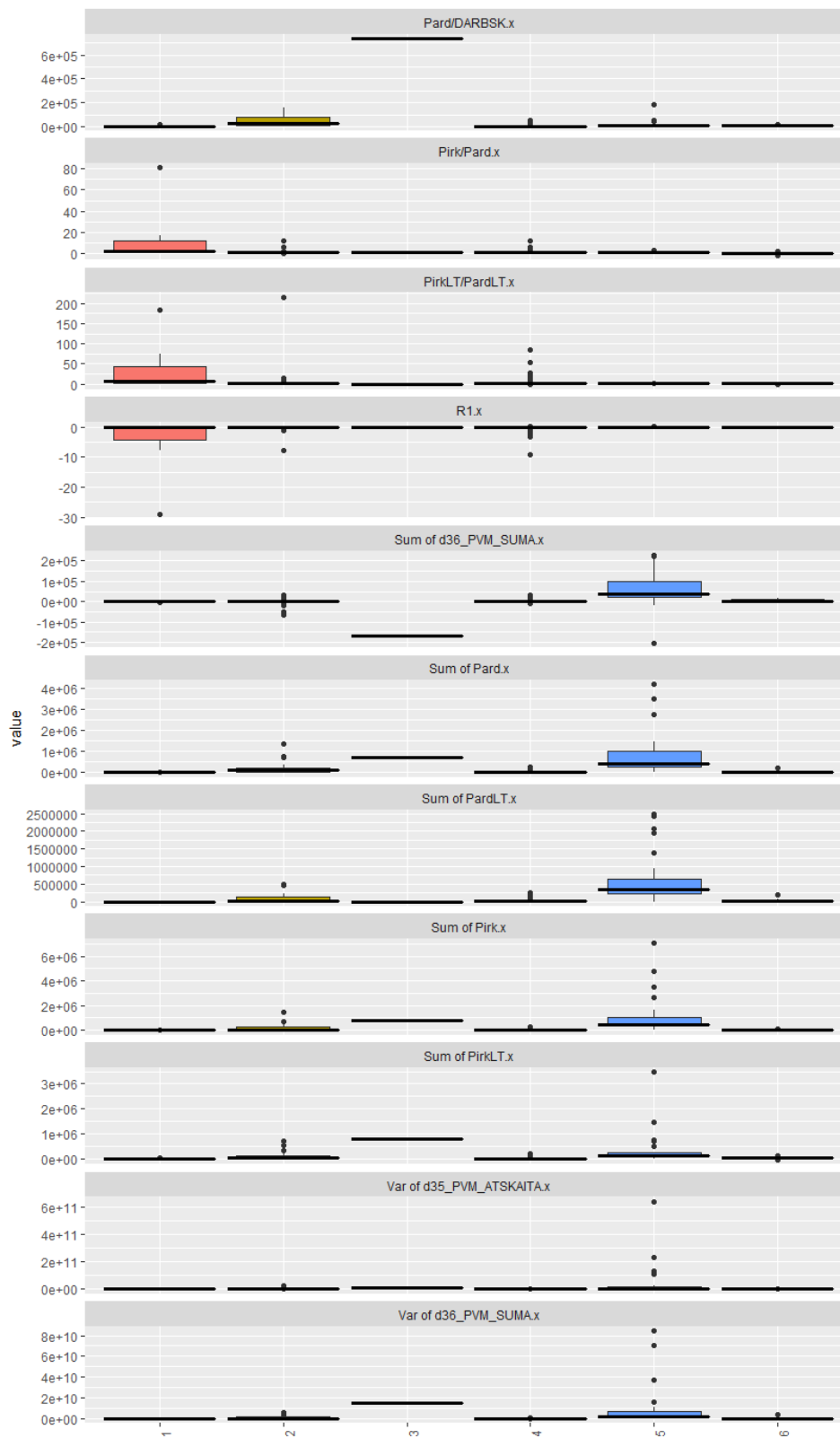
62. MURTAGH, F.; HERNÁNDEZ-PAJARES, M. The Kohonen self-organizing map method: an assessment. *Journal of Classification*, 1995, 12.2: 165-190.
63. MANGIAMELI, Paul; CHEN, Shaw K.; WEST, David. A comparison of SOM neural network and hierarchical clustering methods. *European Journal of Operational Research*, 1996, 93.2: 402-417.
64. LAMPINEN, Jouko; OJA, Erkki. Clustering properties of hierarchical self-organizing maps. *Journal of Mathematical Imaging and Vision*, 1992, 2.2-3: 261-272.
65. THERNEAU, Terry M., et al. An introduction to recursive partitioning using the RPART routines. 1997.
66. TODOROVSKI, Ljupco; BLOCKEEL, Hendrik; DZEROSKI, Saso. Ranking with predictive clustering trees. In: *European Conference on Machine Learning*. Springer, Berlin, Heidelberg, 2002. p. 444-455.
67. BASAK, Jayanta; KRISHNAPURAM, Raghu. Interpretable hierarchical clustering by constructing an unsupervised decision tree. *IEEE transactions on knowledge and data engineering*, 2005, 17.1: 121-132.
68. YOOK, Dongsuk. Decision tree based clustering. In: *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, Berlin, Heidelberg, 2002. p. 487-492.
69. LANGLEY, Pat. *Elements of machine learning*. Morgan Kaufmann, 1996.
70. FISHER, Douglas H. Knowledge acquisition via incremental conceptual clustering. *Machine learning*, 1987, 2.2: 139-172.
71. KHANBABAEI, Mohammad; ALBORZI, Mahmood. The use of genetic algorithm, clustering and feature selection techniques in construction of decision tree models for credit scoring. *International Journal of Managing Information Technology*, 2013, 5.4: 13-32.
72. BECKER, R. A.; CHAMBERS, J. M.; WILKS, A. R. The New S Language. Wadsworth & Brooks/Cole. *Computer Science Series, Pacific Grove, CA*, 1988.
73. TEAM, R. Core, et al. R: A language and environment for statistical computing. 2013.
74. TEAM, R. Core. R language definition. Vienna, Austria: R foundation for statistical computing, 2000.
75. WEHRENS, Ron; KRUISSELBRINK, Johannes. Flexible Self-Organising Maps in kohonen 3.0. *Journal of Statis*, 2018.
76. WICKHAM, Hadley, et al. Welcome to the Tidyverse. *Journal of Open Source Software*, 2019, 4.43: 1686.
77. BACHE, Stefan Milton; WICKHAM, Hadley. magrittr: a forward-pipe operator for R. *R package version*, 2014, 1.1.
78. NEUWIRTH, Erich; BREWER, R. Color. ColorBrewer palettes. R package version, 2014, 1.1-2.
79. WICKHAM, H., et al. Readr: read rectangular text data. R package version 1.1. 1; 2017. 2018.
80. VALSTYBINĖ MOKESČIŲ INSPEKCIJA, prieiga per: <https://www.vmi.lt/cms/i.saf>.
81. VALSTYBINĖ MOKESČIŲ INSPEKCIJA, prieiga per: <https://www.vmi.lt/cms/documents/10174/8274962/KD-8401/a212463b-9199-4822-89e5-ca23d5588d56>
82. MAHALANOBIS, Prasanta Chandra. On the generalized distance in statistics, 1936.

83. VARMUZA, Kurt; FILZMOSER, Peter; DEHMER, Matthias. Multivariate linear QSPR/QSAR models: Rigorous evaluation of variable selection for PLS. *Computational and structural biotechnology journal*, 2013, 5.6: e201302007.

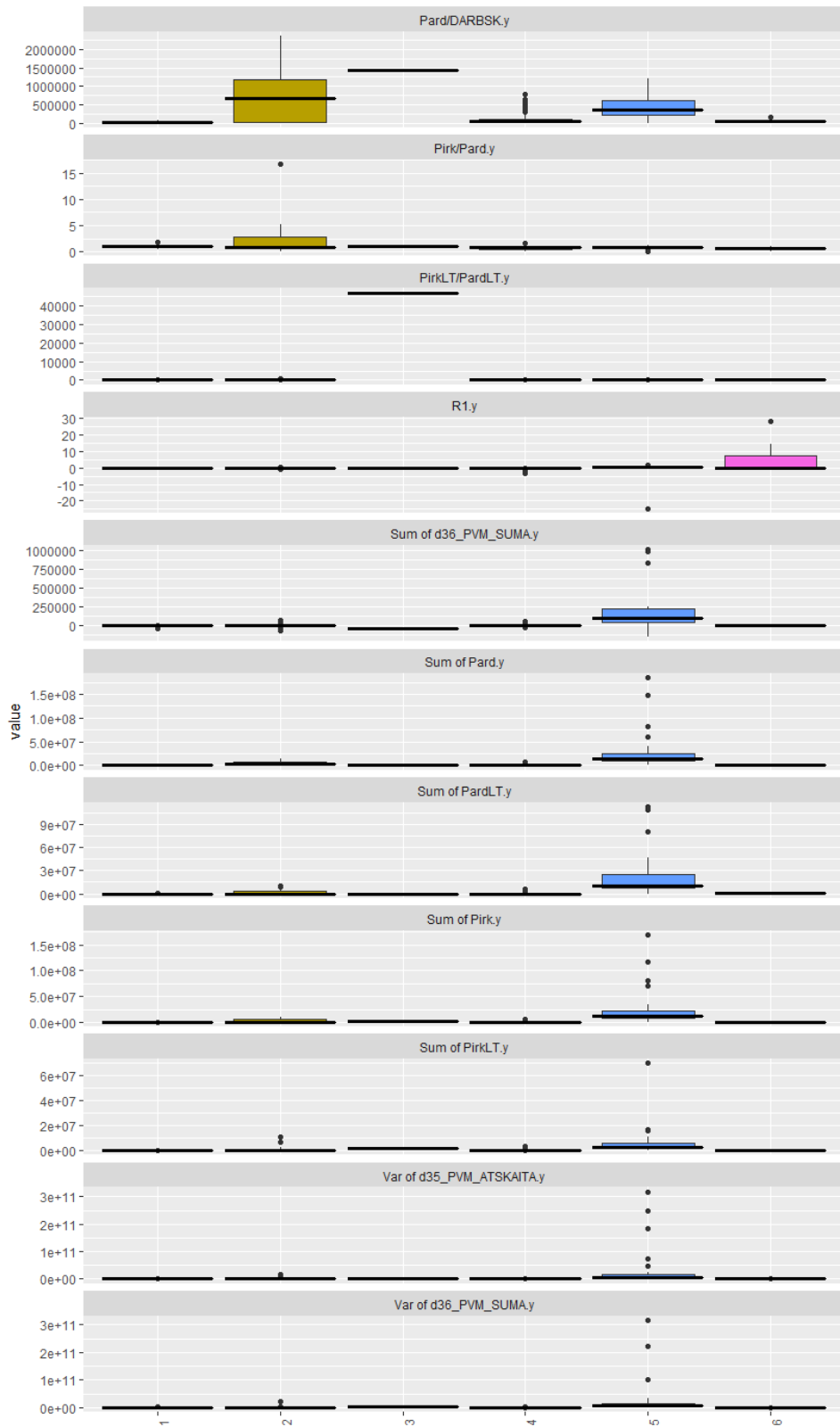
Priedai

1 priedas. Žemės ūkio pardavimų veiklos kintamųjų stačiakampės diagramos, suskirstytos klasteriais ir ketvirčiais.

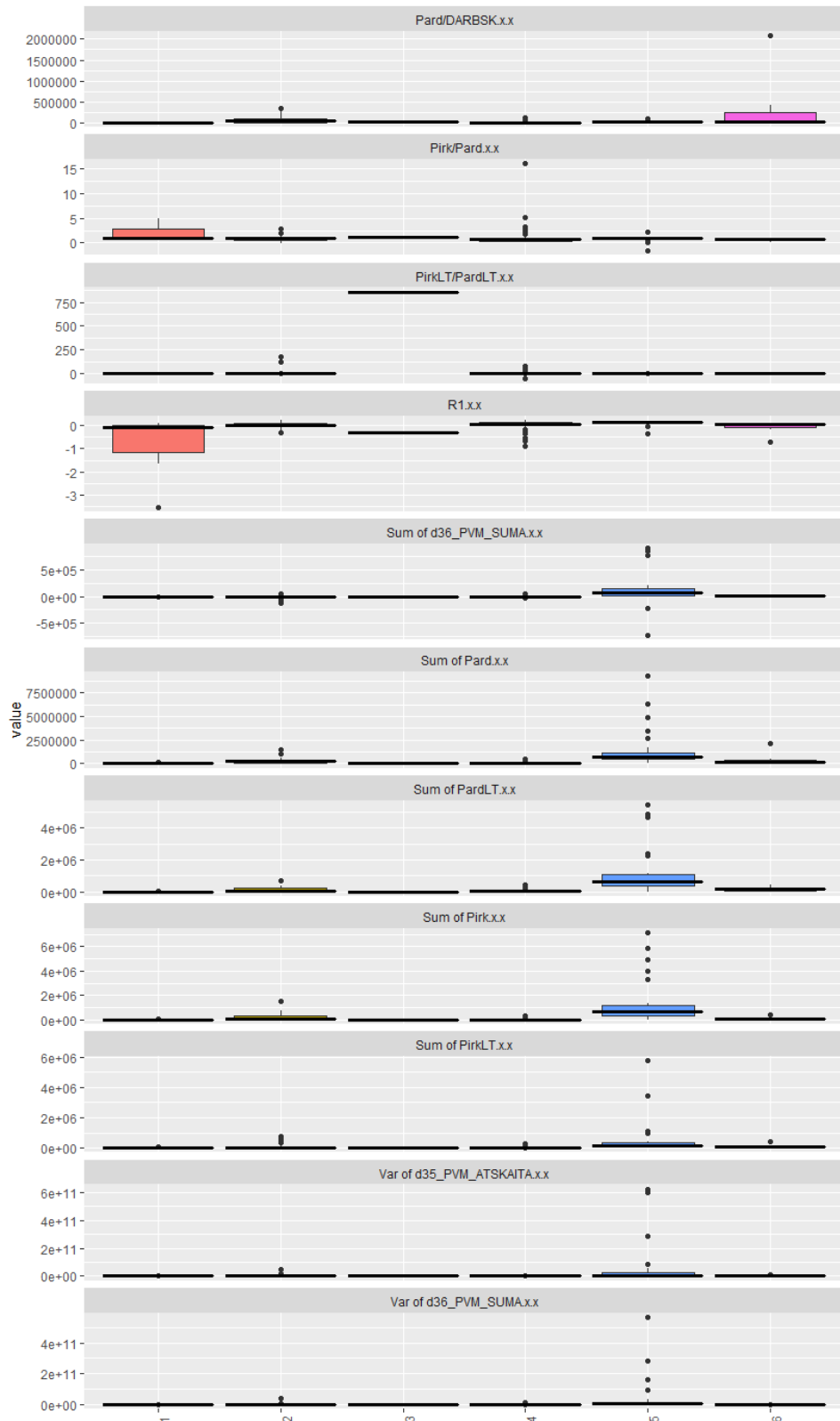
QTR1



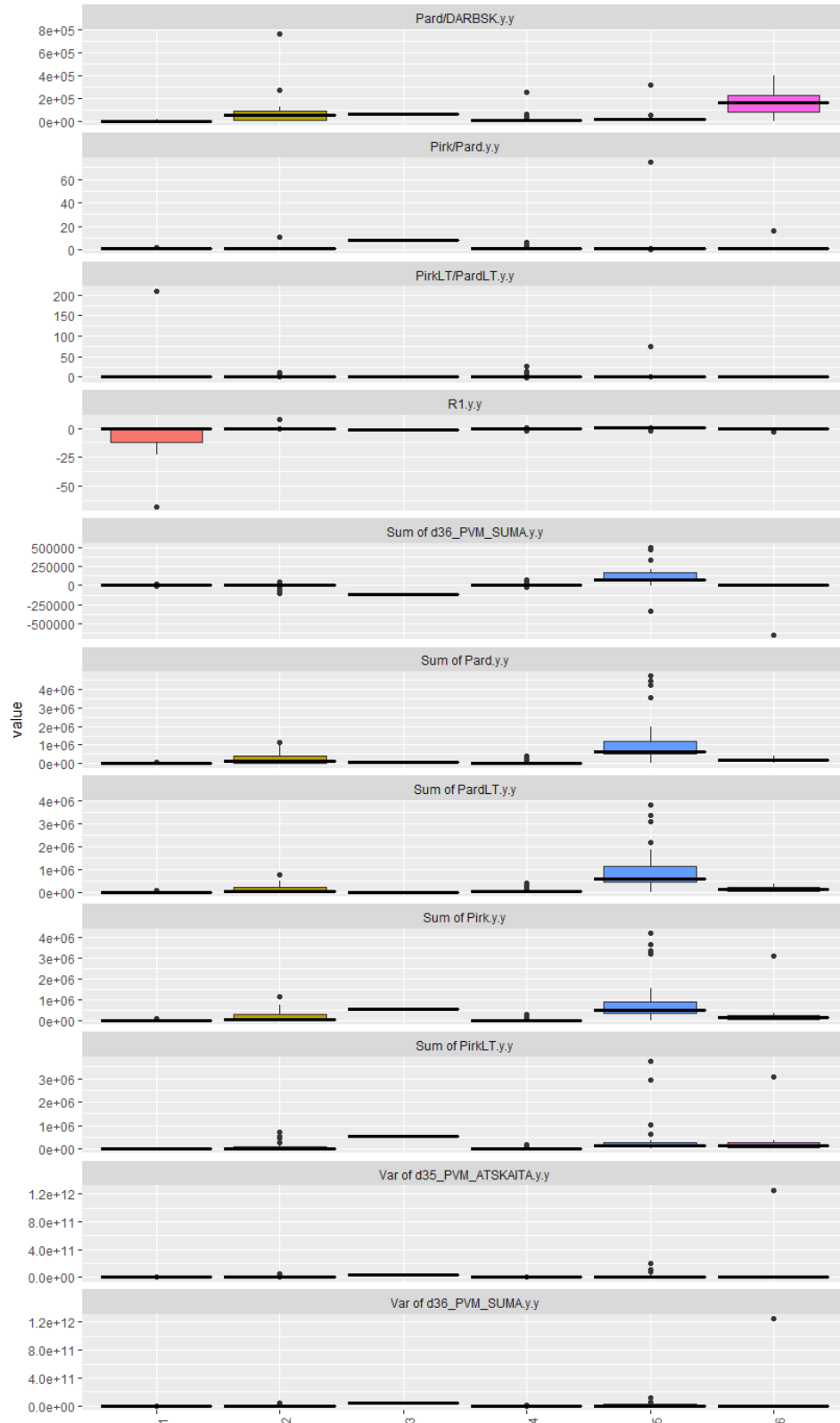
QTR2



QTR3

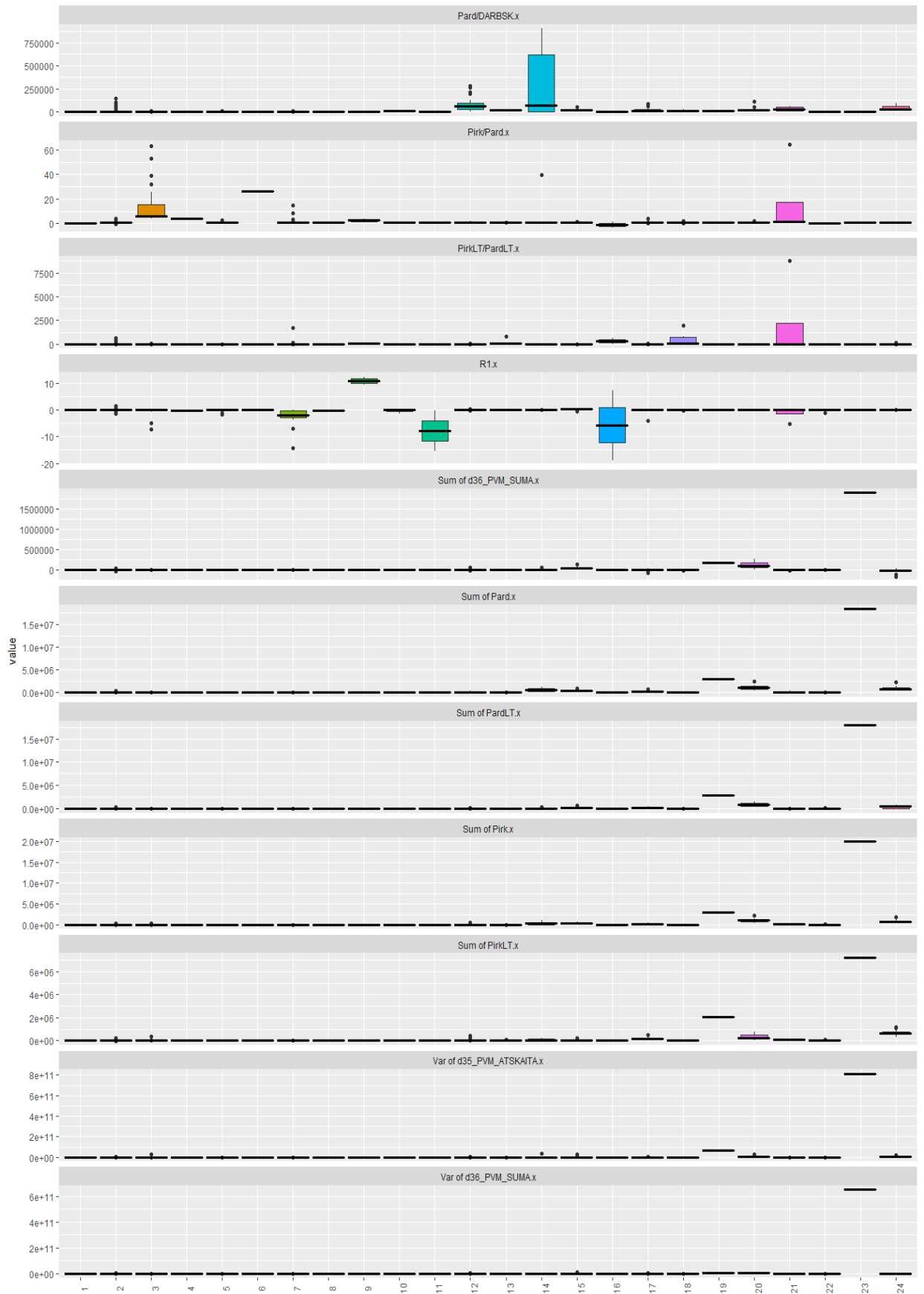


QTR4

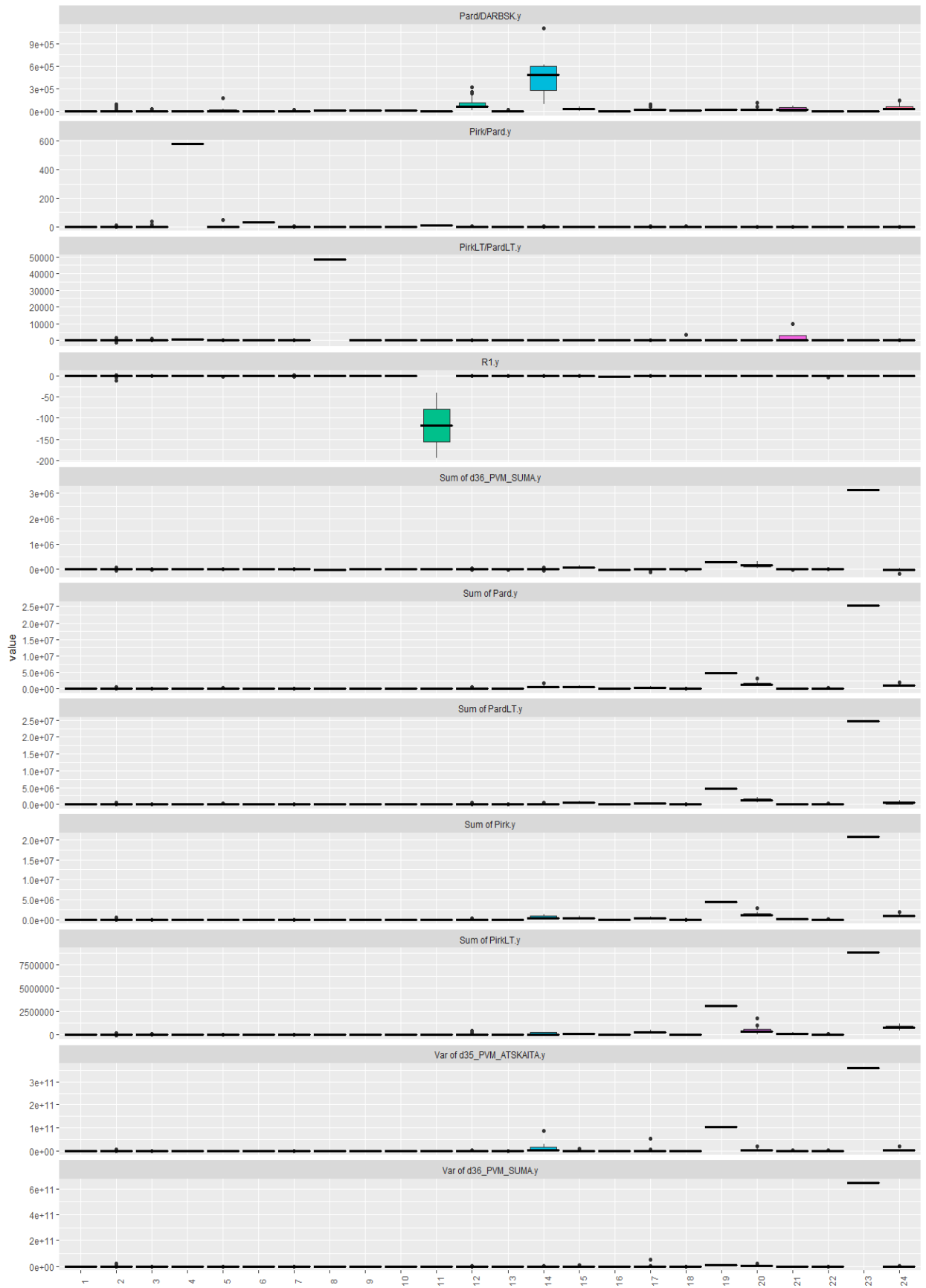


2 priedas. Su statybų sektoriumi susijusių pardavimų veiklos tiriamų kintamųjų suvestinė suskirstyta ketvirčiais.

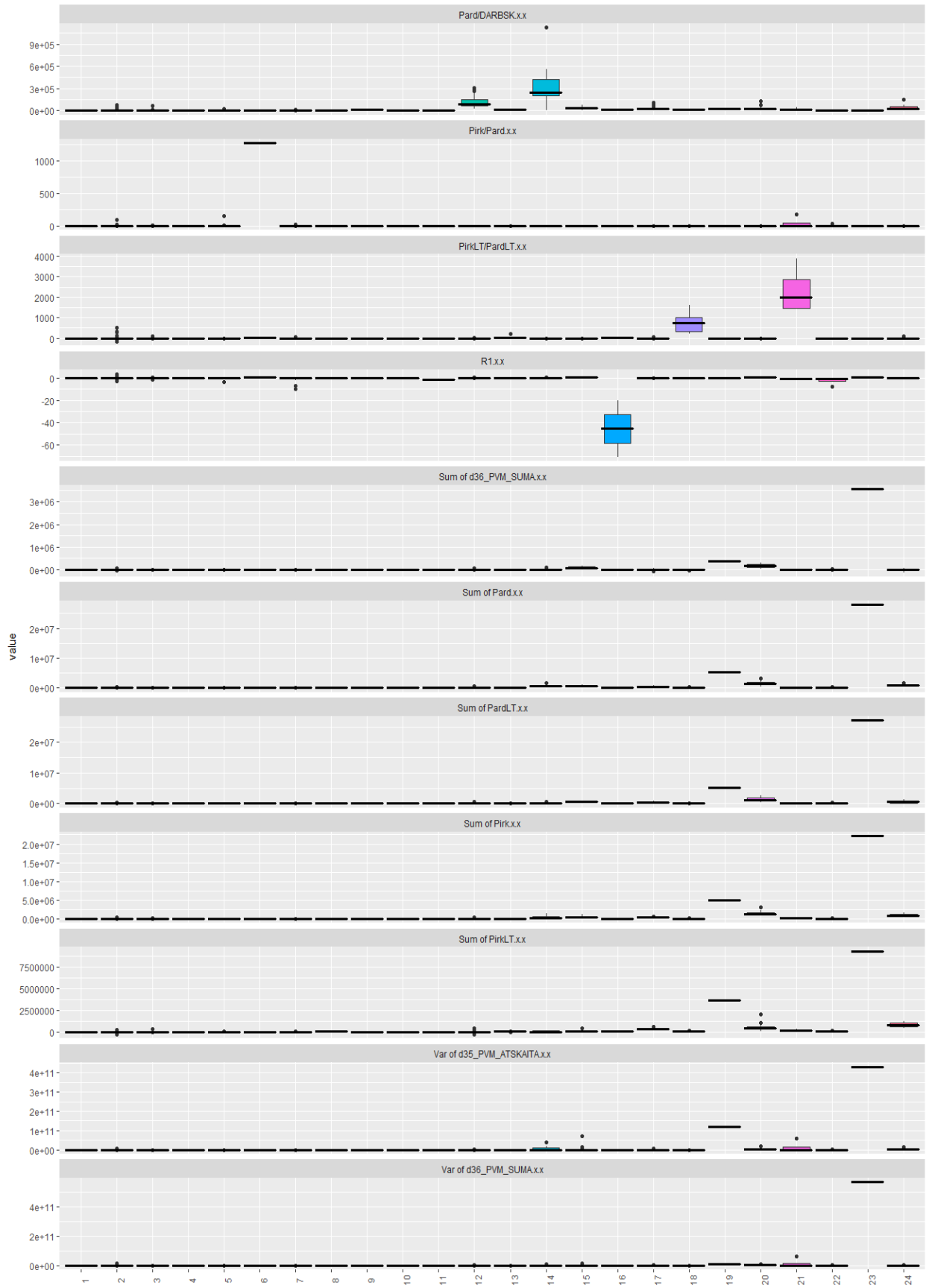
QTR1



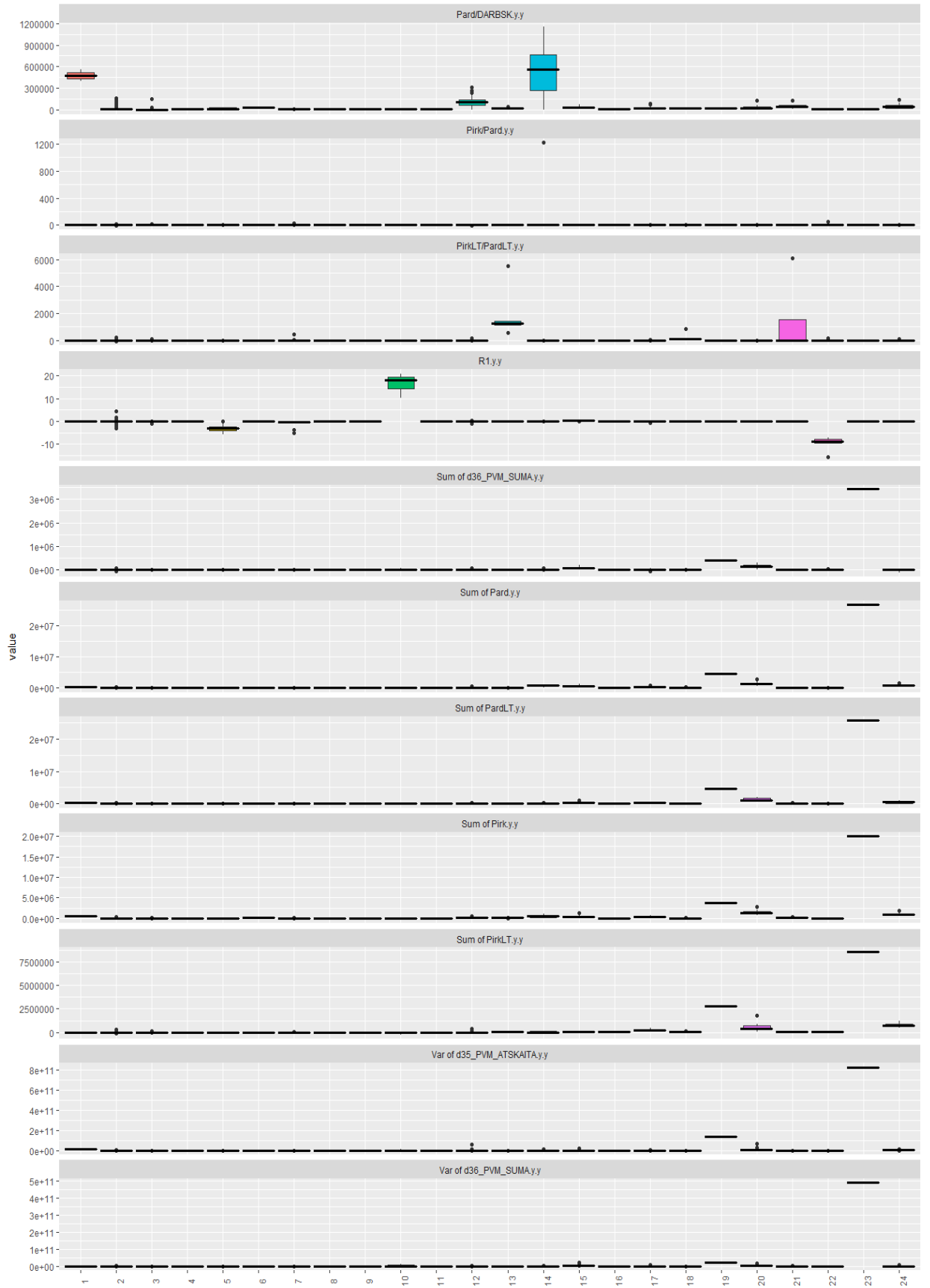
QTR2



QTR3

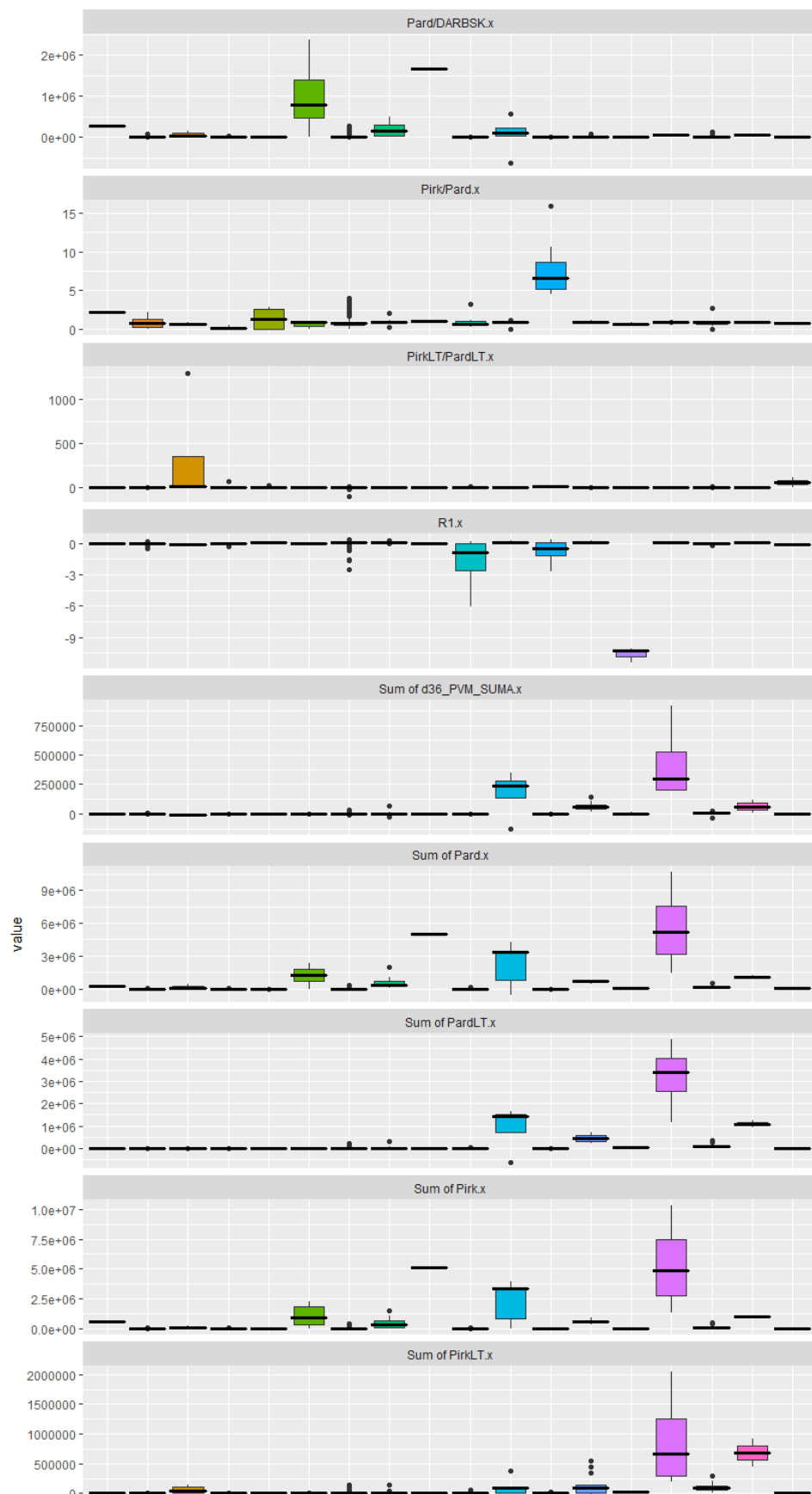


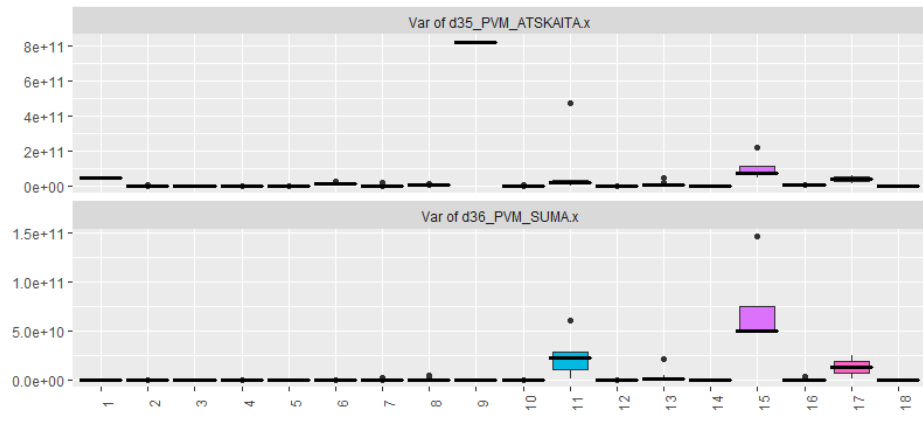
QTR4



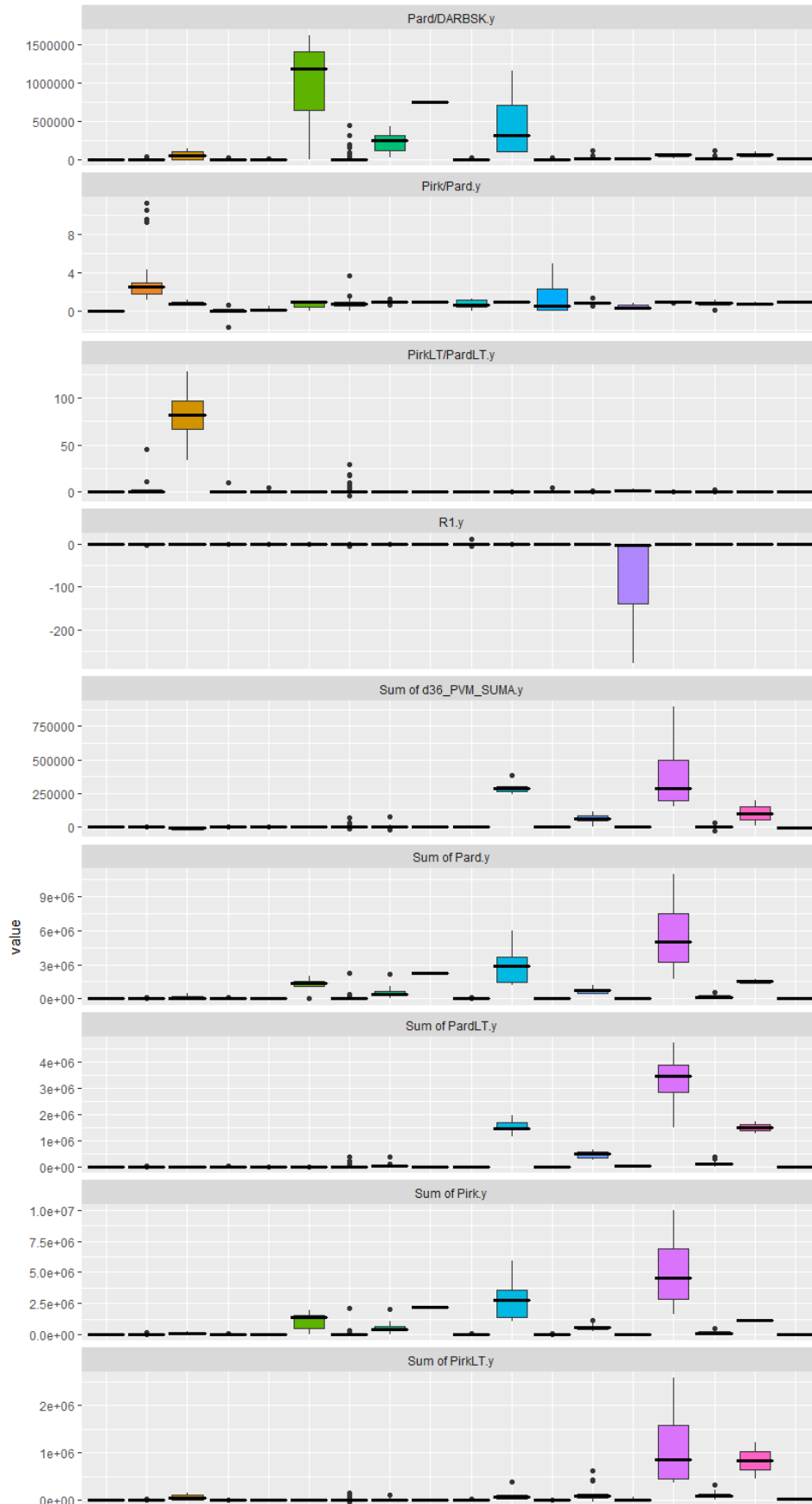
3 priedas. Elektronikos prekių pardavimų veiklos kintamųjų stačiakampės diagramos, suskirstytos klasteriais ir ketvirčiais.

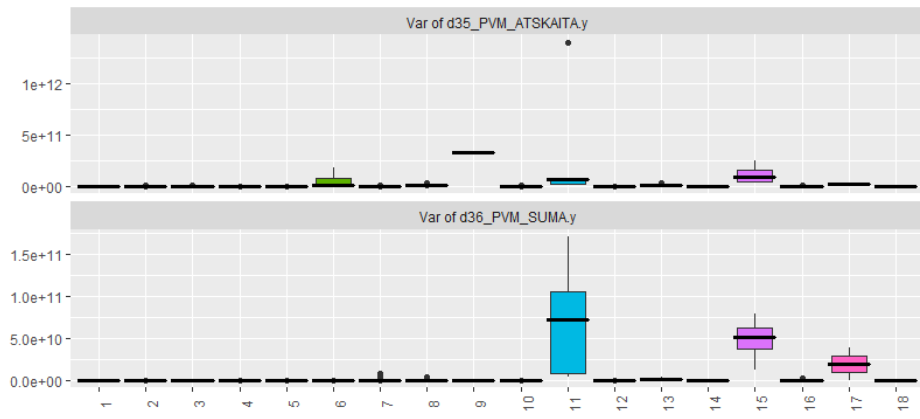
QTR1



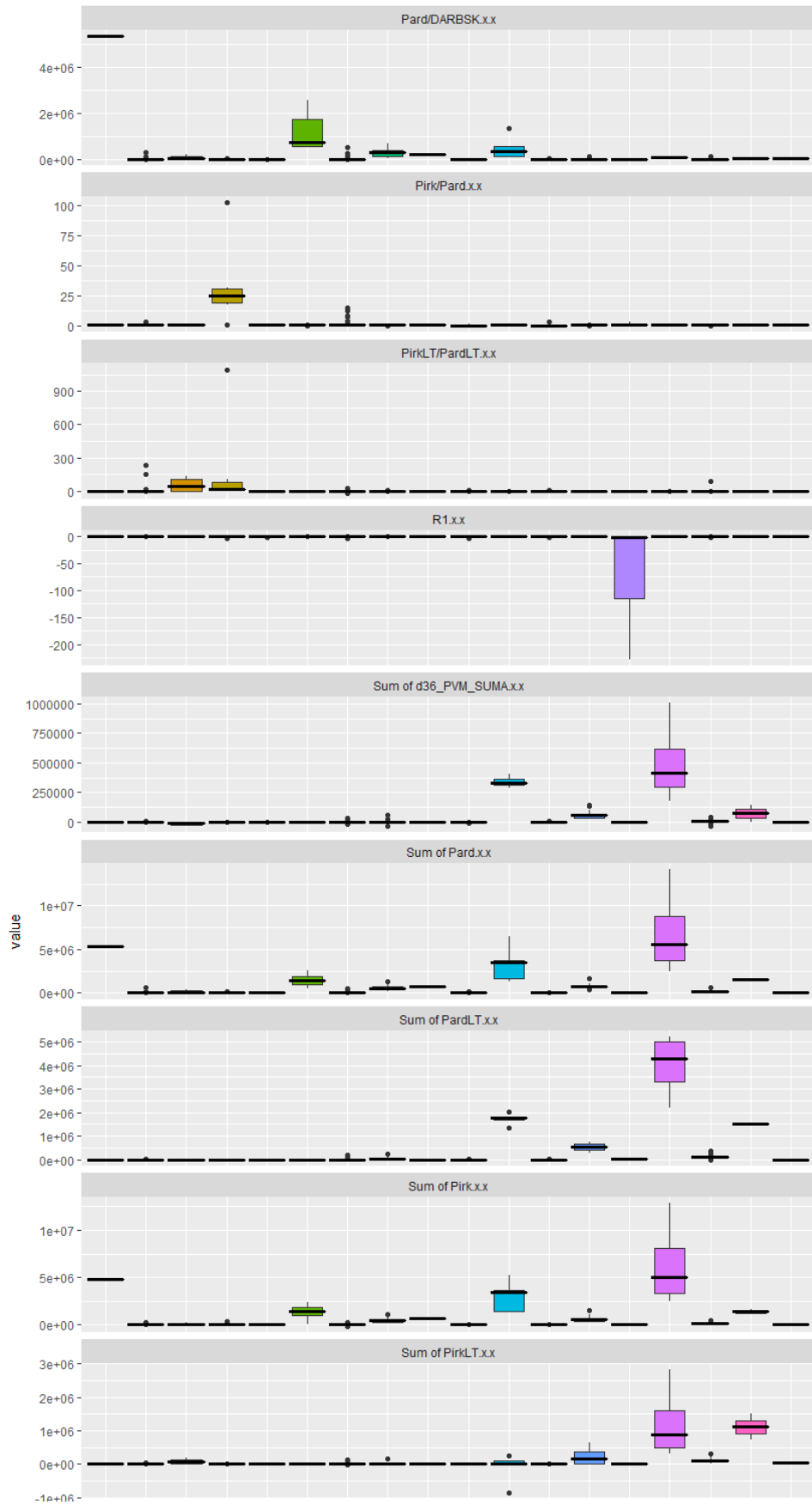


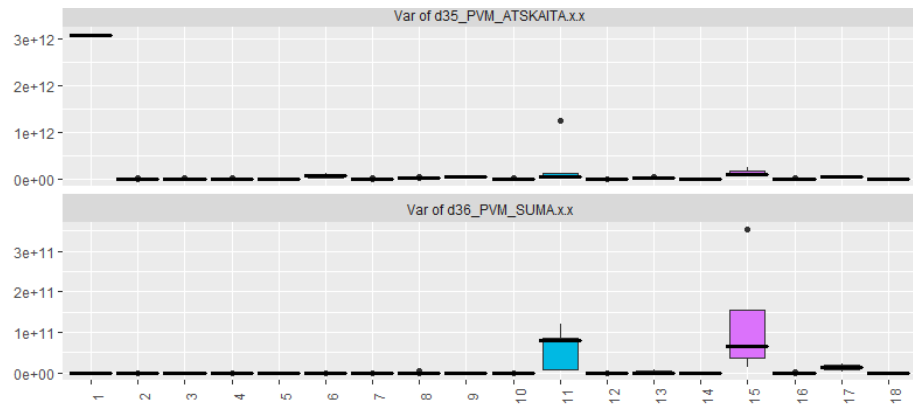
QTR2



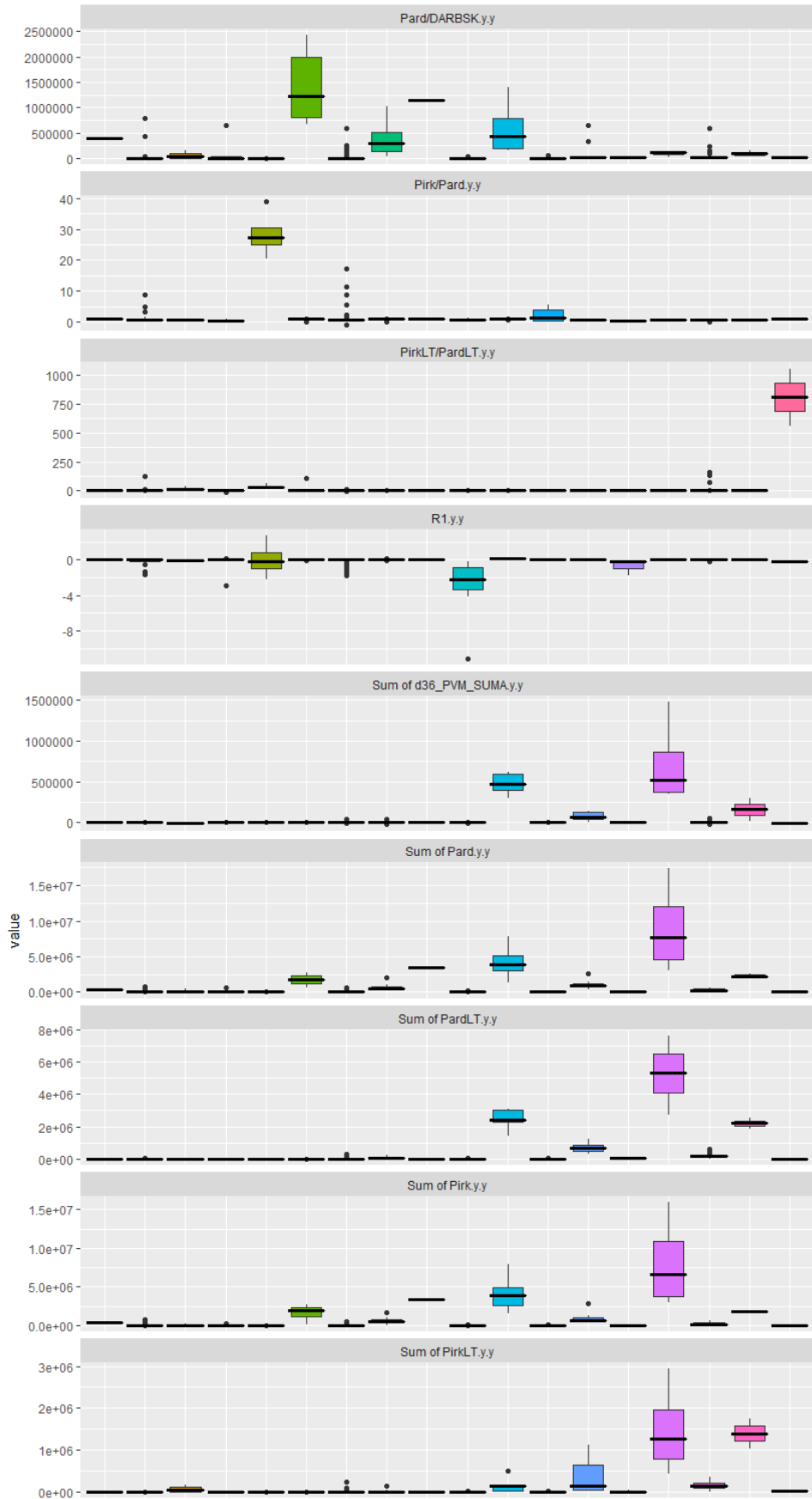


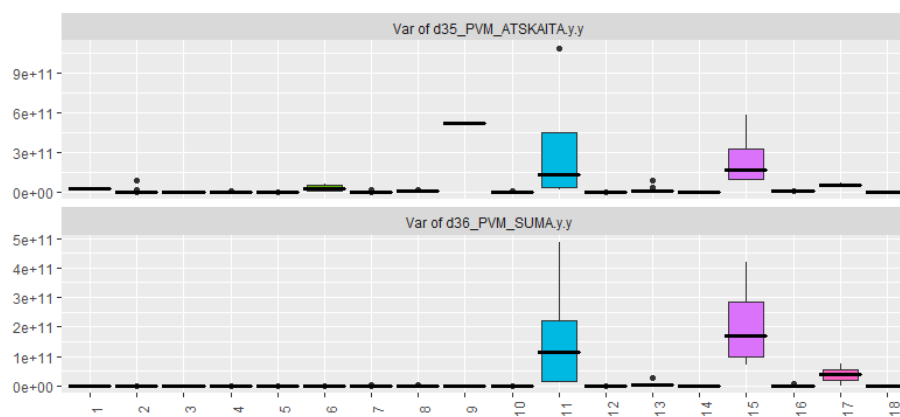
QTR3





QTR4





4 priedas. Žemės ūkio pardavimų veiklos netipinių įmonių sąrašas, jų klasteriai ir priskyrimas prie išskirčių pirminio išskirčių tyrimo metu. Paryškintos įmonės yra priskirtos prie rizikingų sukčiauti.

Įmonė	Klasteris	Išskirtis_QTR1	Išskirtis_QTR2	Išskirtis_QTR3	Išskirtis_QTR4
1	1	ne	ne	ne	išskirtis
2	1	išskirtis	ne	ne	ne
3	1	ne	ne	ne	išskirtis
4	1	ne	ne	išskirtis	ne
5	1	ne	ne	išskirtis	ne
6	1	išskirtis	ne	ne	išskirtis
7	1	išskirtis	ne	ne	ne
8	2	išskirtis	išskirtis	išskirtis	išskirtis
9	2	ne	išskirtis	ne	ne
10	2	išskirtis	išskirtis	išskirtis	išskirtis
11	2	išskirtis	ne	išskirtis	išskirtis
12	2	išskirtis	išskirtis	išskirtis	išskirtis
13	2	ne	ne	ne	ne
14	2	išskirtis	išskirtis	išskirtis	ne
15	2	išskirtis	išskirtis	išskirtis	išskirtis
16	2	ne	išskirtis	ne	ne
17	2	ne	ne	ne	išskirtis
18	2	ne	išskirtis	ne	ne
19	2	ne	ne	ne	ne
20	2	ne	ne	ne	ne
21	2	ne	išskirtis	ne	ne
22	2	ne	išskirtis	ne	ne
23	2	ne	ne	ne	ne
24	2	ne	ne	ne	išskirtis
25	2	ne	ne	ne	ne
26	2	ne	ne	ne	ne

67	6	ne	ne	išskirtis	išskirtis
68	6	ne	ne	ne	ne
69	6	ne	ne	išskirtis	išskirtis
70	4	(Išskirtis, tiriant didžiausio klasterio R1)			
71	4	(Išskirtis, tiriant didžiausio klasterio R1)			
72	4	(Išskirtis, tiriant didžiausio klasterio R1)			
73	4	(Išskirtis, tiriant didžiausio klasterio R1)			
74	4	(Išskirtis, tiriant didžiausio klasterio R1)			
75	4	(Išskirtis, tiriant didžiausio klasterio R1)			
76	4	(Išskirtis, tiriant didžiausio klasterio R1)			
77	4	(Išskirtis, tiriant didžiausio klasterio R1)			
78	4	(Išskirtis, tiriant didžiausio klasterio R1)			
79	4	(Išskirtis, tiriant didžiausio klasterio R1)			
80	4	(Išskirtis, tiriant didžiausio klasterio R1)			
81	4	(Išskirtis, tiriant didžiausio klasterio R1)			
82	4	(Išskirtis, tiriant didžiausio klasterio R1)			
83	4	(Išskirtis, tiriant didžiausio klasterio R1)			
84	4	(Išskirtis, tiriant didžiausio klasterio R1)			
85	4	(Išskirtis, tiriant didžiausio klasterio R1)			
86	4	(Išskirtis, tiriant didžiausio klasterio R1)			
87	4	(Išskirtis, tiriant didžiausio klasterio R1)			
88	4	(Išskirtis, tiriant didžiausio klasterio R1)			

5 priedas. Su statybų sektoriumi susijusių pardavimų veiklos netipinių įmonių sąrašas, jų klasteriai ir priskyrimas prie išskirčių pirminio išskirčių tyrimo metu. Paryškintos įmonės yra priskirtos prie rizikingų sukčiauti.

Įmonė	Klasteris	Išskirtis_QTR1	Išskirtis_QTR2	Išskirtis_QTR3	Išskirtis_QTR4
1	1	ne	ne	ne	ne
2	1	ne	ne	ne	ne
3	3	ne	ne	ne	ne
4	3	ne	ne	ne	ne
5	3	ne	ne	ne	ne
6	3	ne	ne	ne	ne
7	3	ne	ne	ne	ne
8	3	ne	ne	ne	ne
9	3	ne	ne	ne	ne
10	3	ne	ne	ne	ne
11	3	ne	ne	ne	ne
12	3	ne	ne	ne	ne

13	3	ne	ne	ne	ne
14	3	ne	ne	ne	ne
15	3	ne	ne	ne	ne
16	3	ne	ne	ne	ne
17	3	ne	ne	ne	ne
18	3	ne	ne	ne	ne
19	3	ne	ne	ne	ne
20	3	ne	ne	ne	ne
21	3	ne	ne	ne	ne
22	3	ne	ne	ne	ne
23	3	ne	ne	ne	ne
24	3	ne	ne	ne	ne
25	3	ne	ne	ne	ne
26	3	ne	ne	ne	ne
27	3	ne	ne	ne	ne
28	3	ne	ne	ne	ne
29	3	ne	ne	ne	ne
30	3	ne	ne	ne	ne
31	3	ne	ne	ne	ne
32	4	ne	ne	ne	ne
33	5	ne	ne	ne	ne
34	5	ne	išķirtis	ne	ne
35	5	ne	ne	ne	ne
36	5	ne	ne	ne	ne
37	5	ne	ne	ne	ne
38	5	ne	ne	ne	ne
39	5	ne	ne	ne	išķirtis
40	5	ne	ne	ne	ne
41	5	ne	ne	ne	ne
42	5	ne	ne	ne	ne
43	5	ne	ne	ne	ne
44	6	ne	ne	ne	ne
45	7	ne	ne	ne	ne
46	7	ne	ne	ne	ne
47	7	ne	ne	ne	ne
48	7	ne	ne	ne	ne
49	7	ne	ne	ne	ne
50	7	ne	ne	ne	ne
51	7	ne	ne	ne	ne
52	7	išķirtis	ne	ne	ne

53	7	ne	ne	ne	ne
54	7	ne	ne	ne	ne
55	7	ne	ne	ne	ne
56	7	ne	ne	ne	ne
57	7	ne	ne	ne	ne
58	7	ne	ne	ne	ne
59	7	ne	ne	ne	ne
60	7	ne	ne	ne	ne
61	7	ne	ne	ne	ne
62	7	ne	ne	ne	ne
63	7	ne	ne	ne	ne
64	7	ne	ne	ne	ne
65	7	ne	ne	ne	ne
66	7	ne	ne	ne	ne
67	7	ne	ne	ne	ne
68	7	ne	ne	ne	ne
69	7	ne	ne	ne	ne
70	7	ne	ne	ne	ne
71	7	ne	ne	ne	ne
72	7	ne	ne	ne	ne
73	7	ne	ne	ne	ne
74	8	ne	ne	ne	ne
75	9	išķirtis	ne	ne	ne
76	9	ne	ne	ne	ne
77	10	ne	ne	ne	ne
78	10	ne	ne	ne	ne
79	10	ne	ne	ne	ne
80	11	ne	ne	ne	ne
81	11	ne	ne	ne	ne
82	12	ne	išķirtis	išķirtis	išķirtis
83	12	ne	ne	ne	ne
84	12	ne	ne	ne	ne
85	12	ne	ne	ne	ne
86	12	ne	ne	ne	ne
87	12	ne	ne	ne	ne
88	12	išķirtis	išķirtis	išķirtis	išķirtis
89	12	ne	ne	ne	ne
90	12	ne	ne	ne	ne
91	12	ne	ne	ne	ne
92	12	ne	ne	ne	ne

93	12	ne	ne	ne	ne
94	12	ne	ne	ne	ne
95	12	ne	ne	ne	ne
96	12	ne	ne	ne	ne
97	12	ne	ne	ne	ne
98	12	ne	ne	ne	ne
99	12	ne	ne	ne	ne
100	12	ne	ne	ne	ne
101	12	ne	ne	ne	ne
102	12	ne	ne	ne	ne
103	12	ne	ne	ne	ne
104	12	ne	ne	ne	ne
105	12	ne	ne	ne	ne
106	12	ne	ne	ne	ne
107	12	ne	ne	ne	ne
108	12	ne	ne	ne	ne
109	12	ne	ne	ne	ne
110	12	ne	ne	ne	ne
111	12	ne	ne	ne	ne
112	12	ne	ne	ne	ne
113	12	ne	ne	ne	ne
114	12	ne	ne	ne	ne
115	12	ne	ne	ne	ne
116	12	ne	ne	ne	ne
117	12	ne	ne	ne	ne
118	12	ne	ne	ne	ne
119	12	ne	ne	ne	ne
120	12	ne	ne	ne	ne
121	12	ne	ne	ne	ne
122	12	ne	ne	ne	ne
123	13	ne	ne	ne	ne
124	13	ne	ne	ne	išķirtis
125	13	ne	ne	ne	ne
126	13	ne	ne	ne	ne
127	13	ne	ne	ne	ne
128	14	ne	ne	ne	ne
129	14	išķirtis	išķirtis	išķirtis	išķirtis
130	14	ne	ne	ne	ne
131	14	ne	ne	ne	ne
132	14	ne	ne	ne	ne

133	14	ne	ne	ne	ne
134	14	ne	ne	ne	ne
135	15	ne	ne	ne	ne
136	15	ne	ne	ne	ne
137	15	ne	ne	ne	ne
138	15	ne	ne	ne	ne
139	15	ne	ne	ne	ne
140	15	ne	ne	ne	ne
141	15	ne	ne	ne	ne
142	15	ne	ne	ne	ne
143	15	ne	ne	ne	ne
144	15	ne	ne	ne	išķirtis
145	15	ne	ne	ne	ne
146	15	ne	ne	ne	ne
147	15	ne	ne	ne	ne
148	15	išķirtis	išķirtis	išķirtis	ne
149	15	išķirtis	išķirtis	išķirtis	išķirtis
150	15	ne	ne	ne	ne
151	15	ne	išķirtis	išķirtis	išķirtis
152	15	ne	ne	ne	ne
153	15	ne	ne	ne	ne
154	15	ne	ne	ne	ne
155	15	ne	ne	ne	ne
156	15	ne	ne	ne	ne
157	15	ne	ne	ne	ne
158	15	ne	ne	ne	ne
159	15	ne	ne	ne	ne
160	16	išķirtis	ne	išķirtis	ne
161	16	išķirtis	ne	išķirtis	ne
162	17	ne	ne	ne	ne
163	17	ne	ne	ne	ne
164	17	ne	ne	ne	ne
165	17	ne	ne	ne	ne
166	17	išķirtis	išķirtis	išķirtis	išķirtis
167	17	ne	ne	ne	ne
168	17	ne	ne	ne	ne
169	17	ne	ne	ne	ne
170	17	išķirtis	išķirtis	išķirtis	išķirtis
171	17	ne	ne	ne	ne
172	17	ne	ne	ne	ne

173	17	ne	ne	ne	ne
174	17	ne	ne	ne	ne
175	17	ne	ne	ne	ne
176	17	ne	ne	ne	ne
177	17	išķirtis	išķirtis	išķirtis	išķirtis
178	17	ne	ne	ne	ne
179	17	ne	ne	ne	ne
180	17	ne	ne	ne	ne
181	17	ne	ne	ne	ne
182	17	ne	ne	ne	ne
183	17	ne	ne	ne	ne
184	17	ne	ne	ne	ne
185	17	ne	ne	ne	ne
186	17	ne	ne	ne	ne
187	17	ne	ne	ne	ne
188	17	ne	ne	ne	ne
189	17	ne	ne	ne	ne
190	17	ne	ne	ne	ne
191	17	ne	ne	išķirtis	išķirtis
192	17	ne	ne	ne	ne
193	17	ne	ne	ne	ne
194	17	ne	ne	ne	ne
195	17	ne	ne	ne	ne
196	17	ne	ne	ne	ne
197	17	ne	ne	ne	ne
198	17	ne	ne	ne	ne
199	17	ne	ne	ne	ne
200	17	ne	ne	ne	ne
201	17	ne	ne	ne	ne
202	18	ne	ne	ne	ne
203	18	ne	ne	ne	ne
204	18	ne	ne	ne	ne
205	18	ne	ne	ne	ne
206	18	ne	ne	ne	ne
207	18	ne	ne	ne	ne
208	18	ne	ne	ne	ne
209	19	išķirtis	išķirtis	išķirtis	išķirtis
210	20	ne	ne	ne	ne
211	20	ne	ne	ne	ne
212	20	išķirtis	išķirtis	išķirtis	išķirtis

213	20	ne	ne	ne	ne
214	20	ne	ne	ne	ne
215	20	ne	ne	ne	ne
216	20	išskirtis	išskirtis	išskirtis	išskirtis
217	20	ne	ne	ne	ne
218	20	ne	ne	ne	ne
219	20	išskirtis	išskirtis	išskirtis	išskirtis
220	20	išskirtis	išskirtis	išskirtis	išskirtis
221	20	išskirtis	išskirtis	išskirtis	išskirtis
222	20	išskirtis	išskirtis	išskirtis	išskirtis
223	20	išskirtis	išskirtis	išskirtis	išskirtis
224	20	ne	ne	ne	ne
225	20	ne	išskirtis	išskirtis	išskirtis
226	20	ne	ne	ne	ne
227	20	ne	ne	ne	ne
228	21	ne	ne	ne	ne
229	21	išskirtis	išskirtis	išskirtis	ne
230	21	ne	ne	ne	ne
231	21	ne	ne	ne	ne
232	22	ne	ne	ne	ne
233	22	ne	ne	ne	išskirtis
234	22	ne	ne	ne	išskirtis
235	22	ne	ne	ne	ne
236	22	ne	ne	ne	ne
237	22	ne	ne	ne	ne
238	23	išskirtis	išskirtis	išskirtis	išskirtis
239	24	išskirtis	išskirtis	išskirtis	išskirtis
240	24	ne	ne	ne	ne
241	24	išskirtis	išskirtis	išskirtis	išskirtis
242	24	ne	ne	ne	ne
243	24	išskirtis	išskirtis	išskirtis	išskirtis
244	24	ne	ne	ne	ne
245	24	išskirtis	išskirtis	išskirtis	išskirtis
246	24	išskirtis	išskirtis	išskirtis	išskirtis
247	24	ne	ne	ne	ne
248	24	ne	ne	ne	ne
249	24	ne	ne	ne	ne
250	24	ne	ne	ne	ne
251	24	ne	ne	ne	ne
252	2	(Išskirtis, tiriant didžiausio klasterio R1)			

293	2	(Išskirtis, tiriant didžiausio klasterio R1)
294	2	(Išskirtis, tiriant didžiausio klasterio R1)
295	2	(Išskirtis, tiriant didžiausio klasterio R1)
296	2	(Išskirtis, tiriant didžiausio klasterio R1)
297	2	(Išskirtis, tiriant didžiausio klasterio R1)
298	2	(Išskirtis, tiriant didžiausio klasterio R1)
299	2	(Išskirtis, tiriant didžiausio klasterio R1)
300	2	(Išskirtis, tiriant didžiausio klasterio R1)
301	2	(Išskirtis, tiriant didžiausio klasterio R1)
302	2	(Išskirtis, tiriant didžiausio klasterio R1)
303	2	(Išskirtis, tiriant didžiausio klasterio R1)
304	2	(Išskirtis, tiriant didžiausio klasterio R1)
305	2	(Išskirtis, tiriant didžiausio klasterio R1)
306	2	(Išskirtis, tiriant didžiausio klasterio R1)
307	2	(Išskirtis, tiriant didžiausio klasterio R1)
308	2	(Išskirtis, tiriant didžiausio klasterio R1)
309	2	(Išskirtis, tiriant didžiausio klasterio R1)
310	2	(Išskirtis, tiriant didžiausio klasterio R1)
311	2	(Išskirtis, tiriant didžiausio klasterio R1)
312	2	(Išskirtis, tiriant didžiausio klasterio R1)
313	2	(Išskirtis, tiriant didžiausio klasterio R1)
314	2	(Išskirtis, tiriant didžiausio klasterio R1)
315	2	(Išskirtis, tiriant didžiausio klasterio R1)
316	2	(Išskirtis, tiriant didžiausio klasterio R1)
317	2	(Išskirtis, tiriant didžiausio klasterio R1)
318	2	(Išskirtis, tiriant didžiausio klasterio R1)
319	2	(Išskirtis, tiriant didžiausio klasterio R1)
320	2	(Išskirtis, tiriant didžiausio klasterio R1)
321	2	(Išskirtis, tiriant didžiausio klasterio R1)
322	2	(Išskirtis, tiriant didžiausio klasterio R1)
323	2	(Išskirtis, tiriant didžiausio klasterio R1)
324	2	(Išskirtis, tiriant didžiausio klasterio R1)
325	2	(Išskirtis, tiriant didžiausio klasterio R1)
326	2	(Išskirtis, tiriant didžiausio klasterio R1)
327	2	(Išskirtis, tiriant didžiausio klasterio R1)
328	2	(Išskirtis, tiriant didžiausio klasterio R1)

6 priedas. Elektronikos prekių pardavimų veiklos netipinių įmonių sąrašas, jų klasteriai ir priskyrimas prie išskirčių pirminio išskirčių tyrimo metu. Paryškintos įmonės yra priskirtos prie rizikingų sukčiauti.

Įmonė	Klasteris	Išskirtis_QTR1	Išskirtis_QTR2	Išskirtis_QTR3	Išskirtis_QTR4
1	1	ne	ne	ne	ne
2	2	ne	ne	ne	ne
3	2	ne	ne	ne	ne
4	2	ne	ne	ne	ne
5	2	ne	išskirtis	išskirtis	ne
6	2	ne	ne	ne	ne
7	2	ne	ne	ne	ne
8	2	ne	ne	ne	ne
9	2	ne	ne	ne	ne
10	2	ne	ne	ne	ne
11	2	ne	ne	ne	ne
12	2	ne	ne	ne	ne
13	2	ne	ne	ne	ne
14	2	ne	ne	ne	ne
15	2	ne	ne	ne	ne
16	2	ne	ne	ne	ne
17	2	ne	ne	ne	ne
18	2	ne	ne	ne	ne
19	2	ne	ne	ne	ne
20	2	ne	ne	ne	ne
21	2	ne	ne	ne	ne
22	2	ne	ne	ne	ne
23	2	ne	ne	ne	ne
24	2	ne	ne	ne	ne
25	2	ne	ne	ne	ne
26	2	ne	ne	ne	ne
27	2	ne	ne	ne	ne
28	2	ne	ne	ne	ne
29	3	ne	išskirtis	ne	ne
30	3	ne	ne	ne	ne
31	3	ne	ne	ne	ne
32	3	ne	ne	ne	ne
33	4	ne	ne	išskirtis	ne
34	4	ne	ne	išskirtis	išskirtis
35	4	ne	ne	išskirtis	ne
36	4	ne	ne	išskirtis	ne

37	4	ne	ne	ne	ne
38	4	ne	ne	ne	ne
39	5	ne	ne	ne	ne
40	5	ne	ne	ne	ne
41	5	ne	ne	ne	ne
42	5	ne	ne	ne	ne
43	6	ne	išskirtis	išskirtis	išskirtis
44	6	ne	ne	ne	ne
45	6	ne	ne	ne	ne
46	6	ne	ne	ne	ne
47	6	išskirtis	išskirtis	išskirtis	išskirtis
48	6	ne	ne	ne	ne
49	7	(Išskirtis, tiriant didžiausio klasterio R1)			
50	7	(Išskirtis, tiriant didžiausio klasterio R1)			
51	7	(Išskirtis, tiriant didžiausio klasterio R1)			
52	7	(Išskirtis, tiriant didžiausio klasterio R1)			
53	7	(Išskirtis, tiriant didžiausio klasterio R1)			
54	7	(Išskirtis, tiriant didžiausio klasterio R1)			
55	7	(Išskirtis, tiriant didžiausio klasterio R1)			
56	7	(Išskirtis, tiriant didžiausio klasterio R1)			
57	7	(Išskirtis, tiriant didžiausio klasterio R1)			
58	7	(Išskirtis, tiriant didžiausio klasterio R1)			
59	7	(Išskirtis, tiriant didžiausio klasterio R1)			
60	7	(Išskirtis, tiriant didžiausio klasterio R1)			
61	7	(Išskirtis, tiriant didžiausio klasterio R1)			
62	7	(Išskirtis, tiriant didžiausio klasterio R1)			
63	7	(Išskirtis, tiriant didžiausio klasterio R1)			
64	7	(Išskirtis, tiriant didžiausio klasterio R1)			
65	7	(Išskirtis, tiriant didžiausio klasterio R1)			
66	7	(Išskirtis, tiriant didžiausio klasterio R1)			
67	7	(Išskirtis, tiriant didžiausio klasterio R1)			
68	7	(Išskirtis, tiriant didžiausio klasterio R1)			
69	7	(Išskirtis, tiriant didžiausio klasterio R1)			
70	7	(Išskirtis, tiriant didžiausio klasterio R1)			
71	7	(Išskirtis, tiriant didžiausio klasterio R1)			
72	7	(Išskirtis, tiriant didžiausio klasterio R1)			
73	7	(Išskirtis, tiriant didžiausio klasterio R1)			
74	7	(Išskirtis, tiriant didžiausio klasterio R1)			
75	7	(Išskirtis, tiriant didžiausio klasterio R1)			
76	7	(Išskirtis, tiriant didžiausio klasterio R1)			

77	7	(Išskirtis, tiriant didžiausio klasterio R1)			
78	7	(Išskirtis, tiriant didžiausio klasterio R1)			
79	8	ne	ne	ne	ne
80	8	išskirtis	ne	išskirtis	išskirtis
81	8	ne	ne	ne	ne
82	8	ne	ne	ne	ne
83	8	ne	ne	ne	ne
84	8	ne	ne	ne	ne
85	8	ne	ne	ne	ne
86	8	ne	ne	ne	ne
87	8	ne	ne	ne	ne
88	8	ne	ne	ne	ne
89	8	ne	ne	ne	ne
90	8	ne	ne	ne	ne
91	9	ne	ne	ne	ne
92	10	ne	ne	ne	išskirtis
93	10	išskirtis	ne	ne	ne
94	10	ne	ne	ne	ne
95	10	išskirtis	ne	ne	ne
96	10	ne	ne	ne	ne
97	10	ne	ne	ne	išskirtis
98	10	ne	ne	ne	išskirtis
99	10	ne	ne	ne	ne
100	10	ne	ne	ne	ne
101	10	ne	ne	ne	ne
102	11	išskirtis	išskirtis	išskirtis	išskirtis
103	11	išskirtis	išskirtis	išskirtis	išskirtis
104	11	išskirtis	išskirtis	išskirtis	išskirtis
105	11	išskirtis	išskirtis	išskirtis	išskirtis
106	11	ne	ne	ne	ne
107	12	išskirtis	ne	ne	ne
108	12	ne	ne	ne	ne
109	12	išskirtis	ne	ne	ne
110	12	ne	ne	ne	ne
111	12	ne	ne	ne	ne
112	12	išskirtis	ne	ne	ne
113	12	ne	ne	ne	ne
114	12	ne	ne	ne	ne
115	13	ne	išskirtis	ne	išskirtis
116	13	ne	ne	ne	ne

117	13	ne	išķirtis	išķirtis	išķirtis
118	13	ne	ne	ne	ne
119	13	išķirtis	išķirtis	ne	ne
120	13	ne	ne	ne	išķirtis
121	13	ne	ne	ne	ne
122	13	išķirtis	išķirtis	išķirtis	ne
123	13	išķirtis	išķirtis	išķirtis	išķirtis
124	13	ne	ne	ne	išķirtis
125	13	ne	ne	ne	ne
126	13	išķirtis	ne	ne	ne
127	13	ne	ne	ne	ne
128	14	išķirtis	išķirtis	išķirtis	ne
129	14	ne	ne	ne	ne
130	14	ne	ne	ne	ne
131	15	ne	ne	ne	ne
132	15	išķirtis	išķirtis	išķirtis	išķirtis
133	15	išķirtis	išķirtis	išķirtis	išķirtis
134	15	išķirtis	išķirtis	išķirtis	išķirtis
135	16	ne	ne	ne	ne
136	16	ne	ne	ne	ne
137	16	ne	ne	ne	ne
138	16	ne	ne	ne	ne
139	16	ne	ne	ne	ne
140	16	išķirtis	išķirtis	išķirtis	išķirtis
141	16	ne	ne	ne	ne
142	16	ne	ne	ne	ne
143	16	ne	ne	ne	ne
144	16	ne	ne	ne	išķirtis
145	16	ne	ne	ne	ne
146	16	ne	ne	ne	ne
147	16	ne	ne	ne	ne
148	16	ne	ne	ne	ne
149	16	ne	ne	ne	ne
150	16	ne	ne	ne	ne
151	16	ne	ne	ne	ne
152	16	ne	ne	ne	ne
153	16	ne	ne	ne	ne
154	16	ne	ne	ne	ne
155	16	ne	ne	ne	ne
156	16	ne	ne	ne	ne

157	16	ne	ne	ne	ne
158	16	ne	ne	ne	ne
159	16	ne	ne	ne	ne
160	16	ne	ne	ne	ne
161	16	ne	ne	ne	ne
162	17	išķirtis	išķirtis	išķirtis	išķirtis
163	17	ne	ne	ne	ne
164	18	ne	ne	ne	ne
165	18	ne	ne	ne	ne