KAUNO TECHNOLOGIJOS UNIVERSITETAS
MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS
MATEMATINIO MODELIAVIMO KATEDRA

Laura Viselgaitė

VILNIAUS MIESTO BUTŲ VIENO KVADRATINIO METRO KAINOS PROGNOZAVIMAS

Magistro darbas

Vadovas
doc. dr. A. Kabašinskas

KAUNAS, 2014

KAUNO TECHNOLOGIJOS UNIVERSITETAS

MATEMATIKOS IR GAMTOS MOKSLŲ FAKULTETAS

MATEMATINIO MODELIAVIMO KATEDRA

TVIRTINU
Katedros vedėjas
prof.dr. E. Valakevičius
2014 06 02

VILNIAUS MIESTO BUTŲ VIENO KVADRATINIO METRO KAINOS PROGNOZAVIMAS

Taikomosios matematikos magistro baigiamasis darbas

Vadovas
doc. dr. A. Kabašinskas
2014 06 01

Recenzentas
dr. J. Petrauskienė
2014 06 01

Atliko
FMMM 2 gr. stud.
Laura Viselgaitė
2014 05 30

KAUNAS, 2014

KVALIFIKACINĖ KOMISIJA

Pirmininkas:	Juozas Augutis, profesorius (VDU)
Sekretorius:	Eimutis Valakevičius, profesorius (KTU)
Nariai:	Arūnas Barauskas, dr., direktoriaus pavaduotojas (UAB „Danet Baltic“)
	Vytautas Janilionis, docentas (KTU)
	Zenonas Navickas, profesorius (KTU)
	Kristina Šutienė, docentė (KTU)
	Jonas Valantinas, profesorius (KTU)

Viselgaitė L., Vilniaus miesto butų vieno kvadratinio metro kainos prognozavimas: taikomosios matematikos magistro darbas / vadovas doc. dr. A. Kabašinskas; Matematinio modeliavimo katedra, Matematikos ir gamtos mokslų fakultetas, Kauno technologijos universitetas. – Kaunas, 2014. – 56 psl.

SANTRAUKA

Tyrimo tikslas – išsiaiškinti, kokios būsto charakteristikos turi didžiausią įtaką Vilniaus miesto butų vieno kvadratinio metro kainai bei sudaryti keletą ją prognozuojančių modelių.

Prognozių modeliams sudaryti buvo panaudoti 2008–2012 metais valstybės įmonės Registrų centro Nekilnojamojo turto sandorių duomenų bazėje užregistruoti Vilniaus miesto butų pirkimo-pardavimo sandoriai, kai pirkėjas yra Lietuvos arba užsienio valstybės fizinis asmuo. Vertinant butų vieno kvadratinio metro kainą Vilniaus miesto savivaldybė padalinta į keturias rajonų zonas: centro, prestižinius, miegamuosius ir kitus rajonus. Be šių kriterijų taip pat atsižvelgėme į buto vertę, objektų skaičių sutartyje, kambarių skaičių, sienų tipą, rūsio buvimą bei ar butas yra naujos statybos.

Vilniaus butų vieno kvadratinio metro kaina prognozuojama tiesinės, kvantlinės regresijos ir neuroninių tinklų modeliais. Kvantilinei regresijai pasirinkti 0,05; 0,25; 0,5; 0,75 ir 0,95 kvantiliai. Visi gauti modeliai tarpusavyje palyginti pagal jų prognozavimo tikslumą 2008–2012 metų duomenims ir 2013 metų duomenims, kurie nebuvo įtraukti į modelių sudarymą. Modelių palyginimui naudojome koreguotą determinacijos koeficientą ir paklaidas MAPE, MSE, MAE, MPE. Nustatyta, jog Vilniaus miesto butų vieno kvadratinio metro kainą tiksliausiai prognozuoja medianos kvantilinės ir tiesinės regresijos modeliai. Tačiau tiesinės regresijos modelis netenkina liekanų normalumo ir homoskedastiškumo prielaidų.

Prasminiai žodžiai: kvantilinė regresija, tiesinė regresija, neuroniniai tinklai, prognozė, kvantilis, butas.

Viselgaitė L., Forecast of One Square Meter Price of Apartments in Vilnius Municipality: Master‘s work in applied mathematics / supervisor assoc. prof. dr. A. Kabašinskas; Department of Mathematical Modelling, Faculty of Mathematics and Nature Sciences, Kaunas University of Technology. – Kaunas, 2014. – 56 p.

SUMMARY

The aim of the research is to find out which dwelling characteristics are significant to one square meter price of apartments in Vilnius city municipality and prepare a few different forecasting models.

For forecast models we have used 2008–2012 years Real Estate Transactions Database records of State Enterprise Centre of Registers of Vilnius city apartments' purchases transactions, where the buyer is Lithuanian or foreign natural person. During evaluation of one square meter price of apartments, Vilnius city municipality was divided in four zones: center, prestige, living and other districts. Additionally, the value of the apartment, number of objects in the contract, number of rooms, wall type, presence of a basement and whether the building is newly built are also taken into account as a criteria.

One square meter price of apartments in Vilnius city municipality is forecasted using linear, quantile regression and neural network models. Quantile regression was made for 0,05; 0,25; 0,5; 0,75 and 0,95 quantiles. All models were compared to each other according to their predictive accuracy for the year 2008–2012 data and year 2013 data, which were not included in the training set for the model. Comparison was made using adjusted determination coefficient and errors – MAPE, MSE, MAE, MPE. One square meter price of apartments in Vilnius city municipality is most accurately predicted by the median quantile and linear regression models. However it was found that the linear regression model does not meet the assumptions of residuals normality and homoscedasticity.

Keywords: quantile regression, linear regression, neural network, prediction, quantile, apartment.

TURINYS

Įvadas ... 9
1. Būsto rinkos analizė ... 11
1.1. Būsto rinka Lietuvoje ... 11
1.2. Būsto vertinimo metodai ... 13
1.3. Būsto prognozavimo analizė ... 17
1.4. Kvantilinės regresijos taikymas ... 18
1.5. Neuroninių tinklų taikymas ... 20
2. Tyrimo metodai ... 22
2.1. Pradinė duomenų struktūra ... 22
2.2. Daugialypė tiesinė regresija ... 24
2.2.1. Modelio tinkamumo įvertinimas ... 25
2.2.2. Išskirčių tyrimas ... 27
2.3. Kvantilinė regresija ... 28
2.3.1. Kvantilinės regresijos sudarymas ... 29
2.4. Dirbtiniai neuroniniai tinklai ... 29
2.5. Sudarytų modelių tikslumo įvertinimas ... 31
2.6. Vilniaus miesto butų duomenų analizės modelis ... 32
2.7. Programinė įranga ... 33
3. Tiriamoji dalis ... 34
3.1. Duomenų parengimas tyrimui ... 34
3.2. Daugialypės tiesinės regresijos modelis ... 35
3.3. Kvantilinės regresijos modeliai ... 38
3.3.1. 0,05 kvantilio regresijos modelis ... 38
3.3.2. 0,25 kvantilio regresijos modelis ... 39
3.3.3. Medianos kvantilinės regresijos modelis ... 40
3.3.4. 0,75 kvantilio regresijos modelis ... 42
3.3.5. 0,95 kvantilio regresijos modelis ... 43

3.3.6. Kvantilinės regresijos apibendrinimas 44
3.4. Neuroninių tinklų modelis 46
3.5. Metodų palyginimas 47
Diskusija 52
Išvados 53
Rekomendacijos 54
Literatūra 55
Priedai 57
Priedas A 57
Priedas B 62
Priedas C 63
Priedas D 65
Priedas E 67

LENTELIŲ SĄRAŠAS

2.1 lentelė Tyrime naudojamos charakteristikos ir jų paaiškinimai 22
3.1 lentelė Vilniaus miesto butams taikomi atrankos kriterijai 34
3.2 lentelė Vilniaus miesto zonų analizė 35
3.3 Lentelė Tiesinės regresijos parametrų įverčiai ir reikšmingumas 36
3.4 Lentelė 0,05 kvantilinės regresijos parametrų įverčiai ir reikšmingumas 38
3.5 Lentelė 0,25 kvantilinės regresijos parametrų įverčiai ir reikšmingumas 39
3.6 Lentelė 0,5 kvantilinės regresijos parametrų įverčiai ir reikšmingumas 40
3.7 Lentelė 0,75 kvantilinės regresijos parametrų įverčiai ir reikšmingumas 42
3.8 Lentelė 0,95 kvantilinės regresijos parametrų įverčiai ir reikšmingumas 43
3.9 lentelė Prognozuojamų bei tikrųjų kainų tarpusavio koreliacijų koeficientai 47
3.10 lentelė Prognozių tikslumas modelio sudarymui naudotiems 2008–2012 m. pirkimo-pardavimo sandoriams...48
3.11 lentelė Prognozių tikslumas 2013 m. pirkimo-pardavimo sandoriams............................48

PAVEIKSLŲ SĄRAŠAS

1.1 pav. Kriterijai, lemiantys būsto įsigyjimą 13
1.2 pav. Būsto kainų indeksai, palyginti su indekso baziniu laikotarpiu (2010 m.=100).........14
1.3 pav. Būsto pokyčiai, palyginti su indekso baziniu laikotarpiu (1998 m. IV ketv.=100) 15
1.4 pav. Realusis ir nominalusis butų kainų indeksas...16
1.5 pav. „Swedbank“ būsto įperkamumo indeksas 17
2.1 pav. Neuroninio tinklo struktūra 30
3.1 pav. Kitamųjų įtaka tiesinės regresijos modelyje 37
3.2 pav. Liekanų normalumo tyrimas 37
3.3 pav. 2008–2012 m. duomenų prognozavimas tiesinės regresijos modeliu.........................38
3.4 pav. 2008–2012 m. duomenų prognozavimas 0,05 kvantilinės regresijos modeliu 39
3.5 pav. 2008–2012 m. duomenų prognozavimas 0,25 kvantilinės regresijos modeliu 40
3.6 pav. 2008–2012 m. duomenų prognozavimas medianos kvantilinės regresijos modeliu...41
3.7 pav. 2008–2012 m. duomenų prognozavimas 0,75 kvantilinės regresijos modeliu 42
3.8 pav. 2008–2012 m. duomenų prognozavimas 0,95 kvantilinės regresijos modeliu 43
3.9 pav. Kintamųjų įverčių kitimas skirtinguose kvantiliuose...44
3.10 pav. Kintamųjų įverčių kitimas skirtinguose kvantiliuose (tęsinys)................................45
3.11 pav. Rūsio ir naujos statybos kintamųjų įverčių kitimas skirtinguose kvantiliuose 46
3.12 pav. 2008–2012 m. duomenų prognozavimas neuroniniais tinklais................................47
3.13 pav. 90 proc. 2013 metų pirkimo-pardavimo sandorių prognozavimas tiesinės ir kvantilinės regresijos lygtimis..50

ĮVADAS

Nuosavas būstas yra vienas svarbiausių pirkinių žmogaus gyvenime. Prieš pasiryžtant įsigyti būstą, kiekvienas susikuria svajonių namų viziją, kurią siekia perteikti realybėje. Tačiau ne visuomet tai padaryti yra įmanoma – daugelį vizijų apriboja finansinė padėtis. Būsto kainą dažniausiai lemia keletas pagrindinių charakteristikų – vietovė, statybos pabaigos metai, plotas ir t.t. Remiantis valstybės įmonės Registrų centro duomenimis, Lietuvoje daugiausiai gyvenamojo nekilnojamojo turto sandorių yra įvykdoma Vilniaus miesto savivaldybėje. Šiuo darbu sieksime išsiaiškini, kokios būsto charakteristikos turi įtakos Vilniaus miesto butų vieno kvadratinio metro kainai bei sudarysime keletą ją prognozuojančių modelių.

Prognozių modeliams sudaryti buvo panaudoti 2008–2012 metais valstybės įmonės Registrų centro Nekilnojamojo turto sandorių duomenų bazėje užregistruoti Vilniaus miesto butų pirkimo-pardavimo sandoriai, kai pirkėjas yra Lietuvos arba užsienio valstybės fizinis asmuo. Vertinant butų vieno kvadratinio metro kainą Vilniaus miesto savivaldybė padalinta į keturias rajonų zonas: centro, prestižinius, miegamuosius ir kitus rajonus. Be šių kriterijų taip pat atsižvelgėme į buto vertę, objektų skaičių sutartyje (pvz. tuo pačiu sandoriu įsigyjamą garažą ir pan.), kambarių skaičių, sienų tipą, rūsio buvimą bei ar butas yra naujos statybos.

Magistro baigiamasis darbas susideda iš trijų dalių. Analitinėje dalyje plačiau susipažįstama su Lietuvos būsto rinka ir joje vyraujančiomis tendencijomis, apžvelgiama keletas pagrindinių būsto rinkos analizių rengėjų bei jų taikomi kriterijai. Taip pat analizuojame įvairius su būsto rinka susijusius mokslinės literatūros straipsnius, tyrėjų taikomus kriterijus bei analizės metodus. Metodologinėje dalyje detaliai aprašome tiriamų duomenų struktūrą, vieno kvadratinio metro kainos prognozavimui taikomus metodus: tiesinę ir kvantilinę regresijas bei neuroninius tinklus, šių metodų sudarymo ir vertinimo kriterijus. Tiriamojoje dalyje apžvelgiame atlikto tyrimo eigą ir gautus rezultatus. Detaliai aprašome ir analizuojame gautą tiesinės regresijos modelį, sudarytas kvantilinės regresijos lygtis 0,05; 0,25; 0,5; 0,75 ir 0,95 kvantiliams bei neuroninius tinklus. Tuomet visus sudarytus modelius lyginame pagal jų prognozavimo tikslumą 2008–2012 metų duomenims ir 2013 metų duomenims, kurie nebuvo įtraukti į modelių sudarymą. Šiame darbe gauti rezultatai apžvelgiami išvadose, taip pat pateikiamos rekomendacijos tolesniam Vilniaus miesto butų vieno kvadratinio metro kainos prognozavimo tyrimui.

Dalyvavimas konferencijose:

- Tarptautinė konferencija „Taikomoji matematika“. Pranešimas „Vilniaus miesto butų vieno kvadratinio metro kainos prognozavimas“ (2014 m.).
- Lietuvos jaunųjų mokslininkų konferencija „Operacijų tyrimas ir taikymai“. Pranešimas „Kvantilinės regresijos taikymas KTU modulio Matematika 1 rezultatų prognozei“ (2013 m.).
- Tarptautinė konferencija „Taikomoji matematika“. Pranešimas „Modulio Matematika 1 semestro įvertinimo prognozavimas“ (2013 m.).

Straipsniai ir publikacijos:

- Straipsnis „Application of Quantile Regression for Prediction of KTU Module Mathematics 1 Results“ žurnale „Jaunųjų mokslininkų darbai“ (2014 m.).
- Publikacija „Vilniaus miesto butų vieno kvadratinio metro kainos prognozavimas“ tarptautinės konferencijos „Taikomoji matematika“ leidinyje (2014 m.).
- Publikacija „Modulio Matematika 1 semestro įvertinimo prognozavimas“ tarptautinės konferencijos „Taikomoji matematika“ leidinyje (2013 m.).

1. BŪSTO RINKOS ANALIZĖ

Nekilnojamojo turto rinkoje galime išskirti tris pagrindines sritis: gyvenamos ir negyvenamos paskirties patalpas bei žemės įsigijimus. Šiame darbe tiriama gyvenamosios paskirties patalpų (toliau būstų) rinka. Apibrėžkime pagrindines sąvokas:

- **Būstas (gyvenamosios patalpos)** – vienbutis gyvenamasis namas, jo dalis, butas ar kitos gyvenamosios patalpos arba jų dalys, tinkamos asmeniui arba šeimai gyventi.
- **Būsto kaina** – nekilnojamojo turto pirkimo-pardavimo sandorio metu perkamo ar parduodamo būsto kaina su pridėtinės vertės mokesčiu, bet be žemės ir kitų tuo pačiu sandoriu perkamų ar parduodamų nekilnojamojo turto objektų vertės.
- **Naujas būstas** – būstas, kuris iki sudarant pirkimo-pardavimo sandorį nebuvo naudotas (nuo kurio pastatymo praėjo ne daugiau kaip dveji metai).
- **Esamas būstas** – būstas, kuris iki sudarant pirkimo-pardavimo sandorį jau buvo naudotas (nuo kurio pastatymo praėjo daugiau kaip dveji metai) [18].

Būsto rinkos analizė yra skirta matuoti infliaciniams procesams šioje srityje, formuoti pinigų politikai, vertinti finansiniam stabilumui, skaičiuoti būsto kainų pokyčius, būsto įperkamumo rodiklius. Būsto kainų indeksai dažnai naudojami infliacijos prognozėms ir konvergencijos ataskaitoms rengti, tendencijų prognozėms sudaryti bei vietiniams ir tarptautiniams palyginimams atlikti. Įprasta, jog būsto kainų analizės yra skelbiamos kas ketvirtį, kadangi trijų mėnesių laikotarpyje yra surenkamas pakankamas duomenų kiekis bei galima įžvelgti tam tikras tendencijas. Duomenys, apie atliktus pirkimo-pardavimo sandorius, yra registruojami valstybės įmonės Registrų centro Nekilnojamojo turto sandorių duomenų bazėje. Verta pastebėti, jog VĮ Registrų centras registruoja tik tuos pirkimo-pardavimo sandorius, kurie jau yra patvirtinti notarine sutartimi. Todėl statomuose namuose įsigyti būstai yra užregistruojami tik pabaigus statybas. Tačiau ne vien tik VĮ Registrų centro duomenys yra naudojami būsto rinkos analizėje. Dažniausiai bankai ir nekilnojamojo turto agentūros turi sukaupę savo duomenų bazes, kurias gali papildyti VĮ Registrų centro duomenimis.

1.1. BŪSTO RINKA LIETUVOJE

Lietuvoje gyvenamojo būsto rinką galima suskirstyti pagal kelis kriterijus. Atsižvelgiant į teritoriją yra išskiriamos Vilniaus miesto, didžiųjų miestų (Kauno, Klaipėdos, Šiaulių, Panevėžio) bei kitos šalies teritorijos grupės (šie skirstymai gali keistis skirtinguose šaltiniuose). Į dvi atskiras grupes išskiriami gyvenamieji namai (kuriuose gyvena 1–2 šeimos) ir butai. Būsto amžius taip pat yra labai svarbus kriterijus būsto rinkoje. Dažniausiai būstas grupuojamas į naujų ir esamų būstų grupes [18]. Šios grupės yra sudarytos atsižvelgiant į vidutines vieno kvadratinio metro kainas, pirkimo-pardavimo sandorių skaičių (kiekvienai grupei reikia reprezentatyvaus sandorių skaičiaus) bei kainų kitimo tendencijas. Verta pastebėti, jog būstas gali būti perkamas nuolatiniam ar sezoniniam gyvenimui ir investiciniais tikslais.

Būsto rinkos analizes bei apskaičiuotus būsto kainų ar būsto įperkamumo indeksus periodiškai skelbia:

- Lietuvos statistikos departamentas;
- VĮ Registrų centras;
- Nekilnojamojo turto agentūra „Ober-Haus“;
- Lietuvos bankas;
- Komercinis bankas „Swedbank“.

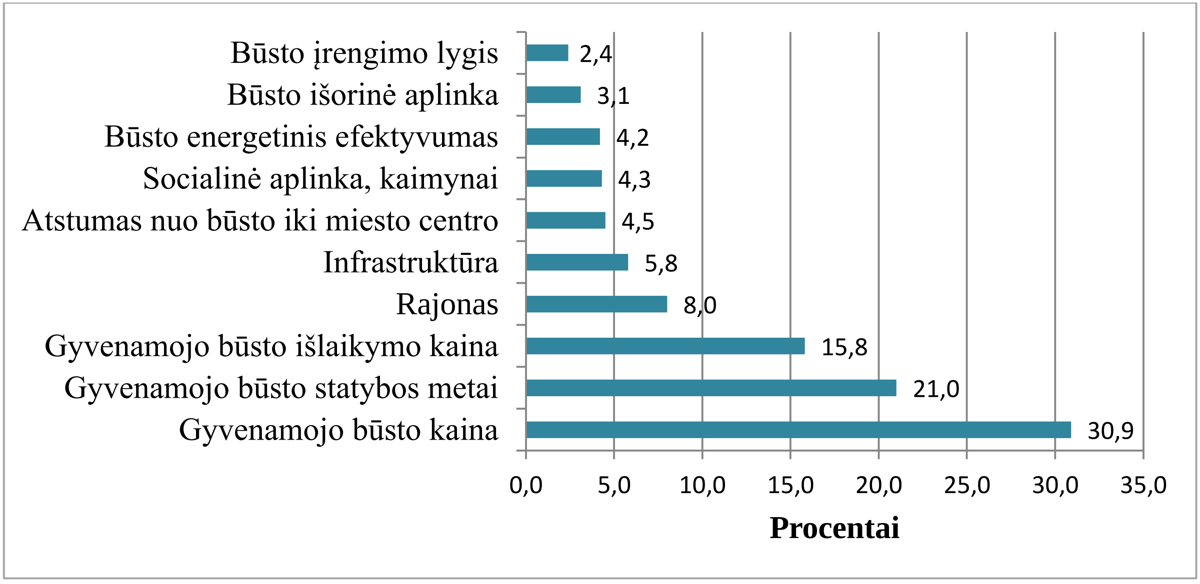
„Swedbank“ 2012 metais atliktoje Asmeninių finansų instituto būsto įsigijimo ir nuomos tendencijų analizėje teigiama, jog 85,3% lietuvių gyvena būste įsigytame už savo lėšas, 7,8% būste įsigytame su paskola ir tik 6,9% nuomojasi gyvenamąjį būstą. Tai atskleidžia Lietuvoje vyraujantį požiūrį, jog būstas turi būti nuosavas. Palyginus, Europos Sąjungos valstybėse būstą vidutiniškai nuomojasi 29,3% gyventojų, t.y. keturis kartus daugiau nei Lietuvoje. Taip pat, ES valstybėse daugiau būstų yra įsigyjama su paskola – 27,9%, o už nuosavas lėšas nuperkama tik 42,9% būstų [7]. Išskirtiniu galima pavadinti Vokietijos pavyzdį, kur 2013 metų duomenimis tik 43% gyventojų turėjo nuosavo nekilnojamojo turto, šis rodiklis Šveicarijoje yra dar mažesnis ir svyruoja apie 40% (2007 metų duomenimis – 38,4%).

Šioje analizėje taip pat pateikiama „Sprinter tyrimų“ užsakytos apklausos rezultatai: net 96% šalies gyventojų norėtų gyventi nuosavame būste, 2% sutiktų nuomotis ir 2% neturi nuomonės. Statistinis nuomininko portretas atitiktų šiuos kriterijus:

- 18–35 m. žmogus gyvenantis didmiestyje.
- Dar nesukūręs šeimos.
- Studentas arba įgijęs aukštąjį/aukštesnįjį išsilavinimą.
- Dažniausiai pajamos siekia iki 1000 Lt/mėn. [26].

Matome, jog būsto nuomą renkasi žmonės, kurie tik pradeda sėslų, atskirą nuo tėvų gyvenimą. Daugeliui nuoma yra laikinas gyvenamosios vietos sprendimas, kol nuspręs kur nori gyventi bei sukaups tam reikiamą dalį lėšų. Taigi būsto rinka yra aktuali didžiajai daliai Lietuvos gyventojų.

Apklausus 1006 respondentus, „Swedbank“ analizavo, kokie kriterijai yra svarbiausi, nusprendus įsigyti būstą:



1.1 pav. Kriterijai, lemiantys būsto įsigyjimą

Kaip matome 1.1 pav., svarbiausias kriterijus yra gyvenamojo būsto kaina, į kurią labiausiai atsižvelgtų 30,9% respondentų. Antroje vietoje yra būsto statybos metai su 21,0%, o trečioje – būsto išlaikymo kaina (15,8%). Respondentams mažiausiai rūpėjo būsto įrengimo lygis (2,4%) ir būsto išorinė aplinka (3,1%).

Remiantis „Sprinter tyrimų“ apklausa, kurioje Lietuvos gyventojai buvo apklausiami, kokiame būste jie gyvena, išsiaiškinta, jog:

- 54% žmonių gyvena daugiabučiuose, pastatytuose daugiau nei prieš 20 metų.
- 20% gyvena nuosavuose mūriniuose, blokiniuose, karkasiniuose ar monolitiniuose namuose.
- 4% įsikūrę nuosavuose mediniuose namuose.
- 12% apklaustųjų nuosavo būsto neturėjo ir tuo metu jį nuomojo.
- 10% gyventojų turėjo kitokio tipo būstą.

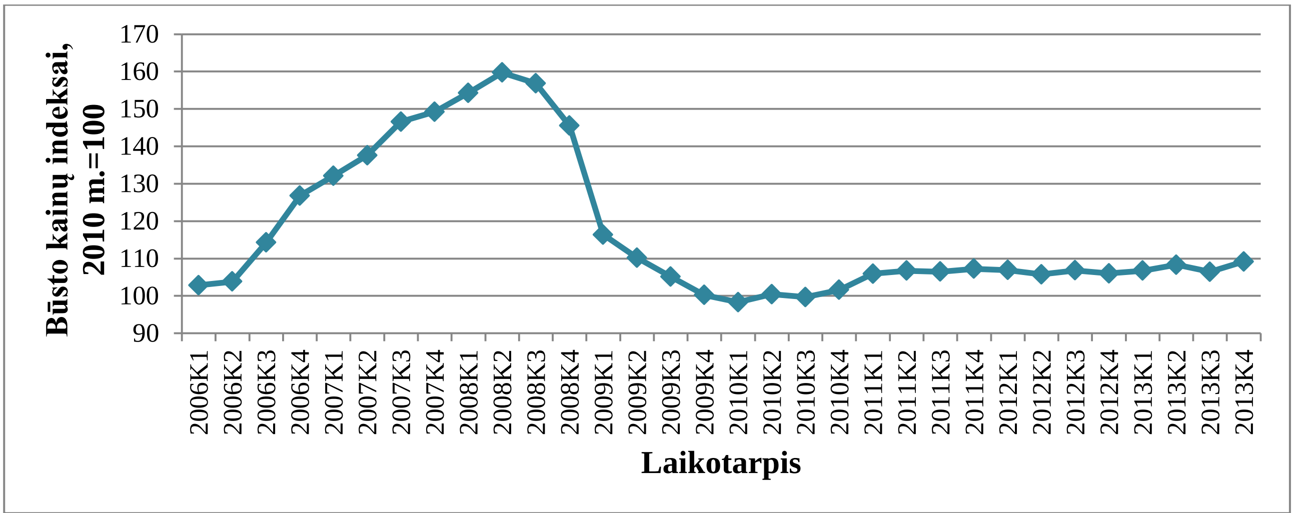
Taigi matome, jog didžioji gyventojų dalis gyvena esamos statybos butuose, kurių tam tikrai daliai yra reikalinga renovacija. Beveik ketvirtadalis (24 proc.) šalies gyventojų turi nuosavus namus. Pagal šį tyrimą 12 proc. gyventojų nuomojasi būstą, nors pagal „Swedbank“ statistiką, būstą nuomojasi 6,9 proc. gyventojų. Šie neatitikimai galėjo kilti dėl skirtingų duomenų šaltinių naudojimo. Likęs dešimtadalis gyventojų turi naujos statybos būstą arba yra įsikūrę „loftuose“ ar kitokiose gyvenamosiose patalpose.

1.2. BŪSTO VERTINIMO METODAI

Dažnam vartotojui perskaičius periodines būsto rinkos analizes iškyla klausimas, kuriuo šaltiniu pasikliauti. Taip įvyksta todėl, kad būsto rinkos analizei yra naudojamos skirtingos metodikos ir duomenų šaltiniai. Apžvelkime keletą pagrindinių būsto rinkos analizės skelbėjų bei jų taikomus metodus.

- **Lietuvos statistikos departamento Būsto kainų indeksas (BKI) ir pokyčiai**

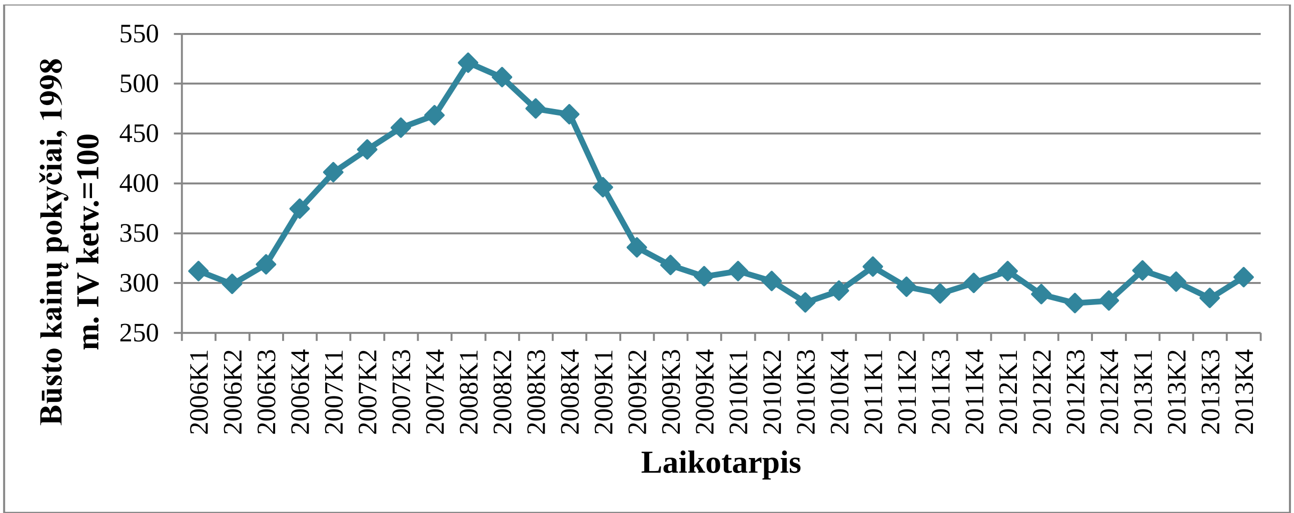
Lietuvos statistikos departamentas kiekvieną ketvirtį skelbia Būsto kainų indeksus (BKI) ir jų pokyčius. Šiems skaičiavimams atlikti yra naudojama VĮ Registrų centro Nekilnojamojo turto registro objektų sandorių duomenų bazė. BKI rodo namų ūkių šalies ekonominėje teritorijoje įsigyto būsto rinkos kainų pokyčius, neatsižvelgiant į tai, iš kokio savininko (fizinio ar juridinio asmens, valstybės, savivaldybės) ir kokiam tikslui (gyvenimui ar investicijai) tas būstas įsigyjamas. Į būsto kainą, sudarant BKI, įtraukiama žemės kaina. BKI neaprėpia kitų su būsto įsigijimu ir nuosavybe susijusių prekių ir paslaugų ir neatspindi jų kainų pokyčių [18]. Sudarant BKI yra atrenkami tie sandoriai, kurių pirkėjas yra fizinis asmuo (šalies arba užsienio šalies pilietis), būsto baigtumo lygis >50 proc., atliekama klaidingai įvestų sandorių paieška. Tuomet sandoriai yra suskirstomi į elementariąsias visumas (vienarūšių būstų grupes), kuriose suskaičiuojamas geometrinis vidurkis. Taikant Laspeireso formulę [18] elementarieji kainų indeksai yra agreguojami į aukštesnio lygmens kainų indeksus. Kiekvienai elementariajai visumai kasmet yra skaičiuojamas svoris. Lietuvos statistikos departamentas skelbia BKI ir BKI pokyčius (palyginti su ankstesniu ketvirčiu, ankstesnių metų atitinkamu ketvirčiu, keturi paskutiniai ketvirčiai palyginti su ankstesniais keturiais ketvirčiais) šalies, Vilniaus miesto savivaldybės ir šalies teritorijos be Vilniaus miesto savivaldybės lygmenimis. BKI yra klasifikuojamas į tris išlaidų kategorijas: būsto, naujo būsto ir esamo būsto pirkimus. Šiuo metu skaičiuojamo BKI bazinis laikotarpis yra 2010 m.=100. Žemiau pateikiame BKI kitimo grafiką [20]:



1.2 pav. Būsto kainų indeksai (BKI), palyginti su indekso baziniu laikotarpiu (2010 m.=100)

- **VĮ Registrų centro Būsto kainų pokyčiai**

VĮ Registrų centro būsto kainų pokyčių skaičiavimui yra naudojama jų turima sandorių duomenų bazė. Būsto pirkėjas ar pardavėjas gali būti fizinis ar juridinis asmuo. Įtraukiami sandoriai, kai būstas yra perkamas ne tik iškart, bet ir su lizingu ar išsimokėtinai. Be to, sandoryje turi būti perkamas tik būstas, kitu atveju jis yra neįtraukiamas į tyrimą, būsto baigtumas turi būti >50 proc., individualių gyvenamųjų namų plotas tarp 60–500 kv. metrų su žemės sklypu iki 2 ha, o butų plotas tarp 20–300 kv. metrų [29]. Pirkimo-pardavimo sandoriai analizuojami pagal vietovės požymį: didžiuosius miestus (Alytus, Kaunas, Klaipėda, Panevėžys, Šiauliai, Vilnius bei Palanga ir Neringa, kuriuose būstų kainos panašios į didžiųjų miestų kainas), rajonų centrus ir likusią teritoriją – kaimų, miestų ir miestelių. Gauti rezultatai skelbiami šalies, Vilniaus m. savivaldybės ir kitos šalies teritorijos be Vilniaus m. savivaldybės lygmenimis. Taip pat grupės yra sudaromos naujos ir ankstesnės statybos butams ir individualiems gyvenamiesiems namas. Butai ir namai pagal vietovę turi nustatytus įtraukiamų sandorių kainų intervalus. Naudojant šiuos intervalus atmetama iki 20 proc. sandorių, o likusiems taikomas svertinis vidurkis (sumarinė sandorių kaina padalinta iš sumarinio parduoto ploto). Būsto kainų pokyčiai skaičiuojami palyginti su ankstesniu ketvirčiu ir su 1998 m. IV ketvirčiu. Būsto kainų pokyčiai skelbiami kiekvieną ketvirtį su baziniu laikotarpiu 1998 m. IV ketv. =100. Pateikiame parduotų būstų kainų pokyčių grafiką:



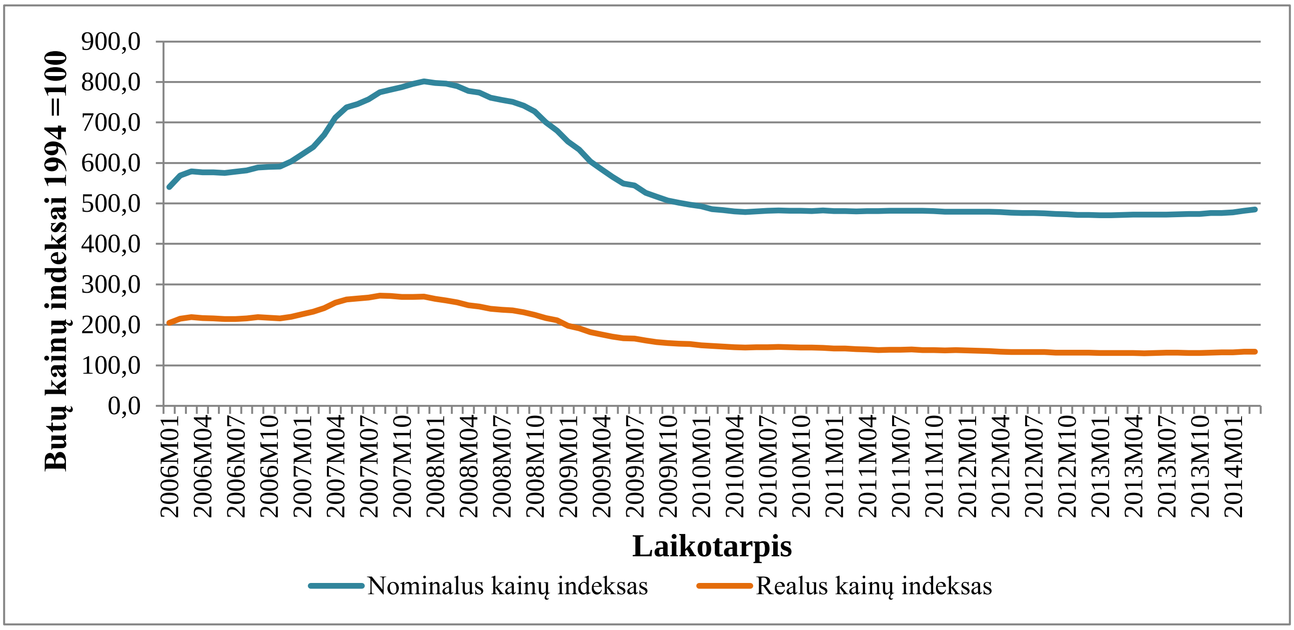
1.3 pav. Būsto pokyčiai, palyginti su indekso baziniu laikotarpiu (1998 m. IV ketv.=100)

- **„Ober-Haus“ gyvenamojo nekilnojamojo turto rinkos komentaras ir Lietuvos butų kainų indeksas**

Nekilnojamojo turto agentūra „Ober-Haus“ kiekvieną ketvirtį nuo 2008 m. IV ketvirčio skelbia „Gyvenamojo nekilnojamojo turto rinkos komentarą“. Šioje apžvalgoje yra analizuojamos didžiųjų Lietuvos miestų (Vilniaus, Kauno, Klaipėdos, Šiaulių ir Panevėžio) būsto rinkos tendencijos. Analizėje yra apžvelgiamos vidutinės būsto (butų ir namų) kainos ir jų pokyčiai naujos ir senos statybos būste. „Ober-Haus“ taip pat kiekvieną mėnesį skelbia ir Lietuvos butų kainų indeksą (OHBI), kuris parodo apibendrintą butų kainų pokytį penkiuose didžiausiuose Lietuvos miestuose. Šiam indeksui apskaičiuoti yra naudojami svoriai. OHBI bazinis laikotarpis yra 1994 m. sausio mėn.=100. OHBI indekso skaičiavimams naudojami įvairūs informacijos ir duomenų šaltiniai: „Ober-Haus“ sukaupta informacija ir duomenų bazės, valstybinių ir privačių institucijų duomenys bei kiti informacijos ir duomenų šaltiniai, kurie gali turėti įtakos galutiniams rezultatams. OHBI yra dviejų rūšių:

Nominalusis butų kainų indeksas parodo butų nominaliųjų kainų pokyčius laike, palyginus su baziniu indekso laikotarpiu (1994–01).

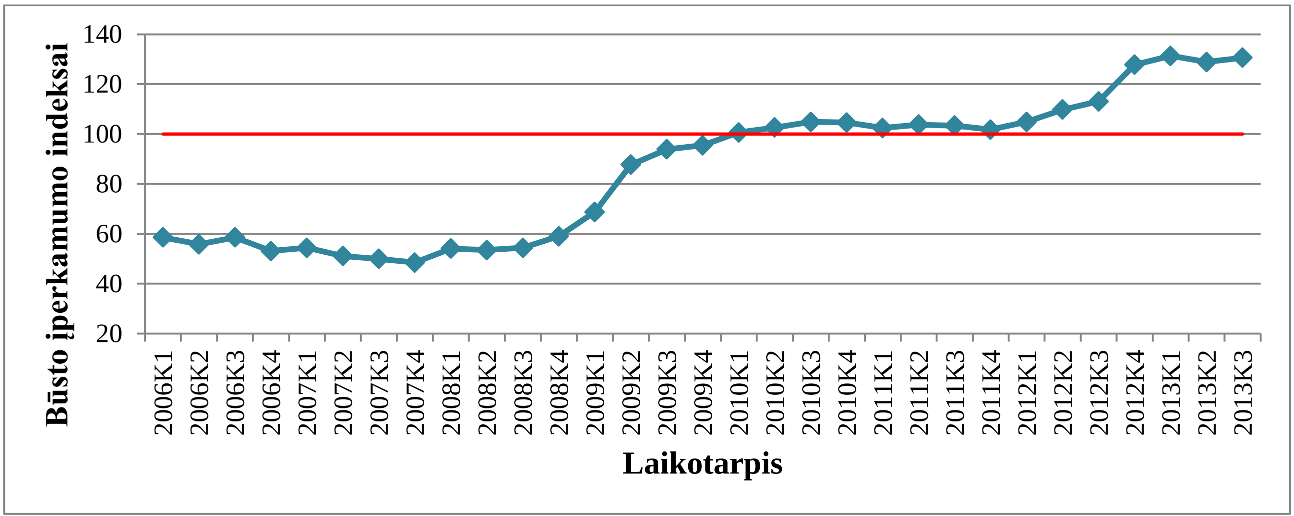
Realusis butų kainų indeksas parodo butų realiųjų kainų pokyčius laike, palyginus su baziniu indekso laikotarpiu (1994–01). Realusis kainų indeksas apskaičiuojamas eliminavus infliacijos įtaką nominaliosioms indekso reikšmėms. Realiojo butų kainų indekso apskaičiavimui naudojamas Lietuvos statistikos departamento sudaromas ir viešai skelbiamas suderintas vartotojų kainų indeksas (SVKI) [28].



1.4 pav. Realusis ir nominalusis butų kainų indeksas

- **„Swedbank“ būsto įperkamumo indeksas**

Būsto įperkamumo indeksas yra vienas iš finansinio stabilumo rodiklių. Jis įvertina palūkanas, butų kainas ir atlyginimus (didesnės palūkanos ir butų kainos mažina būsto įperkamumo indeksą; didesnis atlyginimas didina ir atvirkščiai) [27]. Plačiau paanalizuokime kiekvieną ketvirtį „Swedbank“ skelbiamą būsto įperkamumo indeksą. Šis indeksas skaičiuojamas visoms Baltijos šalims bei vertina namų ūkio, kurio atlyginimas lygus 1,5 vidutinio neto (atskaičius mokesčius) atlyginimo, galimybes įsigyti 55 kv. metrų butą Vilniaus mieste tuometinėmis vidutinėmis palūkanomis (apimančiomis ir kitus su būsto paskola susijusius mokesčius) pasiskolinant 85 proc. būsto vertės ir paskolą atiduodant per 30 metų [27]. Jeigu indeksas = 100 – atlyginimas pakankamas (įmoka už būsto paskolą sudaro 30 proc. namų ūkio pajamų); indeksas >100 – namų ūkio atlyginimai yra didesni nei pakankami (įmoka už būsto paskolą sudaro mažiau 30 proc. namų ūkio pajamų); jei indeksas <100 – namų ūkio atlyginimai mažesni nei pakankami (įmoka už būsto paskolą sudaro daugiau 30 proc. namų ūkio pajamų).



1.5 pav. „Swedbank“ būsto įperkamumo indeksas

1.3. BŪSTO PROGNOZAVIMO ANALIZĖ

Būsto analizėje dažniausiai yra naudojamos kelių rūšių charakteristikos: struktūrinės (charakteristikos apie būsto dydį, naudotas statybines medžiagas, aukštų skaičių ir t.t.), vietos (miestas, gyvenamasis rajonas, gatvė) ir kaimynystės (netoliese esančios parduotuvės, viešojo transporto stotelės, mokyklos ir darželiai). Galima teigti, jog trokštamos charakteristikos nulemia didesnę būsto kainą. Struktūrinės charakteristikos, tokios kaip aukštų skaičius ir kambarių skaičius, turi didžiausią įtaką būsto kainai [2]. Būsto duomenys yra gan heterogeniški, todėl analizei stengiamasi surinkti kuo daugiau duomenų. Esant galimybei patartina įtraukti duomenis apie rūsio ir garažo egzistavimą, įdiegtas šildymo, dujų bei vandens sistemas.

Galima išskirti keletą būsto pirkimo dilemos sprendimo etapų:

- Nustatomi ir išsirenkami alternatyvūs variantai.
- Analizuojamos būsto charakteristikos ir visos veiklos, kurias reiktų atlikti jį įsigijus.
- Vykdomas būsto vertės kitimo bei likvidumo įvertinimas.
- Priimamas sprendimas ir įsigyjamas išsirinktas būstas.

Šie etapai gali būti taikomi norint įsigyti būstą ar įvertinti planuojamas nekilnojamojo turto investicijas [11].

Kai kurie autoriai į būsto prognozavimo analizę siūlo įtraukti suderintą vartotojų kainų indeksą (SVKI), kuris rodo vidutinį kainų pokytį per tam tikrą laiką, taip pat bendruosius populiacijos rodiklius, nedarbo lygį, nuomojamo būsto rodiklius, mokesčių dydžius, jei jie skiriasi skirtinguose tiriamos vietovės regionuose (pavyzdžiui JAV valstijose) [21].

Prognozuojant būsto kainų tendencijas dažniausiai yra pasirenkami hedoninės regresijos modeliai. Hedoninės regresijos modeliuose nekilnojamojo turto kaina yra išreikšta kaip funkcija, į kurią įtraukiamos charakteristikos apie būsto kokybę (plotas, kambarių skaičius ir t.t.) ir kiti faktoriai (atstumas iki mokyklos ir t.t.). Sudarant šiuos modelius stengiamasi, kad kaina būtų prognozuojama kuo tiksliau bei, kad pavyktų užfiksuoti besikuriančius nekilnojamojo turto „burbulus“. J. M. Quigley atliktoje „Nekilnojamojo turto kainų ir ekonominių ciklų“ analizėje [21] buvo surinkti devynerių metų JAV didmiesčių duomenys ir bandoma prognozuoti didesnius kainų pokyčių ekstremumus. Tačiau naudojami regresijos modeliai dar negali tiksliai nustatyti nekilnojamo turto vertės šuolių. Geriausias minėtame J. M. Quigley tyrime naudotas regresijos modelis teisingai įvertino tik apie pusę ekstremumų.

Būsto kainos prognozavimu domisi ir jį tiria ne tik mokslininkai tačiau ir nekilnojamojo turto agentūros, kitos statistinius duomenis ruošiančios įstaigos. Nekilnojamojo turto brokeriams yra itin svarbu kuo tiksliau nustatyti būsto kainą, kadangi nuo jos priklauso jų gaunamas atlygis. Nekilnojamojo turto agentūros „Re/max“ brokeris R. Hand savo pateiktoje analizėje teigia, jog nekilnojamojo turto kainų įvertinimo prielaidos turi daug trūkumų, kadangi kiekvienas pirkėjas turi skirtingus poreikius [10]. Todėl jis rekomenduoja naudoti daugialypę regresinę analizę, įvardindamas ją kaip modernesnį kainos apskaičiavimo metodą. Pasak šio autoriaus, svarbiausi kriterijai yra perkamos žemės bei nekilnojamojo turto plotas.

Europos Sąjungos statistikos tarnyba (Eurostatas) taip pat rekomenduoja taikyti hedoninę regresiją, kaip skaičiavimų kokybės koregavimo procedūrą. Šis metodas yra siūlomas dviem būsto kainų tyrimams: būsto kainų indeksui (BKI) ir savininkų užimtų būstų kainų indeksui (SUBKI) [8]. Hedoninėje regresijoje būsto charakteristikos yra individualios ir susijusios su vietove. Naudojant regresijos metodus, sudaroma hedoninės regresijos funkcija, kurioje kiekvienai charakteristikai yra priskiriamas ribinis įvertis (įvertis įgyja tam tikras ribas dėl pačios charakteristikos apribojimų, pvz. kambarių skaičiaus ir t.t.). Remiantis apskaičiuota ribine būsto kaina, visos kainos gali būti koreguojamos siekiant pašalinti heterogeniškų būstų kainų įtaką. Atliekamas konkretus kainų tikslinimas priklauso nuo hedoninės funkcijos formos (tiesinė, logaritminė), indekso apskaičiavimo tipo (aritmetinis, geometrinis vidurkis) ir nuo pasirinktos kokybės koregavimo procedūros (kainų perskaičiavimo, kainų charakteristikos ar fiktyvaus laiko kintamojo įvedimo).

1.4. KVANTILINĖS REGRESIJOS TAIKYMAS

Kvantilinė regresinė analizė gali būti taikoma daugelyje sričių: prognozavime, vartojimo rinkos, ugdymo, rizikos vertinimo ir finansų tyrimuose. Pastaraisiais metais šis metodas yra vis plačiau taikomas ir dažniau analizuojamas mokslinėje literatūroje: įvairius praktinio pritaikymo straipsnius yra paskelbę R. Koenker, G. Basset, G. Li ir kiti mokslininkai. Šiame skyrelyje analizuosime keletą kvantilinės regresijos taikymo pavyzdžių.

Galima pastebėti, jog kvantilinės regresijos modeliai papildo daugialypės tiesinės regresijos modelius, kuriuose yra susiduriama su heteroskedastiškumu. Taipogi, analizuojant gautus rezultatus naudojant kvantilinės ir tiesinės regresijos modelius, galima sužinoti daugiau informacijos apie pačius duomenis. G. Li su kolegomis atliko pastarųjų dešimties metų Kinijos žemės ūkio produktų (kiaulienos, kiaušinių, vištų bei pieno) kainų tyrimą. Analizuojant surinktus duomenis tiesinės daugialypės ir nuo 0,05 iki 0,95 kvantilio kvantilinės regresijos metodais galima plačiau išanalizuoti kiekvieno lygties kintamojo įverčių kitimo tendencijas, prognozavimo tikslumą. Pagal gautus rezultatus visuose modeliuose pastebėta, jog 0,05 ir 0,95 kvantilinės regresijos įverčių koeficientai skiriasi gan stipriai: pasikeičia jų ženklas arba pastebimas žymus padidėjimas ar sumažėjimas. Šių dviejų kvantilinių lygčių pateikti rezultatai gali padėti prognozuoti kainų kitimo intervalus, įvertinti galimą paklausos kitimo riziką [17]. Š. Kovač ir T. Želinsky įmonių pelningumo tyrime taip pat pritaikė kvantilinės regresijos metodus. Pasirinkę dvylika kintamųjų, susijusių su įmonės finansine situacija (ilgalaikiai ir trumpalaikiai įsiskolinimai, turto grąža, įmonės įsikūrimo regionas ir kt.), autoriai tyrė įmonių pateiktus duomenis 2001, 2006 ir 2011 metais. Buvo sudarytos septynios kvantilinės regresijos lygtys: 0,05; 0,1; 0,25; 0,5; 0,75; 0,9 ir 0,95 kvantiliams. Sudarius modelius buvo tikrinimas kintamųjų reikšmingumas bei multikolinearumas, jų koeficientų įverčių kitimas. Trumpalaikio ir ilgalaikio įsiskolinimo kintamųjų įverčiai yra neigiami, t.y. jie turi neigiamą įtaką įmonės pelningumui. Tačiau jų įtaka yra didesnė esant mažiems kvantiliams. Kintamojo *regionas* įverčiai iki 0,5 kvantilio rodė, jog pelningesnės įmonės yra kituose šalies regionuose (išskyrus Bratislavos regioną), o didesniuose už 0,5 kvantilio modeliuose Bratislavos regionas buvo įvertintas kaip pelningesnis. Gauti rezultatai parodė, jog atsiradusiems pokyčiams taip pat įtakos turėjo ir duomenų surinkimo metai [16]. Kvantilinė regresija taip pat yra taikoma ir medicinos srityje. K. Kan ir W. Tsai savo tyrime analizavo Taivane atliktos apklausos rezultatus apie kūno masės indeksą (KMI) ir žinias apie nutukimo sukeliamas rizikas. Tyrimo metu nustatyta, jog vyrai susidomi nutikimo keliamomis rizikomis tik tuomet, kai jų svoris tampa ekstremaliai didelis. Tačiau moterų žinios, apie nutikimo sukeliamas rizikas, visiškai neturi įtakos KMI dydžiui. Taip pat, naudojant socialinio gyvenimo kriterijus yra daug sunkiau nustatyti moterų KMI nei vyrų [13]. D. F. Benoit ir D. van der Poel pirmieji panaudojo kvantilinės regresijos analizę tiriant vartotojų pasitenkinimo lygį. Kadangi tyrimo duomenys nebuvo pilnai simetriški, kvantilinė regresinė analizė pateikė daug tikslesnes prognozes negu tiesinė regresinė analizė. Autoriai pastebi, kad kvantilinę regresinę analizę yra ypač patogu naudoti, jeigu norima detaliau paanalizuoti tam tikrą vartotojų segmentą [4].

Kvantilinės regresinės analizės aktualumą bei svarbumą parodo nuolat besiplečiančios jos taikymo sritys bei sparčiai didėjantis mokslinių darbų skaičius. Deja, mokslinių tyrimų su nekilnojamojo turto duomenimis surasti nepavyko.

1.5. NEURONINIŲ TINKLŲ TAIKYMAS

Neuroniniai tinklai yra plačiai taikomi įvairiose srityse. Galime išskirti pagrindines sritis, kurių prognozavimui yra naudojamas šis metodas: pirkimai-pardavimai, gamybos procesų analizė, pirkėjų poreikių analizė, rizikos valdymas, tikslinis marketingas. Taip pat, neuroniniai tinklai yra pritaikomi tam tikroms specifinėms sritims: veidų, rašysenos, 3D dimensijos objektų atpažinime, tekstūros analizėje, keletą prasmių turinčių kinų kalbos žodžių interpretacijose, hepatito diagnostikoje ir t.t.

Tikimasi, jog neuroniniai tinklai keleto metų bėgyje bus plačiai taikomi biomedicinoje. Neuroniniai tinklai yra naudojami ligų atpažinime iš žmogaus kūno požymių ir įvairių testų (kardiogramų, ultragarsų ir t.t.). Įvairių testų rezultatams neuroninius tinklus yra nesudėtinga pritaikyti, kadangi nereikia kurti naujų algoritmų ligos atpažinimui. Tačiau šiuo atveju reikia labai gerai atrinkti pateikiamus stebėjimų duomenis. Tinkamai pritaikius neuroninius tinklus įvairių ligų atpažinime, jos turėtų būti aptinkamos net pačiose ankstyviausiose stadijose. Pavyzdžiui, širdies stimuliatoriaus atveju planuojama neuroninius tinklus taikyti kiekvienam pacientui atskirai, iš dažnai surinktų jų pačių stebėjimų. Tokiu būdu stimuliatoriai sugebėtų patys prisitaikyti prie individualių žmogaus savybių ir tam nereikėtų gydytojų pagalbos [25].

Neuroninių tinklų modeliai dažnai yra lyginami su kitais prognozavimo metodais (laiko eilučių, regresinės analizės ir t.t.). Trumpai apžvelkime atliktus tyrimus ir jų rezultatus, kuomet neuroninių tinklų prognozavimo rezultatai buvo lyginami su daugialypės regresinės analizės pateikiamais rezultatais. A. Comrie atliko tyrimą apie ozono sluoksnio storį aštuoniuose JAV didmiesčiuose. Šio tyrimo metu nustatyta, jog neuroninių tinklų pateikiami rezultatai yra tik šiek tiek geresni negu regresinės analizės. Kiek didesnis skirtumas atsiranda tuomet, kai prisideda pavėluoti duomenys. Tuomet regresinės analizės prognozių pokyčiai yra didesni negu neuroninių tinklų [5]. S. Badran ir O. Abouellata analizavo elektros apkrovos duomenis ir sudarinėjo prognozes keletui metų į priekį. Tyrime buvo naudoti neuroniniai tinklai valandinėms prognozėms ir tiesinės regresijos modeliai, kurie prognozavo paklausos rezultatus skirtingais laiko intervalais (valandomis, dienomis ir t.t.). Tyrimas buvo atliekamas šiais etapais: pirmiausiai apmokomi neuroniniai tinklai ir gaunamos jų prognozės, tuomet sudaromi keli skirtingi regresijos modeliai ir galiausiai gautos neuroninių tinklų prognozės yra naudojamos regresijos lygtyse. Taip gaunama prognozuojamos vidutinė elektros apkrovos prognozė [3]. R. G. Ahangar su kolegomis tyrė 100 Teherano akcijų biržos įmonių. Tyrimo metu sudarytos prognozės tiesinės regresijos ir neuroninių tinklų metodais. Gauti rezultatai parodė, jog koreliacija tarp tikrosios ir prognozuojamos akcijų kainos tiesinės regresijos atveju yra dvigubai mažesnė nei neuroninių tinklų atveju. Taigi šie autoriai taip pat rekomendavo naudoti neuroninius tinklus, tačiau pabrėžė, jog norint gauti optimalius rezultatus turi būti kruopščiai atrinkta tiriama imtis [1].

Neuroninių tinklų savybė mokytis iš pateiktų pavyzdžių suteikia jiems lankstumo ir galios, kadangi tokiu būdu juos galima labai plačiai pritaikyti. Taip pat, siekiant atlikti konkrečią užduotį su neuroniniais tinklais, nereikia tam sukurti algoritmo. O greitas neuroninių tinklų apskaičiavimas ir reagavimas, suteikia galimybę juos greitai ir efektyviai pritaikyti realaus laiko sistemoms.

2. TYRIMO METODAI

Šiame skyrelyje apžvelgime pagrindines Vilniaus miesto butų vieno kvadratinio metro kainos prognozavimui naudojamas charakteristikas bei aptarsime naudojamus metodus. Prognozavimui naudosime tris skirtingus metodus: tiesinę regresiją, kvantilinę regresiją ir neuroninius tinklus. Po to, visi gauti rezultatai palyginami tarpusavyje, įvertinamas jų tikslumas bei taikymo praktiškumas.

2.1. PRADINĖ DUOMENŲ STRUKTŪRA

VĮ Registrų centras Nekilnojamojo turto sandorių duomenų bazėje (toliau Registrų centro duomenų bazė) pateikia apie 100 charakteristikų apie kiekvieną įvykusį sandorį. Šios charakteristikos apima duomenis apie pirkėją ir pardavėją (fiziniai, juridiniai Lietuvos Respublikos arba užsienio asmenys, savivaldybės ir t.t.), sandorio užregistravimo datą (metai, mėnuo) ir su įsigyjamu būstu susijusius duomenis. Visų pirma, nurodoma informacija apie vietą (savivaldybė ar rajonas, miestas, gyvenvietė, gyvenamoji zona, gatvė ir t.t.), nekilnojamąjį turtą (statybos ir rekonstrukcijos metai, aukštų, butų skaičius, sienų tipas ir t.t.) ir galiausiai paties būsto struktūrines charakteristikas (plotas, kambarių skaičius, aukštas, tūris, baigtumo lygis ir t.t.).

Šiame darbe prognozuosime vieno kvadratinio metro kainą (įskaitant patį nekilnojamąjį turtą bei žemę, jei ji buvo kartu perkama). Žemiau pateikiame 2.1 lentelę, kurioje aprašytos tyrime naudojamos charakteristikos.

2.1 lentelė

Tyrime naudojamos charakteristikos ir jų paaiškinimai

Žymėjimas		Charakteristika	Aprašymas
Y	Kaina	Vieno kvadratinio metro kaina, tolydusis kintamasis matuojamas Lt/m^2.	Logaritmuota vieno kvadratinio metro kaina, kuri yra apskaičiuota iš kintamojo „Paskaičiuota suma su žeme“.
–	Paskaičiuota suma su žeme	Paskaičiuota suma su žeme, tolydusis kintamasis matuojamas litais (Lt).	Nurodo viso būsto kainą, įskaitant ir žemės kainą.
X_1	Rinkos vertė	Rinkos vertė, tolydusis kintamasis matuojamas litais (Lt).	Logaritmuota vidutinė būsto rinkos vertė, įrašyta Registrų centro duomenų bazėje.
X_2	Centras	Centras, diskretusis pseudo kintamasis įgyjantis 0 arba 1	Pseudokintamasis, įgyjantis reikšmę 1, jei butas perkamas centre, senamiestyje, Užupyje. Kitu

		reikšmes.	atveju įgyjama 0 reikšmė.
X_3	Prestižiniai rajonai	Prestižiniai rajonai, diskretusis pseudo kintamasis įgyjantis 0 arba 1 reikšmes.	Pseudokintamasis, įgyjantis reikšmę 1, jei butas perkamas Antakalnyje, Turniškėse, Valakupiuose, Žvėryne, dalyje Naujamiesčio. Kitu atveju įgyjama 0 reikšmė.
X_4	Miegamieji rajonai	Miegamieji rajonai, diskretusis pseudo kintamasis įgyjantis 0 arba 1 reikšmes.	Pseudokintamasis, įgyjantis reikšmę 1, jei butas perkamas Baltupiuose, Bukčiuose, Dvarčionyse, Fabijoniškėse, Jeruzalėje, Justiniškėse, Karoliniškėse, Lazdynuose, Lazdynėliuose, Naujininkuose, Pašilaičiuose, Pilaitėje, Santariškėse, Saulėtekyje, Stotyje, Šeškinėje, Šnipiškėse, Vilkpėdėje, Viršuliškėse, Žirmūnuose, dalyje Naujamiesčio. Kitu atveju įgyjama 0 reikšmė.
X_5	Kiti rajonai	Kiti Vilniaus m. rajonai, diskretusis pseudo kintamasis įgyjantis 0 arba 1 reikšmes.	Pseudokintamasis, įgyjantis reikšmę 1, jei butas perkamas Balsiuose, Grigiškėse, Kirtimuose, Naujaneriuose, Naujojoje Vilnioje, Ožkiniuose, Paneriuose, Rokantiškėse, Rudaminoje, Salininkuose, Užusienyje. Kitu atveju įgyjama 0 reikšmė.
X_6	Objektų skaičius	Objektų skaičius sutartyje, diskretusis kiekybinis kintamasis įgyjantis reikšmes intervale [1;3].	Sutartyje nurodytas objektų skaičius. Būstas gali būti perkamas su garažu, rūsiu ir t.t. Tačiau į „Paskaičiuota suma su žeme“ papildomų objektų kainą nėra įtraukta.
X_7	Kambarių skaičius	Kambarių skaičius, diskretusis kiekybinis kintamasis įgyjantis reikšmes intervale [1;8].	Būste esantis kambarių skaičius.
X_8	Sienos	Trečiasis sienų tipas, diskretusis pseudo kintamasis įgyjantis 0 arba 1 reikšmes.	Pseudokintamasis, įgyjantis reikšmę 1, jei sienos yra iš plytų arba blokelių. Kitu atveju įgyjama 0 reikšmė.
X_9	Rūsys	Rūsys, diskretusis pseudo kintamasis įgyjantis 0 arba 1	Pseudokintamasis, įgyjantis reikšmę 1, jei buvo įsigyjamas rūsys. Kitu atveju įgyjama 0

		reikšmes.	reikšmė.
X_{10}	Nauja statyba	Naujos statybos būstas, diskretusis pseudo kintamasis įgyjantis 0 arba 1 reikšmes.	Pseudokintamasis, įgyjantis reikšmę 1, jei įsigyjamas naujos statybos butas. Kitu atveju įgyjama 0 reikšmė. Naujos statybos būstas yra toks, nuo kurio pastatymo praėjo ne daugiau kaip dveji metai.

Prognozių modeliams sudaryti buvo panaudoti 2008–2012 metais užregistruoti Vilniaus miesto butų pirkimo-pardavimo sandoriai (toliau sandoriai), kai pirkėjas yra Lietuvos arba užsienio valstybės fizinis asmuo. Pradinis turimų duomenų skaičius – 26 403 sandoriai. Prognozių modelių tinkamumo įvertinimui buvo naudojami 2013 metais užregistruoti 7 724 Vilniaus m. butų sandoriai.

2.2. DAUGIALYPĖ TIESINĖ REGRESIJA

Tiesinė daugialypė regresija nustato, kokia yra statistinė priklausomybė tarp nepriklausomų kintamųjų $X_1, X_2 \dots, X_k$ ir priklausomo kintamojo Y. Vienas iš jos tikslų – priklausomo kintamojo reikšmių prognozavimas. Tiesinės regresijos modelis užrašomas:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.1)$$

čia Y – priklausomas kintamasis, $X_1, X_2 \dots, X_k$ – nepriklausomi kintamieji, ε – atsitiktinė paklaida, $\beta_0, \dots, \beta_k$ – modelio koeficientai. Kadangi duomenų tiesinimo metu priklausomas kintamasis *Persk_vnt_vertė_log* ir nepriklausomas kintamasis *Rinkos_vertė* buvo logaritmuoti, šiuo atveju daugialypės regresijos modelis išreiškiamas:

$$\ln(Y) = \beta_0 + \beta_1 \ln(X_1) + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.2)$$

Sudarant daugialypės regresijos modelį, pirmiausia yra tikrinama hipotezė apie regresijos modelio koeficientus:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$
$$H_1: Bent\ vienas\ \beta_i\ nelygus\ nuliui \quad (2.3)$$

Kriterijaus statistika:

$$F = \frac{\overline{SS_A}}{\overline{SS_e}} \quad (2.4)$$

Kur: $\overline{SS_A}$ ir $\overline{SS_e}$ – nuokrypių kvadratų vidurkiai, SS_A – nuokrypių kvadratų suma, apibūdinanti faktoriaus A poveikį stebimo atsitiktinio dydžio Y vidurkiui, SS_e – nuokrypių kvadratų suma, apibūdinanti atsitiktinių klaidų faktoriaus E poveikį stebimo a. d. Y vidurkiui, kurį modelyje nusako atsitiktinis dydis e_{ij}, I – imčių skaičius, kurios kiekvienos dydis yra n_i, $i = 1, \dots, I$.

$$SS_A = \sum_{i=1}^{I} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 n_i \quad (2.5)$$

$$SS_e = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2 \quad (2.6)$$

Jeigu hipotezė H_0 teisinga, tai $F \sim F(\nu_1, \nu_2)$, čia $\nu_1 = I - 1$ ir $\nu_2 = n - I$. Tuomet regresijos modelis su duomenimis yra nesuderintas (visi modelio koeficientai β_k yra lygūs nuliui). Atmetus H_0 hipotezę yra sudaromas daugialypės regresijos modelis. Sudarytas modelis yra įvertinamas remiantis keletu kriterijų.

2.2.1. MODELIO TINKAMUMO ĮVERTINIMAS

Sudarytą daugialypės tiesinės regresijos modelį reikia įvertinti remiantis keliais pagrindiniais kriterijais ir prielaidomis, kuriuos apžvelgsime šiame skyrelyje. Determinacijos koeficientas nurodo kurią atsitiktinio dydžio Y sklaidos dalį galima paaiškinti Y daugialype regresija kintamųjų $X_1, X_2 \dots, X_k$ atžvilgiu. Koreguotas determinacijos koeficientas yra naudojamas tuomet, kai lygtyje yra daug regresorių. Modelis yra atmetamas dėl per mažo tikslumo tuomet, kai $R^2 < 0{,}2$. Determinacijos koeficientai apskaičiuojami:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{2.7}$$

$$R_{adj}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} \tag{2.8}$$

Kur $\hat{Y}_i$ – pagal regresijos lygtį apskaičiuota reikšmė, $\bar{Y}$ – vidurkis, Y_i – stebėta reikšmė, n – stebėjimų skaičius, p – nepriklausomų kintamųjų skaičius lygtyje.

Taip pat sudarytas modelis turi atitikti pagrindines daugialypės regresijos prielaidas. Pirmiausiai yra tikrinamas kiekvieno parametro reikšmingumas. Atliekami *T (Stjudento)* testai atskiriems regresoriams. Kintamasis yra statistiškai reikšmingas, jeigu $p < \alpha$ (šiame darbe $\alpha = 0{,}05$). Jeigu $p \geq \alpha$, kintamasis yra statistiškai nereikšmingas ir modelyje paliekamas tik ypatingais atvejais. Daugialypės regresijos modelis yra teisingai sudarytas tuomet, kai visi nepriklausomi kintamieji stipriai koreliuoja su priklausomu kintamuoju, tačiau nekoreliuoja tarpusavyje. Dispersijos mažėjimo daugiklis (*VIF*) parodo, ar regresoriai tarpusavyje stipriai koreliuoja: ar jie yra multikolinearūs. Multikolinearumas yra tada, kai $VIF > 4$.

$$VIF = \frac{1}{1 - R_j^2} \tag{2.9}$$

Norint nustatyti, ar skirtingų stebėjimų liekamosios paklaidos koreliuoja (ar yra susijusios tarpusavyje), taikomas Durbino-Vatsono testas. Jeigu sudarytame modelyje yra autokoreliacija, siūloma keisti pasirinktus modelio kintamuosius ar matematinę išraišką ir patikrinti duomenų inertiškumą. Durbino-Vatsono testas apskaičiuojamas pagal šią formulę:

$$DW = d = \frac{\sum_{i=2}^{n}(e_t - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \tag{2.10}$$

Kur e_i – paklaida, e_{i-1}– vėluojanti paklaida. Jeigu $d \in [1{,}5; 2{,}0]$, tuomet teigiama, jog autokoreliacijos nėra.

Modelio atsitiktinės paklaidos ε_i $(i = 1 \dots n)$ turi tenkinti tam tikrus reikalavimus, kurie sudaro tiesinės regresinės analizės prielaidas:

- atsitiktinės paklaidos ε_i yra normaliai pasiskirstę atsitiktiniai dydžiai;
- visų ε_i vidurkiai lygus nuliui $E\varepsilon_i = 0$;
- visų ε_i dispersijos turi būti pastovios (homoskedastiškumo prielaida) $D(\varepsilon_i) = \sigma^2$;
- visi ε_i turi būti nepriklausomi;
- duomenyse turi būti pašalintos išskirtys;
- nėra stiprios koreliacijos tarp nepriklausomų kintamųjų (multikolinearumo prielaida).

Sudarius daugialypės regresinės analizės modelį yra tikrinamas ir liekamųjų paklaidų normalumas. Dažniausiai yra atliekamas *Šapiro–Vilko* ir *Kolmogorovo–Smirnovo* testai. *Šapiro–Vilko* statistika apskaičiuojama:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2.11}$$

Kur $\bar{x}$ – imties vidurkis, $x_{(i)}$ – i–tosios eilės statistika, a_i – konstanta,

Kolmogorovo–Smirnovo statistika apskaičiuojama iš empirinės pasiskirstymo funkcijos $F_n(x)$:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \le x} \tag{2.12}$$

Kur $I_{X_i \le x}$ – rodiklinė funkcija, lygi 1, jei $I_{X_i \le x}$, ir 0, kitu atveju. $F(x)$ – kumuliacinė pasiskirstymo funkcija.

$$D_n = \sup_x |F_n(x) - F(x)| \tag{2.13}$$

Jei šiuose testuose p reikšmė $p \ge 0{,}05$ tuomet H_0 hipotezė yra tenkinama ir liekamosios paklaidos yra pasiskirsčiusios pagal normalųjį skirstinį. Taip pat, liekamųjų paklaidų normalumą galima įvertinti ir grafiniu būdu.

Vienas iš pagrindinių regresiniam modeliui keliamų reikalavimų yra dispersijos pastovumas – homoskedastiškumo prielaida. Ši prielaida gali būti tiriama ir grafiškai, ir pagal *Vaito* bei *Breušo–Pagano* testus. Šiems testams apskaičiuoti yra naudojamas Lagranžo daugiklis:

$$LM = \left(\frac{\partial l}{\partial \theta}\right)' \left(-E\left[\frac{\partial^2 l}{\partial \theta \partial \theta}\right]\right)^{-1} \left(\frac{\partial l}{\partial \theta}\right) \tag{2.14}$$

Kur $l(x, \theta)$ – tikslo funkcija. Šis testas yra analogiškas trijų etapų procedūrai:

- Naudojant mažiausių kvadratų metodą (MKM) sudaromas regresijos modelis ir apskaičiuojamos liekanos.
- Skaičiuojama pagalbinė regresija: $e_i^2 = \gamma_1 + \gamma_2 z_{2i} + \dots + \gamma_p z_{pi} + \eta_i$.
- Testo statistikos rezultatas apskaičiuojamas pagal pagalbinės regresijos lygties determinacijos koeficientą, apskaičiuotą antrame žingsnyje, ir imties didumą n:

$$LM = nR^2 \tag{2.15}$$

Testo statistika yra asimptotiškai pasiskirsčiusi pagal $\chi(p-1)$ chi–kvadrato skirstinį. *Vaito* teste naudojamas chi–kvadrato skirstinys, su laisvės laipsnių skaičiumi, lygiu parametrų skaičiumi pagalbinėje regresijos lygtyje.

Verta pastebėti, jog heteroskedastiškumas dažniausiai pasireiškia tuose duomenyse, kuriuose stebimi individualūs žmonės ar namų ūkiai. Tai galima pagrįsti tuo, jog kiekvieno elgesys ir poreikiai skiriasi. Jeigu stebėjimai turi esminių skirtumų, jiems nėra jokios priežasties turėti tokias pačias tendencijas, taigi ir susiduriame su heteroskedastiškumu [6].

2.2.2. IŠSKIRČIŲ TYRIMAS

Duomenų išskirtys[1] gali stipriai pakoreguoti regresijos tiesės lygties parametrų įverčius. Išskirtis – tai nuo kitų duomenų besiskiriantys stebėjimai. Radus išskirtį duomenyse, pirmiausia rekomenduojama patikrinti, ar tai nėra duomenų įvedimo klaida. Kitas variantas – papildyti duomenis naujas stebėjimais ir tuomet kartoti tyrimą. Jeigu nėra galimybės papildyti duomenis, tuomet išskirtis patartina keisti kitų kintamųjų sandauga arba kitu sąryšiu, arba tiesiog šalinti iš duomenų.

Šio tyrimo metu naudojome penkis skirtingus išskirčių nustatymo testus: standartizuotas liekamąsias paklaidas, stjudentizuotas liekamąsias paklaidas, Kuko įtakos matą, įtakos indeksą ir įtakos matą DfFit [23]. Pažymėkime n – stebėjimų skaičius (imties didys) ir K – nepriklausomų kintamųjų skaičius modelyje, liekanos $e_i = y_i - \hat{y}_i$. Trumpai apžvelkime kiekvieną kriterijų:

- Standartizuotos liekanos stebėjimą įvertina kaip iškirtį, jeigu $|e_i^*| > 3$. Standartizuota liekana apskaičiuojama:

$$e_i^* = \frac{e_i}{s_{e_i}} \tag{2.16}$$

- Stjudentizuotos liekanos apskaičiuojama:

$$e_i^s = \frac{e_i}{\sqrt{s_{(i)}^2(1-h_i)}} \tag{2.17}$$

  Kur $s_{(i)}$ yra standartinis nuokrypis, kai *i-tasis* stebėjimas yra pašalintas. Stebėjimą laikome išskirtimi, jeigu $|e_i^s| > 2$.

- Kuko įtakos matas parodo, parodo prognozės pokytį, kai yra pašalinamas *i-tasis* stebėjimas. Turint n dydžio imtį, Kuko matų bus n. Jeigu Kuko įtakos matas viršija $4/n$ reikšmę, tai *i-tasis* stebėjimas laikomas išskirtimi.

---

[1] Kalbame apie porinių stebėjimų išskirtis

- Įtakos indeksas įvertinas *i-tojo* stebėjimo atstumą nuo „centro“. Stebėjimą laikome išskirtimi, jeigu $ch_i > 2(K+1)n$.
- Įtakos matas DfFit parodo *i-tojo* stebėjimo pašalinimo įtaką prognozei $\hat{Y}_i$ ir yra apskaičiuojamas:

$$DfFit_i = \hat{Y}_i - \hat{Y}_{i(i)} \tag{2.18}$$

Kur $\hat{Y}_{i(i)}$ – yra prognozė pagal regresijos lygtį gautą pašalinus *i-tąjį* stebėjimą. Jeigu $|DfFit_i| > 2\sqrt{(K+1)/n}$, stebėjimas laikomas išskirtimi ir jo pašalinimas įtakoja prognozę $\hat{Y}_i$.

Tyrimo metu stebėjimas buvo laikomas išskirtimi, jeigu du iš anksčiau aptartų kriterijų, rodydavo, jog stebėjimas išsiskiria.

2.3. KVANTILINĖ REGRESIJA

Kvantilinės regresijos modelis, pristatytas R. Kroenker ir G. Basset, siekia įvertinti sąlygines kvantilių funkcijas – modelius, kuriuose prognozuojamos sąlyginio pasiskirstymo kvantilių reikšmės būtų išreikštos stebimų kintamųjų lygtimis [15]. F. Mosteller ir J. Tukey (1977 m., 266 psl.) teigė: „Regresijos kreivė pateikia vidurkių suvestinę, priklausančią nuo pasirinktos *x-ų* imties. Norint gilintis plačiau, galima sudaryti keletą skirtingų regresijos kreivių skirtingiems kvantiliams ir tokiu būdu gauti daugiau išsamesnės informacijos apie turimą imtį. Dažniausiai tai nėra atliekama ir regresija nesuteikia pilnos informacijos. Taip kaip vidurkis nepateikia išsamios informacijos apie pasiskirstymą, taip ir regresijos kreivė nepateikia visiškai išsamios informacijos apie imties pasiskirstymą“ [19].

Lyginant kvantilinę ir tiesines regresijas, galima išskirti šiuos itin svarbius skirtumus:

- Kvantilinės regresijos metodas suteikia galimybę ištirti ne tik vidutinį charakteristikų kitimą, tačiau ir išanalizuoti ekstremalius atvejus. Šią regresiją yra patogu naudoti tuomet, kai norima išsiaiškinti didžiausią tikėtiną temperatūrą esant stipriam debesuotumui (pvz. 0,95 kvantilio) ar mažiausią tikėtiną galimą upės vandens lygį (pvz. 0,05 kvantilį).
- Medianos kvantilinė regresija pateikia daug geresnius rezultatus asimetriškoms duomenų imtims negu tiesinė regresija [12].
- Kvantilinės regresijos liekamosioms paklaidoms nėra taikomi pasiskirstymo reikalavimai (kaip tiesinės regresijos atveju, liekanos turi būti pasiskirstę pagal normalųjį skirstinį) [14]. Taip yra todėl, kadangi prognozuojami įverčiai stipriau įvertina pasiskirstymą prie tam tikro nustatyto kvantilio, negu prie visos duomenų imties.
- Kvantilinė regresija yra atspari išskirtims.
- Kvantilinėje regresijoje nėra taikoma homoskedastiškumo prielaida.

2.3.1. KVANTILINĖS REGRESIJOS SUDARYMAS

Apžvelkime, kaip sudaroma kvantilinė regresija. Atsitiktiniam dydžiui (toliau a.d.) Y tikimybinio pasiskirstymo funkciją galima užrašyti:

$$F(x) = P(Y \leq x) \tag{2.19}$$

Tuomet *τ-asis* a.d. kvantilis Y kvantilis užrašomas kaip atvirkštinė funkcija:

$$Q(\tau) = inf\{y: F(y) \geq \tau\} \tag{2.20}$$

kur $\tau \in (0,1)$. Kai $Q(0,5)$ tuomet turime medianos kvantilį. A.d. Y atsitiktinei imčiai $\{y_1, \dots, y_n\}$ imties mediana sumažina nuokrypių modulių sumą:

$$mediana = argmin_{\xi \in R} \sum_{i=1}^{n} |y_i - \xi| \tag{2.21}$$

Tuo tarpu bendra *τ-ojo* imties kvantilio $\xi(\tau)$ išraiška, kuri yra analogiška $Q(\tau)$ išraiškai, užrašoma:

$$\xi(\tau) = argmin_{\xi \in R} \sum_{i=1}^{n} \rho_\tau (y_i - \xi) \tag{2.22}$$

kur $\rho_\tau(z) = z(\tau - I(z < 0))$ ir $I(\cdot)$ – indikatoriaus funkcija. Nuostolių funkcija ρ_τ įgyja τ svorį teigiamoms $y_i - \xi$ liekanoms ir $I - \tau$ svorį neigiamoms liekanoms.

Naudojant šią nuostolių funkciją, sąlyginė tiesinė kvantilinė funkcija priskiria *τ-ojo* imties kvantilio išraišką $\xi(\tau)$ regresijos kriterijams tokiu pačiu būdu kaip ir sąlyginė tiesinė vidurkių funkcija priskiriama imties vidurkiui. Žinant, jog mažiausių kvadratų metodu (MKM) sąlyginės tiesinės regresijos funkcijos $E(Y|X = x) = x'\beta$ vidurkių įverčiai apskaičiuojami taip:

$$\hat{\beta} = argmin_{\beta \in R} \sum_{i=1}^{n} (y_i - x_i'\beta)^2 \tag{2.23}$$

Apskaičiuotas $\hat{\beta}$ parametras minimizuoja liekanų kvadratų sumą tokiu pačiu būdu kaip imties vidurkis $\hat{\mu}$ minimizuoja kvadratų sumą:

$$\hat{\mu} = argmin_{\mu \in R} \sum_{i=1}^{n} (y_i - \mu)^2 \tag{2.24}$$

Taip pat kvantilinė regresija įvertina sąlyginę tiesinę kvantilinę funkciją $Q(\tau|\mathrm{X} = \mathrm{x}) = x'\beta(\tau)$ išsprendžiant

$$\hat{\beta}(\tau) = argmin_{\beta \in R} \sum_{i=1}^{n} \rho_\tau (y_i - x_i'\beta) \tag{2.25}$$

kai $\tau \in (0,1)$. Dydis $\hat{\beta}(\tau)$ yra vadinamas *τ-uoju* regresijos kvantiliu. Regresijos kvantilių rinkinys, kitaip dar vadinamas kvantilių procesu, žymimas:

$$\{\beta(\tau): \tau \in (0,1)\} \tag{2.26}$$

Taigi šiame darbe buvo sudaromos kvantilių funkcijos $Q(\tau|\mathrm{X} = \mathrm{x})$ ir statistiškai apskaičiuojami parametrai $\hat{\beta}(\tau)$ [22].

2.4. DIRBTINIAI NEURONINIAI TINKLAI

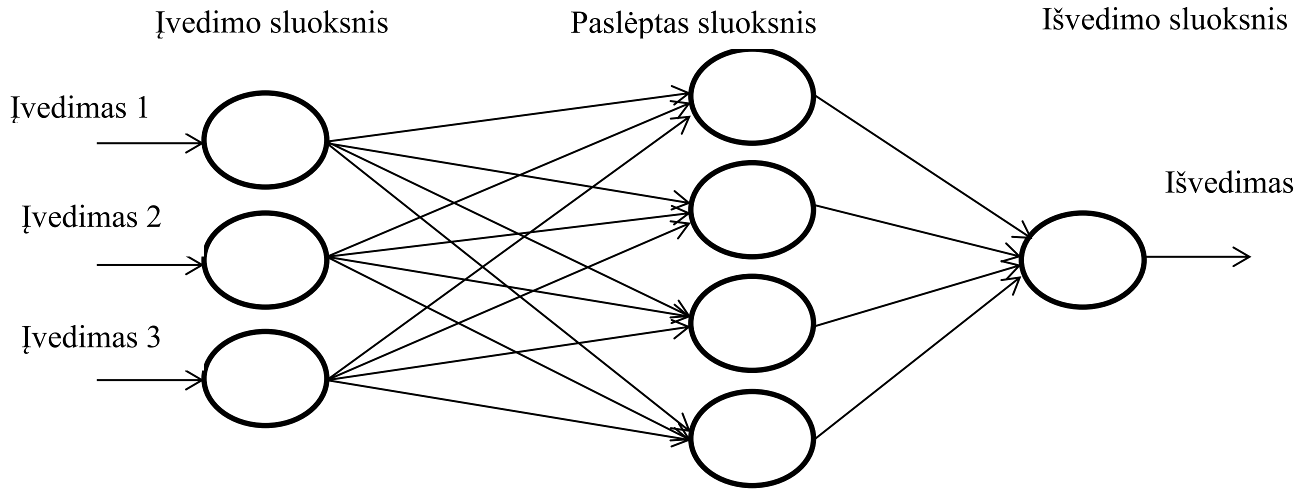
Dirbtiniai neuroniniai tinklai (toliau neuroniniai tinklai) yra sukurti remiantis biologinės nervų sistemos analogu norint išsiaiškinti žmonių atmintį, mokinimosi sugebėjimus ir kognityvius procesus. Neuroninis tinklas yra sudarytas iš daugelio tarpusavyje susijusių skaičiavimus atliekančių komponentų – neuronų. Neuroninis tinklas turi tris rėžimus:

- Apmokymo, kurio metu neuronas keičia su įvedamais duomenimis susijusius svorius taip, kad būtų gautas kuo tikslesnis rezultatas.
- Testavimo, kurio metu sureguliuojami parametrai – nustatomas optimalus neuronų skaičius paslėptame sluoksnyje.
- Naudojimo, kurio metu neuroninis tinklas grąžina su pateiktomis reikšmėmis susijusią prognozuojamą reikšmę.

Galima išskirti vienasluoksnius ir daugiasluoksnius neuroninius tinklus. Pirmuoju atveju, neuronai yra tiesiogiai tarpusavyje susiję, o kitu atveju yra sudaroma hierarchinė sistema, kurioje kiekvieno sluoksnio neuronų įėjimai yra susiję tik su prieš tai buvusio sluoksnio neuronais. Dažniausiai neuroniniai tinklai turi šias sudedamąsias dalis [25]:

- Įvedimo sluoksnį, kuriame duomenys yra įvedami į tinklą.
- Paslėptą neuronų sluoksnį, kuriame remiantis įvestais duomenimis ir apmokymo metu nustatytais kintamųjų svoriais vykdomi uždavinio sprendimo skaičiavimai.
- Išvedimo sluoksnį, kuriame pateikiami apskaičiuoti optimalūs svoriai ir išvedimo reikšmė.

Žemiau pateiktame 2.1 paveiksle matome paprasčiausią neuroninių tinklų modelį, kuris turi įvedimo, paslėptą ir išvesties sluoksnius. Egzistuoja keletas taisyklių apie paslėpto sluoksnio neuronų skaičių [24]: teigiama, jog šio sluoksnio neuronų skaičius turėtų būti intervale tarp įvedamų ir išvedamų kintamųjų skaičiaus sumos, kitais atvejais siūloma, jog šis skaičius neviršytų 2/3 įvedamų ir išvedamų kintamųjų skaičiaus sumos arba paslėpto sluoksnio neuronų skaičius turėtų būti dvigubai mažesnis negu įvedamų kintamųjų skaičius.



2.1 pav. Neuroninio tinklo struktūra

Taip pat galima išskirti keletą neuroninių tinklų skaičiavimo metodų:

- Vienasluoksnis perceptronas – kai tarp įėjimo ir išėjimo sluoksnių nėra paslėptų neuronų sluoksnių;
- Daugiasluoksnis perceptronas – kai tarp įėjimo ir išėjimo sluoksnių yra bent vienas paslėptas neuronų sluoksnis;
- Radialinių bazinių funkcijų tinklai – kai paslėptuose sluoksniuose kaip aktyvavimo funkcija yra naudojamos radialinės bazinės funkcijos.

Paprasčiausias daugiasluoksnio perceptrono modelis turi n kintamųjų įvedimo sluoksnį ir išvedimo sluoksnį su vienu kintamuoju. Šis neuroninis tinklas apskaičiuojamas pagal funkciją:

$$o(x) = f(w_0 + \sum_{i=1}^{n} w_i x_i) = f(w_0 + w^T x) \quad (2.27)$$

kur w_0 ir $w = (w_1, \dots, w_n)$ yra apskaičiuoti svoriai, $x = (x_1, \dots, x_n)$ – kintamųjų vektorius. Norint padidinti modelio lankstumą į neuroninių tinklų sudarymą buvo įtrauktas paslėptas sluoksnis. Tarkime paslėptas sluoksnis turi *J* neuronų, tuomet:

$$o(x) = f\left(w_0 + \sum_{j=1}^{J} w_j \cdot f\left(w_{0j} + \sum_{i=1}^{n} w_{ij} x_i\right)\right) = f\left(w_0 + \sum_{j=1}^{J} w_j \cdot f\left(w_{0j} + w^T x\right)\right) \quad (2.28)$$

Šios funkcijos yra matematiškai ekvivalenčios apibendrinto tiesinio modelio (GLM) procedūra gautoms funkcijoms susiejant per f^{-1}. Šiuo atveju, visi suskaičiuoti svoriai yra ekvivalentūs GLM procedūra gautiems regresijos modelio įverčiams [9].

2.5. SUDARYTŲ MODELIŲ TIKSLUMO ĮVERTINIMAS

Tyrimo metu sudaryti tiesinės ir kvantilinės regresinės analizės modeliai bei neuroninis tinklas buvo analizuojami ne tik atskirai, bet ir lyginami tarpusavyje, kad būtų galima nustatyti tiksliausiai prognozuojantį Vilniaus butų vieno kvadratinio metro kainą. Modeliai tarpusavyje buvo lyginami remiantis šiais kriterijais:

- Vidutinė absoliutinė procentinė paklaida (MAPE):

$$MAPE = \frac{100}{n} \cdot \sum_{i=1}^{n} \left|\frac{f_i - y_i}{y_i}\right| \quad (2.29)$$

Kur y_i – Registrų centre užfiksuota kaina, f_i – prognozuojama kaina, n – kintamųjų skaičius. Vidutinė absoliutinė procentinė paklaida nusako santykinį prognozavimo tikslumą, kuriuo remiantis galima palyginti skirtingų rodiklių prognozes. Jeigu $MAPE < 10\%$, tuomet prognozavimas yra laikomas labai tiksliu.

- Vidutinė procentinė paklaida (MPE):

$$MPE = \frac{100}{n} \cdot \sum_{i=1}^{n} \frac{f_i - y_i}{y_i} \quad (2.30)$$

Vidutinė procentinė paklaida yra santykinis dydis, rodantis prognozės nuokrypį. Esant labai tiksliai prognozei, tiek nukrypimas į viršų, tiek nukrypimas į apačią turi artėti prie nulio. Prognozė yra laikoma tikslia, jeigu $MPE < 5\%$.

- Vidutinė kvadratinė paklaida (MSE):

$$MSE = \frac{1}{n} \cdot \sum_{i=1}^{n} (f_i - y_i)^2 \tag{2.31}$$

Vidutinė kvadratinė paklaida nusako paklaidos dispersiją. Ja remiantis dažnai parenkami optimalūs prognozavimo modelio parametrai.

- Vidutinė absoliutinė paklaida (MAE):

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} |f_i - y_i| = \frac{1}{n} \cdot \sum_{i=1}^{n} |e_i| \tag{2.32}$$

Kur $e_i = f_i - y_i$ yra prognozavimo paklaida (skirtumas tarp prognozuojamos ir tikrosios reikšmės). Ši paklaida įvertina, gautų prognozių artumą tikrosioms reikšmėms.

- Koreguotas determinacijos koeficientas R_{adj}^2:

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1} \tag{2.33}$$

Kur R^2 – determinacijos koeficientas, p – nepriklausomų kintamųjų skaičius lygtyje. Koreguotas determinacijos koeficientas nurodo kurią atsitiktinio dydžio Y sklaidos dalį galima paaiškinti sudarytu prognozavimo modeliu kintamųjų $X_1, X_2 \dots, X_k$ atžvilgiu.

2.6. VILNIAUS MIESTO BUTŲ DUOMENŲ ANALIZĖS MODELIS

Šiame skyrelyje trumpai apžvelgsime atliekamą Vilniaus mieto butų vieno kvadratinio metro kainos prognozavimo tyrimą. Jį galime suskirstyti į šiuos etapus:

- Duomenų parengimas tyrimui: užpildomos trūkstamos reikšmės, atliekama klaidų paieška.
- Pradinė duomenų analizė. Šiame etape sudarytos Vilniaus miesto zonų bei sienų tipų kintamųjų grupės, ištirtos kintamųjų koreliacijos bei atliktas duomenų tiesinimas.
- Pradinio regresijos modelio sudarymas. Turint atrinktus kintamuosius sudaromas pradinis regresijos modelis, išsiaiškinama, kurie kintamieji yra reikšmingi, atrenkama galutinė nepriklausomų kintamųjų imtis bei ištiriamos išskirtys.
- Išskirčių šalinimas. Surandamos išskirtys ir, jeigu jos tenkina nustatytus kriterijus, pašaliname jas iš tyrimo.
- Tiesinės regresijos modelio sudarymas. Sudarome regresijos modelį ir patikriname, ar tenkinamos visos prielaidos ir reikalavimai.
- Kvantilinės regresijos sudarymas. Sudaromi penki skirtingi kvantilinės regresijos modeliai šiems kvantiliams: 0,05; 0,25; 0,5; 0,75 ir 0,95. Analizuojami gauti rezultatai.

- Neuroninių tinklų sudarymas. Paruošiami duomenys ir atliekami skaičiavimai naudojant neuroninius tinklus. Išrenkamas tiksliausiai prognozuojantis neuroninis tinklas.
- Apskaičiuojamos prognozavimo paklaidos visiems atliktiems metodams.
- Analizuojami gauti rezultatai, pateikiamos išvados ir rekomendacijos.

Atlikus visus šiuos žingsnius, kiekvienu metodu gausime skirtingas Vilniaus miesto butų vieno kvadratinio metro kainos prognozes. Gautus rezultatus tarpusavyje palyginsime ir nustatysime labiausiai tinkantį metodą šiam tyrimui.

2.7. PROGRAMINĖ ĮRANGA

Pagrindinei tyrimo daliai atlikti buvo pasirinkta SAS programinė įranga. SAS yra lyderis verslo analitikos programinėje įrangoje ir paslaugose bei didžiausias nepriklausomas pardavėjas verslo intelekto rinkoje. Ši programinė įranga padeda optimizuoti verslo galimybes analizuojant įvairius duomenis. Ji yra naudojama 140-yje pasaulio valstybių. Be to, net 91 įmonė iš geriausių pasaulio įmonių 100 yra pasirinkusi būtent šį tiekėją[2].

Darbui atlikti buvo naudojama SAS 9.2 versija. SAS programinė įranga susideda iš dialoginės sistemos sesijų, programų valdymo, SAS kalbos kompiliatoriaus, vidinio duomenų valdymo, įvedimo/išvedimo galimybių, makro programos bei daugybės SAS procedūrų. Galime išskirti tris pagrindines SAS dalis – SAS Base (duomenų prieiga, transformacijos, ataskaitos), SAS STAT (statistinė analizė), SAS GRAPH (aukštos kokybės grafikai). Ši programa sugeba puikiai atlikti visas užduotis reikalingas tyrimui.

Taip pat tyrimo metu naudojama Statistica 8 programinė įranga ir MS Excel programa. Programinėje įrangoje Statistica buvo naudojamasi Automatizuotų neuroninių tinklų (SANN) sudarymo galimybėmis. MS Excel programoje kaupiami pradiniai duomenys ir gauti rezultatai, braižomi sudarytų modelių lyginimo grafikai bei skaičiuojamos paklaidos.

[2] Pagal 2013 metų Fortune Global 500® sąrašo duomenis.

3. TIRIAMOJI DALIS

Šiame skyrelyje detaliau aptarsime duomenų parengimą tyrimui, tiesinės ir kvantilinės regresijų modelių bei neuroninių tinklų pritaikymą 2008–2012 metais užregistruotiems Vilniaus miesto butų pirkimo-pardavimo sandoriams. Sudarytų prognozavimo metodų tikslumą įvertinsime prognozuojant 2013 metų Vilniaus miesto butų vieno kvadratinio metro kainas. Apžvelgus ir išanalizavus gautus rezultatus, aptarsime, kuris metodas yra tinkamiausias turimų duomenų prognozei.

3.1. DUOMENŲ PARENGIMAS TYRIMUI

Gavus Registrų centro duomenis, pirmiausiai buvo išanalizuoti turimi kintamieji: ištirtos koreliacijos, įgyjamų reikšmių intervalai bei sudarytas grupavimas. Nuspręsta pritaikyti tam tikrus apribojimus įsigyjamų butų plotui, baigtumo lygiui, rinkos vertės ir vieno kvadratinio metro kainai. Vilniaus miesto butams taikomus kriterijus pateikiame 3.1 lentelėje.

3.1 lentelė

Vilniaus miesto butams taikomi atrankos kriterijai

	2008-2012 metai		**2013 metai**	
Kriterijai:	**Sandorių sk.**	**Procentai**	**Sandorių sk.**	**Procentai**
Pradinis sandorių sk.	**26403**	**100**	**7724**	**100**
Įsigyjamas plotas [15 – 200] m^2	172	0,65	76	0,98
Rinkos vertė [1000 – 1.000.000] Lt	159	0,60	21	0,27
1 m^2/Lt: [250 – 20 000][3]	391	1,48	9	0,12
Baigtumas > 50 %	87	0,33	0	0,00
Atmesta po pagrindinių kriterijų	**809**	**3,06**	**106**	**1,37**
Išskirtys	1594	6,04	0,00	0,00
Viso atmesta	2403	9,10	106	1,37
Galutinis sandorių sk.	**24000**	**90,9**	**7618**	**98,63**

Sudaryti Vilniaus miesto butų atrankos kriterijai pritaikyti tiek tiriamai 2008–2012 metų duomenų imčiai, tiek ir 2013 metų duomenims. Ištyrėme, jog dažniausiai Vilniaus mieste butai yra 15–200 m^2 ploto. Šio kriterijaus neatitiko 172 tyrimo imties sandoriai. Buto baigtumas turi būti didesnis >50%, kadangi tokio baigtumo butuose jau galima gyventi (šio kriterijaus neatitiko 87 sandoriai). Taip pat, nustatyti kainų limitai: įvertintus buto rinkos vertes ir vieno kvadratinio metro

[3] Pagal kintamąjį *Kaina* (Perskaičiuota vieneto vertė su žeme)

kainas atmesta dar 550 sandorių. Pradinių kriterijų neatitiko 3,06 % sandorių, o suradus ir pašalinus išskirtis tyrime liko 24 000 sandorių. Taigi iš viso buvo pašalinta 9,1% sandorių. 2013 metų duomenims pritaikyti tik pagrindiniai kriterijai, po kurių atmesta 1,37% sandorių. Išskirčių tyrimas šiems duomenims nebuvo taikomas, nes siekėme prognozuoti visas Registrų centre užregistruotas butų pirkimo-pardavimo kainas.

Sekantis žingsnis po sandorių atrinkimo buvo duomenų grupavimas. Palikti keturi galimi sienų tipai: *sienos_1* – karkasinis (medis su karkasu ir metalas su karkasu), *sienos_2* – monolitinis (gelžbetonio plokštės, monolitinis gelžbetonis ir akmenbetonis), *sienos_3* – plytos/blokeliai (plytų arba blokelių sienos) ir *sienos_4* – kitos medžiagos (rąstai ir kitos medžiagos). Ištyrus šių kintamųjų reikšmingumą, nustatyta, kad reikšmingas yra tik trečiasis sienų tipas *sienos_3* (toliau *sienos*) – plytos/blokeliai. Kiti sienų tipai reikšmingos įtakos butų kainai neturi.

Vilniaus miesto savivaldybės zonos suskirstytos į keturias grupes: *centras, prestižiniai, miegamieji* ir *kiti rajonai* (mikrorajonų priskyrimas kiekvienai grupei pateiktas 2.1. lentelėje). Zonos buvo suskirstytos į grupes atsižvelgiant į minimalias ir maksimalias vieno kvadratinio metro kainas, vidurkį, standartinį nuokrypį ir kitas statistikas, kurių rezultatus pateikiame 3.2 lentelėje:

3.2 lentelė

Vilniaus miesto zonų analizė

Vilniaus miesto zonos	Sandorių skaičius	Minimali kaina	Maksimali kaina	Vidurkis	Mediana	Moda	Stand. nuokrypis
Centras	1067	355,99	15358,69	5876,08	5824,19	9200	2235,68
Prestižiniai raj.	2485	296,52	11014,04	4802,06	4637,54	4900	1862,43
Miegamieji raj.	20386	251,76	14422,27	3928,72	3727,45	4500	1432,45
Kiti raj.	1656	288,18	8904,67	2966,30	2624,25	5000	1423,16

Kaip matome, daugiausiai sandorių atliekama *miegamųjų rajonų* zonoje, mažiausiai – *centre*. Kintamasis *objektų skaičius* taip pat buvo perkoduotas: jeigu perkamas vienas objektas – paliekama reikšmė „1“, jei du objektai – „2“, jei trys ar daugiau – „3“.

Pritaikius pradinius atrankos kriterijus bei sudarius kintamųjų grupes atlikome duomenų tiesinimą. Geriausia duomenų tiesinė priklausomybė buvo nustatyta logaritmavus priklausomąjį kintamąjį *kaina* ir nepriklausomąjį kintamąjį *rinkos vertė*.

3.2. DAUGIALYPĖS TIESINĖS REGRESIJOS MODELIS

Pirmąjį prognozės modelį sudarėme naudodamiesi daugialypės tiesinės regresijos metodu. Modelis sudarytas iš 24 000 pirkimo-pardavimo sandorių, užregistruotų Registrų centro duomenų bazėje 2008–2012 metais. Pagal tiesinės regresijos metodus sudarytas modelis:

$$\ln(Y) = -0,683 + 0,770 \cdot \ln(X_1) + 0,141 \cdot X_2 + 0,118 \cdot X_3 + 0,059 \cdot X_4 - 0,018 \cdot X_6 - 0,250 \cdot X_7 + 0,027 \cdot X_8 + 0,030 \cdot X_9 + 0,021 \cdot X_{10} \quad (3.1)$$

Šiame modelyje kintamasis X_5 (*kiti rajonai)* yra išreikštas per kitus zonų kintamuosius:

$$X_5 = Intercept - X_2 - X_3 - X_4 \quad (3.2)$$

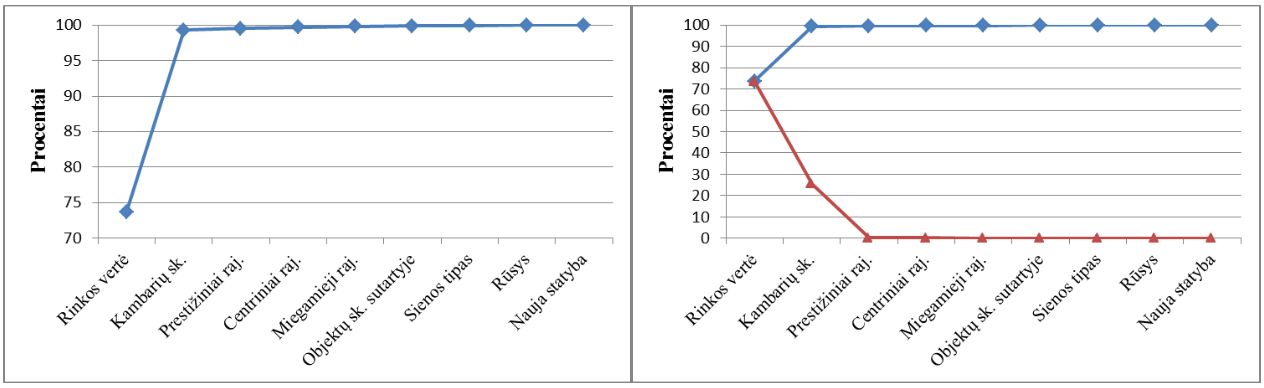
3.3 Lentelė

Tiesinės regresijos parametrų įverčiai ir reikšmingumas

Kintamieji	Parametrų įverčiai	Pr > |t|	VIF
Intercept	-0,683	< ,0001	0
Rinkos vertė	0,770	< ,0001	1,979
Centras	0,141	< ,0001	1,653
Prestižiniai raj.	0,118	< ,0001	2,438
Miegamieji raj.	0,059	< ,0001	2,900
Kiti raj.	0,000	.	.
Objektų skaičius	-0,018	< ,0001	1,416
Kambarių skaičius	-0,250	< ,0001	1,415
Sienos	0,027	< ,0001	1,466
Rūsys	0,030	< ,0001	1,831
Nauja statyba	0,021	< ,0001	1,691

Didžiausią įtaką vieno kvadratinio metro kainai iš Vilniaus miesto zonų turi X_2 (*centras)* – 0,141, tuomet nedaug atsilieka *prestižiniai rajonai* (X_3) – 0,118 ir kiek didesnis skirtumas matomas *miegamuosiuose rajonuose* (X_4) – 0,059. Kintamasis *objektų skaičius* (X_6) įgyja neigiamą reikšmę, kadangi dažniausiai butai su garažu ar kitais papildomais objektas yra perkami su miesto pakraštyje esančiais butais. Kintamasis *kambarių skaičius* (X_7) taip pat įgyja neigiamą reikšmę, kadangi daugiausiai perkami ir parduodami 1–2 kambarių butai, kurių vieno kvadratinio metro kaina yra didžiausia. Daugiau kambarių turintys butai nėra tokie populiarūs ir jų vieno kv. metro kaina dažnai būna mažesnė, kadangi išauga išlaikymo kaštai. Jeigu buto *sienos* yra plytų arba blokelių (X_8), vidutiniškai tokio tipo butas bus šiek tiek brangesnis. *Nauja statyba* (X_{10}) bei *rūsys* (X_9) taip pat turi teigiamą įtaką didesnei buto kainai.

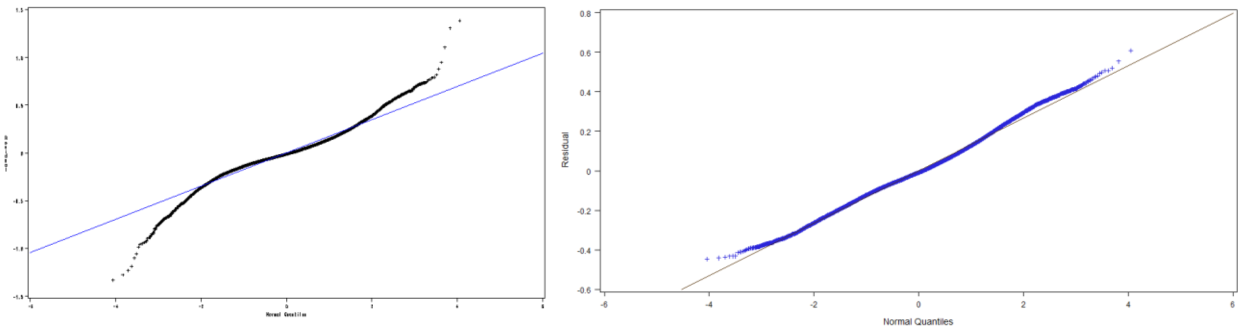
Pateiktame 3.1 pav. Parametrų įverčiai matome, jog visi į lygtį patenkantys kintamieji yra reikšmingi ($p < .0001$). Dispersijos mažėjimo daugiklis rodo, jog su multikolinearumo problema nesusiduriama ($VIF < 4$). Modelio pakoreguotas determinacijos koeficientas R^2_{adj} = 89,28 %. Tai rodo, jog šis modelis yra tikslus ir paaiškina beveik 90 % tiriamos imties.



3.1 pav. Kitamųjų įtaka tiesinės regresijos modelyje

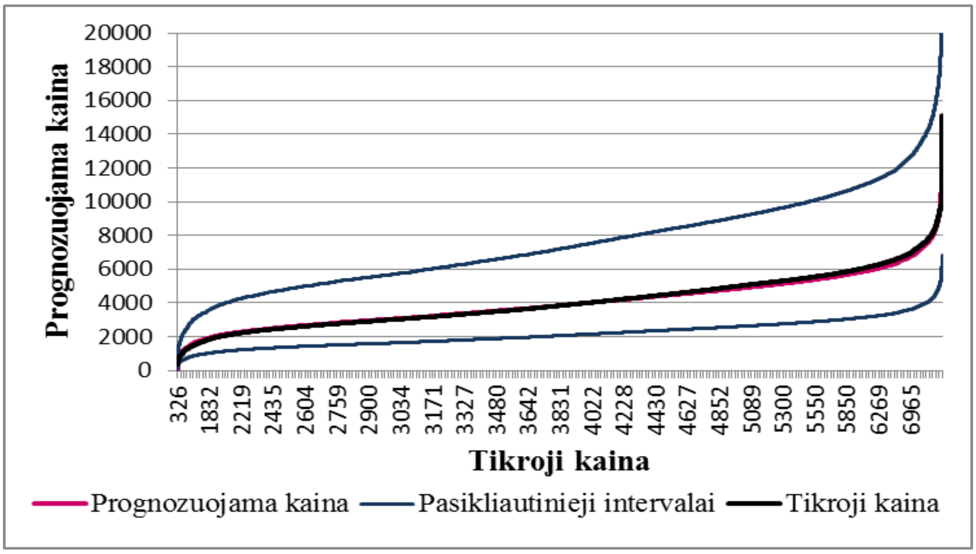
Kintamųjų įtaka tiesinės regresijos modeliui yra pavaizduota 3.1 paveiksle. Kaip matome, didžiausią įtaką modeliui turi kintamasis *rinkos vertė* (X_1), kurio dalinis determinacijos koeficientas $part. R^2 = 65{,}77\ \%$, antroje vietoje lieka kintamasis *kambarių skaičius* (X_7) su $part. R^2 = 22{,}87\ \%$. Likę kintamieji nepasižymi labai aukštu daliniu determinacijos koeficientu, kadangi jie įgyja ir nulines reikšmes, tad nėra pritaikomi kiekvienai lygčiai. Atlikus *Durbino-Vatsono* testą paaiškėja, jog autokoreliacijos nėra – $DW = 1{,}58$.

Atlikus liekanų normalumo testą gauname, jog prielaida apie liekanų normalumą nėra patvirtinama, o *Kolmogorovo–Smirnovo* statistika $D = 0{,}036$. Tačiau paanalizavus grafinius rezultatus prielaidos apie liekanų normalumą atmesti negalime. Žemiau pateiktame paveiksle vaizduojame pradinį ir šio modelio liekanų normalumo grafikus.



3.2 pav. Liekanų normalumo tyrimas

Homoskedastiškumo prielaida buvo tiriama ir grafiškai, ir pagal *Vaito* bei *Breušo–Pagano* testus. Atlikti testai nepatvirtino, jog liekanos yra homoskedastiškos, tačiau grafiniai rezultatai leidžia daryti prielaidą apie dispersijų pastovumą. Šie grafikai yra B Priede.



3.3 pav. 2008–2012 m. duomenų prognozavimas tiesinės regresijos modeliu

3.3 paveiksle pateikiame grafinius rezultatus, kaip tiesinės regresijos modelis prognozuoja 2008–2012 metais Registrų centre užfiksuotas pirkimo-pardavimo sandorių vieno kvadratinio metro kainas su šių prognozių pasikliautinaisiais intervalais (kai $\alpha = 0{,}05$).

3.3. KVANTILINĖS REGRESIJOS MODELIAI

Tyrimo metu buvo nuspręsta sudaryti penkis skirtingus kvantilinės regresijos modelius šiems kvantiliams: 0,05; 0,25; 0,5; 0,75 ir 0,95. Sudarius šias lygtis analizavome kintamųjų parametrų įverčius, jų reikšmingumą bei prognozavimo tikslumą. Trumpai apžvelgime kiekvieną modelį atskirai.

3.3.1. 0,05 KVANTILIO REGRESIJOS MODELIS

Pirmiausiai sudarėme 0,05 kvantilio kvantilinės regresijos modelį. Gavome šį modelį:

$$\mathbf{ln}(Y) = -\mathbf{1{,}454} + \mathbf{0{,}820} \cdot \mathbf{ln}(X_1) + \mathbf{0{,}114} \cdot X_2 + \mathbf{0{,}091} \cdot X_3 + \mathbf{0{,}062} \cdot X_4 - \mathbf{0{,}016} \cdot X_6 - \mathbf{0{,}259} \cdot X_7 - \mathbf{0{,}049} \cdot X_8 + \mathbf{0{,}073} \cdot X_9 + \mathbf{0{,}034} \cdot X_{10} \quad (3.3)$$

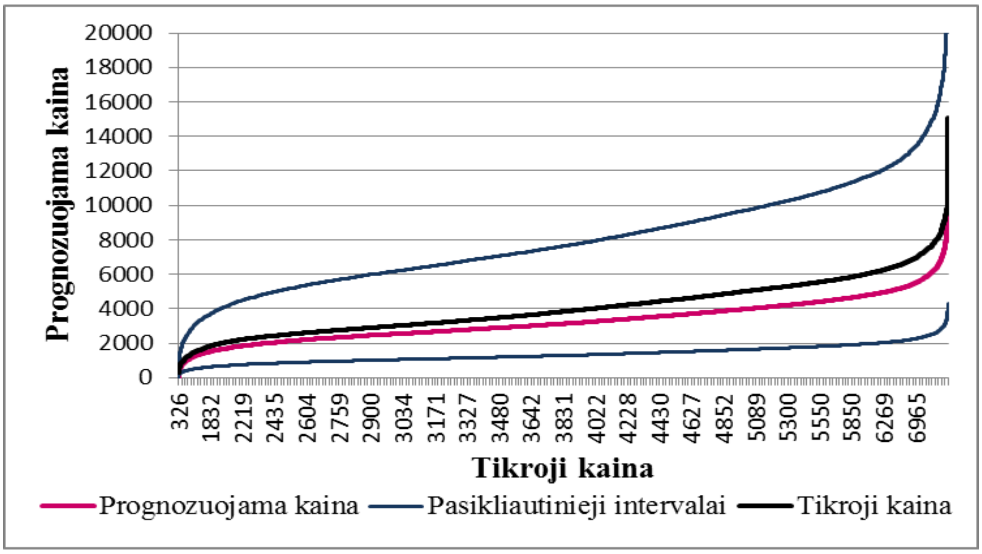
3.4 Lentelė

0,05 kvantilinės regresijos parametrų įverčiai ir reikšmingumas

Kintamieji	Parametrų įverčiai	95% pasikliautinieji intervalai		Pr > |t|
Intercept	-1,454	-1,531	-1,377	< ,0001
Rinkos vertė	0,820	0,813	0,827	< ,0001
Centras	0,114	0,091	0,136	< ,0001
Prestižiniai raj.	0,091	0,075	0,107	< ,0001
Miegamieji raj.	0,062	0,052	0,072	< ,0001
Kiti raj.	0,000	0,000	0,000	.
Objektų skaičius	-0,016	-0,022	-0,010	< ,0001
Kambarių skaičius	-0,259	-0,263	-0,254	< ,0001

Sienos	-0,049	-0,057	-0,042	< ,0001
Rūsys	0,073	0,064	0,082	< ,0001
Nauja statyba	0,034	0,025	0,043	< ,0001

Visi į lygtį įtraukti kintamieji yra reikšmingi. Matome, jog iš Vilniaus miesto zonų kintamųjų didžiausią įtaką turi *centras* (X_2) – 0,114, nuo jo nedaug atsilieka ir *prestižiniai rajonai* (X_3) – 0,091. Gautąjį modelį palyginus su tiesinės regresijos modeliu matome, jog pasikeitė kintamojo *sienos* (X_8) įtaka vieno kvadratinio metro kainai – ji tapo neigiama. 3.4 pav. pateikiame grafinius rezultatus, kaip 0,05 kvantilio regresijos modelis prognozuoja 2008–2012 metais Registrų centre užfiksuotas pirkimo-pardavimo sandorių vieno kvadratinio metro kainas su šių prognozių pasikliautinaisiais intervalais (kai $\alpha = 0{,}05$).



3.4 pav. 2008–2012 m. duomenų prognozavimas 0,05 kvantilinės regresijos modeliu

Iš pateikto 3.4 pav. matome, jog 0,05 kvantilinės regresijos modelis prognozuoja mažesnes vieno kvadratinio metro kainas, negu jos yra iš tikrųjų. Santykiniui tikrųjų ir prognozuojamų kainų skirtumui vieno kvadratinio metro kainos dydis įtakos neturi.

3.3.2. 0,25 KVANTILIO REGRESIJOS MODELIS

Kitas sudarytas regresijos modelis – 0,25 kvantilio. Pateikiame jo išraišką ir parametrų įverčių lentelę:

$$\mathbf{ln}(\boldsymbol{Y}) = -\mathbf{1{,}289} + \mathbf{0{,}818} \cdot \mathbf{ln}(\boldsymbol{X_1}) + \mathbf{0{,}091} \cdot \boldsymbol{X_2} + \mathbf{0{,}085} \cdot \boldsymbol{X_3} + \mathbf{0{,}039} \cdot \boldsymbol{X_4} - \mathbf{0{,}020} \cdot \boldsymbol{X_6} - \mathbf{0{,}262} \cdot \boldsymbol{X_7} - \mathbf{0{,}019} \cdot \boldsymbol{X_8} + \mathbf{0{,}046} \cdot \boldsymbol{X_9} + \mathbf{0{,}023} \cdot \boldsymbol{X_{10}} \tag{3.4}$$

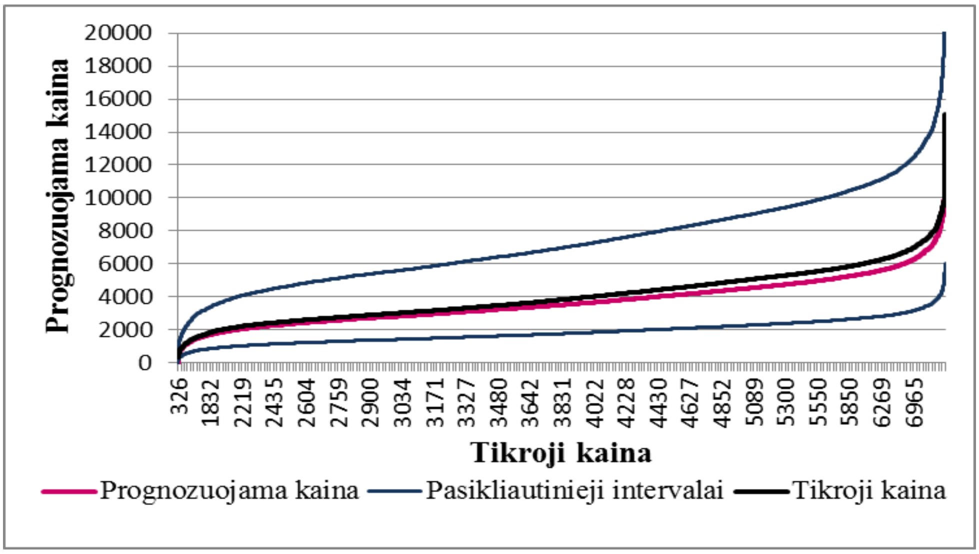
3.5 Lentelė

0,25 kvantilinės regresijos parametrų įverčiai ir reikšmingumas

Kintamieji	Parametrų įverčiai	95% pasikliautinieji intervalai		Pr > |t|
Intercept	-1,289	-1,356	-1,221	< ,0001
Rinkos vertė	0,818	0,812	0,824	< ,0001

Centras	0,091	0,072	0,110	< ,0001
Prestižiniai raj.	0,085	0,070	0,100	< ,0001
Miegamieji raj.	0,039	0,029	0,049	< ,0001
Kiti raj.	0,000	0,000	0,000	.
Objektų skaičius	-0,020	-0,024	-0,015	< ,0001
Kambarių skaičius	-0,262	-0,266	-0,259	< ,0001
Sienos	-0,019	-0,024	-0,015	< ,0001
Rūsys	0,046	0,041	0,052	< ,0001
Nauja statyba	0,023	0,017	0,028	< ,0001

Visi į lygtį įtraukti kintamieji yra reikšmingi. Palyginus su 0,05 kvantilio regresijos modeliu, jų tendencijos nepasikeitė.



3.5 pav. 2008–2012 m. duomenų prognozavimas 0,25 kvantilinės regresijos modeliu

3.5 pav. pateikiame grafinius rezultatus, kaip šis regresijos modelis prognozuoja 2008–2012 metais Registrų centre užfiksuotas pirkimo-pardavimo sandorių vieno kvadratinio metro kainas su šių prognozių pasikliautinaisiais intervalais (kai $\alpha = 0{,}05$). Matome, jog šio modelio prognozės jau yra artimesnės tikrajai kainai, kurią žymi juoda kreivė.

3.3.3. MEDIANOS KVANTILINĖS REGRESIJOS MODELIS

Mediana turimą duomenų imtį dalina į dvi lygias dalis. Pateikiame medianos kvantilinės regresijos modelį:

$$\mathbf{ln}(\boldsymbol{Y}) = -\mathbf{0,949} + \mathbf{0,794} \cdot \mathbf{ln}(\boldsymbol{X_1}) + \mathbf{0,118} \cdot \boldsymbol{X_2} + \mathbf{0,114} \cdot \boldsymbol{X_3} + \mathbf{0,045} \cdot \boldsymbol{X_4} - \mathbf{0,020} \cdot \boldsymbol{X_6} - \mathbf{0,255} \cdot \boldsymbol{X_7} + \mathbf{0,017} \cdot \boldsymbol{X_8} + \mathbf{0,026} \cdot \boldsymbol{X_9} + \mathbf{0,019} \cdot \boldsymbol{X_{10}} \tag{3.5}$$

3.6 Lentelė

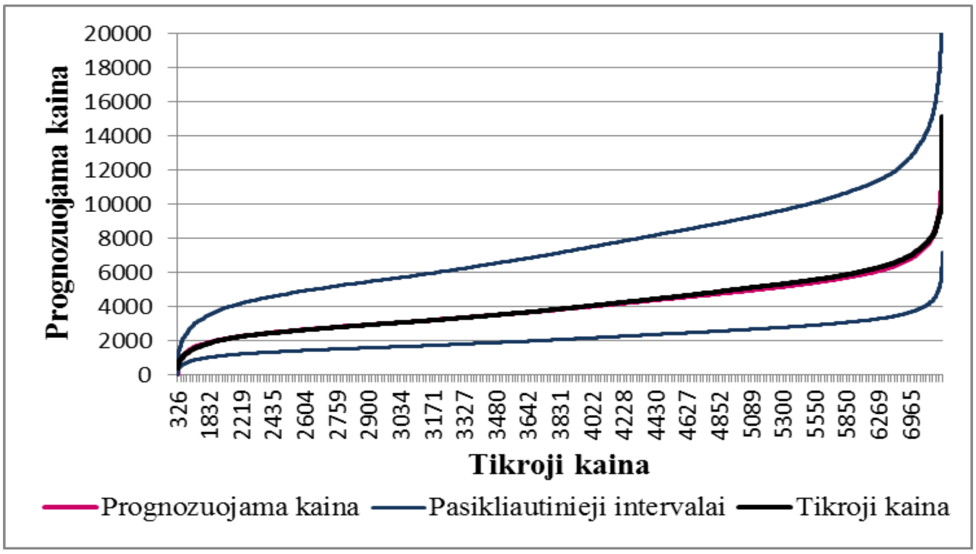
0,5 kvantilinės regresijos parametrų įverčiai ir reikšmingumas

Kintamieji	Parametrų įverčiai	95% pasikliautinieji intervalai		Pr > |t|
Intercept	-0,949	-1,016	-0,883	< ,0001

Rinkos vertė	0,794	0,788	0,800	< ,0001
Centras	0,118	0,102	0,134	< ,0001
Prestižiniai raj.	0,114	0,103	0,125	< ,0001
Miegamieji raj.	0,045	0,037	0,053	< ,0001
Kiti raj.	0,000	0,000	0,000	.
Objektų skaičius	-0,020	-0,023	-0,016	< ,0001
Kambarių skaičius	-0,255	-0,258	-0,252	< ,0001
Sienos	0,017	0,012	0,022	< ,0001
Rūsys	0,026	0,021	0,030	< ,0001
Nauja statyba	0,019	0,013	0,025	< ,0001

Medianos kvantilinės regresijos modelyje kintamasis *sienos* (X_8) jau įgyja teigiamą reikšmę. Vilniaus miesto zonų – *centro* (X_2) ir *prestižinių rajonų* (X_3) įverčių skirtumas dar labiau sumažėja – iki 0,004. *Miegamųjų rajonų* (X_4) įvertis yra daugiau nei du kartus mažesnis, palyginus su *prestižiniais rajonais* (X_2). Tai rodo, jog miegamuosiuose rajonuose butų vieno kvadratinio metro kainos yra mažesnės negu šių zonų, tačiau didesnės už *kitų rajonų* (X_5) butų kainas. Visi į modelį įtraukti kintamieji yra reikšmingi.

3.6 pav. pateikiame grafinius rezultatus, kaip šis regresijos modelis prognozuoja 2008–2012 metais Registrų centre užfiksuotas pirkimo-pardavimo sandorių vieno kvadratinio metro kainas su šių prognozių pasikliautinaisiais intervalais (kai $\alpha = 0{,}05$). Matome, jog šio modelio prognozės yra gan tikslios ir nėra žymesnių skirtumų nuo tikrosios kainos.



3.6 pav. 2008–2012 m. duomenų prognozavimas medianos kvantilinės regresijos modeliu

Kaip jau minėjome 2.3 skyrelyje, kvantilinei regresijai nėra taikomi liekanų normalumo ir homoskedastiškumo kriterijai. Tačiau norėdami atlikti detalesnę analizę šiuos testus atlikome ir C Priede pateikiame gautus rezultatus.

3.3.4. 0,75 KVANTILIO REGRESIJOS MODELIS

Pateikiame sudaryto 0,75 kvantilio regresijos modelio išraišką ir parametrų įverčių lentelę:

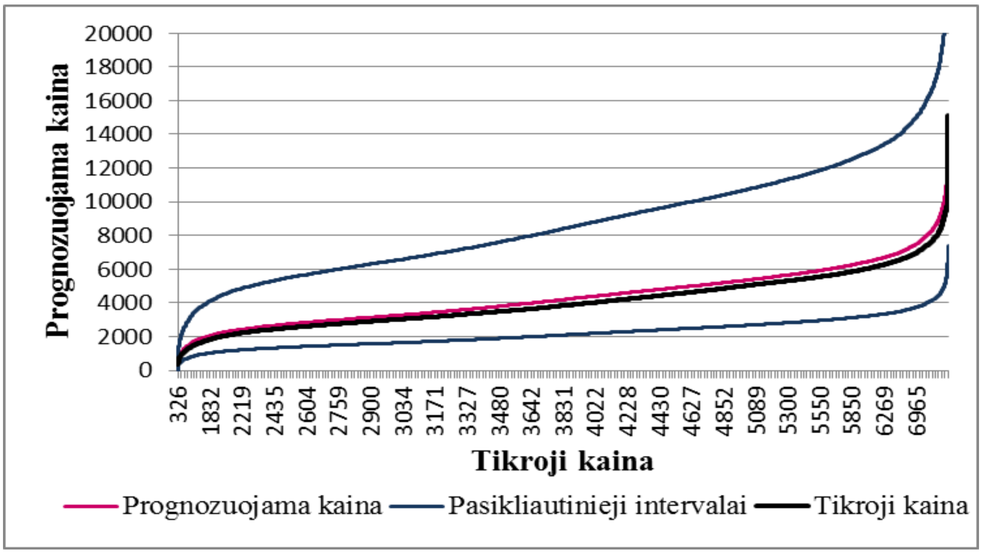
$$\ln(Y) = -0,718 + 0,780 \cdot \ln(X_1) + 0,143 \cdot X_2 + 0,117 \cdot X_3 + 0,058 \cdot X_4 - 0,024 \cdot X_6 - 0,251 \cdot X_7 + 0,059 \cdot X_8 + 0,014 \cdot X_9 + 0,013 \cdot X_{10} \quad (3.6)$$

3.7 Lentelė

0,75 kvantilinės regresijos parametrų įverčiai ir reikšmingumas

Kintamieji	Parametrų įverčiai	95% pasikliautinieji intervalai		Pr > |t|
Intercept	-0,718	-0,782	-0,653	< ,0001
Rinkos vertė	0,780	0,774	0,785	< ,0001
Centras	0,143	0,121	0,166	< ,0001
Prestižiniai raj.	0,117	0,101	0,133	< ,0001
Miegamieji raj.	0,058	0,045	0,070	< ,0001
Kiti raj.	0,000	0,000	0,000	.
Objektų skaičius	-0,024	-0,029	-0,020	< ,0001
Kambarių skaičius	-0,251	-0,254	-0,248	< ,0001
Sienos	0,059	0,053	0,065	< ,0001
Rūsys	0,014	0,007	0,020	< ,0001
Nauja statyba	0,013	0,006	0,019	< ,0001

Šiame kvantilinės regresijos modelyje jau galime įžvelgti didesnį skirtumą tarp Vilniaus miesto zonų: palyginus su medianos kvantiline regresija, *centro* (X_2) įtaka stipriai padidėjo, kai tuo tarpu *prestižinių rajonų* (X_3) – išliko beveik nepakitusi. *Miegamųjų rajonų* (X_4) įvertis, kaip ir visose lygtyse, yra mažesnis. Taip pat galime pastebėti, jog didėjant kvantiliui mažėja *naujos statybos* (X_{10}) butų kintamojo įvertis, kuris šiame modelyje yra lygus 0,013.



3.7 pav. 2008–2012 m. duomenų prognozavimas 0,75 kvantilinės regresijos modeliu

3.7 paveiksle matome, jog 0,75 kvantilinės regresijos lygtis prognozuoja šiek tiek didesnes nei tikrosios kainas. Grafike taip pat pavaizduoti ir pasikliautinieji intervalai (kai $\alpha = 0,05$).

3.3.5. 0,95 KVANTILIO REGRESIJOS MODELIS

Paskutinysis buvo sudarytas 0,95 kvantilio regresijos modelis. Pateikiame gautą modelio išraišką ir parametrų įverčių lentelę:

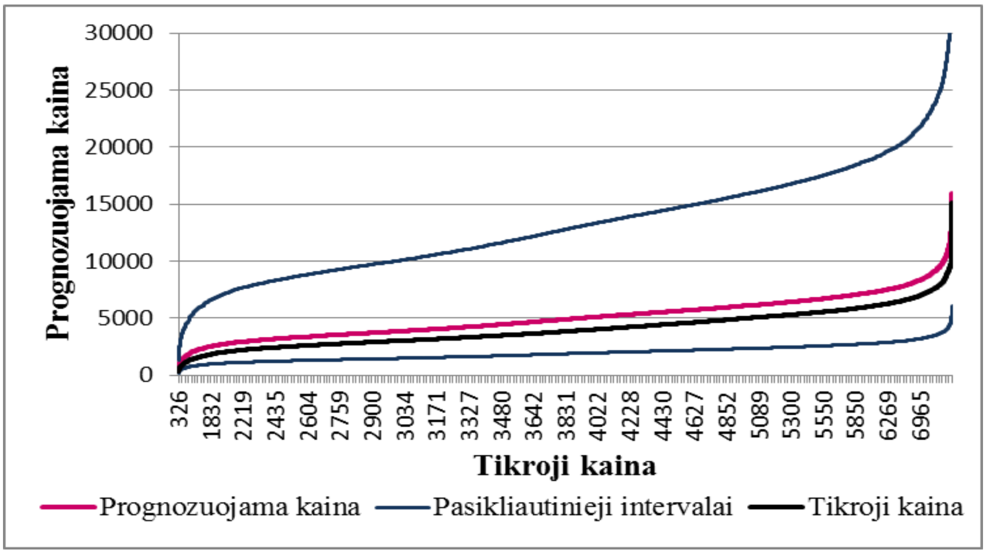
$$\mathbf{ln}(Y) = \mathbf{0,266} + \mathbf{0,704} \cdot \mathbf{ln}(X_1) + \mathbf{0,182} \cdot X_2 + \mathbf{0,135} \cdot X_3 + \mathbf{0,087} \cdot X_4 - \mathbf{0,014} \cdot X_6 - \mathbf{0,237} \cdot X_7 + \mathbf{0,094} \cdot X_8 + \mathbf{0,015} \cdot X_9 - \mathbf{0,010} \cdot X_{10} \tag{3.7}$$

3.8 Lentelė

0,95 kvantilinės regresijos parametrų įverčiai ir reikšmingumas

Kintamieji	Parametrų įverčiai	95% pasikliautinieji intervalai		Pr > |t|
Intercept	0,266	0,185	0,347	< ,0001
Rinkos vertė	0,704	0,698	0,711	< ,0001
Centras	0,182	0,156	0,209	< ,0001
Prestižiniai raj.	0,135	0,118	0,152	< ,0001
Miegamieji raj.	0,087	0,072	0,102	< ,0001
Kiti raj.	0,000	0,000	0,000	.
Objektų skaičius	-0,014	-0,020	-0,008	< ,0001
Kambarių skaičius	-0,237	-0,241	-0,233	< ,0001
Sienos	0,094	0,086	0,101	< ,0001
Rūsys	0,015	0,006	0,024	0,002
Nauja statyba	-0,010	-0,019	0,000	0,048

Paskutiniame – 0,95 kvantilio – kvantilinės regresijos modelyje matome šiek tiek žymesnius įverčių pokyčius: laisvojo nario įvertis tapo teigiamas, sumažėjo *rinkos vertės* (X_1) kintamojo įtaka, o *naujos statybos* (X_{10}) butai pirmą kartą turi neigiamą įtaką vieno kvadratinio metro kainai (-0,010). Nors visi kintamieji yra reikšmingi, tačiau galime pastebėti, jog sumažėjo kintamųjų *rūsys* (X_9) 0,0016 ir *nauja statyba* (X_{10}) 0,0475 reikšmingumas.



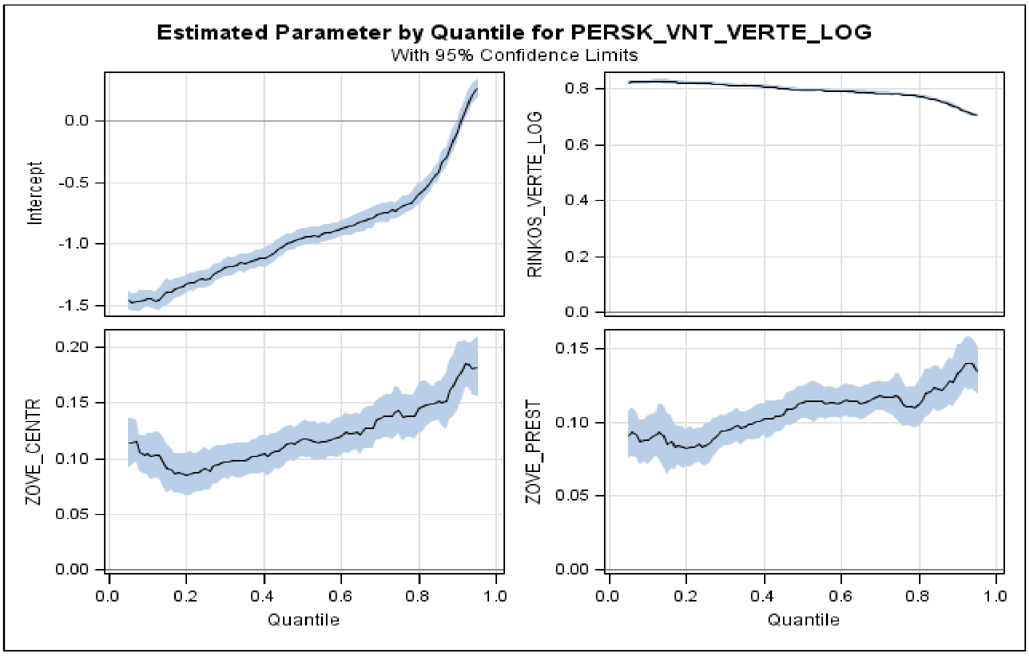
3.8 pav. 2008–2012 m. duomenų prognozavimas 0,95 kvantilinės regresijos modeliu

0,95 kvantilinės regresijos modelis prognozuoja didesnes vieno kvadratinio metro kainas negu jos yra iš tikrųjų. Šį kainų modelį būtų galima taikyti praktikoje, jeigu pirkėjas ruošiasi pirkti brangesnį negu vidutinis butą arba nori nusistatyti galimą brangesnę vieno kvadratinio metro kainą.

3.3.6. KVANTILINĖS REGRESIJOS APIBENDRINIMAS

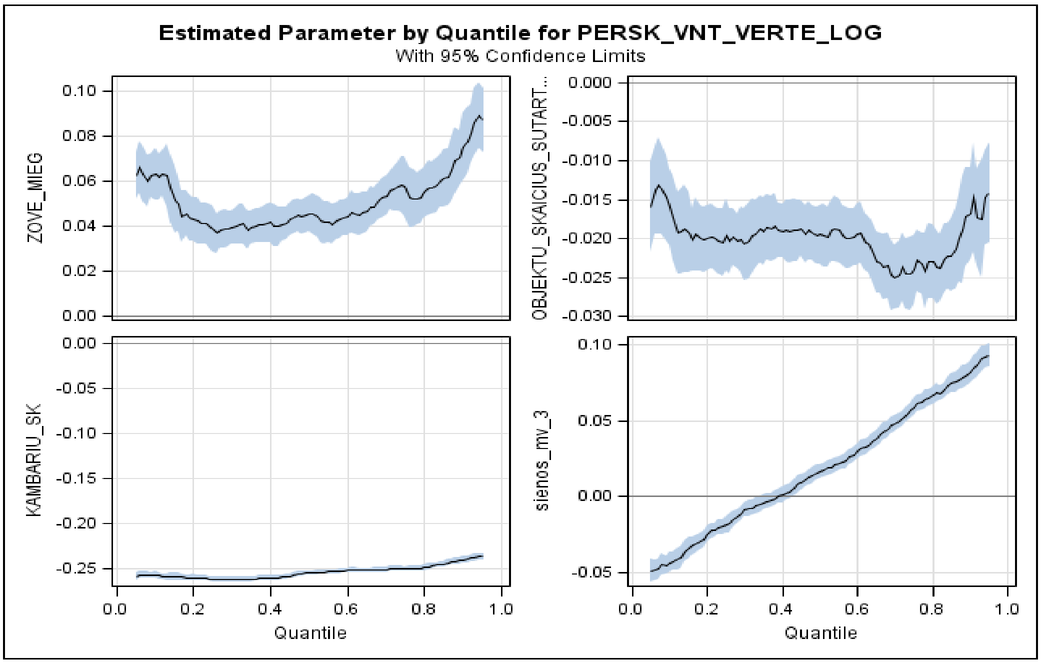
Sudarius kvantilinės regresijos lygtis galime pastebėti, jog nuo pasirinkto kvantilio priklauso ir vieno kvadratinio metro kainos prognozavimas: mažesnių kvantilių modeliai prognozuoja žemesnes kainas ir atvirkščiai. Tačiau, žinant šią kvantilinės regresijos modelių savybę, ją galima puikiai išnaudoti. Jeigu pirkėjas planuoja pirkti butą centre ar prestižiniame rajone, jis iškarto gali pasirinkti 0,75 ar 0,95 kvantilių regresijos modelius, nes tokių butų kainos būna didesnės už vidutines. Taigi skirtingos kvantilinės lygtys gali padėti įvertinti galimus kainos kitimo intervalus.

Apžvelkime modelių kintamųjų įverčių kitimo tendencijas su 95% pasikliautinaisiais intervalais kvantilinės regresijos grafikuose.



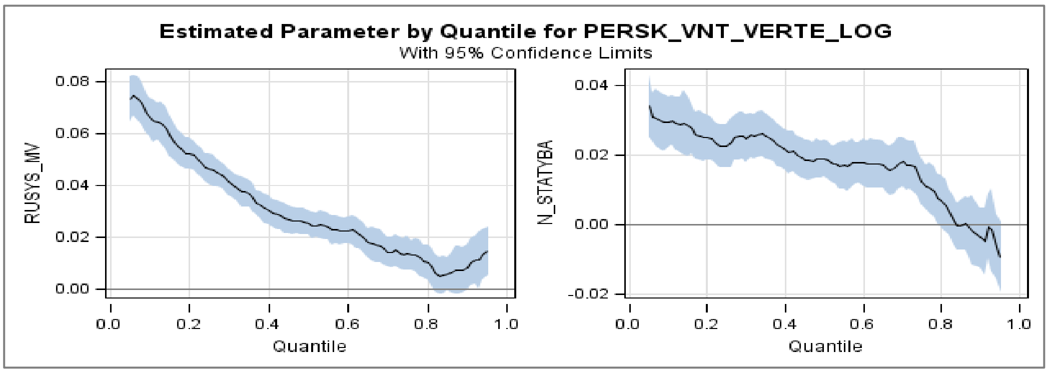
3.9 pav. Kintamųjų įverčių kitimas skirtinguose kvantiliuose

3.9 paveiksle juoda kreivė žymi kiekvieno kintamojo koeficientų įverčių kitimą esant skirtingiems kvantiliams (nuo 0,05 iki 0,95 kvantilio). Mėlyna spalva pažymėti 95 procentų pasikliautinieji intervalai. Laisvojo nario parametras (β_0) yra neigiamas ir tik nuo 0,9 kvantilio įgyja teigiamas reikšmes. Kintamojo *rinkos vertė* (X_1) įgyjamos reikšmės yra stabilios su labai mažais pasikliautinaisiais intervalais. Galima pastebėti, jog nuo 0,8 kvantilio kreivė gan stipriai pradeda leistis, tačiau tam įtakos gali turėti kitų kintamųjų kreivių kilimas paskutiniuosiuose kvantiliuose. *Centro* (X_2) kreivė visuomet turi teigiamą įtaką buto vieno kvadratinio metro kainai, tačiau ties 0,2 jos įtaka stipriai sumažėja ir po to stabiliai kyla. Šį sumažėjimą gali lemti mažas sandorių patekimas į šių kvantilių modelius. *Prestižiniuose rajonuose* (X_3) taip pat ties 0,2 kvantiliu pastebimas sumažėjimas, kitą sumažėjimą galime įžvelgti ties 0,8 kvantiliu. Palyginus galime pastebėti, jog *centro* (X_2) įtaka kainai labiau svyruoja negu *prestižinių rajonų* (X_3).



3.10 pav. Kintamųjų įverčių kitimas skirtinguose kvantiliuose (tęsinys)

3.10 paveiksle matome kitų keturių kintamųjų kvantilių grafikus. *Miegamųjų rajonų* (X_4) kreivė pasižymi aukštesne įtaka vieno kvadratinio metro kainai esant žemiems kvantiliams ir staigiu kainos kilimu nuo 0,8 kvantilio. Nuo 0,2 iki 0,8 kvantilio kintamojo įvertis didėja. *Miegamųjų rajonų* (X_4) kreivė, palyginus su kitų zonų kreivėmis, turi staigiausius įverčio kitimus pradžioje ir pabaigoje. Tam įtaką gali turėti šių rajonų butų imtis, kurioje nėra labai daug išskirtinių butų. Kintamasis *objektų skaičius* (X_6) visais atvejais turi nedidelę neigiamą įtaką butų kainai. Šio kintamojo pasikliautinieji intervalai rodo didžiausius svyravimus. *Kambarių skaičius* (X_7) įgyja neigiamus įverčius su labai mažais pasikliautinaisiais intervalais. Nuo 0,8 kvantilio neigiama *kambarių skaičiaus* (X_7) įtaka buto vieno kvadratinio metro kainai sumažėja. Neigiamas įverčio reikšmė atspindi butų paklausą – brangiausi ir paklausiausi yra 1–2 kambarių butai. Tuo tarpu 4–5 kambarių ar didesniems butams yra daug sunkiau surasti pirkėją. *Sienų* (X_8) įtaka butų vieno kvadratinio metro kainai stipriai svyruoja – įvertis iki 0,4 kvantilio įgyja neigiamas reikšmes, o po to teigiamas. Tokiems dideliems kintamojo įverčių svyravimams įtakos gali turėti tai, jog plytų/blokelių sienas turintys butai nėra homogeniški.



3.11 pav. Rūsio ir naujos statybos kintamųjų įverčių kitimas skirtinguose kvantiliuose

Rūsio (X_9) įsigijimas perkant butą lemia didesnę vieno kvadratinio metro kainą. Didžiausia įtaka yra pastebima mažiausiuose kvantiliuose ir po to ji mažėja. *Naujos statybos* (X_{10}) kintamasis teigiamą įtaką buto kainai turi iki 0,8 kvantilio, vėliau ji tampa neigiama. *Nauja statyba* (X_{10}) gali įgyti neigiamas reikšmes, kadangi centre ir prestižiniuose rajonuose daugiausiai yra esamos statybos namų, o jų kainos – vienos didžiausių.

3.4. NEURONINIŲ TINKLŲ MODELIS

Tiesinės ir kvantilinės regresijų prognozių palyginimui pasirinkome sudaryti dirbtinius neuroninius tinklus. Neuroniniai tinklai buvo sudaryti naudojant daugiasluoksnio perceptrono metodą. Sudarant neuroninius tinklus duomenų imtis padalinama į tris dalis: apmokymo, testavimo ir naudojimo. Apmokymo imtis visuomet sudaro didžiausią dalį. Tyrimo duomenys buvo padalinti taip: apmokymo rėžimui priskirta 80 proc. 2008–2012 metų pirkimo-pardavimo sandorių, tuomet testavimo rėžimui – likę 20 proc. sandorių. Naudojimo rėžimui, kurio metu jau yra išbandomas sudarytas neuroninis tinklas, priskirti 2 679 pirkimo-pardavimo sandoriai iš 2013 metų. Naudojimo rėžimas sudaro 10 proc. bendros sandorių imties, todėl visų 2013 metais įvykusių prikimo-pardavimo sandorių jam priskirti nepavyko. Neuroninius tinklus sudarėme tokiu pačiu principu kaip ir regresijos modelius – neuroniniai tinklai apmokomi su 2008–2012 m. imtimi, o naudojami 2013 m. duomenims. Taigi iš viso buvo ištirta 34,68 proc. 2013 metų imties.

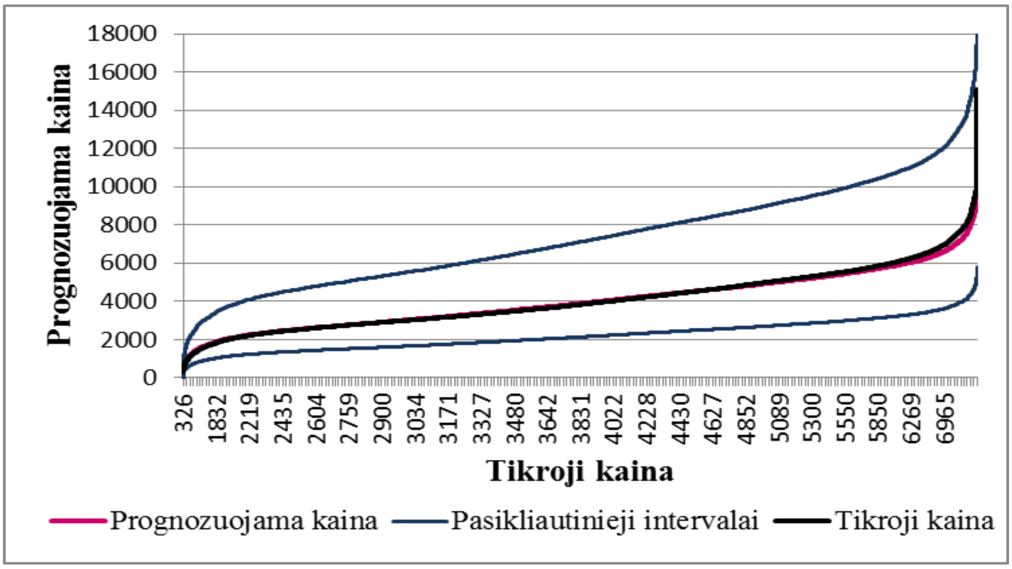
Geriausiai prognozavo daugiasluoksnio perceptrono metodu sudarytas neuroninis tinklas, kuris įvedimo sluoksnyje turėjo 10 kintamųjų (nepriklausomi kintamieji naudojami ir regresijos lygtyse), paslėptų neuronų – 8, išvedimo sluoksnyje 1 kintamąjį (prognozuojama vieno kvadratinio metro kaina). Sutrumpintai šis neuroninis tinklas užsirašo MLP 10-8-1. Paslėptų neuronų skaičius yra mažesnis už įvedimo ir išvedimo sluoksnių kintamųjų skaičių (8<11), tad galime teigti, jog neuroninis tinklas yra sudarytas teisingai. Tiksliausio neuroninio tinklo sudarymui paslėptų neuronų sluoksnyje buvo pasirinkta tangento funkcija, o išvedimo sluoksnyje – logistinė funkcija. Nuokrypių kvadratų suma SS_e buvo naudojama kaip klaidos funkcija.

3.9 lentelė

Prognozuojamų bei tikrųjų kainų tarpusavio koreliacijų koeficientai

Imtys:	Apmokymo	Testavimo	Naudojimo
MLP 10-8-1	0,952	0,949	0,860

Pateiktoje lentelėje matome prognozuojamų ir tikrųjų kainų tarpusavio koreliacijas. Geriausiai koreliuoja apmokymo rėžimo kainos, o mažiausiai – naudojimo. Tačiau visi koreliacijos koeficientai yra stiprūs. 3.12 pav. pavaizduota, kaip neuroniniai tinklai prognozuoja 2008–2012 metų kainas su pasikliautinaisiais intervalais. Neuroninių tinklų prognozė yra tiksli, daugiau skirtumų nuo tikrosios kainos atsiranda didžiausiose vieno kvadratinio metro kainose.



3.12 pav. 2008–2012 m. duomenų prognozavimas neuroniniais tinklais

D Priede yra pateikiamas prognozuojamų ir tikrųjų kainų grafikas ir kita detali analizė, susijusi su neuroniniais tinklais.

3.5. METODŲ PALYGINIMAS

Šiame skyrelyje palyginsime sudarytų tiesinės ir kvantilinės regresijos metodų bei neuroninių tinklų vieno kvadratinio metro kainos prognozes 2008–2012 ir 2013 metais užregistruotiems pirkimo-pardavimo sandoriams. Metodų tikslumą lyginsime pagal penkis kriterijus: vidutinę absoliutinę procentinę paklaidą (MAPE), vidutinę procentinę paklaidą (MPE), vidutinę kvadratinę paklaidą (MSE), vidutinę absoliutinę paklaidą (MAE) ir koreguotą determinacijos koeficientą R^2_{adj}.

3.10 lentelė

Prognozių tikslumas modelio sudarymui naudotiems 2008–2012 m. pirkimo-pardavimo sandoriams

	MAPE,%	MSE	MPE,%	MAE	R^2_{adj},%
Neuroniniai tinklai	**1,15**	0,02	**0,02**	**0,09**	89,06
Tiesinė regresija	1,25	0,02	0,03	0,10	**89,28**
0,05 kvantilinė regresija	2,50	0,06	-2,43	0,21	63,35
0,25 kvantilinė regresija	1,48	0,03	-1,03	0,12	84,25
0,5 kvantilinė regresija	1,24	0,02	-0,08	0,10	89,16
0,75 kvantilinė regresija	1,54	0,02	1,00	0,13	85,05
0,95 kvantilinė regresija	2,96	0,07	2,89	0,24	54,90

3.4 lentelėje pateikiamos modelio sudarymui naudotų 2008–2012 metų pirkimo-pardavimo sandorių prognozavimo paklaidos. Matome, jog pagal daugumą kriterijų geriausius rezultatus gauname naudojant neuroninius tinklus, kurių $MAPE = 1{,}15\%$, o koreguotas determinacijos koeficientas $R^2_{adj} = 89{,}06\%$. Taip pat galėtume išskirti tiesinės ir medianos regresijų modelius, kurių vidutinės absoliutinės paklaidos lygios $1{,}25\%$ ir $1{,}24\%$, o koreguoti determinacijos koeficientai $R^2_{adj} = 89{,}28\%$ ir $R^2_{adj} = 89{,}16\%$. Daugiausiai netikslumų yra gaunama naudojant 0,05 ir 0,95 kvantilių regresijos modelius.

3.11 lentelė

Prognozių tikslumas 2013 m. pirkimo-pardavimo sandoriams

	MAPE,%	MSE	MPE,%	MAE	R^2_{adj},%
Neuroniniai tinklai[4]	2,04	0,06	-0,49	0,16	68,12
Tiesinė regresija	1,80	**0,05**	0,23	0,14	**72,61**
0,05 kvantilinė regresija	2,75	0,09	-2,36	0,22	48,50
0,25 kvantilinė regresija	1,91	0,06	-0,97	0,15	66,64
0,5 kvantilinė regresija	**1,78**	**0,05**	**0,07**	**0,14**	71,53
0,75 kvantilinė regresija	2,17	0,06	1,20	0,17	66,26
0,95 kvantilinė regresija	3,60	0,11	3,27	0,29	34,57

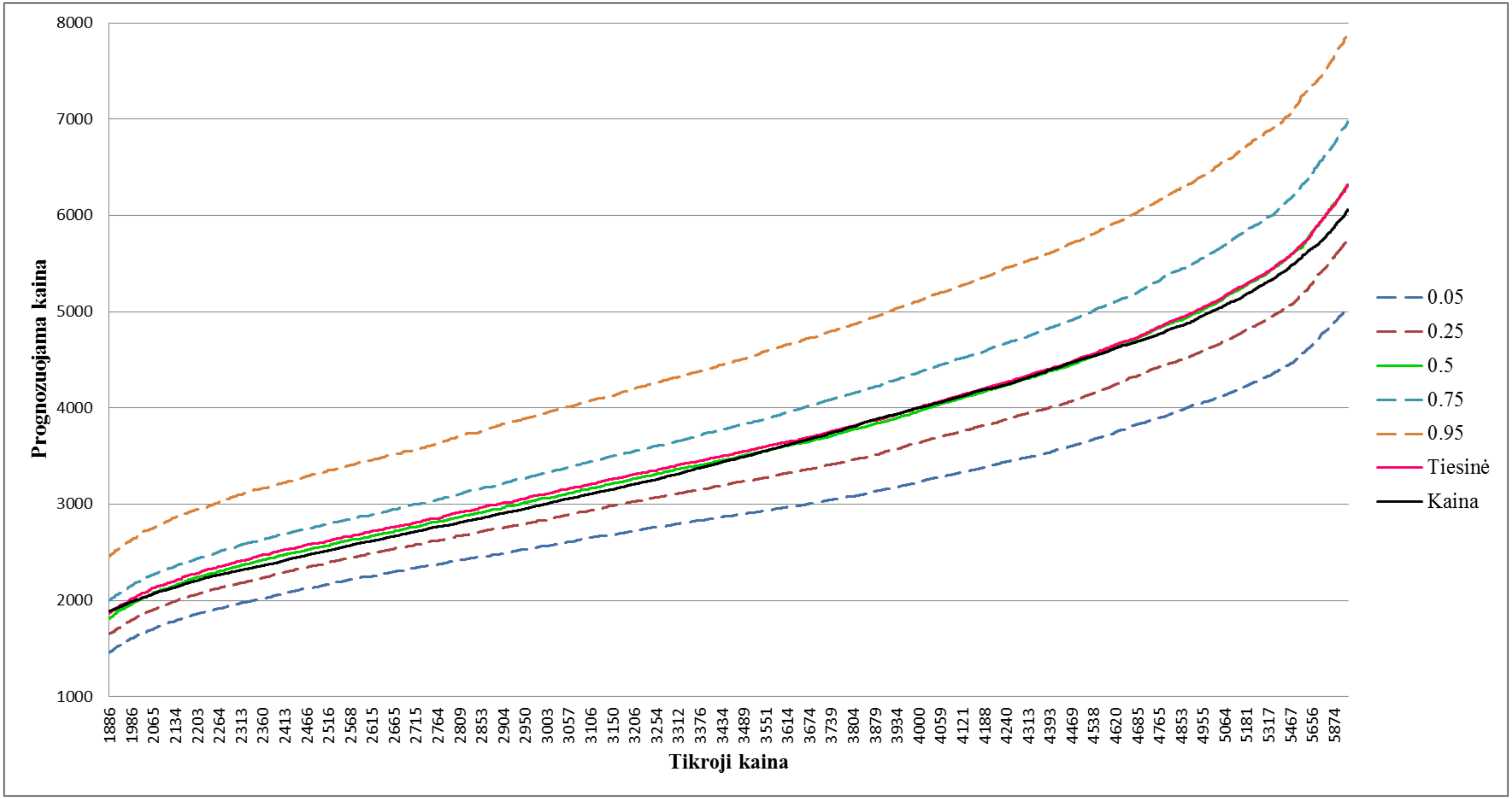
Kiek kitokius rezultatus gauname išanalizavę prognozavimo tikslumą 2013 metų pirkimo-pardavimo sandoriams. Priminsime, jog šie sandoriai nebuvo įtraukti į modelių sudarymą. Šiuo atveju pagal daugelį kriterijų tiksliausia prognozė gaunama naudojant medianos kvantilinės

[4] Neuroniniais tinklais buvo prognozuojami 2 679 pirkimo-pardavimo sandoriai (34,68% nuo bendros 2013 metų imties).

regresijos modelį $(MAPE = 1{,}78\%,\ R^2_{adj} = 71{,}53\%)$. Antroje vietoje lieka tiesinės regresijos modelis, kurio $MAPE = 1{,}80\%$, o koreguotas determinacijos koeficientas $R^2_{adj} = 72{,}61\%$. Palyginus šiuos du modelius gauname, jog medianos kvantilinės regresijos modelio MAPE yra tik $0{,}02\%$ tikslesnė. Kaip ir pirmu atveju, netiksliausias prognozes pateikia 0,95 kvantilinės regresijos modelis, kurio $MAPE = 3{,}60\%$, o koreguotas determinacijos koeficientas yra tik $R^2_{adj} = 34{,}57\%$. Galime pastebėti, jog visų modelių paklaidos yra mažos ir tenkina keliamus reikalavimus: $MAPE < 10\%$ ir $MPE < 5\%$.

Kadangi medianos kvantilinės regresijos modelis tiksliausiai prognozuoja Vilniaus miesto butų vieno kvadratinio metro kainą, ištyrėme ir 0,4–0,6 kvantilių regresijos lygtis 0,025 žingsniu. Gauti rezultatai parodė, jog 0,525 kvantilio regresijos lygtis tiksliau prognozuoja modelio sudarymui naudotus 2008–2012 metų pirkimo-pardavimo sandorių vieno kvadratinio metro kainas: koreguotas determinacijos koeficientas $R^2_{adj} = 89{,}19\%$, o vidutinė absoliutinė procentinė paklaida $MAPE = 1{,}24\%$ yra tokia pati kaip ir medianos kvantilinės regresijos lygties. Palyginus 2013 metų pirkimo-pardavimo sandorių prognozavimą, 0,525 kvantilio lygtis su $MAPE = 1{,}80\%$ ir $R^2_{adj} = 71{,}47\%$ yra ne tokia tiksli kaip medianos kvantilinės regresijos lygtis, kurios $MAPE = 1{,}78\%$ ir $R^2_{adj} = 71{,}53\%$. Taigi įsitikinome, jog medianos kvantilinės regresijos modelis iš tiesų tiksliausiai prognozuoja Vilniaus miesto butų vieno kvadratinio metro kainas.

Verta pastebėti, jog 0,05 ir 0,95 kvantilių lygtis patartina naudoti tuo atveju, jeigu planuojama pirkti atitinkamai prastos kokybės arba prabangų butą. Prognozuojant 2013 metų 10 procentų žemiausias kainas turinčius pirkimo-pardavimo sandorius su 0,05 kvantilinės regresijos lygtimi, koreguotas determinacijos koeficientas $R^2_{adj} = 76{,}07\%$. Jeigu su 0,95 kvantilio lygtimi prognozuosime 10 procentų brangiausių butų, koreguotas determinacijos koeficientas yra lygus $R^2_{adj} = 81{,}27\%$. Šie rezultatai rodo, jog žemesnių ir aukštesnių kvantilių regresijos lygtys turėtų būti naudojamos tuomet, kai yra ieškoma išskirtinio kainos atžvilgiu buto.



3.13 pav. 90 proc. 2013 metų pirkimo-pardavimo sandorių prognozavimas tiesinės ir kvantilinės regresijos lygtimis

Šiame grafike (3.13 pav.) matome, jog arčiausiai tikrosios vieno kvadratinio metro kainos kreivės, kuri pažymėta juoda spalva, yra tiesinės ir medianos regresijos kreivės. Kiek didesnis skirtumas tarp prognozių ir tikrosios kainos išryškėja pasiekus 4 500 Lt kainą už vieną kvadratinį metrą. Tuomet tiek medianos, tiek tiesinės regresijos modeliai prognozuoja didesnes kainas. 0,05 ir 0,25 kvantilinės regresijos modeliai prognozuoja žemesnes kainas, negu jos yra iš tikrųjų, o 0,75 ir 0,95 atitinkamai didesnes. Kol vieno kvadratinio metro kaina yra gan maža, prognozės yra tikslesnės, tačiau didėjant kainai didėja ir atotrūkis tarp tikrosios ir prognozuojamos kainos.

DISKUSIJA

Vertinant tiesinės, kvantilinės regresijų ir neuroninių tinklų modelių naudojimo patogumą, regresijos modelius išskirčiau kaip patogesnius, kadangi juos sudarius yra gaunama tam tikra išraiška. Sudarytuose modeliuose galime peržvelgti ir išanalizuoti kintamųjų įverčius, jų dydį bei ženklus, patikrinti, ar tokie rezultatai neprieštarauja loginėms prielaidoms. Taip pat, turint regresijos modelių lygtis, vieno kvadratinio metro kainas galima prognozuoti ir nenaudojant specialių matematinių programų. Kvantilinės regresijos modeliai yra pranašesni už tiesinės regresijos modelius, kadangi jie gali būti sudaromi nepriklausomai nuo liekanų normalumo bei homoskedastiškumo prielaidų tenkinimo. Atsižvelgiant, jog ekonomikoje ir kitose srityse duomenys dažnai nėra homoskedastiški, tai palengvina šio metodo taikymą. Taip pat, verta pabrėžti ir kvantilinės regresinės analizės lankstumą – tyrėjas gali sudaryti keletą lygčių pasirinktiems kvantiliams ir išsirinkti geriausią variantą. Neuroninių tinklų sudarymas dar yra pakankamai naujas, bet jau labai plačiai taikomas metodas. Šio metodo privalumas yra tas, jog neuroniniai tinklai greitai apmoko imtį ir pateikia tikslias prognozes. Tačiau tyrėjas turi labai kruopščiai atrinkti tyrimui pateikiamą imtį, kadangi dėl keleto klaidingų stebėjimų, gali būti gaunamos netikslios prognozės.

Naudojant turimus pirkimo-pardavimo sandorių duomenis būtų galima kartoti J. M. Quigley atliktą „Nekilnojamojo turo kainų ir ekonominių ciklų“ analizę [21] ir bandyti prognozuoti Vilniaus miesto butų kainų ekstremumus naudojant skirtingus kvantilinės regresinės analizės modelius. Taip pat, Vilniaus miesto butų pirkimo-pardavimo sandorių duomenis galima suskirstyti į keturias grupes kiekvienai zonai: centrui, prestižiniams, miegamiesiems ir kitiems rajonams. Tuomet sudaryti keturias skirtingas lygtis, kuriose tiriami sandoriai būtų homogeniškesni bei galbūt tenkintų tiesinės regresijos modelių reikalavimus.

Remiantis A. C. Comrie atlikto tyrimo patirtimi, kurio metu ozono sluoksnio storiui virš aštuonių JAV didmiesčių ištirti buvo sudaryti atskiri neuroninių tinklų modeliai [5], sudarytą Vilniaus miesto butų prognozavimo modelį būtų galima lyginti su Kauno, Klaipėdos ar kitų Baltijos šalių sostinių (Rygos ir Talino) modeliais, analizuoti susidariusius skirtumus bei esamas tendencijas.

IŠVADOS

1. Pašalinus iš duomenų klaidas ir netinkamai įvestus duomenis (pvz. 40 m^2 butas su 12 kambarių) buvo suformuota tiriamų duomenų imtis iš 26 403 ir rezultatų testavimo imtis iš 7 724 pirkimo-pardavimo sandorių.
2. Ištyrus požymių išsibarstymo grafikus nustatyta, jog geriausia tiesinė duomenų priklausomybė atsiranda juos logaritmavus. Taigi buvo atlikta transformacija, kurios metu kintamieji *kaina* ir *rinkos vertė* buvo logaritmuoti. Šių kintamųjų Pirsono koreliacijos koeficientas pakilo iki 0,81.
3. Naudojant tiesinės regresijos metodą paaiškėjo, jog sudaryto modelio tikslumui daugiausiai įtakos turi kintamieji *rinkos vertė* su daliniu determinacijos koeficientu $part. R^2 = 65{,}77$ % ir *kambarių skaičius*, kurio $part. R^2 = 22{,}87$ %.
4. Sudarytas tiesinės regresijos modelis $\ln(Y) = -0{,}683 + 0{,}770 \cdot \ln(X_1) + 0{,}141 \cdot X_2 + 0{,}118 \cdot X_3 + 0{,}059 \cdot X_4 - 0{,}018 \cdot X_6 - 0{,}250 \cdot X_7 + 0{,}027 \cdot X_8 + 0{,}030 \cdot X_9 + 0{,}021 \cdot X_{10}$ yra tikslus ir paaiškina 89,28% visų tiriamų reikšmių. Tačiau skaitiniai testai nepatvirtino liekanų normalumo ir homoskedastiškumo prielaidų.
5. Sudaryti penki skirtingi kvantilinės regresijos modeliai (0,05; 0,25; 0,5; 0,75 ir 0,95). Tiksliausias prognozes gavome su medianos (0,5) kvantilinės regresijos modeliu, kurio vidutinė absoliutinė procentinė paklaida modelio sudarymui naudotiems 2008–2012 metų duomenims $MAPE = 1{,}24\%$, o koreguotas determinacijos koeficientas $R^2_{adj} = 89{,}16\%$.
6. Tarpusavyje palyginus visus sudarytus Vilniaus miesto butų vieno kvadratinio metro kainos prognozavimo modelius nustatyta, jog tiksliausiai prognozuoja medianos (0,5) kvantilinės regresijos modelis, kurio vidutinė absoliutinė procentinė paklaida ir koreguotas determinacijos koeficientas 2013 metų pirkimo-pardavimo sandorių kainoms yra $MAPE = 1{,}78\%$ ir $R^2_{adj} = 71{,}53\%$.
7. Panaudojus daugiasluoksnio perceptrono neuroninių tinklų metodą, gautas neuroninis tinklas su aštuoniais paslėptais neuronais: MLP 10-8-1. Sudaryto modelio 2013 metų duomenų prognozių tikslumas ($MAPE = 2{,}04\%$ ir $R^2_{adj} = 68{,}12\%$) daugeliu atžvilgių yra prastesnis už tiesinės ir kvantilinės regresijos modelius.
8. Prognozuojant 2013 metų 10 procentų žemiausias kainas turinčius pirkimo-pardavimo sandorius su 0,05 kvantilinės regresijos lygtimi, koreguotas determinacijos koeficientas $R^2_{adj} = 76{,}07\%$. Naudojant 0,95 kvantilio lygtį 10 procentų brangiausių butų prognozavimui, koreguotas determinacijos koeficientas yra lygus $R^2_{adj} = 81{,}27\%$.

REKOMENDACIJOS

Sudarytas Vilniaus miesto butų vieno kvadratinio metro kainos medianos kvantilinės regresijos prognozavimo modelis galėtų būti pritaikytas Lietuvos statistikos departamento Būsto kainų indekso (BKI) skaičiavimui. Šio metodo pagalba įtraukiamos į analizę pirkimo-pardavimo sandorių kainos taptų homogeniškesnės. Sudarytus modelius patartina atnaujinti kiekvienais metais, papildant duomenis naujais metais ir pašalinant seniausius. Tokiu būdu prognozavimo modeliai atspindėtų vyraujančias tendencijas. Taip, pat kiekvieną kartą atnaujinant modelius patartina peržiūrėti ir dar kartą apsvarstyti kintamųjų reikšmingumą ir įtaką modeliui. Vilniaus miestas nuolat plečiasi, jame kiekvienais metais pabaigiama nemažai naujų statybos projektų, tad Vilniaus miesto savivaldybės zonų peržiūrėjimas taip pat yra stipriai rekomenduotinas.

LITERATŪRA

1. Ahangar R. G., Yahyazadehfar M., Pournaghshband H., „The Comparison of Methods Artificial Neural Network with Linear Regression Using Specific Variables for Prediction Stock Price in Tehran Stock Exchange“, International Journal of Computer Science and Information Security, Vol. 7, No. 2, 28-46 pages, 2010.
2. Ball M. „Recent Empirical Work on the Determinants of Relative House Prices“, Urban Studies vol.10, 213-233 psl., 1973.
3. Bdran S., Abouelatta O., „Neural Network Integrated with Regression Methods to Forecast Electrical Load“, International Conference on Electrical, Electronics and Biomedical Engineering (ICEEBE'2012) Penang (Malaysia) May 19-20, 2012.
4. Benoit D. F., van der Poel, D., „Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services“, Expert systems with Applications (36), 2009.
5. Comrie A. C., „Comparing Neural Networks and Regression Models for Ozone Forecasting“, Journal of the Air & Waiste Management Assocciation 47:653-663 pages, 1997.
6. Conference ECON 370 – Urban and Regional Economics, „Violations of OLS Assumptions: Heteroscedasticity“, Canada, September of 2013.
7. Europos Sąjungos statistikos tarnybos (Eurostat) būsto statistika http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Housing_statistics
8. Eurostat „Detailed Technical manual on Owner-Occupied Housing for Harmonised Index of Consumer Prices“, March 2012, v.2.0., pages 48-57.
9. Gunther F., Fritsch S., „neuralnet: Training of Neural Networks“, The R Journal, Vol.2/1, pages 30-38, June 2010.
10. Hand R., „Forecasting Real Estate Values: Appraisals are Old News, Regression is the New Black“, „Re/max“ Commercial Brokers, Inc., 2011.
11. Yusof, A. M., Ismail, S. „Multiple Regression in Analysing House Price Variations“ IBIMA Publishing Vol. 2012.
12. Jalali N., Babanezhad M., „Quantile Regression due to Skewness and Outliers“, Applied Mathematical Sciences Vol.5, no.39, pages 1947-1951, 2011.
13. Kan K., Tsai W., „Obesity and risk knowledge“, Journal of Health Economics, 2004, 23:907-934 pages.
14. Koenker R., „Short Course on Quantile Regression“, University of Illinois at Urbana-Champaign, 2003.

15. Koenker R., Hallock K. F., „Quantile Regression“, Journal of Economic Perspectives, Vol.15(4), pages 143-156, 2001 Fall.
16. Kovač Š., Želinsky T., „Determinants of the Slovak Enterprise Profitability:Quantile Regression Approach“, Statistica, 2013, 93(3).
17. Li G., Xu S., Li Z., Sun Y., Dong X., „Using Quantile Regression Approach to Analyze Price Movements of Agricultural Products in China“ Journal of Integrative Agriculture 2012, 11(4): 674-683.
18. Lietuvos statistikos departamentas „Būsto kainų indekso sudarymo metodika“, patvirtinta 2012 m. gruodžio 31 d.
19. Mosteller F., Tukey J., „Data Analysis and Regression: A Second Course in Statistics“, 1977.
20. Oficialiosios statistikos portalas http://www.osp.stat.gov.lt/, [žiūrėta 2014.05.03].
21. Quigley J. M., „Real Estate Prices and Economic Cycles“, International Real Estate Review 1999, Vol 2., No 1: pages 1-20.
22. SAS 9.2 Proc Quantreg aprašas, https://support.sas.com/documentation/cdl/en/statugquantreg/, [žiūrėta 2014.05.07].
23. Seber G. A. F., Lee A. J., „Linear Regression Analysis“ Second edition, Chapter 6, pages 89-100, 2012.
24. Sergiu C. „Financial Predictor via Neural Network“ 2011 http://www.codeproject.com/Articles/175777/Financial-predictor-via-neural-network [žiūrėta 2014.05.07].
25. Stergiou C., Siganos D., „Neural Networks“, Imperial College of Science Technology and Medicine http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html [žiūrėta 2014.05.03].
26. „Swedbank“ Asmeninių finansų instituto atliktas tyrimas „Būstas kaip darnių bendruomenių vystymo pagrindas“.
27. „Swedbank“ Būsto įperkamumo indeksas Baltijos šalims http://www.swedbank.lt/lt/previews/privatiems/4/71
28. UAB „Ober-Haus“ Nekilnojamo turto rinkos tyrimai http://www.ober-haus.lt/naujienos/nekilnojamojo-turto-rinkos-tyrimai ir http://www.ober-haus.lt/files/lt/files/apzvalgos/OHBI_apzvalga_2014_kovas.pdf, [žiūrėta 2014.05.01].
29. VĮ Registrų centras skelbiami Būsto kainų pokyčiai ir skaičiavimo metodika http://www.registrucentras.lt/ntr/stat/busto_kainos.php?regionId=1, [žiūrėta 2014.05.01]

PRIEDAI

PRIEDAS A

Šiame priede pateikiame modelių prognozavimui naudotos SAS programos tekstą.

```
/*Duomenų nuskaitymas*/
PROC IMPORT OUT= WORK.VLN_BUTAI
            DATAFILE= "E:\Lauros\Magistrinis\Vilniaus miesto butai\2Vilniaus m._butai_2008-2012.xls"
            DBMS=EXCEL REPLACE;
     SHEET="'T_2008_2012_N$'";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

%MACRO pernumer(duomen= ,kint= );
  proc sort data=&duomen;by &kint;run;
  data nr;set &duomen;run; proc sort nodupkey;by &kint;run;
  data nr;set nr;nr=_N_;keep &kint nr;run;
  data &duomen;merge &duomen nr;by &kint;&kint=nr;drop nr;run;
%MEND;
/*Duomenų perkodavimas*/
DATA isrinktiDuomenys;
SET WORK.VLN_BUTAI;
IF OBJEKTU_SKAICIUS_SUTARTYJE>=3 THEN OBJEKTU_SKAICIUS_SUTARTYJE=3;
IF SIENOS_MV in(6,11) THEN SIENOS_MV=50;
IF SIENOS_MV in(4,20) THEN SIENOS_MV=51;
IF SIENOS_MV in(1,8) THEN SIENOS_MV=52;
IF SIENOS_MV in(2,3,18) THEN SIENOS_MV=53;
IF ISIG_TIPAS=11 THEN ISIG_TIPAS=1;
IF ISIG_TIPAS in(20,22) THEN ISIG_TIPAS=2;
IF ISIG_TIPAS in(12,21) THEN ISIG_TIPAS=3;
IF ISIG_TIPAS=10 THEN ISIG_TIPAS=4;
IF VANDENT_MV=2 THEN VANDENT_MV=1;
IF SILDYMAS_MV in (1,2,3) THEN SILDYMAS_MV=1;
IF DUJOS_MV=2 THEN DUJOS_MV=1;
IF GYV_ID not in(1) THEN GYV_ID=0;
IF ZOVE_NR IN ("57.1" "57.2" "57.3" "57.4" "57.5" "57.6" "57.7" "57.8")THEN DO;
      ZOVE_CENTR = 1;
      ZOVE_PREST = 0;
      ZOVE_MIEG = 0;
      ZOVE_KITI = 0;
END;
ELSE IF ZOVE_NR IN ("57.20" "57.64" "57.65" "57.66" "57.69" "57.11" "57.14" "57.15" "57.16" "57.18" "57.68")THEN DO;
      ZOVE_CENTR = 0;
      ZOVE_PREST = 1;
      ZOVE_MIEG = 0;
      ZOVE_KITI = 0;
END;
ELSE IF ZOVE_NR IN ("57.10" "57.12" "57.13" "57.17" "57.19" "57.36" "57.37" "57.38" "57.39" "57.40" "57.41" "57.42"
"57.43" "57.44" "57.45" "57.46" "57.47" "57.48" "57.49" "57.50" "57.51" "57.52" "57.53" "57.54" "57.55" "57.56"
"57.57" "57.58" "57.59" "5767" "57.70" "57.72" "57.73" "57.77" "57.78" "57.9")THEN DO;
```

```
            ZOVE_CENTR = 0;
            ZOVE_PREST = 0;
            ZOVE_MIEG = 1;
            ZOVE_KITI = 0;
     END;
     ELSE IF ZOVE_NR IN ("57.21" "57.22" "57.23" "57.24" "57.25" "57.26" "57.27"
"57.28" "57.29" "57.30" "57.31" "57.32"
     "57.33"  "57.34"  "57.35"  "57.60"  "57.61"  "57.62"  "57.63"  "57.71"  "57.74"
"57.75" "57.76" "57.80" "57.81")THEN DO;
            ZOVE_CENTR = 0;
            ZOVE_PREST = 0;
            ZOVE_MIEG = 0;
            ZOVE_KITI = 1;
     END;
     PERSK_VNT_VERTE_LOG=LOG(Perskaiciuota_vnt_verte_su_zeme);
     RINKOS_VERTE_LOG=LOG(RINKOS_PASKAICIUOTA_VERTE);
     KEEP Perskaiciuota_vnt_verte_su_zeme PERSK_VNT_VERTE_LOG  N_STATYBA METAI
MENUO SAN_ID DALIS PROCENTAI PASKAICIUOTA_SUMA SAKT_TIPAS ZOVE_CENTR ZOVE_PREST
ZOVE_MIEG  ZOVE_KITI  GYV_ID  OBJEKTU_SKAICIUS_SUTARTYJE  ISIG_PLOTAS  SIENOS_MV
SILDYMAS_MV VANDENT_MV DUJOS_MV RUSYS_MV BUTU_SK KAMBARIU_SK AUKSTAS AUKSTU_SK
AUKST_TIKR  BAIGTUMAS  STATYBOS_PABAIGA1  ISIG_TIPAS  ZEMES_ISIG_PLOTAS  RINK_VERT
RINKOS_VERTE RINKOS_PASKAICIUOTA_VERTE RINKOS_VERTE_LOG SAK_SUMA;
     run;
     /*Duomenų lentelės sudarymas*/
     PROC SQL;
     CREATE TABLE Duomenys
     AS  SELECT  Perskaiciuota_vnt_verte_su_zeme,  SAN_ID,  DALIS,  PROCENTAI,
PASKAICIUOTA_SUMA, SAKT_TIPAS,  ZOVE_CENTR, ZOVE_PREST, ZOVE_MIEG, ZOVE_KITI,
GYV_ID,  OBJEKTU_SKAICIUS_SUTARTYJE,  ISIG_PLOTAS,  SIENOS_MV,  SILDYMAS_MV,
VANDENT_MV,  DUJOS_MV,  RUSYS_MV,  BUTU_SK,  KAMBARIU_SK,  AUKSTAS,  AUKSTU_SK,
AUKST_TIKR,  BAIGTUMAS,  STATYBOS_PABAIGA1,ISIG_TIPAS,  ZEMES_ISIG_PLOTAS,  METAI,
N_STATYBA, RINK_VERT, RINKOS_VERTE, RINKOS_PASKAICIUOTA_VERTE, SAK_SUMA, MENUO,
RINKOS_VERTE_LOG, PERSK_VNT_VERTE_LOG
     FROM isrinktiDuomenys
     where BAIGTUMAS >= 51;
     QUIT;
     /*Pasikartojančių sandorių, jei tokių yra, suradimas*/
     proc sort data=duomenys nodupkey dupout=duplicates; by _all_; run;
     /*Kintamųjų dažnių lentelės*/
     proc freq data=Duomenys;
     tables
     ZOVE_CENTR ZOVE_PREST ZOVE_MIEG ZOVE_KITI GYV_ID OBJEKTU_SKAICIUS_SUTARTYJE
SIENOS_MV SILDYMAS_MV VANDENT_MV DUJOS_MV RUSYS_MV KAMBARIU_SK AUKSTAS AUKSTU_SK
AUKST_TIKR  BAIGTUMAS  STATYBOS_PABAIGA1  ISIG_TIPAS  METAI  DALIS  ISIG_PLOTAS
BUTU_SK;
      run;
     %pernumer(duomen=duomenys, kint=SIENOS_MV);
     %pernumer(duomen=duomenys, kint=ISIG_TIPAS);
     /*Kintamieji po pernumeravimo*/
     proc freq data=Duomenys;
     tables SIENOS_MV ISIG_TIPAS PATOGUMAI;
      run;
     /*Masyvų sukūrimas*/
     data Duomenys;
     set Duomenys;
     ARRAY sakt {*} sakt_tipas_1 - sakt_tipas_2;
       DO i=1 TO 2; sakt(i) = 0; END;
       sakt(sakt_tipas) = 1;
     ARRAY sienos {*} sienos_mv_1 - sienos_mv_4;
       DO i=1 TO 4; sienos(i) = 0; END;
       sienos(sienos_mv) = 1;
     ARRAY isig {*} isig_tipas_1 - isig_tipas_4;
       DO i=1 TO 4; isig(i) = 0; END;
```

```
        isig(isig_tipas) = 1;
     drop i;
     run;
     /*Koreliacinė analizė*/
     proc corr data=Duomenys out=corr;
     var  Perskaiciuota_vnt_verte_su_zeme  DALIS  PASKAICIUOTA_SUMA  SAKT_TIPAS
ZOVE_CENTR  ZOVE_PREST  ZOVE_MIEG  ZOVE_KITI  GYV_ID  OBJEKTU_SKAICIUS_SUTARTYJE
ISIG_PLOTAS  SIENOS_MV  SILDYMAS_MV  VANDENT_MV  DUJOS_MV  RUSYS_MV  BUTU_SK
KAMBARIU_SK AUKSTAS AUKSTU_SK AUKST_TIKR BAIGTUMAS STATYBOS_PABAIGA1 ISIG_TIPAS
ZEMES_ISIG_PLOTAS RINK_VERT RINKOS_VERTE RINKOS_PASKAICIUOTA_VERTE SAK_SUMA;
     run;
     /*Duomenų normalumo tikrinimas*/
     proc univariate data=duomenys noprint;
       var   Perskaiciuota_vnt_verte_su_zeme   PASKAICIUOTA_SUMA   ISIG_PLOTAS
SIENOS_MV  BUTU_SK  KAMBARIU_SK  AUKSTAS  AUKSTU_SK   BAIGTUMAS  STATYBOS_PABAIGA1
ZEMES_ISIG_PLOTAS;
     histogram / cfill=gray normal;
     run;
     /*Kintamųjų etikečių naikinimas*/
     proc datasets library=work nolist;
       modify duomenys;
       attrib _all_ label='';
     quit;

     data duomenys2;
     set duomenys;
     run;
     /*Regresinė analizė*/
     proc reg data=duomenys2;
     model    PERSK_VNT_VERTE_LOG=
     RINKOS_VERTE_LOG      ZOVE_CENTR      ZOVE_PREST      ZOVE_MIEG      ZOVE_KITI
OBJEKTU_SKAICIUS_SUTARTYJE RUSYS_MV BUTU_SK KAMBARIU_SK     AUKSTAS     AUKSTU_SK
BAIGTUMAS N_STATYBA STATYBOS_PABAIGA1 SIENOS_MV_3/ selection=stepwise clm collin
influence spec dw vif R pcorr2;
     output  out=paklaidos  p=yhat  residual=res  stdr=rstd  rstudent=rstu  h=lev
cookd=cd DFFITS=DF;
     plot r.*p.;
     plot r.*PERSK_VNT_VERTE_LOG;
     run;
     quit;
     /*Heteroskedastiškumo tyrimas*/
     proc model data=duomenys2;
     instruments  RINKOS_VERTE_LOG  ZOVE_CENTR  ZOVE_PREST  ZOVE_MIEG  ZOVE_KITI
OBJEKTU_SKAICIUS_SUTARTYJE  RUSYS_MV  BUTU_SK  KAMBARIU_SK  AUKSTAS  AUKSTU_SK
BAIGTUMAS N_STATYBA STATYBOS_PABAIGA1 SIENOS_MV_3;
     Perskaiciuota_vnt_verte_su_zeme=b0+b1*RINKOS_VERTE_LOG+b2*ZOVE_CENTR+b3*ZOV
E_PREST+b4*ZOVE_MIEG+b5*ZOVE_KITI+b6*OBJEKTU_SKAICIUS_SUTARTYJE+b7*RUSYS_MV+b8*B
UTU_SK+b9*KAMBARIU_SK+b10*AUKSTAS+b11*AUKSTU_SK+b12*BAIGTUMAS+b13*N_STATYBA+b14*
STATYBOS_PABAIGA1+b15*SIENOS_MV_3;
     fit  PERSK_VNT_VERTE_LOG  /  white  BREUSCH=(RINKOS_VERTE_LOG  ZOVE_CENTR
ZOVE_PREST  ZOVE_MIEG  ZOVE_KITI  OBJEKTU_SKAICIUS_SUTARTYJE  RUSYS_MV  BUTU_SK
KAMBARIU_SK AUKSTAS AUKSTU_SK BAIGTUMAS N_STATYBA STATYBOS_PABAIGA1 SIENOS_MV_3)
     out=resid outresid;
     run;
     quit;
     /*Liekanų grafikai*/
     proc gplot data=paklaidos;
           plot res*yhat;
           run;
     proc gplot data=paklaidos;
           plot res*PERSK_VNT_VERTE_LOG;
           run;
           quit;
```

```
goptions reset=all;
/*Liekanų normalumo testas*/
proc univariate data=paklaidos normal;
 var res;
 qqplot res / normal(mu=est sigma=est);
run;
/*Išskirčių šalinimas*/
data isskirtys;
set paklaidos;
if abs(r)>3 then r_out=1;
if abs(rstu)>2 then rstu_out=1;
if lev>0.001396 then lev_out=1;
if cd>0.000151 then cd_out=1;
if abs(DFF)>0.052844 then df_out=1;
if sum(of r_out rstu_out lev_out cd_out df_out)>=2 then outlier=1;
run;
data atrinkta;
set isskirtys;
if outlier=1 then delete;
run;
/*Koreliacinė analizė*/
proc corr data=atrinkta out=corr;
var  PERSK_VNT_VERTE_LOG  RINKOS_VERTE_LOG  ZOVE_CENTR  ZOVE_PREST  ZOVE_MIEG
ZOVE_KITI OBJEKTU_SKAICIUS_SUTARTYJE KAMBARIU_SK SIENOS_MV_3 RUSYS_MV N_STATYBA;
run;
/*Regresinė analizė*/
proc reg data=atrinkta;
model    PERSK_VNT_VERTE_LOG=
RINKOS_VERTE_LOG
ZOVE_CENTR
ZOVE_PREST
ZOVE_MIEG
ZOVE_KITI
OBJEKTU_SKAICIUS_SUTARTYJE
KAMBARIU_SK
SIENOS_MV_3
RUSYS_MV
N_STATYBA/vif R dw p;
output  out=paklaidos2  p=yhat2  residual=res2  stdr=rstd2  rstudent=rstu2
h=lev2 cookd=cd2 DFFITS=DF2;
plot r.*p.;
plot r.*PERSK_VNT_VERTE_LOG;
run;
quit;
/*Heteroskedastiškumo tyrimas*/
proc model data=atrinkta;
parms b0 b1 b2 b3 b4 b5 b6 b7 b8 b9;
PERSK_VNT_VERTE_LOG=b0+b1*RINKOS_VERTE_LOG+b2*ZOVE_CENTR+b3*ZOVE_PREST+b4*Z
OVE_MIEG+b5*OBJEKTU_SKAICIUS_SUTARTYJE+b6*KAMBARIU_SK+b7*SIENOS_MV_3+b8*RUSYS_MV
+b9*N_STATYBA;
fit   PERSK_VNT_VERTE_LOG   /   white   BREUSCH=(RINKOS_VERTE_LOG   ZOVE_CENTR
ZOVE_PREST ZOVE_MIEG OBJEKTU_SKAICIUS_SUTARTYJE KAMBARIU_SK SIENOS_MV_3 RUSYS_MV
N_STATYBA)
out=resid outresid;
run;
quit;
/*Liekanų grafikai*/
proc gplot data=paklaidos2;
      plot res2*yhat2;
      run;
proc gplot data=paklaidos2;
      plot res2*Perskaiciuota_vnt_verte_su_zeme;
      run;
```

```
          quit;
     proc gplot data=paklaidos2;
          plot res2*PERSK_VNT_VERTE_LOG;
          run;
          quit;
     goptions reset=all;
     /*Liekanų normalumo testas*/
     proc univariate data=paklaidos2 normal;
      var res2;
      qqplot res2 / normal(mu=est sigma=est);
     run;
     /*Kvantilinė regresija kiekvienam kvantiliui atskirai*/
     ods graphics on;
     proc quantreg data=atrinkta alpha=0.05 ci=resampling plots=(rdplot ddplot
reshistogram);
     model PERSK_VNT_VERTE_LOG=RINKOS_VERTE_LOG ZOVE_CENTR ZOVE_PREST ZOVE_MIEG
ZOVE_KITI OBJEKTU_SKAICIUS_SUTARTYJE KAMBARIU_SK SIENOS_MV_3 RUSYS_MV N_STATYBA/
quantile=0.95 CovB CorrB seed=12345;
     test   RINKOS_VERTE_LOG   ZOVE_CENTR   ZOVE_PREST   ZOVE_MIEG   ZOVE_KITI
OBJEKTU_SKAICIUS_SUTARTYJE KAMBARIU_SK SIENOS_MV_3 RUSYS_MV N_STATYBA/ wald lr;
     output out=kvantiline residual=res predicted=pred;
     run;
     /*Liekanų grafikai*/
     proc gplot data=kvantiline;
          plot res*pred;
          run;
          quit;
     proc gplot data=kvantiline;
          plot res*PERSK_VNT_VERTE_LOG;
          run;
          quit;
     /*Liekanų normalumo testas*/
     proc univariate data=kvantiline normal;
      var res;
      qqplot res / normal(mu=est sigma=est);
     run;
     ods graphics on;
     /*Kvantilinė regresija*/
     proc quantreg data=atrinkta alpha=0.05 ci=resampling plots=(rdplot ddplot
reshistogram) algorithm=SIMPLEX;
     model PERSK_VNT_VERTE_LOG=RINKOS_VERTE_LOG ZOVE_CENTR ZOVE_PREST ZOVE_MIEG
ZOVE_KITI OBJEKTU_SKAICIUS_SUTARTYJE KAMBARIU_SK SIENOS_MV_3 RUSYS_MV N_STATYBA/
quantile=0.05 to 0.95 by 0.01 seed=12345 plot=quantplot;
     test   RINKOS_VERTE_LOG   ZOVE_CENTR   ZOVE_PREST   ZOVE_MIEG   ZOVE_KITI
OBJEKTU_SKAICIUS_SUTARTYJE KAMBARIU_SK SIENOS_MV_3 RUSYS_MV N_STATYBA/ wald lr;
     run;
     ods graphics off;
```

PRIEDAS B

Šiame priede pateikiame išsamias tiesinės regresinės analizės tyrimo suvestines.

Durbin-Watson D	1.577
Number of Observations	24000
1st Order Autocorrelation	0.212

B.1 pav. Autokoreliacijos testas

Durbino-Vatsono testas rodo, jog autokoreliacijos šiame tiesinės regresijos modelyje nėra $(DW = 1{,}577)$.

Heteroscedasticity Test					
Equation	**Test**	**Statistic**	**DF**	**Pr > ChiSq**	**Variables**
PERSK_VNT_VERTE_LOG	White's Test	2079	45	<.0001	Cross of all vars
	Breusch-Pagan	1275	9	<.0001	RINKOS_VERTE_LOG, ZOVE_CENTR, ZOVE_PREST, ZOVE_MIEG, OBJEKTU_SKAICIUS_SUTARTYJE, KAMBARIU_SK, sienos_mv_3, RUSYS_MV, N_STATYBA, 1

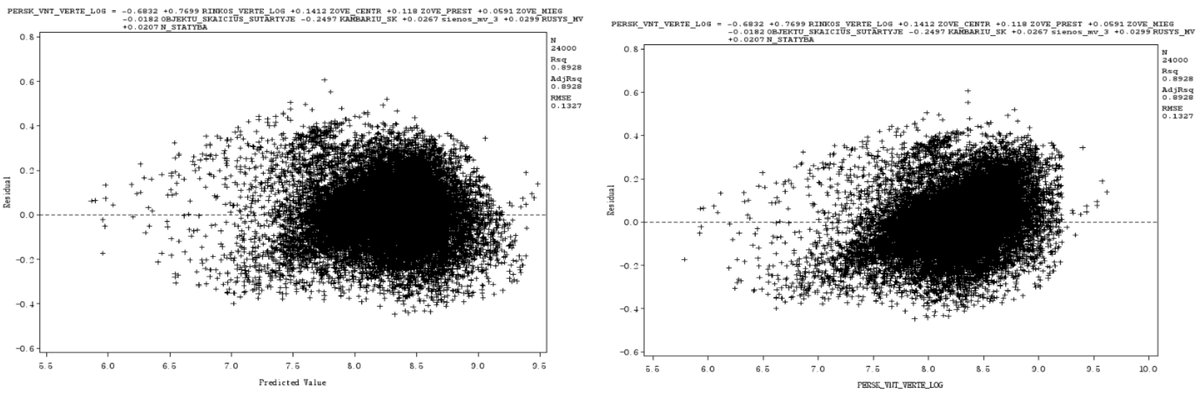
B.2 pav. Heteroskedastiškumo testas

Atlikus *Vaito* ir *Breušo-Pagano* testus liekanų homoskedastiškumo prielaida yra atmetama.

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.036064	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	9.487875	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	52.23579	**Pr > A-Sq**	<0.0050

B.3 pav. Liekanų normalumo testas

Pagal *Kolmogorovo-Smirnovo* ir kitus testus teigiama, jog liekanos nėra pasiskirsčiusios pagal normalųjį skirstinį.

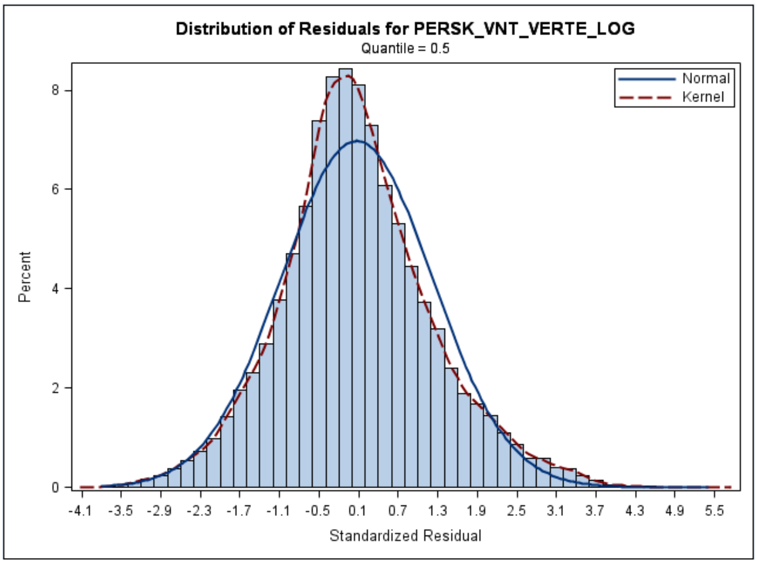


B.4 pav. Liekanų ir prognozuojamos kainos bei liekanų ir tikrosios kainos grafikai

B.4 paveiksle vaizduojamas liekanų ir prognozuojamos kainos grafikas bei liekanų ir vieno kvadratinio metro kainos grafikas. Šiuose grafikuose prognozuojamos ir tikrosios kainos yra logaritmuotos. Matome, jog taškai yra gan simetriškai išsidėstę aplink *x* ašį.

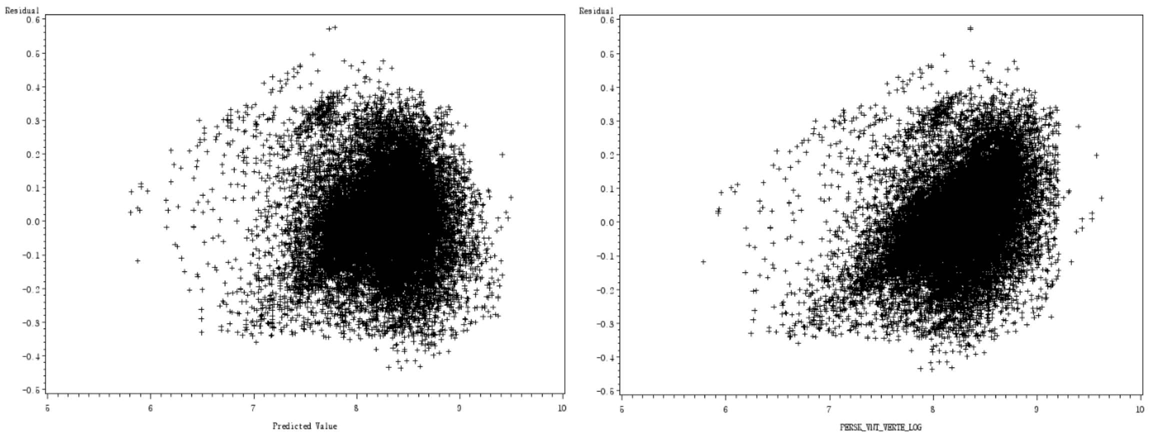
PRIEDAS C

Šiame priede pateikiame medianos kvantilinės regresijos testų rezultatus.



C.1 pav. Liekanų histograma

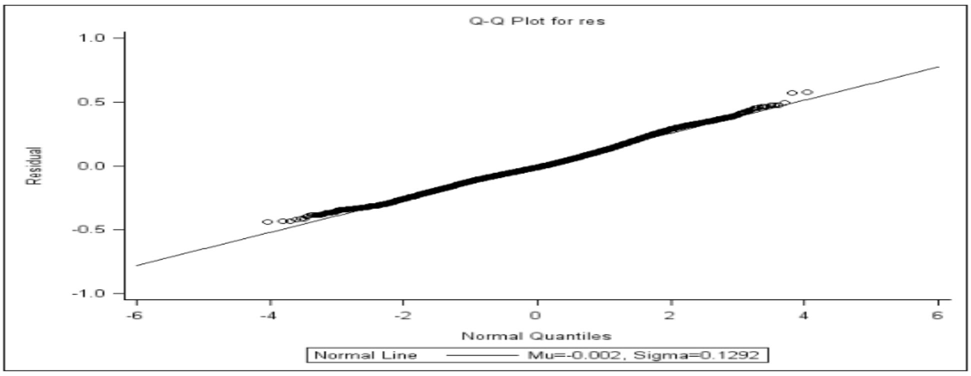
Liekanų histograma parodo priklausomo kintamojo – logaritmuotos vieno kvadratinio metro kainos standartizuotų liekanų pasiskirstymą.



C.2 pav. Liekanų ir prognozuojamos kainos bei liekanų ir tikrosios kainos grafikai

Tests for Normality				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.039101	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	10.20927	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	56.18908	**Pr > A-Sq**	<0.0050

C.3 pav. Liekanų normalumo testas



C.4 Liekanų normalumo kreivė 0,5 kvantilinei regresijai

C.2–C.4 paveiksluose pateikiame liekanų normalumo ir heteroskedastiškumo testų rezultatus, kurie nėra būtini sudarant kvantilinės regresijos modelius. Iš pateiktų rezultatų galime spręsti, jog liekanos nėra pasiskirsčiusios pagal normalųjį skirstinį, nes netenkinamas *Kolmogorovo-Smirnovo* testas. Liekanų normalumo kreivė rodo, jog didžiausi nuokrypiai nuo normalaus skirstinio egzistuoja kreivės pradžioje ir pabaigoje.

PRIEDAS D

Šiame priede pateikiami neuroninių tinklų analizės rezultatai.

Summary of active networks (Duomenys neuroniniams tinklams 2008-2013 po isskirciu 03 31 (B2:ATXE31725))											
Index	Net. name	Training perf.	Test perf.	Validation perf.	Training error	Test error	Validation error	Training algorithm	Error function	Hidden activation	Output activation
1	MLP 10-8-1	0,951769	0,948563	0,860054	0,000564	0,000611	0,002030	BFGS 220	SOS	Tanh	Logistic

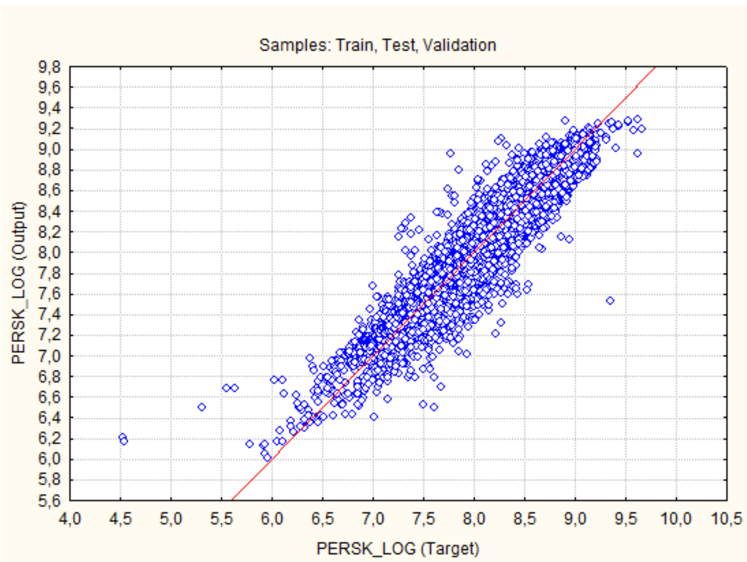
D.1 pav. MLP 10-8-1 apžvalga

Pirmame paveiksle pateikta bendra sudaryto MLP 10-8-1 neuroninio tinklo apžvalga. Sudarymo metu paslėptame sluoksnyje buvo naudojama tangento funkcija, o išvedimo sluoksnyje – logistinė funkcija, Nuokrypių kvadratų suma buvo skaičiuojama klaidos funkcija. Pateikto modelio tikslumas skirtinguose sudarymo etapuose svyruoja nuo 0,86 iki 0,95.

	Pointwise sensitivity analysis for PERSK_LOG (Duomenys neuroniniams tinklams 2008-2013 po isskirciu 03 31 (B2:ATXE31725)) Network: 1.MLP 10-8-1									
Grid points	**ZOVE_CENTR Sensitivity**	ZOVE_PREST Sensitivity	ZOVE_MIEG Sensitivity	ZOVE_KITI Sensitivity	OBJEKTU_SKAICIUS_SUTARTYJE Sensitivity	RUSYS_MV Sensitivity	KAMBARIU_SK Sensitivity	N_STATYBA Sensitivity	sienos_mv_3 Sensitivity	RINKOS_LOG Sensitivity
Minimum	0,161149	0,167316	0,163957	0,065328	-0,111448	0,095891	-0,78123	1,250026	0,189111	0,216031
2	0,155628	0,162086	0,152152	0,054046	-0,124978	0,121859	-0,88117	0,835551	0,221363	0,745429
3	0,149838	0,157348	0,139197	0,043009	-0,137681	0,117275	-1,00940	0,628448	0,227096	1,846840
4	0,143903	0,153209	0,125230	0,032352	-0,149012	0,071451	-1,13693	0,481852	0,214573	3,239180
5	0,137940	0,149762	0,110486	0,022178	-0,158430	-0,019817	-1,23523	0,325811	0,191170	3,013992
6	0,132053	0,147081	0,095309	0,012559	-0,165448	-0,151164	-1,28772	0,135045	0,162854	3,085369
7	0,126332	0,145216	0,080158	0,003543	-0,169690	-0,303654	-1,29518	-0,090580	0,133898	3,564414
8	0,120852	0,144189	0,065601	-0,004845	-0,170939	-0,441743	-1,26961	-0,335063	0,106942	2,594016
9	0,115673	0,143992	0,052285	-0,012591	-0,169161	-0,522349	-1,22361	-0,569533	0,083313	1,761672
Maximum	0,110840	0,144586	0,040896	-0,019699	-0,164515	-0,521339	-1,16444	-0,756173	0,063428	1,309085

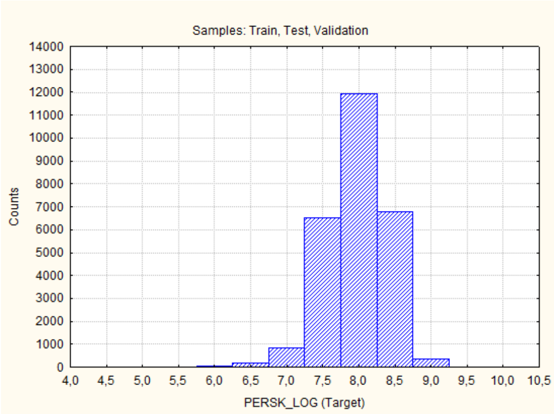
D.2 pav. *Sensitivity* analizė

Sensitivity analizė parodo kiekvieno kintamojo įtaką neuroninio tinklo sudarymui. Kintamųjų įtakos yra parodytos kiekvienam neuronui atskirai. Iš šios lentelės matome, jog kintamieji *objektų skaičius sutartyje* ir *kambarių sk.* turi neigiamą įtaką, o *centro, prestižinių* ir *miegamųjų* rajonų zonos, *rinkos vertė, sienų tipas* – teigiamą.



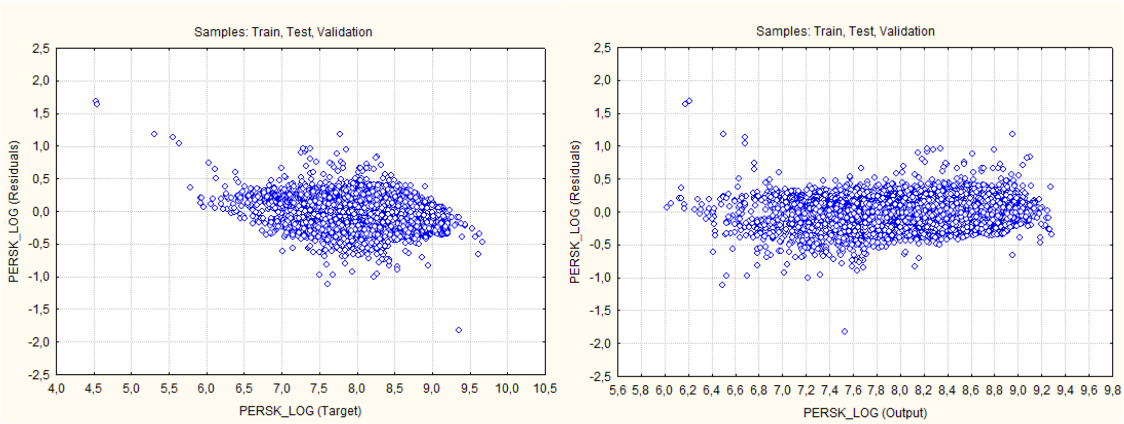
D.3 pav. Prognozuojamų ir tikrųjų kainų priklausomybės grafikas

D.3 paveiksle pateiktas prognozuojamų ir tikrųjų kainų priklausomybės grafikas apmokymo testavimo ir naudojimo imčių duomenims.



D.4 pav. Liekanų normalumo histograma

Liekanų normalumo histograma vaizduoja liekanų pasiskirstymą apmokymo testavimo ir naudojimo imčių duomenyse.

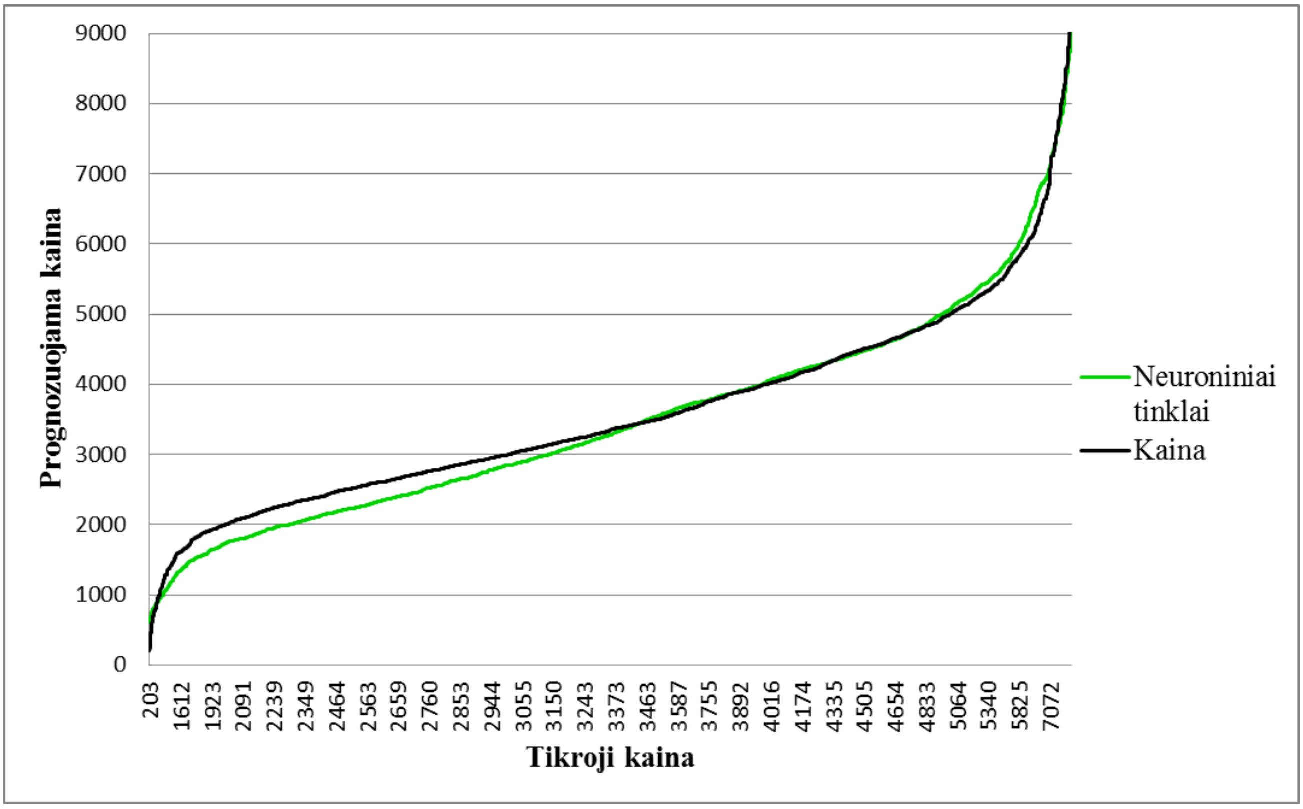


D.5 pav. Liekanų ir prognozuojamos bei liekanų ir tikrosios kainų grafikai.

Liekanų ir prognozuojamos bei liekanų ir tikrosios kainų grafikai gali būti naudojami kaip heteroskedastiškumo testas. D.5 paveiksle matome, jog liekanos yra išsibarstę apie nulį, tačiau yra ir labiau išsiskiriančių reikšmių.

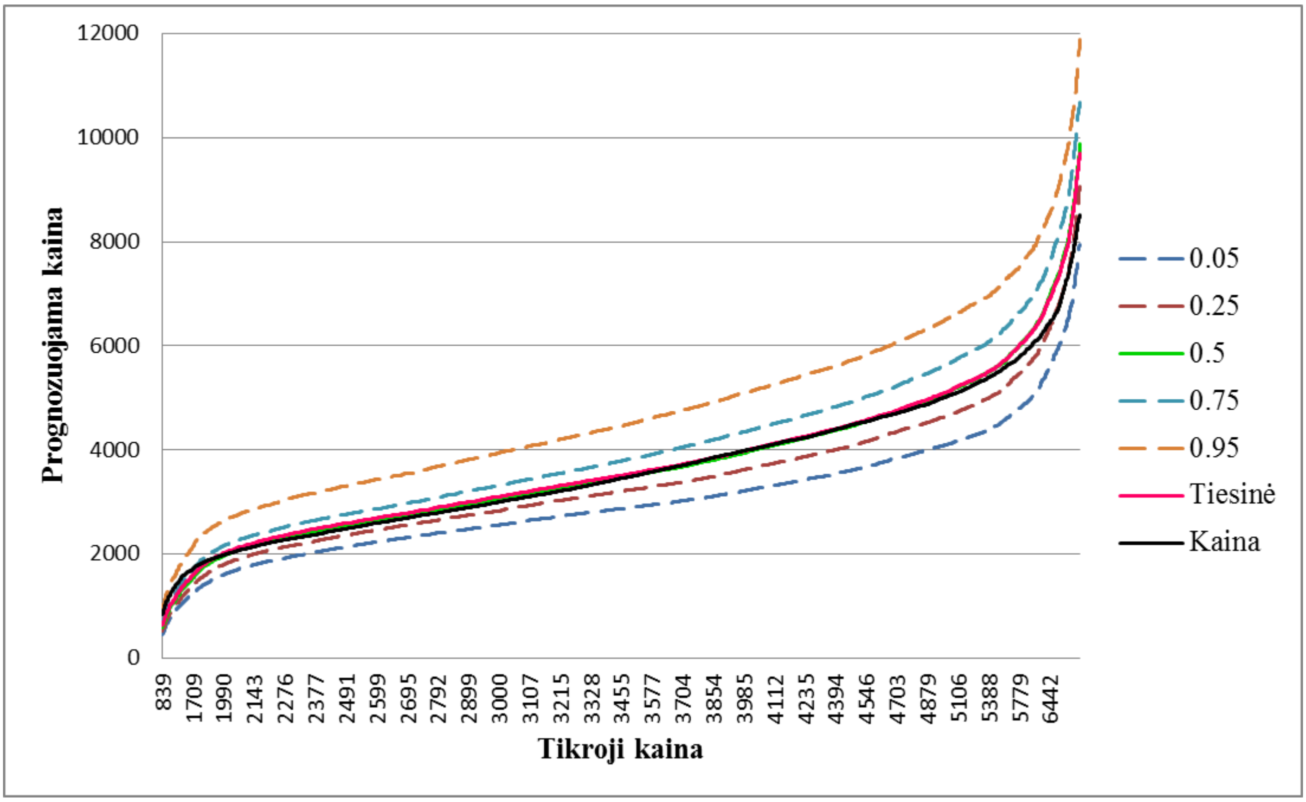
PRIEDAS E

Šiame priede pateikiame 2013 metų pirkimo-pardavimo sandorių prognozių grafikus.



E.1 pav. 2013 metų kainos prognozavimas naudojant neuroninių tinklų modelį

Naudojant neuroninių tinklų metodą buvo prognozuojama 2 679-ių pirkimo-pardavimo sandorių (34,68% visos imties) vieno kvadratinio metro kaina. Matome, jog šis metodas tiksliausiai prognozuoja šiek tiek didesnes nei vidutinės kainas.



E.2 pav. 2013 metų pirkimo-pardavimo sandorių prognozavimas tiesinės ir kvantilinės regresijos lygtimis

E.2 paveiksle pavaizduotos 2013 metų pirkimo-pardavimo sandorių prognozavimo tiesinės ir kvantilinės regresijos kreivės. Grafike vaizduojama visa 2013 metų pirkimo-pardavimo sandorių imtis, tad galime pastebėti esančius kainų ekstremumus. Matome, jog arčiausiai tikrosios kainos yra tiesinės ir medianos kvantilinės regresijos kreivės.