

VILNIAUS UNIVERSITETAS  
MATEMATIKOS IR INFORMATIKOS FAKULTETAS

Projektinis darbas

DISKRETAUS PASIRINKIMO  
MODELIŲ PRITAIKYMAS  
FINANSŲ RINKOJE

VILNIUS 2023

# **EKONOMETRIJOS STUDIJŲ PROGRAMA**

## **III kursas**

Darbą atliko: Rokas Karpavičius,  
Darbo vadovas: Prof., Dr. (HP) Marijus Radavičius

# Turinys

<b>1</b>	<b>Įvadas</b>	<b>2</b>
<b>2</b>	<b>Straipsnių aprašymas</b>	<b>2</b>
2.1	Pirmasis straipsnis . . . . .	2
2.2	Antrasis straipsnis . . . . .	3
<b>3</b>	<b>Duomenys ir metodologija</b>	<b>3</b>
3.1	Metodologija . . . . .	3
3.2	Duomenys . . . . .	7
<b>4</b>	<b>Modelio sudarymas bei duomenų generavimas</b>	<b>7</b>
4.1	Modelio pasirinkimas bei kitos prielaidos . . . . .	7
4.2	Duomenų generavimas . . . . .	8
4.3	Tendencijos . . . . .	9
4.4	Modelio sudarymas . . . . .	11
4.4.1	Pirmasis modelis . . . . .	11
4.4.2	Antrasis modelis . . . . .	12
<b>5</b>	<b>Išvados bei rekomendacijos</b>	<b>13</b>
<b>A</b>	<b>Priedas</b>	<b>14</b>
<b>B</b>	<b>Priedas</b>	<b>16</b>
<b>C</b>	<b>Priedas</b>	<b>19</b>

# 1 Įvadas

*Diskretaus pasirinkimo modelių* dalyke buvo pristatytas vienas iš kurso reikalavimų - projektinio darbo parengimas. Šiam darbui reikėjo pasirinkti bet kokią studentą dominančią (ekonominės) veiklos ar mokslo sritį, visuomenės ar gamtos reiškinių ir susirasti mokslinį tiriamąjį straipsnį, kuriame naudojant diskrečiojo pasirinkimo modelius ir metodologijas būtų sprendžiamas aktualus toje srityje taikomasis uždavinys.

Darbui atlikti buvo pasirinkta nagrinėti finansų rinką. Tai tokia rinka, kuri leidžia investuotojams pirkti bei parduoti, t.y. apsimainyti pinigais, vertybiniais popieriais ir (ar) finansinėmis priemonėmis, tokiomis kaip akcijos, obligacijos, žaliavų – tauriųjų metalų ar grūdų – išankstiniais bei ateities sandoriais.

Bendrasis uždavinys susideda iš straipsnio pasirinkimo, trumpos jo apžvalgos bei „atkartojimo“. Atkartojant reikia sugeneruoti panašius ar identiškus duomenis, kurie buvo naudojami straipsnyje pasitelkiant vieną iš statistinių programinių paketų, skirtų analizuoti duomenims (pvz. R, Python ir kt.). Šiam darbui atlikti buvo išsikelti tokie uždaviniai:

- Straipsnių apžvalga;
- Duomenų bei metodologijos apžvalga;
- Duomenų generavimas (simuliacija);
- Modelio sudarymas;

## 2 Straipsnių aprašymas

### 2.1 Pirmasis straipsnis

Renkantis mokslinį straipsnį akcijų rinkos tematika buvo aptikta daugybė įvairaus tipo darbų. Pvz. sentimentų analizė nuspėti kur judės akcijų rinka, kaip tam tikros naujienos paveikia akcijų kainas ar optimalaus investicinio krepšelio sudarymas. Tokių straipsnių iš tiesų yra apstu, tačiau daugumoje atvejų jie yra mokami arba nesusiję su pasirinkimo modeliavimu.

Viena iš studijų [1] analizuoja kaip galima panaudoti akcijų biržos naujienas bandant nuspėti akcijos kainą. Tikslas yra parodyti kaip sentimentų

analizė (angl. sentiment analysis) (toks analizės tipas, kuris identifikuoja ir charakterizuoja nuomones, išreikštas tekste (pvz. straipsnyje)) gali pagerinti regresinės analizės tikslumą bandant nuspėti akcijos dienos atsidarymo kainą (angl. open price). Problema yra tame, jog šiame straipsnyje nėra užsimeinama apie diskretaus pasirinkimo metodologiją. Taip pat, atlikus straipsnio apžvalgą, buvo pastebėta, jog metodologija yra aprašyta itin lakoniškai, be didesnių paaiškinimų. Taigi, atkartoti tokį tyrimą būtų sudėtinga ne tik dėl išsamesnio aprašo ar patirties stokos (darant tokią analizę), bet tuo pačiu ir dėl tyrimo prasmės. Sentimento analizė yra itin abstraktus dalykas, kurio praktinę naudą būtų sunku įžvelgti bei patikrinti.

## 2.2 Antrasis straipsnis

Pasirinktame straipsnyje [2] yra nagrinėjama vertybinių popierių rinka bei investuotojų pasirinkimo kriterijų svarbumas renkantis įmonių akcijas. Taip pat, kalbama apie optimalaus investicinio krepšelio sudarymą. Tyrime pagrindinis dėmesys yra skiriamas privatiems investuotojams, kurių tikslas yra ilgalaikės investicijos. Naudojama diskretaus pasirinkimo analizė (arba diskretaus pasirinkimo eksperimentas) įvertinti respondentų pageidavimus keturiems akcijos atributams, t.y. pačiai kompanijai, generuojamai grąžai bei dividendams ir akcijos kainos trendui (kitais tariant, rizika, išreikšta per kainos grafiką). Šiam tikslui pasiekti tyrėjai siūlo naudoti daugianarį logit modelį (angl. Multinomial logit model, MNL) arba *Hierarchical Bayes (HB)* metodą.

Siūloma metodologija yra iliustruojama įmonių, kurių akcijomis prekiaujama Belgrado akcijų biržoje (angl. Belgrad Stock Exchange, BELEX). Tyrimo metu renkami respondentų pageidavimai kompanijoms, generuojamai grąžai, dividendams bei kainos trendui.

## 3 Duomenys ir metodologija

### 3.1 Metodologija

Tyrimo metu yra naudojami trys pagrindiniai metodai, pirmiausia, diskretaus pasirinkimo analizė (angl. Discrete Choice Analysis, DCA), kuri yra skirta kiekybiškai įvertinti asmens pasirinkimą esant tam tikram alternatyvų rinkiniui.

Kitas iš metodų yra *Eckel-Grossman (EG)* metodas, kuris yra skirtas pamatuoti individualų norą priimti tam tikrą riziką. Kitaip tariant, jis yra

skirtas įvertinti asmens rizikos prisiėmimo *apetitą*. Teigiama, jog metodas yra patikimas ir tinkamas būdas įvertinti asmens rizikos prisiėmimą.

Taip pat, buvo atsižvelgta ir į modernaus portfelio teoriją (angl. Modern Portfolio Theory, MPT). Pastaroji buvo sukurta siekiant aprašyti optimalų investicinį krepšelį investuotojui tam tikroje rinkoje. Kitaip sakant, parodyti, kad iš ribotos aibės aktyvų visuomet galima sudaryti kiekvienam investuotojui geriausią investicinį krepšelį, atsižvelgiant į jo rizikos apetitą. Vienas iš šios teorijos pamatų yra *Markowitz's mean-variance* modelis. Pastarasis naudojamas kaip pagalba renkantis patį naudingiausią investicinį portfelį iš  $n$  alternatyvų. Tyrėjai šį modelį pritaiko norint įvertinti iš *DCA* gautus rezultatus.

Rezultatai, kurie buvo gauti naudojantis *DCA* metodologija yra vadinami naudingumais. Tai yra skaitinės reikšmės, kurios atspindi kiek kiekvienas atributas ir jo lygis įtakoja individo pasirinkimą. Tarkime, kad  $I$  respondentų renkami iš  $J$  alternatyvų, kur kiekvienas iš jų gauna tam tikrą naudingumą pasirinkus kurią nors alternatyvą. Daroma prielaida, jog asmuo  $i$  ( $i = 1, \dots, I$ ) pasirenks alternatyvą  $j$  ( $j = 1, \dots, J$ ) tada ir tik tada, jeigu jos naudingumas bus didesnis (arba lygus) už kitą, esančią tame pačiame rinkinyje:

$$y_{ij} = \begin{cases} 1, & U_{ij} > U_{im}, \forall j \neq m \\ 0, & \text{kitu atveju} \end{cases}, \text{ kur } y_{ij} \text{ žymi pasirinkimą} \quad (1)$$

Taip pat,  $U_{ij}$  žymi naudingumą, kurį individas gauna pasirinkus alternatyvą  $j$  ir gali būti išskaidytas į sistemine (reprezentatyviąją) komponentę  $V_{ij}$  bei atsitiktinę komponentę  $\varepsilon_{ij}$  ir užrašytas pavidalu:

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \quad (2)$$

Atsitiktinė komponentė  $\varepsilon_{ij}$  žymi nestebimus (neišmatuojamus) naudingumus, kurie gali atsirasti dėl skirtingų pasirinkimų bei matavimo paklaidų, kurie taip pat gali įtakoti asmens pasirinkimą. Tuo tarpu  $V_{ij}$  yra (sistemine) naudingumo komponentė, susijusi su stebimais atributais, turinčiais jam įtakos. Autoriai teigia, jog diskretaus pasirinkimo modelio tikslas yra kiekybiškai įvertinti kiek svorio kiekvienas atributas ir jo lygiai turi asmeniniam naudingumui, t.y. kaip įtakojamas galutinis sprendimas. Tam, jog būtų įmanoma įgyvendinti šį modelį, reikia aprašyti  $V_{ij}$  funkciją:

$$V_{ij} = \sum_{k=1}^K \sum_{l=1}^{L_k} \beta_{ikl} \cdot x_{jkl}, \quad (3)$$

kur  $\beta_{ikl}$  yra naudingumas, kurį individas  $i$  suteikia atributo  $k$ , ( $k = 1, \dots, K$ ) lygiui  $l$  (pvz. atributas - įmonė, lygis - *Jedinstvo a.d.* arba atributas - graža, lygis - 0,15 %). Straipsnyje tai yra įvardinama kaip *part-worth utility*, ką būtų galima versti kaip dalinės vertės naudingumas. Taip pat,  $x_{jkl}$  yra binarusis kintamasis, kurio reikšmė lygi vienetui, jei alternatyva  $j$  savyje turi  $k$  atributo lygį  $l$ , kitu atveju - lygi nuliui.

Kadangi atsitiktinė modelio (2) komponentė ( $\varepsilon$ ) yra nestebima, tai prarandame galimybę tiksliai nuspėti respondento pasirinkimą. Dėl šios priežasties, tyrimo metu yra aprašoma tikimybė  $i$  – *tajam* asmeniui pasirinkti  $j$ -tąją alternatyvą, kuri atrodo taip:

$$\begin{aligned} P_{ij}(y_{ij} = 1) &= P(U_{ij} \geq U_{im} | m \in J, \forall m \neq j) \\ &= P(V_{ij} + \varepsilon_{ij} \geq V_{im} + \varepsilon_{im} | m \in J, \forall m \neq j) \\ &= P(\varepsilon_{ij} - \varepsilon_{im} \geq V_{im} - V_{ij} | m \in J, \forall m \neq j) \\ &= P(\varepsilon_{im} \leq V_{im} + \varepsilon_{ij} | m \in J, \forall m \neq j) \end{aligned} \quad (4)$$

Straipsnyje yra minima, jog išraiška (4) yra įvertinama priklausomai nuo atsitiktinių dydžių (toliau, a.d.) pasiskirstymo prielaidų. Klasikiniu, RU (angl. random utility) modelio atveju, yra žinoma, jog šie dydžiai yra nepriklausomi ir vienodai pasiskirstę (angl. independent and identically distributed). Kitaip tariant, a.d.  $\varepsilon_j \sim Gumbel, j = 1, \dots, n$ ; čia *Gumbel* := *Gumbel*(0, 1) žymi standartinę *Gumbel* skirstinį su vietos (angl. location) arba postūmio (angl. shift) parametru 0 ir mastelio (angl. scale) parametru 1.

Apsibrėžę naudingumo bei sisteminės komponentės funkcijas ir aptarę niuansus, susijusius su nestebimomis paklaidomis, tyrėjai galiausiai pereina prie kito svarbaus žingsnio - modelio pritaikymo. Pirmiausia yra aptariamas *daugianaris logit modelis* (angl. Multinomial logit model, MNL). Pastarasis yra naudojamas įvertinti modelio parametrus, pvz. dalinės vertės naudingumui (turima omenyje  $\beta_{ikl}$ , kuris buvo apžvelgtas (3) išraiškoje). Taip pat, MNL modelio prielaidos buvo aptartos praeitoje pastraipoje.

Tariama, jog tikimybė pasirinkti bet kokią alternatyvą naudojant MNL modelį yra aprašoma taip:

$$P_{ij}(y_{ij} = 1) = \frac{\exp(V_{ij})}{\sum_{m \in J} \exp(V_{im})} = \frac{1}{1 + \sum_{m \in J} \exp(V_{im} - V_{ij})} \quad (5)$$

Paprastai tariant, MNL modelio parametrai arba  $\beta$  reikšmės, kurios yra įvardijamos kaip *part-worths* vektorius, yra įvertinamos visai imčiai pasirinkimų vietoje individualių pasirinkimų. Tai yra todėl, jog šiame modelyje parametrai yra laikomi tokie patys visiems pasirinkimams, nepaisant to, kokia alternatyva buvo pasirinkta. Tam, jog būtų galima įvertinti  $\beta$  reikšmes, yra naudojamas didžiausio tikėtinumo metodas (angl. maximum likelihood estimation), kurio pagalba yra surandamas  $\beta$  reikšmių vektorius. Didžiausio tikėtinumo funkcija atrodo taip:

$$\ln(L(\beta)) = \sum_{i=1}^n \sum_{j=1}^J y_{ij} \cdot \ln(P_{ij}) \quad (6)$$

Nepaisant to, studija pasirinko naudoti *Hierarchical Bayes* (toliau, HB) metodą dėl jo savybės įvertinti individualius pasirinkimus (arba kaip straipsnyje įvardinta, *HB ability to estimate individual-level parameters*). Tai reiškia, jog modelio parametrai yra įvertinami kiekvienam asmeniui atskirai. Teigiama, jog HB metodas naudoja hierarchinius modelius, kurie yra analizuojami pasitelkiant Bajeso metodologiją.

Šiuo atveju tikimybė yra išreiškiama kaip „tikėjimo laipsnis“ (angl. a degree of belief). Taikant šį metodą, respondentai yra laikomi kaip heterogeninė imtis (daroma prielaida, jog asmenys, esantys imtyje, turi skirtingus pageidavimus ar įsitikinimus) ir tariama, jog įvertinti parametrai (*part-worths* arba tiesiog  $\beta$ ) yra pasiskirstę pagal daugiamačią normalųjį skirstinį (turima omenyje, jog yra daugiau nei vienas kintamasis) su vidurkio vektoriumi (angl. mean vector) ir kovariacijų matrica (angl. covariance matrix).

Tikimybė, jog individas pasirinks tam tikrą alternatyvą yra skaičiuojami naudojantis (5) išraiška. Tuo tarpu individualūs dalinių verčių naudingumai (*part-worths* arba  $\beta$ ) yra įvertinami naudojant *Monte Carlo* simuliaciją. Straipsnio autoriai teigia, jog HB modelio pranašumas yra jo gebėjimas įvertinti daugiau parametrų turint mažiau duomenų.

Toliau šio atskiuro modelio nenagrinėsime, kadangi buvo pasirinkta pritaikyti daugianarį logit modelį, kuris buvo apžvelgtas šios pastraipos pradžioje. Taip pat, tai buvo vienas iš modelių, kurį straipsnio autoriai siūlė taikyti.



## 3.2 Duomenys

Duomenys buvo renkami apklausos būdu, kurioje asmenims buvo pateiktas sąrašas (žr. 1 pav.) hipotetinių investavimo pasirinkimų. Kiekvienam iš respondentų buvo pristatytos trys tokios anketos su atsitiktinai parinktomis alternatyvomis. Vienas iš pavyzdžių kaip galėtų atrodyti individui pateikiama anketa:

Įmonė	Gražta, %	Dividendai, %	Trendas
Jedinstvo a.d.	0,15	6,0	2-asis
Nikola Tesla Airport	0,15	0,0	1-asis
Metalac a.d.	0,10	3,0	3-iasis

1 lentelė: Respondentui pateikiamos alternatyvos

Iš šių trijų alternatyvų, atsakantysis turėjo pasirinkti labiausiai priimtina variantą, kuris jo (-os) nuomone turi didžiausią subjektyvų naudingumą. Pažvelgus į pradinę anketą (žr. 1 pav.), matome, jog vienas iš keturių atributų yra kompanija, kurioje turime keturis galimus variantus. Toliau seka tokie atributai kaip akcijos generuojama graža (%), dividendų dydis (% nuo investuotos sumos) bei trendas. Kiekvienas iš šių atributų turi tarp 4 ir 5 alternatyvų (straipsnyje tai yra įvardinta kaip *attributes and their levels*).

Taip pat, buvo renkama informacija apie respondentų demografinius duomenis (žr. 2 pav.), tokius kaip lytis, šeiminė padėtis ir kt. Naudojant *Eckel-Grossman* lošimų žaidimą (angl. *Eckel-Grossman gamble game*) (žr. 3 pav.) buvo tikrinama kiek daug respondentai yra linkę rizikuoti (lyginant lytis).

## 4 Modelio sudarymas bei duomenų generavimas

### 4.1 Modelio pasirinkimas bei kitos prielaidos

Kaip ir buvo minėta anksčiau, straipsnio autoriai pasirinko naudoti *Hierarchical Bayes* metodą dėl jo galimybės įvertinti parametrus individualiai kiekvienam respondentui. Kadangi šis metodas pasirodė ne visai aiškus ir tuo pačiu nepavyko rasti pakankamai informacijos kaip jį atlikti **R** paketo rėmuose, buvo pasirinkta naudoti, ko gero, ne ką prastesnį variantą - daugianarį logit modelį, kuris buvo ne kartą apžvelgtas paskaitų metu [3].

Papildomai prie modelio sudarymo naudojant diskretaus pasirinkimo metodologiją, autoriai taip pat pasirenka padaryti kelis papildomus žingsnius,

tokius kaip *Eckel-Grossman* metodas, kuris yra skirtas įvertinti individo rizikos *apetitą*. Be viso to, yra naudojama *Markowitz's mean-variance* teorema, kuri skirta investicinio portfelio optimizavimui. Kitaip tariant, iš aibės finansinių aktyvų (šiuo atveju, akcijų) yra išrenkamas toks portfelis, kuris maksimizuoja grąžą duotam rizikos lygiui. Kadangi šie du dalykai jau nebėra *Diskretaus pasirinkimo modelių* kurso apimtyje, buvo nuspręsta jų nenagrinėti.

## 4.2 Duomenų generavimas

Vienas iš projekto apribojimų buvo tas, jog duomenys turėjo būti simuliuojami atsižvelgiant į straipsnyje pateiktą informaciją. Šiuo atveju buvo pasitelkta R funkcija `expand.grid` iš bibliotekos `base`, kurios pagalba buvo išrašytos visos įmanomos kompanijų, grąžos, dividendų bei trendo kombinacijos iš duotos apklausos anketos (žr. 1 pav.). Kalbant apie demografinius duomenis, buvo pasirinkta per daug nekomplikuoti modelio ir naudoti tik lyties kintamąjį, kuris buvo generuojamas su `sample` funkcija iš `base` bibliotekos. Kadangi buvo nuspręsta naudoti *MNL* modelį, tai atsitiktinės komponentės generavimui buvo naudojama funkcija `rgumbel` iš `extraDistr` bibliotekos. Galiausiai, į duomenų lentelę (`dataframe`) buvo sudėti tokie stulpeliai (kintamieji):

- ID - respondento nr.;
- alternative - alternatyvos nr.;
- choice - binarusis kint., žymintis kuri iš alternatyvų buvo pasirinkta;
- gender - lytis;
- company - kompanija;
- return\_rate - grąža, %;
- dividend - dividendai, %;
- trend - akcijos kainos trendas;
- epsilon ( $\varepsilon$ ) - at. Gumbel paklaidos;
- V - sisteminis naudingumas;
- U - alternatyvos naudingumas.

Taip pat, teko atlikti tam tikrą kintamųjų manipuliaciją, kurios metu *gender*, *company* ir *trend* buvo transformuoti į kategorinius kintamuosius (faktorius), o tuo tarpu kintamieji *return\_rate* bei *dividend* buvo palikti kaip skaitiniai (angl. numeric). Šių transformacijų mums prireiks skaičiuojant sisteminį naudingumą  $V_{ij}$ . Kaip žinia, norint suskaičiuoti pastarąją vertę, būtina žinoti  $\beta$  reikšmes. Patogumo dėlei, straipsnio autoriai jas taip pat pateikė (žr. 4 pav.), tad papildomai skaičiuoti jų nebereikia (straipsnyje jos įvardijamos kaip *part-worths*). Taigi, indeksų pagalba atitinkamos atributų lygių reikšmės yra susiejamos su atitinkamomis  $\beta$  reikšmėmis ir apskaičiuojama  $V_{ij}$  reikšmė kiekvienai iš alternatyvų. Galiausiai, alternatyvos naudingumas  $U_{ij}$  yra apskaičiuojamas naudojantis (2) išraiška.

Prisimenant 3-iajame skyriuje aptartą metodologiją, tyrimo autoriai padarė prielaidą, jog individas  $i$  renkasi alternatyvą tada ir tik tada, kai jos naudingumas yra didžiausias iš visų trijų alternatyvų. Taigi, šią prielaidą taip pat pritaikome binariajam kintamajam *choice*, esančiam duomenų lentelėje. Jis žymės alternatyvą, turinčią didžiausią naudingumą ir automatiškai tą alternatyvą, kuri buvo pasirinkta.

### 4.3 Tendencijos

Aptarę modelio pasirinkimą bei duomenų generavimą, galime pažvelgti į tendencijas tarp respondentų, renkantis alternatyvą. Pirmiausia, pažvelkime kurios iš įmonių buvo populiariausios:

Įmonė	Metalac a.d.	Nikola Tesla Airport	Jedinstvo a.d.	Energoprojekt holding
Dažnis, %	37	30	16	17

2 lentelė: Įmonės

Pastebime, jog pirmoje vietoje puikuoja *Metalac a.d.* su 37%, toliau seka *Nikola Tesla Airport*, *Jedinstvo a.d.* bei *Energoprojekt holding*, atitinkamai su 30, 16 bei 17 proc.

Taip pat, turime duomenis apie gražos pasirinkimą. Šiuo atveju matome, jog daugiausiai pasirinkta graža buvo 0,15% su 34%. Verta paminėti, jog teigiant *daugiausiai pasirenkama*, neturima omenyje, jog individai turėjo galimybę pabrėžti kuris iš variantų jiems yra geriausias. Tokiu atveju visi trivieniai rinkęsi didžiausią skaičių. Kaip ir buvo minėta anksčiau, buvo pateiktos trys anketos (alternatyvos) kiekvienam iš respondentų. Taigi, šie skaičiai yra paprasčiausios sumos, gautos pagal tai, kur prie *choice* kintamojo buvo vienetasis.

Graža	0	0,05	0,1	0,15
Dažnis, %	21	17	28	34

3 lentelė: Graža, %

Dividendų atveju turime identišką situaciją kaip ir su grąža. Pastebime, jog daugiausiai buvo pasirinkta didžiausias procentas dividendų, t.y. 6% su taip pat 34%.

Dividendai	0	1,5	3,0	4,5	6,0
Dažnis, %	0	18	17	31	34

4 lentelė: Dividendai, %

Galiausiai, priejome prie akcijos kainos trendo. Tai iš esmės yra susiję su norima priimti rizika. Pažvelgę į apklausos anketą (žr. 1 pav.) bei  $\beta$  reikšmių lentelę (žr. 4 pav.), matome, jog autoriai *Trend 1* įvardija kaip nerizikingą (*non-risky*), *Trend 2* kaip vidutiniškai rizikingą (*moderate risky*) bei *Trend 3* kaip rizikingą (*risky*). Taigi, iš to seka išvada, jog rizikingiausią trendą pasirinko mažuma atsakiusiųjų, tik 5%. Tuo tarpu pasirinkimas tarp 1-ojo bei 2-ojo buvo sąlyginai panašus, atitinkamai su 46 bei 49 proc.

Trendas	1-asis	2-asis	3-iasis
Dažnis, %	46	49	5

5 lentelė: Akcijos kainos trendas

Taip pat, turime duomenis apie tikimybes, susietas su kiekviena iš alternatyvų:

Alternatyva	1-oji	2-oji	3-ioji
$\mathbb{P}$ , %	26	44	30

6 lentelė: Alternatyvos pasirinkimo tikimybės

Pastebime, jog respondentai daugiausiai rinkosi 2-ąją alternatyvą su 44%, toliau seka trečioji bei pirmoji alternatyvos, atitinkamai su 30 bei 26 proc.

Apžvelgę dažnių lenteles, galime teigti, jog individai didelę reikšmę suteikia vieniems iš svarbiausių rodiklių renkantis investiciją, t.y. generuojamai grąžai bei dividendams. Taip pat, sugeneruotoje imtyje yra pastebima, jog daugiausiai renkamasi mažiausiai rizikingas investicijas bei vidutiniškai rizikingas. Tuo tarpu polinkis *apsiimti* riziką yra itin mažas. Tai patvirtinta *Eckel-Grossman* lošimų žaidimas (žr. 3 pav.), kuriame galime pastebėti, jog rizikos beieškančių (angl. risk seeking) asmenų procentas yra tik 4,14. Kalbant apie pačias įmones, išvadas daryti būtų tikslinga tik tuo atveju, jei jas išanalizuotume detaliau.

## 4.4 Modelio sudarymas

### 4.4.1 Pirmasis modelis

Tam, jog išsiaiškintume nuo ko priklauso respondento pasirinkimas renkantis vieną iš trijų investavimo alternatyvų, pritaikysime daugianarį logit modelį, kurį jau buvome aptarę ankstesniuose skyriuose. Šiuo atveju buvo sudaryti du modeliai, kur viename iš jų nebuvo įtrauktas lyties efektas, o kitame - buvo. Tiek viename, tiek kitame modelyje bazinė alternatyva yra pirmoji. Prisimename, jog taip pat buvo trys kainos trendai, kur pirmasis iš jų buvo parinktas kaip bazinis. Taigi, pirmasis modelis atrodo taip:

$$\begin{aligned} choice_1 = & \beta_0^{alt2} + \beta_0^{alt3} + \beta_1 \cdot companyNikolaTeslaAirport + \beta_2 \cdot companyJedinstvo + \\ & + \beta_3 \cdot companyEnergoprojekt + \beta_4 \cdot return\_rate + \beta_5 \cdot dividend + \\ & + \beta_6 \cdot (return\_rate \cdot trendTrend2) + \beta_7 \cdot (return\_rate \cdot trendTrend3) \end{aligned}$$

Kaip žinia, pritaikius modelį tam tikri kintamieji, kurių p reikšmė yra mažesnė nei reikšmingumo lygmuo  $\alpha = 0,05$  iš modelio yra išmetami. Šis atvejis nebuvo išimtis ir teko atsisakyti kelių kintamųjų (žr. priedą C). Galutinis modelis atrodo taip:

$$\begin{aligned} choice_1 = & \beta_2 \cdot companyJedinstvo + \\ & + \beta_3 \cdot companyEnergoprojekt + \beta_4 \cdot return\_rate + \beta_5 \cdot dividend + \\ & + \beta_7 \cdot (return\_rate \cdot trendTrend3) \end{aligned}$$

Surašius atitinkamas  $\beta$  koeficientų reikšmes, turime:

$$\begin{aligned} choice_1 = & -5.62423 \cdot companyJedinstvo - \\ & -3.74455 \cdot companyEnergoprojekt + 35.76445 \cdot return\_rate + 2.44322 \cdot dividend - \\ & -119.74976 \cdot (return\_rate \cdot trendTrend3) \end{aligned} \tag{7}$$

Prisiminus prieš tai nagrinėtas dažnių lenteles bei atidžiau pažvelgus į galutinį modelį, pastebime tą pačią tendenciją, jog didžiausią įtaką pasirinkimui daro akcijos generuojama grąža bei dividendai. Tuo tarpu nors likę kintamieji yra statistiškai reikšmingi, jų įtaka pasirinkimui yra neigiamą, t.y. didelės įtakos pasirinkimui jie nedaro. Turint tokį rezultatą, galime teigti, jog respondentai itin kreipia dėmesį į vienus svarbiausių akcijų rodiklių renkantis investiciją. Tuo tarpu, konkreti įmonė ar jos kainos trendas nėra tokie svarbūs.

Verta paminėti, jog naudojant formulę (R pakete)  $choice \sim company + return\_rate + dividend + trend$ , visi kintamieji tampa statistiškai nereikšmingi. Išbandžius visas kitas interakcijas, nebuvo pasiektas toks rezultatas kaip, kad su  $return\_rate \cdot trend$ . Taip pat, R aplinkoje naudojant išraišką „ $x_1 \cdot x_2$ “ apskritai yra gaunami labai prasti rezultatai kur praktiškai visi kintamieji yra statistiškai nereikšmingi. Taigi, buvo pasirinkta naudoti išraišką „ $x_1 : x_2$ “ (dėl detalesnės informacijos žr. priedą C).

#### 4.4.2 Antrasis modelis

Aptarę pirmąjį modelį, pereiname prie antrojo, kuriame buvo įtrauktas lyties efektas. Pabrėžiama, jog *lytis - moteris* buvo paimta kaip bazinė reikšmė. Šiuo atveju turime, jog pilnasis modelis atrodo taip:

$$\begin{aligned} choice = & \beta_0^{alt2} + \beta_0^{alt3} + \beta_1 \cdot companyNikolaTeslaAirport + \beta_2 \cdot companyJedinstvo + \\ & + \beta_3 \cdot companyEnergoprojekt + \beta_4 \cdot return\_rate + \beta_5 \cdot dividend + \\ & + \beta_6 \cdot (return\_rate \cdot trendTrend2) + \beta_7 \cdot (return\_rate \cdot trendTrend3) + \\ & + \beta_8 \cdot (dividend \cdot genderMale) + \beta_9 \cdot (trendTrend2 \cdot genderMale) + \\ & + \beta_{10} \cdot (trendTrend3 \cdot genderMale) \end{aligned}$$

Kaip ir pirmuoju atveju, taip ir šiuo, tam tikri kintamieji nėra statistiškai reikšmingi, tad jų tenka atsisakyti. Galutinis modelis atrodo taip:

$$\begin{aligned} choice_2 = & -11.44165 \cdot companyJedinstvo - 7.27326 \cdot companyEnergoprojekt + \\ & + 61.18836 \cdot return\_rate + 4.62998 \cdot dividend - 170.48160 \cdot (return\_rate \cdot trendTrend3) - \\ & - 23.04389 \cdot (trendTrend3 \cdot genderMale) \end{aligned} \quad (8)$$

Tapačiai, pastebime tendenciją, susijusią su grąža bei dividendais. Kaip ir ankstesniame, taip ir šiame modelyje pastarieji kintamieji daro didžiausią įtaką individo sprendimui. Tuo tarpu likę koeficientai yra neigiami, kas reiškia, jog jie praktiškai neturi įtakos sprendimui. Svarbu paminėti faktą, jog lytis nebuvo įtraukta į alternatyvos naudingumo skaičiavimą. Aptarus šį dalyką su darbo vadovu, buvo paminėta, jog tokiu atveju parametrai, susiję su lytimi, bus statistiškai nereikšmingi. Tam, jog to nebūtų, reikia padaryti keletą tam tikrų prielaidų bei papildomų skaičiavimų. Kadangi tai nebuvo trivialus dalykas, buvo nuspręsta jokių kitų veiksmų nesiimti.




## 5 Išvados bei rekomendacijos

Buvo sudaryti du daugianariai logit modeliai, kurie buvo naudojami nustatyti kaip atributų lygiai (kintamieji) daro įtaką individo pasirinkimui. Pirmasis iš jų buvo be lyties efekto ir parodė, jog pasirinkimą itin įtakoja akcijos generuojama grąža bei dividendai (tą taip pat parodė dažnių lentelės). Taigi, susidaro įspūdis, jog respondentui renkantis vieną iš trijų alternatyvų, sprendimui didelę įtaką darė generuojamos grąžos bei dividendų procentas. Tuo tarpu, pati įmonė ar jos kainos trendo grafikas buvo nustumti šalin.

Kadangi pirmasis modelis neįtraukė lyties efekto, buvo nuspręsta pritaikyti antrąjį modelį, šį kartą - su lyties efektu. Iš esmės, rezultatai buvo panašūs. Visi kintamieji apart grąžos bei dividendų, buvo neigiami. Tai leidžia mums daryti išvadą, jog jų įtaka yra itin maža arba jos nėra apskritai. Kaip ir buvo minėta anksčiau, lytis nebuvo įtraukta skaičiuojant alternatyvų naudingumus. Turint omenyje šį komentarą, antrojo modelio tikslumas yra abejotinas. Norint įvertinti lyties efektą pasirinkimui, būtina atlikti keletą modifikacijų bei papildomų skaičiavimų.

Taigi, aptarus abu pasirinkimo modelius bei įvertinus antrojo modelio sudarymo sistemą, naudingiau būtų rinktis pirmąjį, kadangi jis yra tikslesnis. Žinoma, atlikus reikiamus pakeitimus, antrasis modelis taptų pranašesnis, kadangi būtų galima įvertinti kokią įtaką lytis daro pasirinkimui.

## A Priedas

Attribute	Attribute Levels
Company	Metalac a.d. Nikola Tesla Airport Jedinstvo a.d. Energoprojekt holding
Return rate	0% 0.05% 0.10% 0.15%
Dividend (% of money invested)	0% 1.5% 3.0% 4.5% 6%
The trend of the stock prices over the past 3 years (graphics)	Trend 1 
	Trend 2 
	Trend 3 
Source: Authors.	

1 pav.: Apklausos anketa



Demographic	Category	Percent
Gender	Male	36.69%
	Female	63.31%
Marital status	Single	58.58%
	Married without kids	5.92%
	Married with kids	18.93%
	Other	16.57%
Highest level of completed education	High school	7.69%
	Undergraduate/graduate study	87.57%
	PhD	3.55%
	Other	1.18%
Employment status	Unemployed	21.30%
	Internship/part-time job	13.61%
	Full-time job	56.80%
	Retired	1.18%
	Other	7.10%

Source: Authors calculations.

2 pav.: Demografiniai duomenys

Choice (50/50 Gamble)	Low/High Payoff	Risk Attitude	Female	Male	Sample
Gamble 1	28/28	Extremely risk averse	9.35%	11.29%	10.06%
Gamble 2	24/36	Risk averse	27.1%	25.81%	26.63%
Gamble 3	20/44	Risk averse	38.32%	27.42%	34.32%
Gamble 4	16/52	Risk averse	9.35%	19.35%	13.02%
Gamble 5	12/60	Risk neutral	12.15%	11.29%	11.83%
Gamble 6	2/70	Risk seeking	3.74%	4.84%	4.14%

Source: Authors calculations.

3 pav.: Eckel-Grossman lošimų žaidimas

Attribute	Level	Relative Utility Score (Part-Worths)	Relative Importance Score (Attributes Importance)
Company	Metalac a.d.	3.35	9.49%
	Nikola Tesla Airport	3.00	
	Jedin a.d.	-6.14	
	Energoprojekt holding	-0.21	
Return rate	0%	-4.13	8.93%
	0.05%	-1.57	
	0.1%	0.90	
	0.15%	4.80	
Dividend	0%	-28.14	48.98%
	1.5%	-6.57	
	3%	1.57	
	4.5%	12.30	
	6%	20.84	
Trend	Trend 1 (non-risky)	12.25	32.6%
	Trend 2 (moderate risky)	8.10	
	Trend 3 (risky)	-20.35	

Source: Authors calculations.

4 pav.: Atributų lygių  $\beta$  reikšmės

## B Priedas

```
library(dplyr)
library(extraDistr)
library(mlogit)
library(car)

####                                ####
####   Duomenų generavimas         ####
####                                ####

# Generuojamos visos imanomos atributų lygių kombinacijos

attributes <- expand.grid(
  company = c("Metalac", "NikolaTeslaAirport", "Jedinstvo",
    "Energoprojekt"),
  return_rate = c(0, 0.05, 0.10, 0.15),
  dividend = c(0, 1.5, 3.0, 4.5, 6),
  trend = c("Trend1", "Trend2", "Trend3")
)

set.seed(123) # nustatomas 'seed' tam, jog kas karta generuojant
# alternatyvas bei paklaidas, jie išliktų tokie patys
```

```

# Sukuriamas masyvas ir sudedamos reikšmės

gen <- sample(c("male", "female"), size = 100,
prob = c(0.3669, 0.6331), replace = TRUE)

# Lychiu procentai paimti tiesiogiai iš straipsnio

df <- data.frame(
  ID = rep(1:100, each = 3), # respondentų nr.; kartojama 3x, nes yra
  #3 alternatyvos kiekvienam respondentui
  alternative = as.factor(rep(1:3, times = 100)), # alternatyvos Nr.
  choice = 0, # priskiriama pradine reikšme 0 kiekvienai alternatyvai
  gender = rep(gen, each = 3), # lytis kartojama tris kartus viename
  # respondentui, kadangi yra 3 alternatyvos
  attributes[sample(nrow(attributes), size = 100 * 3, replace = TRUE),],
  epsilon = rgumbel(300) # at. Gumbel paklaidos
)

# Kintamųjų manipuliavimas

df <- df %>% mutate(
  gender = as.factor(gender),
  company = as.factor(company),
  return_rate = as.numeric(return_rate),
  dividend = as.numeric(dividend),
  trend = as.factor(trend)
)

#####

####                                ####
####   Naudingumu skaičiavimas   ####
####                                ####

# Ivertintos 'beta' (part-worths) reikšmės iš straipsnio

# Priskiriami pavadinimai tam, jog algoritmas galėtų susieti reikšmes,
# esančias duomenų lentelės su atitinkamomis 'beta' reikšmėmis
# pagal indeksus

```

```

beta_company <- c("Metalac" = 3.35, "NikolaTeslaAirport" = 3.00,
"Jedinstvo" = -6.14, "Energoprojekt" = -0.21)
beta_return <- c("0" = -4.13, "0.05" = -1.57, "0.1" = 0.90,
"0.15" = 4.80)
beta_dividend <- c("0" = -28.14, "1.5" = -6.57, "3" = 1.57,
"4.5" = 12.30, "6" = 20.84)
beta_trend <- c("Trend1" = 12.25, "Trend2" = 8.10, "Trend3" = -20.35)

v_values <- numeric(nrow(df)) # sukuriamas tuscias vektorius i
# kuri bus saugojamos reiksmes

for (i in 1:nrow(df)) {
  # Istraukiami atributo lygiai alternatyvoms nuo 'i' iki
  # duom. lent. pab.
  company <- df$company[i]
  return_rate <- as.character(df$return_rate[i])
  dividend <- as.character(df$dividend[i])
  trend <- df$trend[i]

  # Apskaiciuojama 'V' reiksme kiekvienai is alternatyvu

  v_values[i] <- beta_company[company] +
    beta_return[return_rate] + beta_dividend[dividend] +
    beta_trend[trend]
}

df$V <- v_values # sukuriamas naujas stulpelis ir pridedamos V
# reiksmes kiekvienai alternatyvai
df$U <- df$V + df$epsilon # apskaiciuojamas kiekvienos alternatyvos
# naudingumas pagal straipsnyje nurodyta f-le

# Jei naudingumas (U) didziausias, priskiriama reiksme 1 kintamajam
# 'choice' (daroma ta pati prielaida kaip ir straipsnyje, jog
# individas alternatyva renkasi tada ir tik tada, jei jos
# naudingumas yra didziausias)

df <- df %>%
  group_by(ID) %>%
  mutate(choice = ifelse(U == max(U), 1, 0))

####
#### Modelio pritaikymas bei aprasomoji statistika ####

```

```
####
```

```
####
```

```
# Dazniu lenteles
```

```
xtabs(choice ~ company, data = df)
```

```
xtabs(choice ~ return_rate, data = df)
```

```
xtabs(choice ~ dividend, data = df)
```

```
xtabs(choice ~ trend, data = df)
```

```
#### Modeliai
```

```
fit <- mlogit(choice ~ company + return_rate + dividend +
```

```
return_rate:trend, data = df)
```

```
summary(fit)
```

```
fit2 <- mlogit(choice ~ company + return_rate + dividend +
```

```
return_rate:trend + dividend:gender +
```

```
trend:gender, data = df)
```

```
summary(fit2)
```

## C Priedas

```
#####
```

```
##### PIRMASIS MODELIS
```

```
#####
```

```
> fit <- mlogit(choice ~ company + return_rate + dividend +
```

```
return_rate:trend, data = df)
```

```
> summary(fit)
```

```
Call:
```

```
mlogit(formula = choice ~ company + return_rate + dividend +
```

```
return_rate:trend, data = df, method = "nr")
```

```
Frequencies of alternatives:choice
```

```
1 2 3
```

```
0.26 0.44 0.30
```

```
nr method
```

```

9 iterations, 0h:0m:0s
g'(-H)^-1g = 9.05E-05
successive function values within tolerance limits

Coefficients :
                Estimate Std. Error z-value Pr(>|z|)
(Intercept):2      0.19922    0.59416   0.3353 0.7374051
(Intercept):3     -0.14054    0.77792  -0.1807 0.8566379
companyNikolaTeslaAirport -0.88887    0.72564  -1.2249 0.2205964
companyJedinstvo     -5.62423    1.52436  -3.6896 0.0002246 ***
companyEnergoprojekt  -3.74455    1.10801  -3.3795 0.0007261 ***
return_rate         35.76445   11.30621   3.1633 0.0015601 **
dividend            2.44322    0.61392   3.9797 6.9e-05 ***
return_rate:trendTrend2 -1.56403    7.86041  -0.1990 0.8422818
return_rate:trendTrend3 -119.74976   31.16840  -3.8420 0.0001220 ***
---
Signif. codes:  0      ***      0.001      **      0.01      *      0.05      .      0.1

Log-Likelihood: -20.126
McFadden R^2: 0.81237
Likelihood ratio test : chisq = 174.28 (p.value = < 2.22e-16)

#####
##### ANTRASIS MODELIS
#####

> fit2 <- mlogit(choice ~ company + return_rate + dividend +
return_rate:trend + dividend:gender + trend:gender, data = df)

> summary(fit2)

Call:
mlogit(formula = choice ~ company + return_rate + dividend +
        return_rate:trend + dividend:gender + trend:gender, data = df,
        method = "nr")

Frequencies of alternatives:choice
      1      2      3
0.26 0.44 0.30

nr method
11 iterations, 0h:0m:0s

```

```
g'(-H)^-1g = 4.49E-08
gradient close to zero
```

Coefficients :

	Estimate	Std. Error	z-value	Pr(> z )	
(Intercept):2	-0.27516	1.00102	-0.2749	0.783408	
(Intercept):3	-0.66191	1.56140	-0.4239	0.671623	
companyNikolaTeslaAirport	-1.86242	1.21665	-1.5308	0.125826	
companyJedinstvo	-11.44165	3.99031	-2.8674	0.004139	**
companyEnergoprojekt	-7.27326	2.63371	-2.7616	0.005752	**
return_rate	61.18836	27.50871	2.2243	0.026126	*
dividend	4.62998	1.73861	2.6630	0.007744	**
return_rate:trendTrend2	-1.80793	14.30272	-0.1264	0.899411	
return_rate:trendTrend3	-170.48160	59.59244	-2.8608	0.004226	**
dividend:gendermale	3.02541	2.08633	1.4501	0.147028	
trendTrend2:gendermale	-1.87496	2.85341	-0.6571	0.511120	
trendTrend3:gendermale	-23.04389	10.16585	-2.2668	0.023403	*
---					
Signif. codes:	0	***	0.001	**	0.01
				*	0.05
				.	0.1

Log-Likelihood: -10.947

McFadden R<sup>2</sup>: 0.89794

Likelihood ratio test : chisq = 192.64 (p.value = < 2.22e-16)

```
# Naudojant choice ~ company + return_rate + dividend + trend -
# visi kint. tampa stat. nereiksm. Isbandzius visas kitas
# interakcijas nepasiekiamas toks rezultatas kaip su
# 'return_rate:trend'. Taip pat, naudojant israiska
# 'x1 * x2' apskritai yra gaunami labai prasti
# rezultatai kur praktiskai visi kint. yra stat. nereiksm.
```

```
# Zymenys:
# 'x1:x2' includes only the interaction term between x1 and x2.
# 'x1*x2' includes the main effects of x1 and x2
# along with their interaction term.
```

```
# Frequencies of alternatives:choice -> alternatyvos pasirinkimo
# tikimybe
```

## Literatūra

- [1] M. P. Cristescu, R. A. Nerisanu, D. A. Mara, and S.-V. Oprea. Using market news sentiment analysis for stock market prediction. *Mathematics*, 10(22):4255, 2022.
- [2] M. Kuzmanovic, D. Makajic-Nikolic, and N. Nikolic. Preference based portfolio for private investors: Discrete choice analysis approach. *Mathematics*, 8(1):30, 2019.
- [3] M. Radavičius. Diskrečiojo pasirinkimo modeliai. 2020.