KAUNO TECHNOLOGIJOS UNIVERSITETAS

FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS

MATEMATINĖS SISTEMOTYROS KATEDRA

Karolis Sokolovas

ARBITRAŽO GALIMYBĖS PANAUDOJIMAS DAUGIALYPĖS REGRESIJOS MODELYJE

Magistro darbas

Vadovas
doc. dr. E. Valakevičius

KAUNAS, 2009

KAUNO TECHNOLOGIJOS UNIVERSITETAS

FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS

MATEMATINĖS SISTEMOTYROS KATEDRA

TVIRTINU
Katedros vedėjas
prof. habil.dr. V.Pekarskas
2009 06 04

ARBITRAŽO GALIMYBĖS PANAUDOJIMAS DAUGIALYPĖS REGRESIJOS MODELYJE

Matematikos magistro baigiamasis darbas

Vadovas
doc. dr. E. Valakevičius
2009 06 02

Recenzentas
Doc. dr. Regina Misevičienė
2009 06 03

Atliko
FMMM 7 gr. stud.
K. Sokolovas
2009 06 01

KAUNAS, 2009

KVALIFIKACINĖ KOMISIJA

Pirmininkas: Leonas Saulis, profesorius (VGTU)

Sekretorius: Eimutis Valakevičius, docentas (KTU)

Nariai: Algimantas Jonas Aksomaitis, profesorius (KTU)

Arūnas Barauskas, Vice-prezidentas projektams
(UAB „Baltic Amadeus“),

Vytautas Janilionis, docentas (KTU)

Zenonas Navickas, profesorius (KTU),

Vidmantas Povilas Pekarskas, profesorius (KTU)

Rimantas Rudzkis, habil. dr., vyriausiasis analitikas (DnB NORD Bankas)

Sokolovas K. Taking arbitrage opportunity in multiple regression model. Master‘s work in applied mathematics / supervisor dr. assoc. prof. E. Valakevičius; Department of mathematical research in systems. Faculty of Fundamental Sciences, Kaunas University of Technology. – Kaunas, 2009. – 84 p.

SUMMARY

Asset pricing problems that occur while choosing economic and social factors that are suitable for the rating of an asset are analysed in the work. In the work, a number of those factors is tried to be reduced not reducing the model accuracy. After accomplishing that, it is tried to choose a set of factors that is suitable for finding the value of the asset. The model that is described in the work selects factors to determine the prices of average type of flats in the Kaunas city centre. In the execution of the work, while avoiding the condition of multicolinearity, the time spent on computer-based calculations was reduced. In the outcome of the work it can be seen that the set of factors that best describes the prices of flats is:

1. Unemployment rate;
2. expectations (expectations is greater, when people want greater sallary);
3. Investment in Lithuania.

Many of the factors was economic factors, but some was social ones. It has been noticed, that it is important to include some social factors to increase the value of coefficient of determination. In a work there were created an algorythm to select every set of factors of different size.

SANTRAUKA

Darbe nagrinėjamos aktyvų įkainojimo problemos, iškylančios parenkant ekonominius ir socialinius veiksnius tinkančius aktyvo įkainojimui. Darbe siekiama sumažinti šių veiksnių skaičių. Tą atlikus siekiama parinkti veiksnių rinkinį, tinkantį nagrinėjamo aktyvo vertės nustatymui. Darbe aprašytas modelis parenka veiksnius butų kainoms, Kauno miesto centre nustatyti. Darbo atlikimo metu, išvengiant multikolinearumo sąlygos, pavyko sumažinti kompiuteriniams skaičiavimams reikalingą laiką. Darbe gauta, kad butų kainas Kauno miesto centre geriausiai apibūdina šių veiksnių rinkinys:

1. Vidutinė bedarbystė per paskutinius 3 mėn;
2. Lūkesčiai (Kuo daugiau žmonių įsidarbindami tikisi gauti didesnį atlyginimą);
3. materialinės investicijos Lietuvoje.

Dauguma įtrauktų veiksnių buvo ekonominiai, tačiau socialiniai veiksniai, kurie yra tiesinėje priklausomybėje su vertinamu aktyvu, neša modeliui didesnę naudą negu ekonominiai veiksniai. Jų įtraukimas į modelį yra svarbus modelio reikšmingumui. Taip pat darbe sukurtas algoritmas, skirtas perrinkti visiems skirtingiems veiksnių variantams, kad būtų įvertinta kiekvieno veiksnio sąveika su kitais veiksniais.

TURINYS

PAVEIKSLŲ SĄRAŠAS

LENTELIŲ SĄRAŠAS

1. ĮVADAS

Kiek dabar verti butai Kaune? Kaip tiksliai galime prognozuoti? Ar pirkti, ar parduoti? Nei vienas matematikas negali žinoti ateities, tačiau kiekvienas turi teisę parašyti tokį matematinį modelį, kuris nurodytų, kokia yra aktyvo vertė, jį vertinant pagal praeities įkainius. Šiame darbe bandysime rasti atsakymą į klausimą, ko reikia geram modeliui? Kokius veiksnius reikia parinkti, kad modelis būtų tikslus ir galėtų būti sudarytas per pakankamai trumpą laiką?

Šaltinyje [7] nurodyta, kad mokslininkai Chen, Roll ir Ross 1986 metais daugialypės regresijos analizės metodu sudaromam arbitražo įkainojimo modeliui pasiūlė taikyti kelių ekonominių veiksnių pokyčius (pavyzdžiui: BVP, infliacijos pokyčiai).Taip pat šie mokslininkai apibūdino, kurie veiksniai galėtų būti geriausi taikant tokius modelius. Kaip viena svarbiausių sąlygų buvo nurodyta, kad tų veiksnių pokyčiai būtų viešai pateikiami dažnai, ir greitai po kainos pokyčio fakto atsiradimo.

Tačiau, suprantama, tuo metu (XX a. 9-ajame dešimtmetyje) kompiuterinis skaičiavimo laikas buvo labai brangus. Todėl dabar atsirado galimybė sudaryti modelį, kuris ne tik nurodytų, kurie veiksniai yra geriausiai tinkami daugialypės regresijos lygtimi apramomam modeliui, tačiau ir padėtų sudaryti patį modelį. Tai yra, parenkant geriausią veiksnių rinkinį dar ir nustatyti, pagal kokius veiksnius kinta Lietuvos aktyvų vertės.

Darbe tikimasi, nustačius geriausią veiksnių rinkinį, parinkti regresijos lygtį, su didžiausiu koreguotu determinacijos koeficientu ir patenkinamu multikolinearumo koeficientu. Taip pat apibūdinti veiksnius, kurie yra turimi, ir kurie duotų didesnį tikslumą, jei egzistuotų. Taip pat darbe aprašomas ir veiksnių parinkimas norimam modeliui sudaryti.

Šiame darbe pirmiausiai aprašoma teorija, reikalinga iškeltas sąlygas atitinkančiam regresijos modeliui sudaryti. Tuo pačiu paminėtos sąlygos, kurias reikia patikrinti, sudarant regresijos modelį. Nurodoma, kaip arbitražo galimybė gali būti išnaudota, taikant regresijos modelį. Aprašyta, kaip parinkti tinkamus aktyvus bei jiems aprašyti gerai tinkančius veiksnius.

Metodinėje dalyje aprašytos ir spręstos problemos, apie tai, kaip galima sudaryti visų veiksnių perrinkimo algoritmą, ir kaip galima sumažinti kompiuterinio skaičiavimo laiką. pradiniams skaičiavimams. Kaip parinkti aktyvo ir veiksnių istorinius duomenis, tinkančius tokiai regresijos lygčiai. Pateikiamos rekomendacijos, kaip pataisyti trūkstamus duomenis. Aprašytas kitas modelis, tokios sąlygos išpildymui. Taip pat šioje dalyje yra aprašomas visas modelio algoritmas ir pateikiamas modelio pavyzdys su mažesniais duomenų masyvais tam, kad skaitytojas galėtų suprasti modelio veikimo principą, o norėdamas – ir pakartoti ekspermentą.

Tiriamojoje dalyje nagrinėjant butų kainas aprašantį regresijos modelį išnagrinėtos kai kurios duomenų transformacijos (duomenų ištiesinimas). Toliau nagrinėjamas visas modelis, pradedant nuo duomenų parinkimo, baigiant parinktų modelių tinkamumo hipotezių patikrinimu. Aprašyta naudota programinė įranga.

Pateikiamos išvados ir siūlymai, bei pastabos apie darbą.

Pastebėtina, kad net jei prisirinktume daug įvairiausių veiksnių iš praeities, mes negalėtume žinoti, kiek kurių iš jų parinkti. Kurie geriausiai veikia tarpusavyje. Tuo tikslu buvo sukurtas algoritmas, skirtas visų galimų sąveikų perrinkimui, ir apskaičiavimui, kuri sąveika geriausiai prognozuoja įvertinamo aktyvo reikšmes.

Šiuo darbu siekta parodyti skaitytojams ir artimiems draugams, nuo ko gali priklausyti kažkurio dalyko kainos. Dėl to darbo metu sukurtas universalus modelis tinkantis veiksnių įtakai tam tikram aktyvui patikrinti. Taip pat pastebėta, kad modelis gali būti nesunkiai pritaikomas ir arbitražo įkainojimo modeliui aprašyti.

Darbe siekiama patikrinti, kaip paprasčiausiai galima parinkti veiksnius parinkimui į modelį..Taip pat siekta įvertinti, kaip patogiau pasirinkti įvairius socialinius veiksnius.

2. TEORINĖ DALIS

2.1. DAUGIALYPĖS REGRESIJOS MODELIS, JO TAIKYMAS FINANSUOSE

Nagrinėkime investicinį portfelį, sudarytą iš vienos rūšies aktyvų, ir valiutos, pagal kurią skaičiuojama tų aktyvų kaina. Tarsime, kad investuojama yra labai daug kartų. Šiame darbe nagrinėsime investicinį aktyvų įvertinimo modelį, kuris aprašomas tiesine daugialype regresijos lygtimi (1), kurios priklausomas kintamasis priklauso nuo m veiksnių.

$$Y_i = a + b_i \cdot x_{1\,i} + b_2 \cdot x_{2\,i} + \ldots + b_m \cdot x_{m\,i} + e_i \qquad \textbf{(1)}$$

šioje regresijos lygtyje Y yra priklausomas kintamasis, kurio i-tąją reikšmę norime prognozuoti esant žinomoms ir pastovioms nepriklausomų kintamųjų $X_{1,} X_2 \ldots , X_m$ reikšmėms, atitinkamai: $x_{1\,i}, x_{2\,i}, \ldots, x_{m\,i}$. Dydis e_i laikomas atsitiktine modelio paklaida. Norint pilnai aprašyti šį modelį, reikia rasti lygties koeficientus $a, b_{1,} b_{2,} \ldots b_m$. Mes, vadovaudamiesi tik kelių periodų statistiniais duomenimis negalime nustatyti, nuo ko tiksliai priklauso priklausomas kintamasis, dėl to ieškosime modelio koeficientų įverčių: $\hat{a}, \hat{b}_{1,} \hat{b}_{2,} \ldots, \hat{b}_m$.

Tiesinės daugialypės regresijos modelio prielaidos, kad modelis galėtų būti taikomas sėkmingai, pagal [2] yra šios:

1) e_i normaliai pasiskirstę atsitiktiniai dydžiai;
2) visų e_i vidurkiai lygūs nuliui;
3) dydžių e_i dispersijos lygios pastoviam skaičiui;
4) modelio liekanas aprašantys nariai, t.y. dydžiai e_i , yra nepriklausomi

Paskaičiuokime priklausomo kintamojo vidurkį:

$$\boldsymbol{E}\,Y_i = a + b_1 x_{1\,i} + b_2 x_{2\,i} + \ldots + b_k x_{k\,i} \qquad \textbf{(2)}$$

iš (2) formulės matome, kad investicijose visuomet pasikliaujant (1) lygtimi – dydis $\boldsymbol{E}\,e_i$ artės prie nulio, arba kitaip sakant: jei investiciniame portfelyje bus n vienodas savybes turinčių aktyvų, tuomet gausime, kad tokiems skaičiavimams galios regresijos tiesės lygtis:

$$\hat{y}(x_{1,} x_{2,} \ldots, x_m) = \hat{a} + \hat{b}_1 \cdot x_1 + \hat{b}_2 \cdot x_2 + \hat{b}_k \cdot x_k \qquad \textbf{(3.1)}$$

Arba panašią lygtį, kai $\hat{a} = 0$:

$$\hat{y}(x_{1,} x_{2,} \ldots, x_m) = \hat{b}_1 \cdot x_1 + \hat{b}_2 \cdot x_2 + \hat{b}_k \cdot x_k \qquad \textbf{(3.2)}$$

Taip pat regresijos lygčiai taikomi tokie papildomi reikalavimai:

1) lygtyje esantys priklausomi kintamieji neturi stipriai koreliuoti tarpusavyje;
2) lygtyje esantys priklausomi kintamieji neturi koreliuoti su liekanomis.

Koeficientų įverčiai turi būti gaunami tokie, kad funkcijos (3.1) reikšmės taškuose ($x_{1\,i}, x_{2\,i}, \ldots, x_{m\,i}$) kuo mažiau skirtųsi nuo y_i verčių. Arba kitais žodžiais: visos liekamosios paklaidos būtų kiek įmanoma mažesnės.

Įverčiai $\hat{a}, \hat{b}_{1,} \hat{b}_{2,} \ldots, \hat{b}_m$ programiniu būdu nesunkiai apskaičiuojami. Apskaičiavimo metodai aprašomi [8] šaltinyje. Tam, kad galėtume apskaičiuoti koeficientus, pirmiausiai suvedame regresijos lygtį į matricinę formą:

$$Y = X \cdot B + \hat{\epsilon} \qquad \textbf{(4)}$$

Čia

$$B = \begin{pmatrix} a & b_1 & \ldots & b_m \end{pmatrix}^T \qquad \textbf{(5)}$$

$$\epsilon_i = \begin{pmatrix} \hat{\epsilon}_{1\,i} & \hat{\epsilon}_{2\,i} & \ldots & \hat{\epsilon}_{n\,i} \end{pmatrix}^T \qquad \textbf{(6)}$$

$$X = \begin{pmatrix} 1 & \hat{x}_{11} & \ldots & \hat{x}_{1m} \\ 1 & \hat{x}_{21} & \ldots & \hat{x}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{x}_{n1} & \ldots & \hat{x}_{nm} \end{pmatrix} \qquad \textbf{(7)}$$

$$Y_i = \begin{pmatrix} Y_{1\,i} & Y_{2\,i} & \ldots & Y_{n\,i} \end{pmatrix}^T \qquad \textbf{(8)}$$

tuomet įverčių matrica:

$$\hat{B} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y \qquad \textbf{(9)}$$

Toliau koeficientą a žymėsime b_0 .

Tokiu būdu gauti lygties koeficientai įrašomi į regresijos lygties modelį. Tačiau tokia lygtis dar gali ir netikti taikymui. Šaltinyje [2] nurodomos tokios galimos kliūtys taikyti regresijos modeliui:

1) Reikia patikrinti hipotezę apie bent vieno koeficiento b_j nelygumą nuliui (tikrinamas F kriterijus). Tuomet statistinės hipotezės formuluotė:

$$\begin{cases} H_0 : b_1 = b_2 = \ldots = b_m = 0 \\ H_1 : bent\ vienas\ b_j \neq 0 \end{cases}, \text{ kai } \quad j = 1 \ldots m \quad .$$

2) Jei 1-osios sąlygos nulinė hipotezė atmetama, tuomet dar reikia patikrinti ir stjudento kriterijų, t.y. šiame žingsnyje tikrinama hipotezė apie kiekvieno koeficiento b_j lygumą nuliui. Tikrinamos hipotezės visiems koeficientams:

$$\begin{cases} H_0 : \ b_j = 0, \\ H_1 : \ b_j \neq 0. \end{cases}, \text{ kai } \quad j = 1 \ldots m$$

3) Jei pirmos dvi sąlygos yra patenkinamos, tuomet modelio koeficientų stabilumui gali trukdyti *multikolinearumo* problema. Jei multikolinearumas yra didelis, tuomet reikia atmesti kai kuriuos tarpusavyje stipriai koreliuojančius veiksnius, atsižvelgiant į jų įtaką modelio determinacijos koeficientui.

4) Ketvirtoji sąlyga reikalauja, kad modelio liekanos būtų pasiskirsčiusios pagal normalųjį skirstinį. Tikrinama modelio liekanų normalumo hipotezė:

$$\begin{cases} H_0 : \ liekamosios\ paklaidos\ nepasiskirsčiusios\ pagal\ normalųjį\ skirstinį, \\ H_1 : \ liekamosios\ paklaidos\ pasiskirsčiusios\ pagal\ normalųjį\ skirstinį. \end{cases}$$

5) Paskutinė sąlyga yra nesunkiai įvertinama grafiškai. Reikia atlikti liekamųjų paklaidų grafinę analizę, ir stebėti, ar tikrai visos modelio liekanos yra išsibarsčiusios apie tiesę y=0 maždaug vienodai ir ar nėra nuo šios tiesės per daug nutolusios. Jei sąlygos tenkinamos, tuomet regresijos modelis tinka.

Pirmajai sąlygai patikrinti skaičiuojamos trys sumos:

$$SST = SSR + SSE \tag{10}$$

$$SST = \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2 \tag{11}$$

$$SSR = \sum_{i=1}^{n} \left(\hat{y}\left(\boldsymbol{x_i} \right) - \bar{y} \right)^2 \tag{12}$$

$$SSE = \sum_{i=1}^{n} \hat{e}_i^2 \tag{13}$$

čia y_i - turėta priklausomo kintamojo i-toji reikšmė, $\hat{y}\left(\boldsymbol{x_i}\right)$ - pagal modelį gauta priklausomojo kintamojo reikšmė pagal i-tąsias visų nepriklausomų kintamųjų reikšmes, $\bar{y}$ - visų priklausomo kintamojo reikšmių vidurkis, o $\hat{e}_i = \hat{y}\left(\boldsymbol{x_i}\right) - y_i$ - regresijos lygties liekamosios paklaidos (liekanos).Dydis SST yra vadinamas *visa kvadratų suma*, SSR – *regresijos kvadratų suma*, SSE – *liekamųjų paklaidų kvadratų suma*.

Tuomet yra normuojamos sumos SSR ir SSE:

$$MSR=\frac{SSR}{m} \tag{12.1}$$

$$MSE=\frac{SSE}{n-m-1} \tag{13.1}$$

kriterijaus statistika:

$$F=\frac{MSR}{MSE} \tag{14}$$

Hipotezė H_0 neatmetama, jeigu $F \leq F_{\alpha}(m, n-m-1)$.

Antrajai sąlygai tikrinti yra sudaromas naujas regresijos modelis, kuriame nepriklausomi kintamieji yra $X_{1,} \dots, X_{j-1}, X_{j+1}, \dots, X_n$ o kintamasis X_j yra šio modelio priklausomas kintamasis. Apskaičiuojama tokio modelio liekamųjų kvadratų suma: SSE_j . Tokiu atveju Stjudento kriterijaus statistika yra:

$$t=\frac{\hat{b}_j}{\sqrt{MSE/SSE_j}} \tag{15}$$

Tokiu atveju hipotezė H_0 atmetama, jei su reikšmingumo lygmeniu α gauname, jog

$$|t|>t_{\alpha/2}(n-m-1) \tag{15.1}$$

Čia $(n-m-1)$ – Stjudento kriterijaus laisvės laispniai, o $\alpha/2$ - kritinė reikšmė.

Trečioji sąlyga patikrinama paskaičiuojant *dispersijos mažėjimo daugiklį* VIF (angl. „Variance Inflation Factor“):

$$VIF=\frac{1}{1-r_j^2} \tag{16}$$

čia r_j^2 yra tokio regresijos modelio, kuriame X_j yra priklausomas kintamasis, o $X_{1,.}.., X_{j-1}, X_{j+1}, \dots, X_m$ - nepriklausomi kintamieji, determinacijos koeficientas (žr. 20 formulę). Literatūroje [2] laikoma, kad „kintamasis X_j yra per daug multikolinearus, jeigu VIF > 4“. Šis koeficientas interpretuojamas, kaip regresijos lygties koeficiento įverčio $\hat{b}_j$ dispersijos santykis su ta dispersija, kurią šis koeficientas turėtų, jei veiksnys X_j nekoreliuotų su likusiais veiksniais. Trečiąja sąlyga darbe ir remsimės, atskirdami tinkamus veiksnius nuo per daug multikolinearių.

Ketvirtoji sąlyga darbe liekamųjų paklaidų pasiskirstymo pagal normalųjį skirstinį hipotezė yra tikrinama Kolmogorovo kriterijumi pagal [9] šaltinį. Tikrinama tokia suderinamumo hipotezė:

H_0: $F_f(x) \equiv F_0(x)$. $F_f(x)$ - tam tikro, stebėto, tolydaus skirstinio funkcija. $F_0(x)$ - nepriklausantis nuo nežinomų parametrų skirstinys. Nagrinėjamo modelio atveju – Normaliojo skirstinio funkcija. Hipotezė tikrinama tokiu būdu: paimamas visų liekanų vektorius $\{e_i\}$, surikiavus didėjimo tvarka gaunama variacinė seka $\{e_{(i)}\}$. Tuomet paskaičiuojamos funkcijos reikšmės gautuose variacinės sekos reikšmėse: $\{F_0(e_{(i)})\}$. Šios gautos reikšmės reikalingos dydžio D_m įvertinimui, kai

$$D_m = max\left(D_m^a, D_m^b\right) \quad \textbf{(17)}$$

Be to:

$$D_m^a = \max_{1\le k\le m}\left(\frac{k}{m} - F_0(e_{(k)})\right) \quad \textbf{(18.1)}$$

ir

$$D_m^b = \max_{1\le k\le m}\left(F_0\left(e_{(k)}\right) - \frac{k-1}{n}\right) \quad \textbf{(18.2)}$$

čia $e_{(k)}$ – k-tasis variacinės eilutės narys. Tuomet jei tenkinama sąlyga:

$k_{1-\alpha} > D_n$ tuomet hipotezė H_0 yra atmetama, t.y. skirstinys normalusis su tikėtinumu $1-\alpha$.

$$k_{1-\alpha} \approx \sqrt{\frac{-\ln\left(\frac{\alpha}{2}\right)}{2\cdot m}} - \frac{1}{6\cdot m} \quad \textbf{(19)}$$

kai $n > 50$. Čia α - reikšmingumo lygmuo, m – liekamųjų paklaidų imties dydis.

Tikrinant **penktąją sąlygą**, tam, kad pamatytume, kaip priklausomas kintamasis Y priklauso nuo konkretaus veiksnio X_j - atliekami tokie veiksmai:

– pirmiausiai sudaromas regresijos modelis su veiksnių rinkiniu, neįtraukiant to veiksnio, nuo kurio priklausomybė yra nagrinėjama, ir tuo pačiu priklausomu kintamuoju. Randamos šio modelio liekamosios paklaidos $e_k\left(y, j\right)$.

– Po to sudaromas regresijos modelis su veiksnių rinkinių, neįtraukiant to veiksnio, kurio priklausomybė yra nagrinėjama, priklausomu kintamuoju naudojant tą veiksnį, kuris išrinktas iš veiksnių rinkinio. Randamos šio modelio liekamosios paklaidos $e_k\left(x_j, j\right)$.

– Braižomas grafikas iš reikšmių $\left(e_k\left(x_j, j\right); e_j\left(y, j\right)\right)$. Jei priklausomybė yra artima tiesinei, tuomet X_j tinka modeliui.

Kai tampa aišku, kuris modelis yra priimtinas, o kuris ne – reikia atsirinkti, kuris modelis yra geresnis, o kuris prastesnis. Tam literatūroje [1], [2] yra rekomenduojama naudoti determinacijos koeficientą: r^2 . Jį galima apskaičiuoti žinant (11) ir (12) sumų reikšmes nagrinėjamam modeliui. Šio koeficiento išraiška (20):

$$r^2 = \frac{SSR}{SST} \quad \textbf{(20)}$$

determinacijos koeficientas nurodo, kurią priklausomo kintamojo pokyčių dalį paaiškina regresijos modelis. Nagrinėjamame modelyje yra naudojamas *koreguotasis determinacijos koeficientas* (21).

$$r_{adj}^2 = 1 - \frac{n-1}{n-m-1} \cdot \left(1 - r^2\right) \quad \textbf{(21)}$$

Šis koeficientas yra naudojamas tuomet, kai kintamųjų skaičius m nedaug skiriasi nuo stebėjimų skaičiaus n. Šis koeficientas, kaip ir aprašytas (20) formule, interpretuojamas taip pat: kuo jis artimesnis vienetui, tuo geriau priklausomojo kintamojo Y reikšmes aprašo regresijos modelyje esančių nepriklausomų kintamųjų pokyčiai.

Tokiu būdu yra parenkamas modelis, turintis mums tinkamos vertės multikolinearumą ir didžiausią determinacijos koeficientą.

Pastebėkime, kad nagrinėtame modelyje visus nepriklausomus kintamuosius pakeitę į ekonominius ar socialinius veiksnius gauname sąryšį, kuriame aktyvo grąžą galime išreikšti per m veiksnių: $(\hat{F}_1, \hat{F}_{2,} ... \hat{F}_m)$. Tokiu būdu yra sudaroma regresijos funkcija, įvertinti aktyvo vertei, esant tam tikroms nagrinėjamų veiksnių reikšmėms.

2.2. ARBITRAŽO GALIMYBĖ

Arbitražo galimybe yra laikoma tokia situacija, kai dvejose vietose vienu metu aktyvas yra įvertintas nevienodai. Pats arbitražas (angl. - „arbitrage“) literatūroje [7] apibrėžiamas taip:

„Arbitražas yra galimybė, pasinaudojant kainų skirtumu tarp dviejų arba daugiau rinkų, atsidurti patogesnėje prekybos pozicijoje: tokiu atveju pelnas yra aktyvo verčių skirtumas tarp rinkos, kur yra didžiausia ir mažiausia kaina“. Žinodami determinacijos koeficientą pagal jį galime įvertinti, kiek rizikinga yra pasinaudoti arbitražo galimybe. Tai yra – modelio determinacijos koeficientas padeda mums apsispręsti, kurį metodą taikyti ieškant arbitražo.

Literatūroje [1] ir [5] nurodoma, kad arbitražo galimybę galima taikyti dviem skirtingais atvejais: Pasinaudojant arbitražo įkainojimo teorija (AĮT - angl. „Arbitrage Pricing Theory – APT“) arba siekiant didesnio vertės nustatymo tikslumo. Nagrinėjant pastarąjį modelį galima įvertinti, kiek pinigų reikia siūlyti akcijų biržoj už vertybinius popierius, ūkininkui už mėsą, ar pardavėjui už parduodamą butą pirmiau, negu rinka spėja sureaguoti į kainų pokyčius. Taip pat, jei į lygtį įtraukti nepriklausomi priežastiniai veiksniai, tuomet pagal naudingą ekonominę informaciją investuotojas gali spėti apie vieno ar kito veiksnio pokyčius. Tuomet pasinaudodamas lygtimi investuotojas gali spėti, kiek daug keisis aktyvo kaina, ir tuo pačiu gali spręsti, ar verta pradėti pirkimo/pardavimo veiksmus, ar geriau dar kurį laiką palaukti.

Darbe aprašytas ir programiškai sukurtas modelis gali būti naudojamas išnaudoti arbitražo galimybę pelnui gauti tiek daugialypės regresijos metodu, tiek naudojant arbitražo įkainojimo teorijos pagrindus. Tačiau toliau bus daugiau gilinamasi į daugialypės regresijos metodu įkainotus aktyvus, nes darbe daugiau nagrinėjamos rinkų tendencijos, o ne didelis skaičius aktyvų, kuris yra reikalingas arbitražo įkainojimo modeliui. Tačiau, jei teigsime, kad visi aktyvai yra perkami vienoje rinkoje – tuomet rinkos tendencijas galima aprašyti kaip vieno investicinio portfelio vertės kitimą. Tokia galimybė detaliau aprašoma šaltinyje [1].

Kaip nurodyta [3] – arbitražu įvertinti galima vieną aktyvą, kai pastebima, jog pagal modelį nustatyta aktyvo vertė statistiškai reikšmingai skiriasi nuo dabartinės rinkos vertės. Tai yra – jei aktyvas realybėje kainuoja pastebimai brangiau, negu apskaičiuojama pagal didelio patikimumo modelį, tuomet aktyvą reikia pasiskolinti ir parduoti (angl – „short sell“). Po to, aktyvo kainai rinkoje sumažėjus – aktyvas nuperkamas pigiau, negu buvo parduotas ankstesniu laiko momentu. Kitu atveju, jei aktyvo kaina rinkoje yra mažesnė negu apskaičiuota pagal modelį - nusiperkamas aktyvas, ir rinkoje jo kainai padidėjus iki realios kainos – aktyvas parduodamas brangiau, negu buvo nupirktas. Kaip nurodoma [6] – Arbitražo įkainojimas, skirtingai negu daugelis kitų modelių, yra laikojimas vieno periodo modeliu.

2.3. AKTYVŲ IR VEIKSNIŲ PARINKIMAS INVESTICIJOMS

Bendru atveju literatūroje [1] laikoma, jog aktyvas gali būti bet koks vertybinis popierius, ar investicinis portfelis, sudarytas iš vertybinių popierių. Tokiu atveju laikoma, kad vertybiniai popieriai gali būti:

- buto nuosavybės teisės;
- automobilio techninis pasas;
- metalo nuosavybės teisės;
- kiti dokumentai, nurodantys teisę į tam tikrą vertę turintį turtą.

Darbe nagrinėjamas butų vertės pokytis bėgant laikui: tiriamas vidutinio būsto, esančio Kauno miesto centre, kvadratinio metro kainos pokytis. Taip pat analizuojamos kitų aktyvų kainos:

- vidutinės apvaliosios medienos kainos Kauno rajono urėdijoje (litais už kubinį metrą);
- vidutinis butų kainos pagal S&P indeksą Los Andželo mieste, JAV;
- vištienos kainos pokytis JAV.

Nagrinėti įvairių veiksnių įtaką labai skirtingiems aktyvams yra gerai tuo, kad skirtingu laikotarpiu gali būti skirtinga paklausa ir pasiūla skirtingose rinkose.

Literatūroje [1] ir [3] nurodomi ekonomistų Chen, Roll ir Ross siūlomi veiksniai:

Daugialypiams modeliams geriausiai tinka tie veiksniai, apie kurių pokyčius galima sužinoti beveik iš karto, tai gali būti tokie ekonominiai veiksniai, kaip:

- infliacijos pokyčiai;
- nagrinėjamos šalies BVP pokyčiai;
- ilgalaikių vyriausybinių obligacijų pokyčiai;

– investuotojų pasitikėjimas investicijomis, priklausantis nuo grąžos, gaunamos iš turimų aktyvų.

Pavyzdžiui infliaciją šie ekonomistai pasiūlė aprašyti tokiu būdu: „infliacijos dydis, tikėtinas mėnesio pradžioje, minus infliacijos dydis, išmatuotas mėnesio pabaigoje.“ (pagal [1]).

Sukurtoje programoje yra aprašomi du modeliai, pirmasis nagrinėja visus galimus veiksnių rinkinius. Antrasis modelis leidžia pasirinkti tuos veiksnių rinkinius, dėl kurių didelės tarpusavio koreliacijos mes esame užtikrinti; tai yra investuotojas yra užtikrintas, kad dėl multikolinearumo sąlygos nepatenkinimo tie du veiksniai tikrai nebus parinkti viename modelyje. Tokiu būdu galime atrinkti daug daugiau įvairių veiksnių rinkinių per daug trumpesnį laiką.

2.4. INVESTICINIŲ MODELIŲ PARINKIMO PROBLEMOS IR DARBE SPRENDŽIAMI UŽDAVINIAI

Kaip nurodyta [1] šaltinyje – galima kurti ir maišytus modelius (angl. „mixed-model“). Juose galima apjungti skirtingus modelius, gaunant iš jų norimus rezultatus. Savo darbe taikau arbitražo įkainojimo teorijos ir daugialypės regresijos modelius. Pirmiausiai parenkami ekonominiai ir socialiniai veiksniai, kurie, tikimasi, turės įtaką priklausomam kintamajam. Tuomet parenkamas geriausias regresijos modelis. Galiausiai įvertinamas modelio tikslumas, ir tokiu būdu galime įvertinti, ar galime pasinaudoti arbitražo galimybe konkrečiam aktyvui, ar bandymas pasinaudoti arbitražo galimybe yra per daug rizikingas dėl prekybos mokesčių ir mums nežinomų veiksnių įtakos. Laikoma, kad nagrinėjama visa rinka, iš kurios parinka daug tos pačios rūšies aktyvų, tokiu būdu sumažinant diversifikuojamą riziką.

Apibūdindami daugialypės regresijos su ekonominiais ir socialiniais veiksniais aktyvų įkainojimo modelį, paprasčiausiais žodžiais galime sakyti, kad šio modelio pagalba galime lengviau apsispręsti, ar verta pirkti aktyvą už mums pasiūlytą kainą. Arba kitu atveju: ar verta parduoti aktyvą už tą kainą, kurią pasiūlė galimas pirkėjas. Ir tuo pačiu galime sužinoti, kokio pelno tikėtis tą aktyvą pardavus.

Trumpai aprašysime arbitražo įkainojimo teoriją. APT modelį aprašo toks daugialypės regresijos lygties atskiras atvejis:

$$E\left[\hat{R}_i\right]-R_f=\lambda_1 b_{i1}+\lambda_2 b_{i2}+\ldots+\lambda_m b_{im} \qquad \textbf{(22)}$$

Čia R_f - nerizikinga palūkanų norma nagrinėjamu momentu. Darbe laikysime, $\bar{R}_i$ - priklausomojo kintamojo grąža. $\lambda_1, \ldots, \lambda_m$ yra ekonominiai veiksniai sunormuoti taip, kad vidurkis būtų lygus nuliui, ir jie būtų paiskirstę pagal normalųjį skirstinį. Vidurkio lygumas nuliui yra reikalingas todėl, kad modelyje neatsirastų konstanta. Akivaizdu, kad skaičiuojant vidutinę tokio modelio reikšmę ir esant lygties konstantai, nelygiai nuliui, bus gautas garantuotas pelnas, nepriklausantis nuo veiksnių. O tai reiškia, kad galima arbitražo sąlyga pasinaudoti bet kuriuo momentu. Gauta sąlyga leistų gauti neribotą pelną. O tai prieštarauja arbitražo sąlygos apibrėžimui. Taip pat kitos sąlygos galioja tokios pačios, kaip ir daugialypės regresijos lygčiai. Kairėje (22) lygybės pusėje esantis skirtumas skaičiavimo metu gali būti suvedamas į vieną veiksnį; iš priklausomojo kintamojo i-tosios reikšmės atimama nerizikinga palūkanų norma tuo momentu. Toliau lygtis nagrinėjama kaip atskiras (3.2) funkcijos atvejis. Šio modelio rizika yra tame, jog esant mažam determinacijos koeficientui, reikia imti daugiau aktyvų, kad būtų galima išlyginti nesisteminę (diversifikuojamą) riziką. Dėl to arbitražo įkainojimo modelis geriau tinka aktyvams, kurie gali būti perkami didesniais kiekiais ir kurių kaina yra vienareikšmė. Mūsų darbe yra nagrinėjami butai, kurie dažniausiai kainuoja didesnius pinigus negu paprastos akcijos. Todėl šiuo atveju mums reikia taikyti modelį, kuris turi didesnį determinacijos koeficientą.

Darbe siekiama sukurti modelį, kurio pagalba būtų galima parinkti veiksnių rinkinį, tiksliausiai prognozuojantį aktyvo kainas. Gautų veiksnių rinkinio tiesinio darinio pokyčiai turėtų gerai paaiškinti priklausomojo kintamojo pokyčius, kad pasinaudojus šiuo modeliu, būtų galima pasinaudoti arbitražo galimybe, esant palankiai situacijai. Sukūrus tokį modelį siekiama išnagrinėti veiksnių rinkinius geriausiai apibūdinančius Kauno miesto centro būstų kainų kitimą, ar bet kokius kitus aktyvus. Taip pat parinkti, kurį modelį kurio aktyvo vertinimui yra geriau taikyti.

3. METODINĖ DALIS

3.1. GERIAUSIO VEIKSNIŲ RINKINIO PARINKIMAS MODELYJE

Darbo užduoties vykdymo eigoje buvo spręsta problema – kaip rasti geriausią veiksnių rinkinį daugialypės regresijos lygtyje. Pirmiausiai didesnio aiškumo dėlei buvo išsikelta sąlyga rezultatus pateikti kompiuterinėje rezultatų tekstinėje byloje. Dėstoma tokiu būdu: pradžioje išrenkamas geriausias rinkinys iš tų, kurie turi mažiausią galimą veiksnių skaičių. Geriausiu rinkiniu laikomas tas, kurio multikolinearumą aprašantis dispersijos mažėjimo daugiklis VIF nei vienam veiksniui neviršija 4, ir kurio determinacijos koeficientas yra didžiausias tarp tą patį veiksnių skaičių turinčių rinkinių. Po to rašomas to modelio rezultatas, kuris duoda tiksliausius rezultatus, turėdamas vienu veiksniu daugiau. Ir taip toliau, kol galiausiai įrašomas visus veiksnius apimantis modelis. Vartotojas pasižiūrėjęs į šią bylą gali nuspręsti, kiek veiksnių turintis modelis jam yra priimtiniausias.

Aišku, kadnagrinėsime modelį, priklausantį nuo m skirtingų veiksnių, tuomet programiniu būdu turės būti išnagrinėta iš viso $K=\sum_{i=1}^{m} C_m^i = 2^m - 1$ galimybių. Pavyzdžiui, kai nagrinėjama 15 veiksnių, tai yra nagrinėjama 32767 skirtingų rinkinių, kai 17 veiksnių – 131071 veiksnių, o kai 20 – tuomet jau nagrinėjami 1048575 veiksniai. Turint 1.6GHz spartos procesorių, buvo atliktas testas, skirtas nustatyti laikui, reikalingam rasti optimalų veiksnių rinkinį, atitinkantį nustatytus optimalumo kriterijus. Laikas, per kurį parenkamas geriausių veiksnių rinkinys, ir kiti dydžiai, gauti eksperimente yra nurodyti 1 lentelėje.

1. lentelė: Resursai, reikalingi optimaliam veiksnių rinkiniui rasti

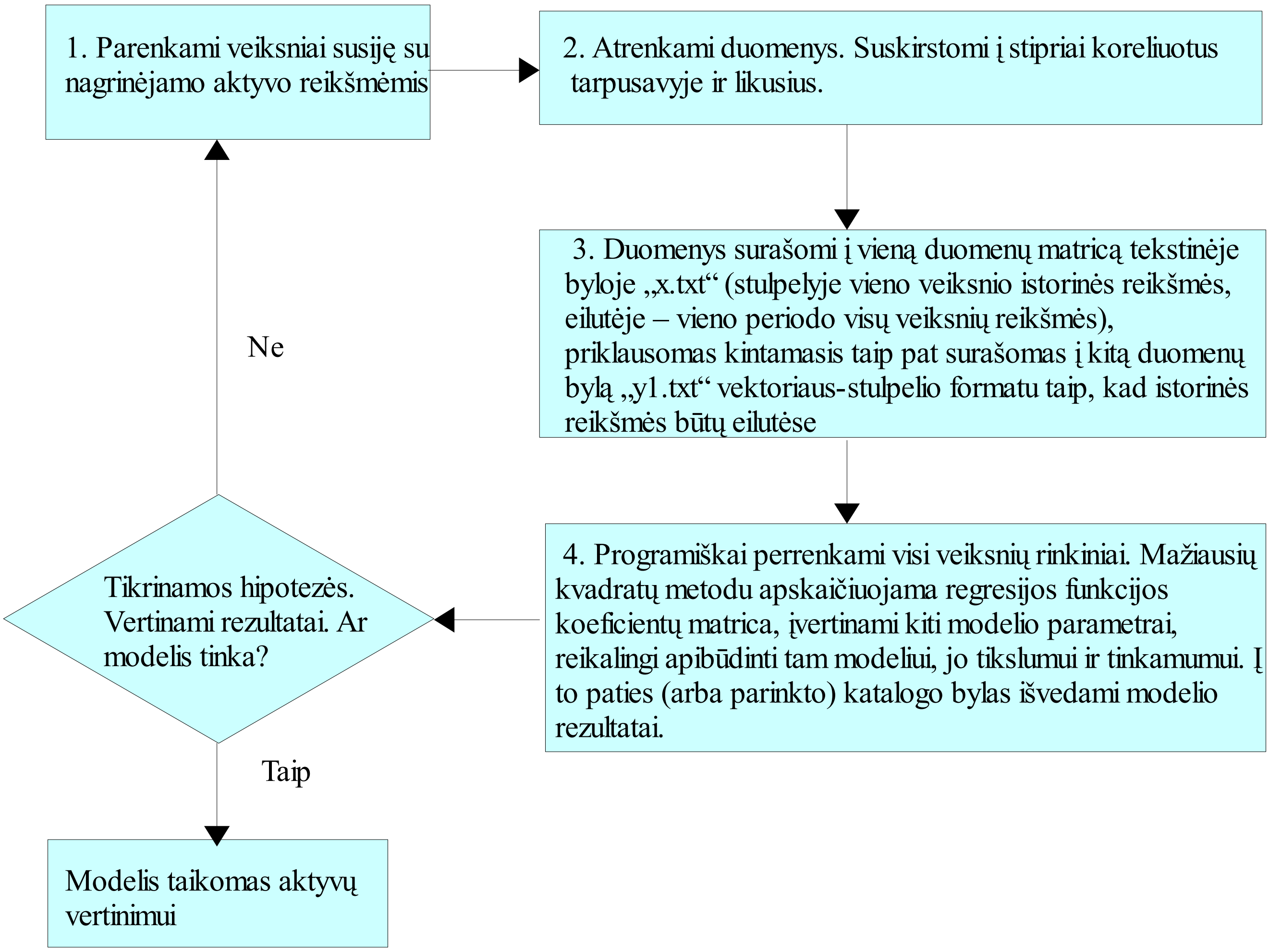
Veiksnių sk. (m)	Kompiuterio laikas (sekundėm)	Galimybių skaičius ($2^m - 1$)	Vidutinis laikas, skirtas patikrinti vienai galimybei
9	10.39	511	0.020
10	24.14	1023	0.023
11	51.39	2047	0.025
12	116.81	4095	0.029
13	330.16	8191	0.040
14	1651.3	16383	0.1

Iš 1 lentelės matome, kad kompiuteriu tiesiogiai parenkant optimalų modelį iš, sakykime, 20 veiksnių ($2^{20} - 1 = 1048575$), reikėtų nagrinėti ilgą laiką (net jei laikytume, kad vienai galimybei apskaičiuoti reikės vidutiniškai 0.11 sekundės – apdoroti 20 veiksnių duomenų bazei reikėtų apie 32 valandų. Nors aišku tai, kad su kiekvienu papildomu veiksniu vidutinis laikas, reikalingas apdoroti vienai galimybei, didėja). Darbe spręsime ir šią problemą: kaip sutrumpinti optimalių veiksnių rinkinių parinkimui reikalingą laiką.

3.2. MODELIO SUDARYMAS, ALGORITMAI, SUDĖTINGUMAS

Modeliui sudaryti buvo sukurti keli skaičiavimo algoritmai (interpoliavimui, rezultatų apskaičiavimui ir išvedimui) su matematinio skaičiavimo įranga SCILAB, o pagalbiniai skaičiavimai (duomenų parinkimas, suskirstymas, hipotezių patikrinimas) su Open Office „Spreadsheet“ programa. Programa yra panaši į paketo MS Office skaičiuoklę „Excel“.

1 pav. Programos veikimo algoritmas



Rezultatuose pateikiamas tiek modelio determinacijos koeficientas, tiek kereguotasis determinacijos koeficientas. Į koreguotąjį determinacijos koeficientą reikia atsižvelgti tuomet, kai koeficientų skaičius k nedaug skiriasi nuo matavimų skaičiaus n. Toliau darbe remsimės tik koreguotuoju determinacijos koeficientu, nes didėjant stebėjimų skaičiui n šis koeficientas artėja prie determinacijos koeficiento.

Regresijos funkcija yra tinkama naudoti tuomet, kai prognozuojamos reikšmės nėra didesnės už priklausomojo kintamojo, pagal kurį prognozė atlikta, didžiausią reikšmę, ir nėra mažesnė už to kintamojo mažiausią reikšmę.

3.3. DUOMENŲ SUDARYMAS IR PARINKIMAS MODELIUI

Duomenys nepriklausomiems kintamiesiems parenkami taip, kad būtų bendruose perioduose su priklausomu kintamuoju. Tuo atveju, kai veiksniui ankstesnio periodo koreliacija su priklausomu kintamuoju yra didesnė, ar pakankamai didelė, skaičiavimuose galima naudoti keliais periodais paslinktus duomenis. Toks būdas leistų įvertinti galima kainų pokytį ateityje, arba tiesiog lengviau gauti ankstesnių periodų duomenis. Jei duomenys nėra pakankami, tuomet užpildyti trūkstamas spragas galima naudojant glodinantį splainą su vienodais arba skirtingais svoriais prie įverčių, jei duomenų skaičius skirtingiems periodams yra nevienodai patikimas.

Jau turint visus duomenis ir priklausomojo kintamojo statistines reikšmės reikia atlikti tų duomenų pradinę analizę – suskirstyti veiksnius į stipriai koreliuotus tarpusavyje, ir į tuos, tarp kurių koreliacija nėra didelė. Šį didumą turėtų nustatyti pats investuotojas, priklausomai nuo jo kompiuterio galimybių. Parinkęs mažesnį dydį, investuotojas rizikuoja netekti kai kurių veiksnių rinkinių tolesnėje analizėje, tačiau tokiu atveju išlošia programinio laiko. Į rinkinį, kur kintamieji yra stipriai koreliuoti tarpusavyje, parenkami tie veiksniai, kurie tikrai neįeis į vieną geriausią rinkinį dėl multikolinearumo sąlygos (žr. (16) ir (23) formules). Dėl to į koreliuotų veiksnių rinkinį rekomenduojame parinkti tuos kintamuosius, kurių j-tojo veiksnio determinacijos koeficientas, kai jis yra priklausomas kintamasis regresijos lygtyje nuo kitų veiksnių, yra didesnis negu 0.75. Iš (16) sąlygos turime, kad jei egzistuoja multikolinearumas, tuomet veiksnio determinacijos koeficientas anksčiau minėtame modelyje turės būti ne mažesnis, negu:

$$r_j^2 = \lim_{VIF \to 4} \frac{(VIF - 1)}{VIF} = 0.75 \tag{23}$$

Kai $VIF \to +\infty$, tuomet $r_j^2 \to 1$.

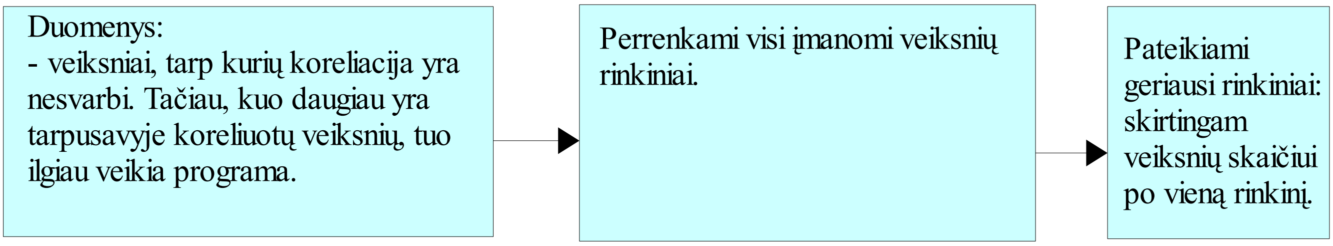
Į geriausią parinktą modelį gali įeiti vienas veiksnys iš gerai koreliuotų tarpusavyje, ir keli veiksniai iš kito veiksnių rinkinio. Suprantama, kad tuomet, kai yra labai daug ekonominių veiksnių (tiek, kad modelio įvykdymas kompiuteriu užtruktų ilgiau, negu investuotojas tikisi), prioritetu reikėtų patinkti tuos, kurių koreliacija su priklausomu kintamuoju statistiškai reikšmingai skiriasi nuo nulio.

Pasinaudojant modeliu ir parinkus geriausią veiksnių rinkinį reikia patikrinti hipotezes, apie to rinkinio tinkamumą regresijos lygčiai. Tuo atveju, kai bent viena hipotezė nėra atmetama (nepriimama bent viena alternatyvi hipotezė), reikia išsiaiškinti, kas tą įtakojo. Šias kliūtis pašalinus (pavyzdžiui vieną periodą, buvo išskirtis ar pan.) modelį reikėtų perskaičiuoti iš naujo, taip parenkant optimalų veiksnių rinkinį, tokį, kurį naudojant būtų tenkinamos visos statistinės hipotezės, naudojamos apibūdinti regresijos lygčiai.

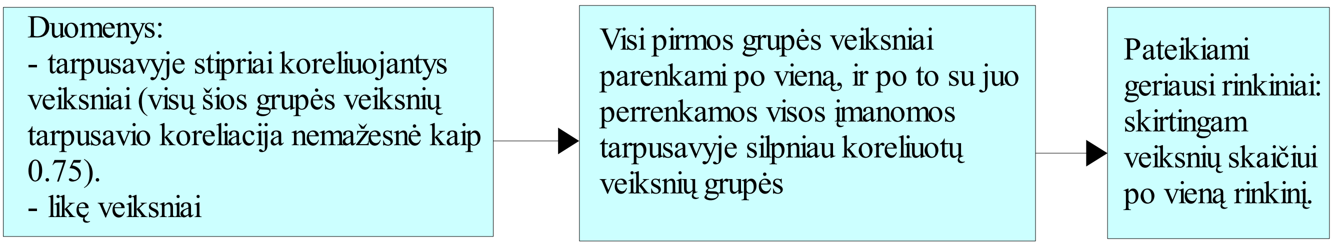
3.4. AKTYVŲ ĮKAINOJIMO STATISTINIS MODELIS

Yra sukurti du modeliai, vienas skirtas greitesniam skaičiavimui, kai keli veiksniai yra stipriai koreliuoti tarpusavyje, kitas – visų galimų veiksnių grupių perrinkimui, kai koreliacijos yra įvairios. Suprantama, kad kuo daugiau yra veiksnių, tuo daugiau skiriasi šių modelių veikimo laikas (iki 1000 kartų ir daugiau). Trumpai šie du modeliai apibūdinti 2. ir 3. paveiksluose.

2 pav: Pirmasis aktyvų įkainojimo statistinis modelis



3 pav: Antrasis aktyvų įkainojimo statistinis modelis



3.5. MODELIO SUDARYMO PAVYZDYS

Sakykime, kad turime dviejų metų sklypų kainas už arą ketvirčių pabaigoje. Regionas yra vienas ir tas pats, o kainos, tūkstančiais litų, yra pateiktos 2 lentelėje. Toliau, sakykime, kad mums reikia parinkti įvairius veiksnius, galbūt įtakosiančius sklypų kainas (žr. 2 lentelė).

2. lentelė: Glodinančio arba interpoliacinio splaino skaičiavimo metodo pavyzdys: duomenys ir rezultatai

Periodas	Vieno aro kaina (tūkst. LT)	Svoriai (svorių failas)	Aritmetinis vidurkis (duomenų failas)	Prognozuotos reikšmės (tt =2) „smoothingy.txt“ (PPzingsn = 1)
I	10 12 13 9	4	11	11
II	-	0	0	9.12336
III	11 9.5	2	10.25	10.25
IV	13	1	13.5	12.50164
V	14 12 16	3	14	14
VI	20	1	20	13.52632
VII	13 12	2	12.5	12.5
VIII	13 11 15	3	13	13
IX	11 9 10	3	10	10
X	10 8.5 8.5	3	9	9
XI	9 10.5 9	3	9.5	9.5
XII	9.3 9.5 10	3	9.6	9.6

Sakykime, pasirenkame keturių veiksnių istorinius duomenis, apibūdinančios pačių periodų, kaip ir nagrinėjamos sklypų kainos. Duomenys pateikti 3 lentelėje.

3. lentelė: Veiksnių parinkimo pavyzdys. Pradiniai ir papildyti duomenys

Periodas	Veiksnys nr. 1	Veiksnys nr. 2	Veiksnys nr. 3	Veiksnys nr. 4
	Pradiniai	Pradiniai	Pradiniai	Pradiniai
I	10	2.5	1	1.8
II	11	3	0.9	2.1
III	11.2	2.8	0.8	1.9
IV	12.8	3.1	1.1	1.65
V	12.2	3.8	1.4	1.6
VI	12.6	3.3	1.3	1.45
VII	11.5	2.9	1.2	1.6
VIII	9.9	3.2	1.1	1.5
IX	9.2	3.5	1	1.9
X	9.3	3.2	0.7	2.1
XI	9.4	2.8	0.8	2
XII	9.1	2.6	0.7	1.95

Jei kurio nors veiksnio vienas duomuo yra nežinomas, tuomet jį galima prognozuoti apskaičiuojant splainą su programiniu kodu „spline.sci“. Tuomet tam, kad žinotume, ar modelyje nebus multikolinearumo, paskaičiuojame koreliacijas tarp visų veiksnių. Gauta koreliacinė matrica pateikta 4 lentelėje.

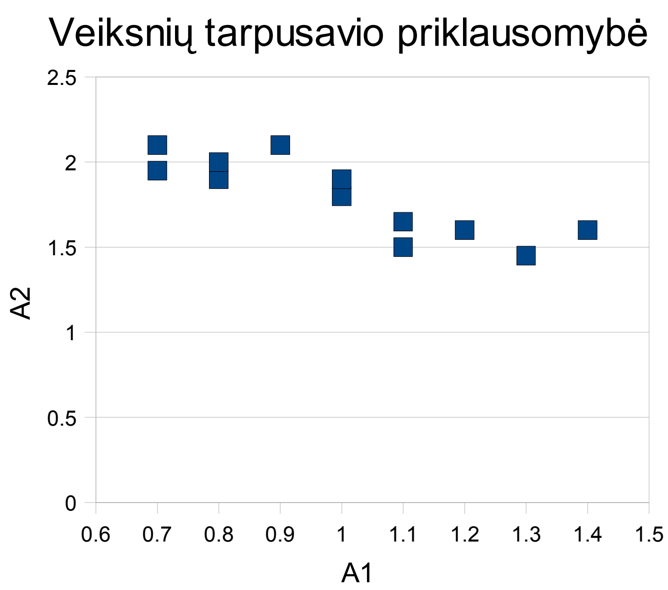
4. lentelė: Veiksnių tarpusavio koreliacija. Koreliacija su priklausomu kintamuoju. Pavyzdys.

veiksnys	**1/B1**	**2/B2**	**3/A1**	**4/A2**
1/B1	1	0.33	0.72	-0.61
2/B2	0.33	1	0.59	-0.35
3/A1	0.72	0.59	1	-0.86
4/A2	-0.61	-0.35	-0.86	1
sklypai	**0.71**	**0.47**	**0.92**	**-0.96**

Taip pat parenkant veiksnius priklausomam kintamajam būtina atlikti porinio stebėjimo grafikų analizę, o tikrinant konkretaus metodo tinkamumą – liekamųjų paklaidų arba standartizuotųjų liekanų grafikų analizę. Porinio stebėjimo grafikas parodo, ar dviejų kintamųjų priklausomybė yra tiesinė. Tuo tarpu liekamųjų paklaidų grafikas nurodo, kad regresios modelis tinka, jei visos liekanos yra išsibarsčiusios apie tiesę y=0 maždaug vienodai nelabai dideliu atstumu.

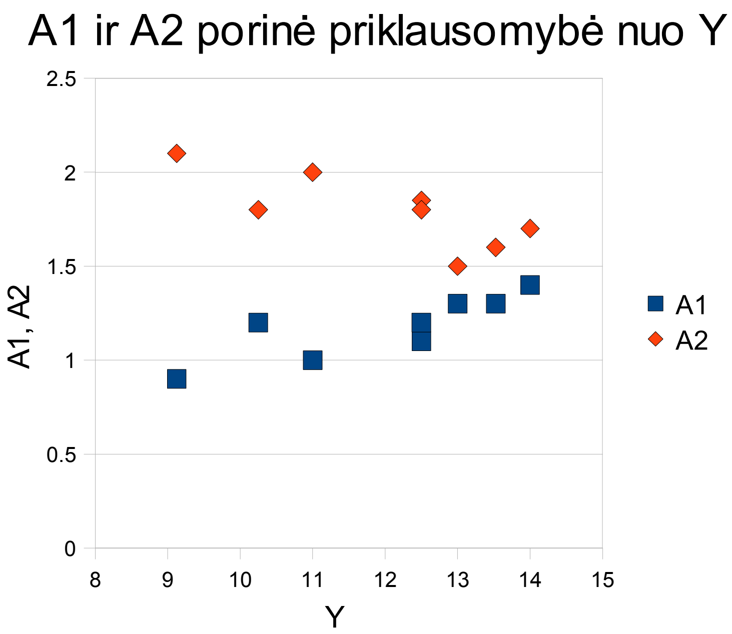
Prieš toliau nagrinėdami modelį patikrinkime, ar visi veiksniai su priklausomu kintamuoju susiję tiesiniu ryšiu.

4 pav. Veiksnių tarpusavio priklausomybė. Porinio stebėjimo grafikas. Pavyzdys

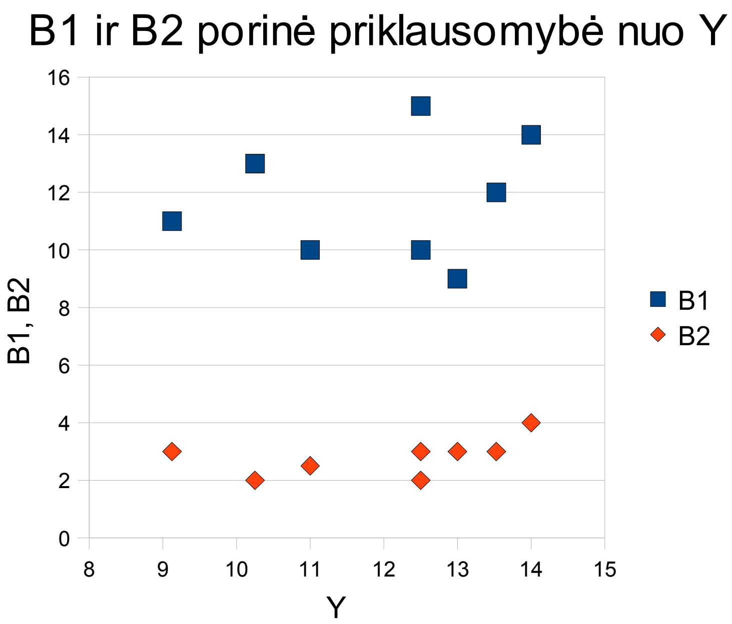


Iš 4 lentelės ir 4 pav. galime matyti, kad 3 ir 4 veiksnys yra stipriai koreliuoti ir tarpusavyje tiesiškai priklausomi, todėl aišku, kad jie į vieną modelį neįeis. Šiuos veiksnius galėtume atskirti. Veiksnį nr. 1 pavadinkime B1, veiksnį nr. 2 – B2, veiksnį nr. 3 – A1, veiksnį nr 4 – A2. Tuomet Galutinius veiksnių duomenis iš 3 lentelės išdėliokime duomenų byloje „x.txt“ stulpeliuose tokia tvarka: A1 A2 B1 B2 **1**, kai „1“ reiškia stulpelį iš vienetų, apibrėžiantį konstantą. Nubraižę veiksnių porines priklausomybes nuo priklausomojo kintamojo (žr 5 ir 6 pav.), pastebime, kad A1 ir A2 veiksnių tiesinė priklausomybė nuo priklausomo kintamojo yra geresnė. Byloje „y1.txt“ tegul būna surašytos atitinkamų periodų sklypų vertės reikšmės.

5 pav. Veiksnių A1 ir A2 porinės priklausomybės. Pavyzdys



6 pav. Veiksnių B1 ir B2 porinės priklausomybės. Pavyzdys



Toliau per Scilab programą įvykdome pograminį kodą iš bylos „koreliuotiems.sci“. Pasirenkame katalogą, kuriame yra dvi ką tik sukurtos bylos, ir programos prašymu įvedame „2“. Tai reiškia, kad pirmi du veiksniai yra stipriai koreliuoti, ir kad jų nebus viename modelyje. Suprantama, jei padarysime klaidą – silpną priklausomybę įvertinsime kaip reikšmingą – ši programa gali ir nepateikti geriausio rezultato. Gautą rezultatą galime patikrinti programa „nekoreliuotiems.sci“, tačiau įvykdymas užtruks ilgiau (žr 6 lentelę).

Duomenų bylų „x.txt“ ir „y1.txt“ turinys pateiktas 5 lentelėje.

5. lentelė: Duomenų apie veiksnius, bylos „x.txt“ ir priklausomojo kintamojo bylos „y1.txt“, pavyzdžiai

„x.txt“					„y1.txt“
10	2.5	1	1.8	1	11
11	3	0.9	2.1	1	9.12
11.2	2.8	0.8	1.9	1	10.25
12.8	3.1	1.1	1.65	1	12.5
12.2	3.8	1.4	1.6	1	14
12.6	3.3	1.3	1.45	1	13.53
11.5	2.9	1.2	1.6	1	12.5
9.9	3.2	1.1	1.5	1	13
9.2	3.5	1	1.9	1	10
9.3	3.2	0.7	2.1	1	9
9.4	2.8	0.8	2	1	9.5
9.1	2.6	0.7	1.95	1	9.6

Programos kodo „koreliuotiems.sci“ ir programos kodo „nekoreliuotiems.sci“ įvykdymu gauti rezultatai palyginti 6 lentelėje.

6. lentelė: Pavyzdžio rezultatų palyginimas, įvykdant du skirtingus programos kodus

Pavadinimas	„koreliuotiems.sci“		„nekoreliuotiems.sci“	
Kai kurie rezultatai: r^2_{Adj}, koef. b_j.	0.92 0.95 0.95	0 0 0 -7.71 25.01 0 0 2.89 -5.23 17.66 0.13 0 2.35 -5.22 16.83	0.6 0.92 0.95 0.95 0	0 0 10.99 0 0 0 0 0 -7.71 25.01 0 0 2.89 -5.23 17.66 0.13 0 2.35 -5.22 16.83 0 0 0 0 0
Laikas (sek.)	0.16		0.39	

Iš 6 lentelės matome, kad pavykus nustatyti veiksnius, kurie gerai koreliuoja tarpusavyje ir turi stiprią tiesinę priklausomybę, galima išlošti daug kompiuterinio laiko.

Gavę atsakymus turime patikrinti modelio tinkamumą. Tai yra, turime patikrinti anksčiau aprašytas penkias modelio tinkamumo sąlygas. Tuo tikslu pirmiausiai pasirenkame mums priimtiną modelį. Šiame pavyzdyje sakykime, kad visi duomenys yra vienodai prieinami ir įmanomi gauti vienodu dažniu. Tuomet pasirenkame modelį iš tirjų veiksnių: B1 ir B2 ir konstantos. Įvykdytos programos kataloge esančioje byloje „SeSrR2adj.txt“ koeficientai išdėstyti tokia tvarka: Koreguotasis determinacijos koeficientas, liekamųjų paklaidų kvadratų suma (13), regresijos kvadratų suma (12). Paskutiniai du skaičiai reikalingi hipotezei patikrinti. Be to, prisiminkime, kad pasirinkome modelį su k=3 ir naudojome duomenis su n=12. Tuomet, kai SSE = 1.25, o SSR = 35.62 gauname:

7 pav. Hipotezės, apie bent vieno koeficiento nelygumą nuliui, tikrinimas. Pavyzdys

3	K-lygties koeficientų skaičius				
12	N – imties dydis				
35.62	SSR				
1.25	SSE	***0.05***	alfa(reikšmingumo lygmuo)		
ANOVA	kvad.suma	laisv.laipsniai	Disp. Įverčiai	F	F(alfa)
regresijos	35.62	3	11.87	75.98	3.59
paklaidų	1.25	8	0.16		
viso	36.87	11			

Čia F – gautos statistikos reikšmė, o F(alfa) – Fišerio statistikos reikšmė su užsiduotu lygmeniu. Pastebime, kad hipotezė apie visų koeficientų lygumą nuliui yra atmetama, tai yra, bent vienas koeficientas

yra statistiškai reikšmingas su tikimybe 0.95.

Tuomet kiekvienam koeficientui atskirai tikrinama hipotezė apie konkretaus koeficiento nelygumą nuliui. Rezultatų byloje „issirink.txt“ randame pasirinktą modelį ir pasirenkame reikiamus dydžius. Eilutėje SSE(k) yra nurodyta liekanų suma, kai priklausomu kintamuoju naudojamas k-tasis pasirinktas koeficientas. Byloje „Bmatrix.txt“ yra nurodyti visų modelių b koeficientai. Tuomet pirmam koeficientui gauname:

8 pav. Hipotezių, apie visų koeficientų nelygumą nuliui, tikrinimas. Pirmasis koeficientas.

0.05			Bj įvertis	2.89	
2.31	<?>	2.89	SSE	***1.25***	0.16
Jei < tai H0 atmetama			SSEj	0.16	
	t.y. regr. Koeficientas tinka		k	3	
			n	12	

9 pav. Hipotezių, apie visų koeficientų nelygumą nuliui, tikrinimas. Antrasis koeficientas.

0.05			Bj įvertis	-5.23	
2.31	<?>	5.21	SSE	***1.25***	0.16
Jei < tai H0 atmetama			SSEj	0.16	
	t.y. regr. Koeficientas tinka		k	3	
			n	12	

10 pav. Hipotezių, apie visų koeficientų nelygumą nuliui tikrinimas. Trečiasis koeficientas.

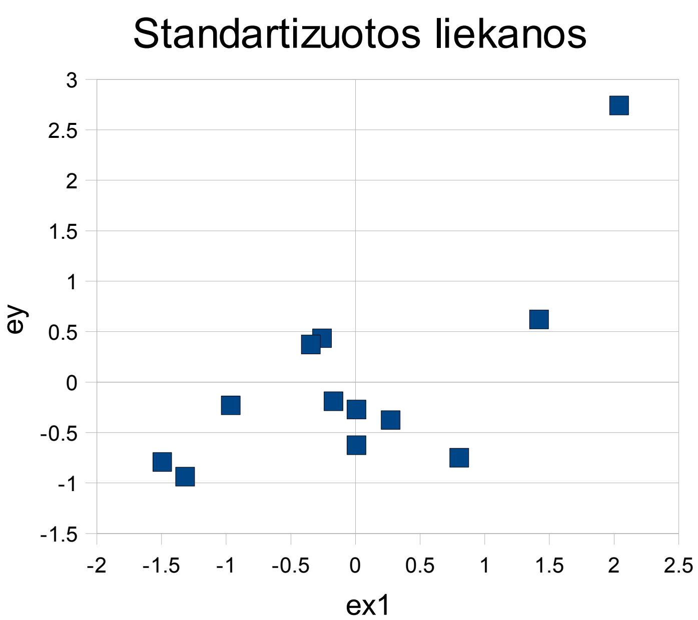
0.05			Bj įvertis	17.66	
2.31	<?>	6.47	SSE	***1.25***	0.16
Jei < tai H0 atmetama			SSEj	0.02	
	t.y. regr. Koeficientas tinka		k	3	
			n	12	

Taigi nulinė hipotezė visiems koeficientams yra atmetama (žr 8, 9 ir 10 pav.). Tai yra, visi koeficientai su 0.95 tikimybe statistiškai reikšmingai skiriasi nuo nulio.

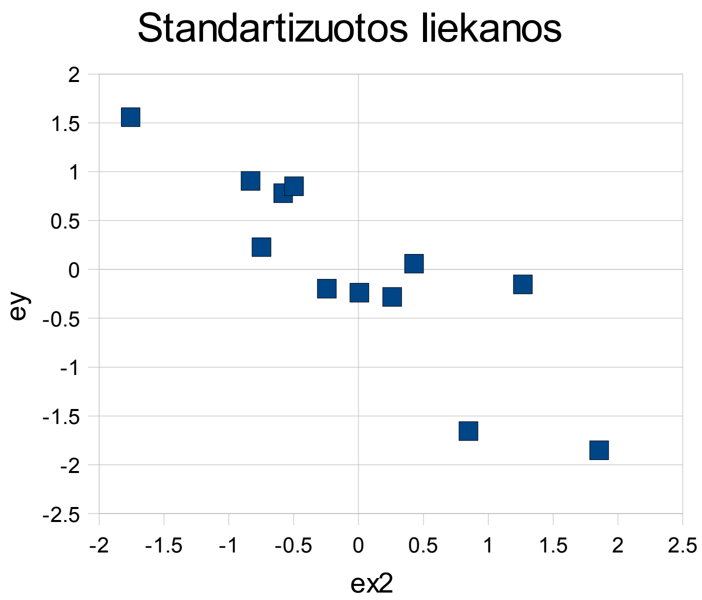
Modelio tinkamumo trečioji sąlyga yra automatiškai patikrinama programos skaičiavimo metu. Taigi, tikriname ketvirtąją ir penktąją sąlygas. Ketvirtąją sąlygą, apie liekanų pasiskirstymą pagal normalųjį skirstinį, tikrinsime Kolmogorovo dėsniu. Byloje „liekanos.txt“ yra pateiktos pasirinkto modelio liekanos. Patikrinę hipotezę gauname, jog skirstinys yra normalusis su tikimybe 0.99 (Nulinė hipotezė yra atmetama).

Tikriname penktąją sąlygą:

11 pav. Standartizuotųjų liekanų sklaidos diagrama. Pirmojo kintamojo liekanoms



12 pav. Standartizuotųjų liekanų sklaidos diagrama. Antrojo kintamojo liekanoms



Pastebėkime, kad antroji priklausomybė yra panaši į tiesinę (žr. 12 pav.), o pirmojoje yra kelios liekanos, kurios šią priklausomybę daro neaiškią (žr. 11 pav.). Tačiau kitos prielaidos yra geros, dėl to sakysime, kad modelis tinka. Tokį modelį aprašys regresijos funkcija:

$$\hat{y}(B_{1,} B_2) = 17.661 + 2.895 \cdot B_1 - 5.228 \cdot B_2 \quad \textbf{(24)}$$

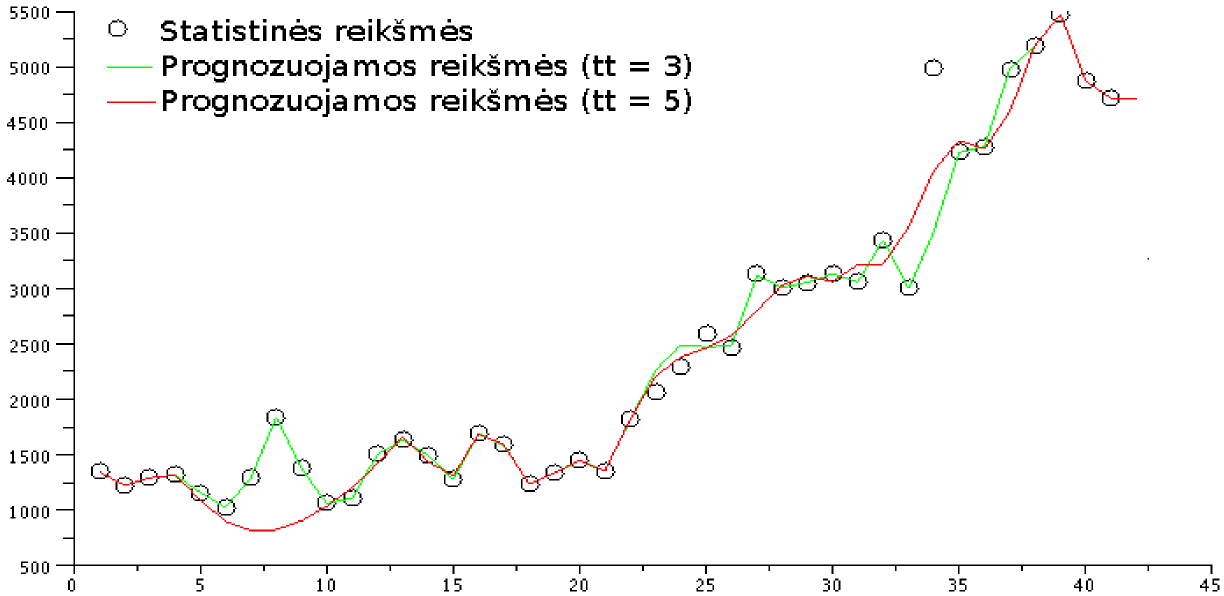
4. TIRIAMOJI DALIS

4.1. KAUNO MIESTO CENTRE ESANČIŲ BUTŲ KAINŲ ĮVERTINIMAS

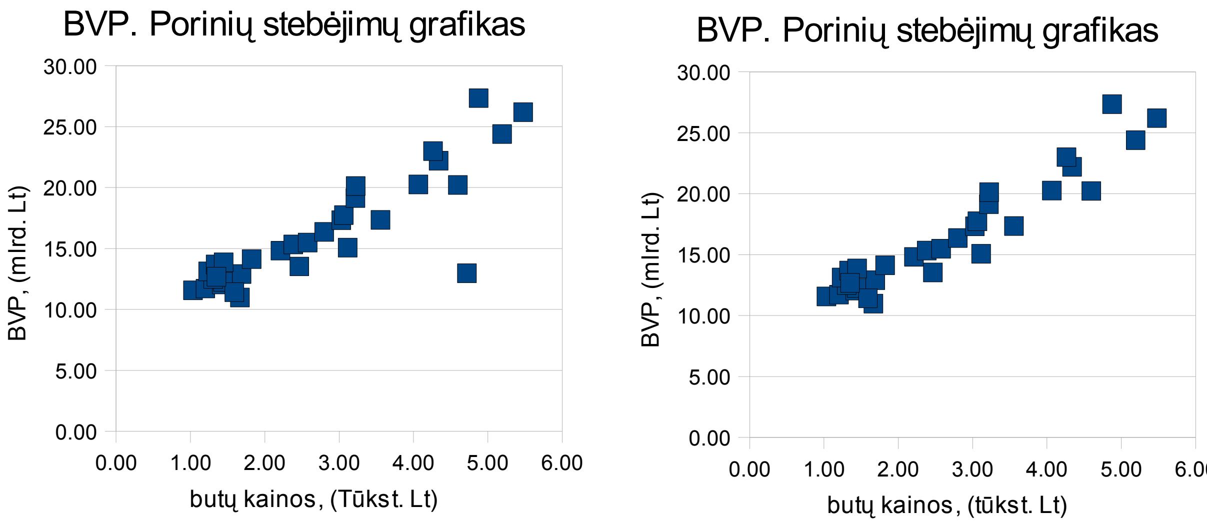
Toliau ieškosime regresijos modelio, geriausiai aprašančio butų kainas Kauno miesto centre. Turime vidutines būstų kainas Kauno miesto centre (16 lentelė, 1 priedas). Pirmiausiai turime kelis istorinius duomenis. Paskaičiuojame kiekvieno periodo vidurkį. Tuomet pagal duomenų kiekį periode, braižome interpoliacinį splainą, ieškodami trūkstamų reikšmių. Mažiausių kvadratų metodu gauname trūkstamas reikšmes. Prie kiekvieno glodinančio splaino nario yra nurodomas koeficientas, atitinkantis duomenų kiekį tame periode. Jei duomenų nėra – koeficientas nurodomas nulis, o prie duomenų įrašomas bet koks skaičius. Tai yra svoriai nurodo, kiek duomenų iš to periodo buvo gauta. Prognozuodami darbo duomenis (16 lentelė, 1 priedas) pastebėsime, kad jautresni gauti tie rezultatai, kur mažesnis svoris buvo nurodytas kaip svarbesnis. Tačiau, pagal aktyvo likvidumą investuotojas turėtų nustatyti šį dydį pats. Suprantama, kad kuo aktyvas daugiau likvidus, tuo greičiau gali keistis jo kaina. Toliau nagrinėsime tą prognozavimo atvejį, kai kintamasis „tt“ buvo parinktas lygus 5. Taip pasirinkome dėl to, kad būstai yra mažiau likvidūs, ir dėl to mažesnis pardavimų skaičius taip pat nurodo ir paklausos sumažėjimą, kas reiškia ir kainos mažesnį pokytį (13 pav).

Pastebėsime, kad daugelis veiksnių yra nurodomi tik nuo 2000 metų 2 ketvirčio. Todėl tolesnė analizė bus atliekama periodui nuo 2000 metų antro ketvirčio iki 2007 metų paskutinio ketvirčio.Pirmą 2008 metų ketvirtį jau buvo pakitę daug veiksnių, tačiau į juos nebuvo spėjusios sureaguoti būsto kainos, ir dėl to atsirado išskirtys tuo periodu (žr. 14 pav.).

13 pav. Vidutinių butų kainų, tūkst. LTL už kvadratinį metrą, kitimas Kauno miesto centre 1999-2008m. Ketvirčiais.



14 pav. Pastebėta ir panaikinta išskirtis paskutiniame periode



Veiksnius parenkame visus tuos, kurie koreliuoja su butų kainomis. Jei veiksnio koreliacija yra silpna – tikriname, ar veiksnio ankstesnių periodų pokyčiai nekoreliuoja statistiškai reikšmingai stipriau su butų kainų pokyčiais tarp periodų. Jei egzistuoja statistiškai reikšminga priklausomybė, tuomet toliau parenkame kelias (dažniausiai dvi) to veiksnio reikšmes skirtingiems periodams, kad programa galėtų parinkti, kuris veiksnys, sąveikaudamas su kitais, duoda geresnius rezultatus prognozuojamam kintamajam. Parinkti veiksniai yra nurodyti pirmame priede.

Tuomet, kad būtų kuo daugiau kartų sutrumpintas kompiuterinio skaičiavimo laikas, reikia pastebėti, kurie kintamieji seka vieną nežinomą veiksnį (yra smarkiai koreliuoti tarpusavyje) ir kurie keičiasi kitaip. Tie, kurie yra stipriai koreliuoti tarpusavyje – priskiriami vienai grupei. Toliau darbe tokie veiksniai bus priskiriami „A" grupei. O patys veiksniai žymimi A raide su prie jos esančiu veiksnį identifikuojančiu numeriu. Kiti veiksniai priskiriami grupei B, ir žymimi B raide, su prie jos esančiu veiksnį identifikuojančiu numeriu.

Stebint koreliaciją tarp veiksnių, tuo pačiu metu verta nusibrėžti porinių stebėjimų grafikus. Porinių stebėjimų grafikai daugialypės regresijos lygties modeliuose yra taikomi, kai norima paviršutiniškai įvertinti, ar kintamieji yra tiesiškai priklausomi, ar nėra. O jei yra – tuomet, kiek tiesiškai priklausomi. Nubraižę šiuos grafikus (žr. 2 priedą) galime pastebėti, kad kai kurie veiksniai yra netinkami tolesnei regresinei analizei, o kai kurie gali būti tinkami panaudojus duomenų ištiesinimo galimybę. Ištiesinus kai kuriuos duomenis, apskaičiuojama „darbinė" koreliacinė matrica (žr. 3 priedą).

Pagal ją suskirstome kintamuosius, ir juos pažymime raidėmis. Toliau darbe visi veiksniai bus žymimi tokiais žymėjimais, kaip parodyta 7 ir 8 lentelėse.

7. lentelė: Veiksnių suskirstymas, apibūdinimas. Veiksniai, stipriai koreliuoti tarpusavyje

Pavadinimas	Apibūdinimas
A1	Bendrasis vidaus produktas, sukurtas per paskutinius 12 mėnesių, milijardais litų..
A2	Vilniaus vertybinių popierių biržos indeksas (OMXV)
A3	Vidutinis mėnesinis atlyginimas (NETO), tūkstančiais litų.
A4	Eksportas per paskutitnius 12 mėnesių, milijardais litų.
A5	Importas per paskutinius 12 mėnesių, su metiniu vėlavimu, milijardais litų.
A6	Importas per paskutinius 12 mėnesių, milijardais litų.
A7	Bedarbystė einamu momentu, procentais.
A8	Žaliavinės naftos kainos, doleriais už barelį.
A9	Statybos leidimų, išduotų per paskutinius 12 mėnesių, skaičius, vienetais.
A10	Statybos sąnaudų indeksas
A11	Cukraus importas į Europos Sąjungą, doleriais už svarą.
A12	Bedarbystės lygis miestuose, buvęs prieš pusmetį, procentais.
A13	Bedarbystės lygis miestuose, esantis vertinimo momentu, procentais.
A14	Bedarbystės lygis miestuose, buvęs prieš 9 mėn, procentais.
A15	Gimstamumas, buvęs prieš 22 metus, gimusių skaičius per tris mėnesius, tūkst.

8. lentelė: Veiksnių suskirstymas, pažymėjimas, apibūdinimas. Kiti veiksniai.

Pavadinimas	Apibūdinimas
B1	Eksporto ir importo skirtumas, milijardais litų, per paskutinius 12 mėn.
B2	Emigracija per paskutinius 3 mėn, tūkstančiais litų.
B3	EURIBOR palūkanų indeksas 3 mėn laikotarpiui, buvęs prieš 3 mėn nuo skaičiavimo momento, procentais.
B4	Ištuokų kiekis per paskutinius 3 mėnesius, tūkstančiais žmonių
B5	Lūkesčiai (kiek tūkstančių bedarbių tikisi gauti daugiau negu 1500LTL atlyginimą)
B6	Valiutų USD ir EUR vertės santykis vertinimo momentu
B7	Skirtumas tarp VILIBOR ir EURIBOR palūkanų normų
B8	Materialinės investicijos į šalį, per paskutinius 12 mėn. Milijardais litų.
B9	EURIBOR palūkanos 3 mėn laikotarpiui vertinimo momentu, procentais.
B10	Santuokos, per paskutinius tris mėn, tūkstančiais.
B11	Valiutų USD ir EUR vertės santykis, buvęs prieš mėnesį, nuo vertinimo momento
B12	Valiutų USD ir EUR vertės santykis, buvęs prieš tris ketvirčius nuo vertinimo momento
B13	Valiutų USD ir EUR vertės santykis vertinimo momentu. Ištiesinti duomenys (laipsnyje 1/3)
B14	Vidutinės paskutinių 3 mėnesių apvaliosios medienos kainos Kauno rajono urėdijoje, litais už kubinį metrą.

Kai jau suskirstyti kintamieji, tuomet porinių stebėjimo grafikų pagalba galima nustatyti, kurie kintamieji tikrai nėra tiesinėje priklausomybėje su priklausomu kintamuoju. Dėl netiesinės priklausomybės atmetame šiuos veiksnius:

A2, A7, A8, A11, B2, B3, B4, B6, B7, B9, B10, B11, B12

taigi lieka veiksniai:

A1, A3, A4, A5, A6, A9, A10, A12, A13, A14, A15

ir iš kitos grupės:

B1, B5, B8, B13, B14.

Perskaičiavę visus galimus variantus, gauname, jog geriausias veiksnių rinkinys yra:

A13, B5, B8 ir konstanta. Koreguotasis determinacijos koeficientas šiam modeliui yra 0.96102

kiti duomenys pateikti 9 lentelėje.

9. lentelė: Optimalių parametrų modelio rezultatai

Parametras	reikšmės
Regresijos lygties koeficientai:A13; B5; B8; konstanta.	-0.204909411839991 0.104229326752084 -0.132663212937029 4.661788669925179
Koreguotasis determinacijos koeficientas	0.961019751189
SSE(j) - modelio, kuriame j-tasis veiksnys yra priklausomas kintamasis, liekamųjų paklaidų kvadratų suma.	296.036 265.278 14.476 0.642
Multikolinearumas tarp koeficientų	2.404 2.584 3.632 0.900
Modelio liekanos (iš kairės į dešinę ir žemyn)	-0.36 0.32 0.15 0.62 0.45 -0.03 -0.36 0.14 -0.06 -0.18 -0.43 -0.28 -0.36 -0.24 -0.11 -0.38 -0.37 0.28 -0.36 -0.03 -0.45 0.02 -0.81 0.26 0.77 0.98 -0.04 0.24 0.53 0.51 -0.41
Modelio liekamųjų paklaidų kvadratų suma	1.790465146280
Modelio regresijos kvadratų suma	51.208717404688

Tikriname hipotezes apie gauto modelio tinkamumą.

1) Hipotezės, apie bent vieno koeficiento nelygumą nuliui, tikrinimas (15 pav):

15 pav. Hipotezės, apie bent vieno koeficiento nelygumą nuliui, tikrinimas

4	K-lygties koeficientų skaičius				
31	N – imties dydis				
51.21	SSR				
1.79	SSE	***0.01***	alfa(reikšmingumo lygmuo)		
ANOVA	kvad.suma	laisv.laipsniai	Disp. Įverčiai	F	F(alfa)
regresijos	51.21	4	12.8	185.91	4.02
paklaidų	1.79	26	0.07		
viso	53	30			

2) Hipotezių tikrinimas apie kiekvieno koeficieno lygumą nuliui.

16 pav. Hipotezės tikrinimas apie A13 veiksnį aprašančio koeficiento lygybę nuliui

0.01			Bj įvertis	-0.2		
2.78	<?>	13.44	SSE	***1.79***	0.07	MSE
	Jei < tai H0 atmetama		SSEj	296.04		
		t.y. regr. Koeficientas tinka	k	4		
			n	31		

17 pav. Hipotezės tikrinimas apie B5 veiksnį aprašančio koeficiento lygybę nuliui

0.01			Bj įvertis	0.1		
2.78	<?>	6.47	SSE	***1.79***	0.07	MSE
Jei < tai H0 atmetama			SSEj	265.28		
t.y. regr. Koeficientas tinka			k	4		
			n	31		

18 pav. Hipotezės tikrinimas apie B8 veiksnį aprašančio koeficiento lygybę nuliui

0.07			Bj įvertis	-0.13		
1.89	<?>	1.92	SSE	***1.79***	0.07	MSE
Jei < tai H0 atmetama			SSEj	14.48		
t.y. regr. Koeficientas tinka			k	4		
			n	31		

19 pav. Hipotezės apie konstantos lygumą nuliui tikrinimas

0.01			Bj įvertis	4.66		
2.78	<?>	14.23	SSE	***1.79***	0.07	MSE
Jei < tai H0 atmetama			SSEj	0.64		
t.y. regr. Koeficientas tinka			k	4		
			n	31		

Pastebėsime, kad trečiojo koeficiento nelygumą nuliui galime garantuoti su 0.93 tikimybe (žr. 18 pav.), kai, tuo tarpu, kitų koeficientų nelygumą nuliui galime garantuoti su 0.99 tikimybe (žr. 16, 17, 19 pav.).

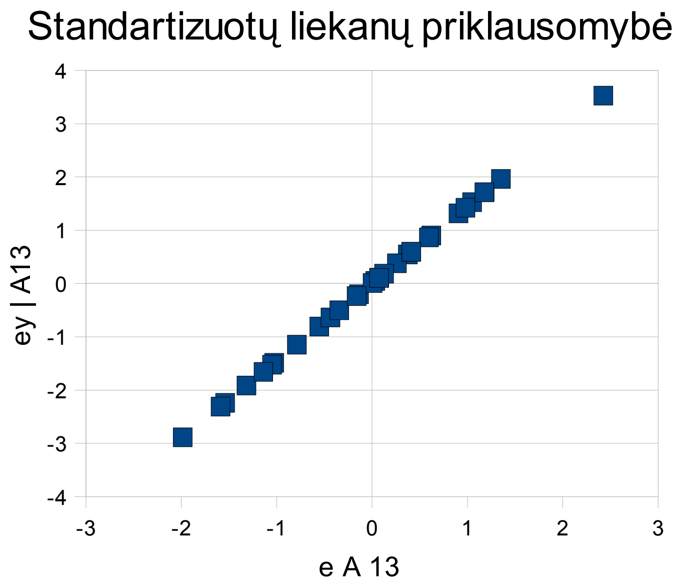
3) Multikolinearumo sąlyga (patikrinama modelio parinkimo metu);

4) Hipotezė apie modelio liekanų pasiskirstymą pagal normalųjį skirstinį:

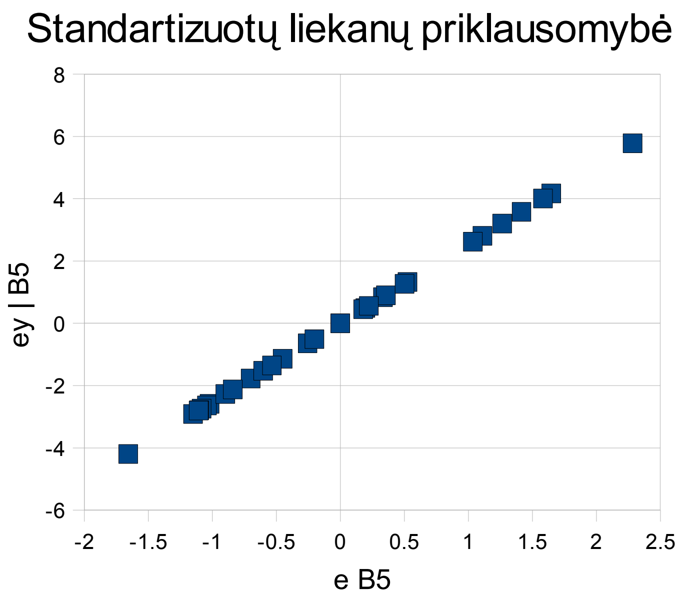
Pagal Kolmogorovo kriterijų liekanos yra pasiskirsčiusios pagal normalųjį skirstinį su 0.99 tikimybe.

5) Tiesiškos prieklausomybės tikrinimas, braižant standartizuotų liekanų sklaidos diagramas (20, 21, 22 pav)

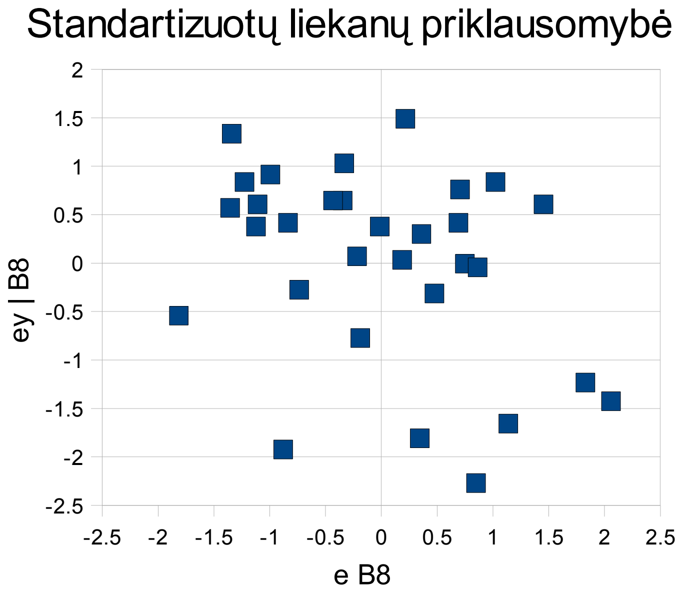
20 pav. Standartizuotųjų liekanų priklausomybė, parodanti veiksnio A13 tiesiškumą su priklausomu kintamuoju



21 pav. Standartizuotųjų liekanų priklausomybė, parodanti veiksnio B5 tiesiškumą su priklausomu kintamuoju



22 pav. Standartizuotųjų liekanų priklausomybė, parodanti veiksnio B8 tiesiškumą su priklausomu kintamuoju



Galime pastebėti, kad koeficientas B8 tiek antrame tiek penktame modelio patikimumo testuose yra mažiau patikimas. Dėl to galime daryti išvadą, kad koeficientas B8 iš tolesnės analizės gali būti pašalintas, jei reikalaujama didelio modelio tikslumo. Modelio tinkamumas, pagal antrosios tinkamumo savybės tikrinimą, yra 0.93. Toks skaičius parinktas pagal mažiausio patikimumo koeficientą.

Šį modelį aprašanti regresijos funkcija:

$$\hat{y}\left(a_{13,}b_{5,}b_8\right)=4.66179-0.20491\cdot a_{13}+0.10423\cdot b_5-0.13266\cdot b_8 \quad \textbf{(25)}$$

čia a_{13} , b_5 ir b_8 yra reikšmės, atitinkamai veiksnių: A13, B5 ir B8 tuo periodu, kai vertinamos kainos.

Pastebėsime, kad nors dispersijos mažėjimo daugiklis yra mažesnis už 4, tačiau prie veiksnio B8 atsirado neigiamas daugiklis, nors šio veiksnio koreliacija su priklausomu kintamuoju yra teigiama (~= 0.8). Tačiau, kadangi multikolinearumas nėra didelis, galima įtarti, kad šis koeficientas atimamas dėl to, kad jo padidėjimą silpnai jau aprašo kitas koeficientas. Tai yra, šis veiksnys didėja kartu su kokiu nors kitu lygtyje esančiu veiksniu su dideliu patikimumu.

Toliau teigsime, kad mums yra reikalingas modelis su tinkamumu, lygiu 0.99. Iš tolesnės analizės pašaliname koeficientą B8. Tuomet geriausias rezultatas apskaičiuojamas iš šių veiksnių:

A13, B5 ir konstanta. Kiti šio modelio parametrai nurodyti 10 lentelėje

10. lentelė: Patikslinto modelio parametrai

Parametras	reikšmės
Regresijos lygties koeficientai:A13; B5; konstanta.	-0.188804570051413 0.085893291220251 4.242638287762211
Koreguotasis determinacijos koeficientas	0.957122268280
SSE(j) - modelio, kuriame j-tasis veiksnys yra priklausomas kintamasis, liekamųjų paklaidų kvadratų suma.	423.728 408.177 1.152
Multikolinearumas tarp koeficientų	1.742 1.742 0.933
Modelio liekanos (iš kairės į dešinę ir žemyn)	-0.42 0.49 0.28 0.47 0.53 -0.05 -0.23 -0.07 -0.03 -0.15 -0.29 -0.43 -0.42 -0.27 0 -0.74 -0.48 0.4 -0.3 -0.29 -0.64 0.02 -0.67 -0.02 0.76 1.19 0.3 0 0.5 0.55 -0.01
Modelio liekamųjų paklaidų kvadratų suma	2.045236257730
Modelio regresijos kvadratų suma	50.953946293237

Tikriname hipotezes apie gauto modelio tinkamumą.

1) Hipotezės, apie bent vieno koeficiento nelygumą nuliui, tikrinimas (23 pav):

23 pav. Hipotezės apie visų koeficientų lygumą nuliui tikrinimas

3	K-lygties koeficientų skaičius				
31	N – imties dydis				
50.95	SSR				
2.05	SSE	0.01	alfa(reikšmingumo lygmuo)		
ANOVA	kvad.suma	laisv.laipsniai	Disp. Įverčiai	F	F(alfa)
regresijos	50.95	3	16.98	224.22	4.51
paklaidų	2.05	27	0.08		
viso	53	30			

Matome, kad nulinė hipotezė, apie visų koeficientų lygumą nuliui, yra atmetama.

2) Hipotezių, apie kiekvieno koeficiento nelygumą nuliui, tikrinimas (24, 25, 26 pav.):

24 pav. Hipotezės tikrinimas apie veiksnį A13 apibūdinančio koeficiento lygybę nuliui

0.01			Bj įvertis	-0.19		
2.77	<?>	14.12	SSE	2.05	0.08	MSE
	Jei < tai H0 atmetama		SSEj	423.73		
	t.y. regr. Koeficientas tinka		k	3		
			n	31		

25 pav. Hipotezės tikrinimas apie veiksnį B5 apibūdinančio koeficiento lygybę nuliui

0.01			Bj įvertis	0.09		
2.77	<?>	6.31	SSE	***2.05***	0.08	MSE
Jei < tai H0 atmetama			SSEj	408.18		
t.y. regr. Koeficientas tinka			k	3		
			n	31		

26 pav. Hipotezės tikrinimas apie konstantos lygymą nuliui

0.01			Bj įvertis	4.24		
2.77	<?>	16.55	SSE	***2.05***	0.08	MSE
Jei < tai H0 atmetama			SSEj	1.15		
t.y. regr. Koeficientas tinka			k	3		
			n	31		

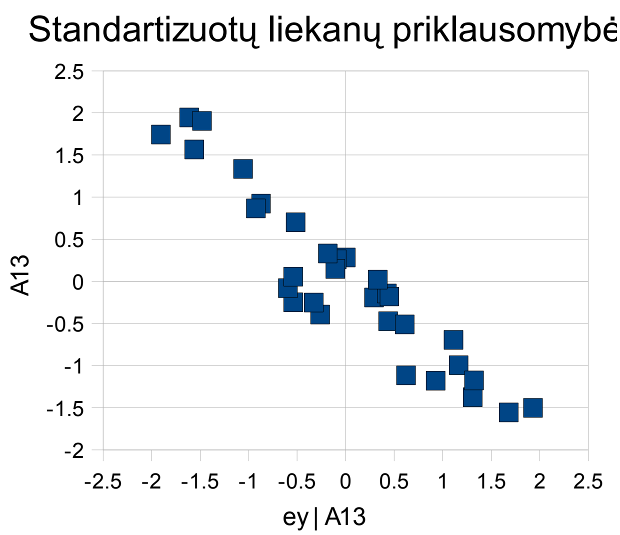
3) Multikolinearumo sąlyga (patikrinama modelio parinkimo metu);

4) Hipotezė apie modelio liekanų pasiskirstymą pagal normalųjį skirstinį:

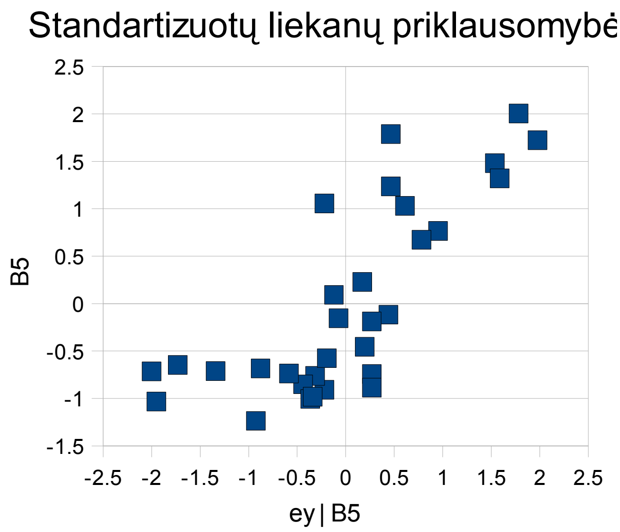
Tikriname pagal Kolmogorovo kriterijų. Pagal šį kriterijų liekanos pasiskirsčiusios pagal normalųjį skirstinį su tikimybe 0.99.

5) Tiesiškos prieklausomybės tikrinimas, braižant standartizuotų liekanų sklaidos diagramas (27, 28 pav.).

27 pav. Standartizuotųjų liekanų priklausomybė, parodanti priklausomo kintamojo tiesinę priklausomybę nuo veiksnio A13



28 pav. Standartizuotųjų liekanų priklausomybė, parodanti priklausomo kintamojo tiesinę priklausomybę nuo veiksnio B5



Matome,kad šiame modelyje penktasis testas veiksniams A13 ir B5 duoda ne tokius gerus rezultatus, kurie buvo duoti ankstesnio testo metu. Tačiau tiesiškumas, nors ir neakivaizdus, išlieka. Visgi, jei norima gauti didesnį tiesiškumą veiksnių A13 ir B5 atžvilgiu, reikėtų naudoti ankstesnį modelį. Nes veiksnys B8, nors ir neturi geros tiesinio prognozavimo savybės butų kainoms, tačiau sąveikoje su šiais dviem veiksniais duoda labai gerą rezultatą.

Aprašytą modelį apibendrinanti regresijos funkcijos forma yra tokia:

$$\hat{y}(a_{13,} a_5) = 4.24264 - 0.1888 \cdot a_{13} + 0.08589 \cdot b_5 \qquad \textbf{(26)}$$

čia a_{13} ir b_5 yra reikšmės, atitinkamai veiksnių: A13 ir B5 prognozuojamu periodu.

Darbe buvo minėta, kad kai kurių veiksnių reikšmės yra sunkiai įvertinamos praktikoje. Todėl dabar atliksime lyginamąją analizę, siekdami pastebėti, kaip keičiasi modelio tikslumas, jei parenkami veiksniai, kurių reikšmes galime įvertinti dažniau ir papraščiau. Tokiai analizei tinka veiksniai:

A5, A12, A14, A15, B13, B14. Tiesą sakant yra sunku objektyviai nustatyti lūkesčių veiksnį, ar jam ekvivalentų. Nes naudojant duomenis iš Lietuvos statistikos departamento yra tikėtinas labai stiprus duomenų vėlavimas, netinkamas momentinei analizei.

Naudojant šiuos veiksnius gautas, didžiausią determinacijos koeficientą turintis modelis, turintis tokį nepriklausomų kintamųjų rinkinį:

A14, B14 ir konstantą. Šio modelio parametrai nurodyti 11 lentelėje

11. lentelė: Greito įvertinimo modelio parametrų lentelė

Parametras	reikšmės
Regresijos lygties koeficientai:A14; B14; konstanta.	-0.226598979414066 0.013305856383248 4.212165860046144
Koreguotasis determinacijos koeficientas	0.945394352102
SSE(j) - modelio, kuriame j-tasis veiksnys yra priklausomas kintamasis, liekamųjų paklaidų kvadratų suma.	187.299 3694.861 0.135
Multikolinearumas tarp koeficientų	3.243 3.243 0.933
Modelio liekanos (iš kairės į dešinę ir žemyn)	-0.42 0.49 0.28 0.47 0.53 -0.05 -0.23 -0.07 -0.03 -0.15 -0.29 -0.43 -0.42 -0.27 0 -0.74 -0.48 0.4 -0.3 -0.29 -0.64 0.02 -0.67 -0.02 0.76 1.19 0.3 0 0.5 0.55 -0.01
Modelio liekamųjų paklaidų kvadratų suma	2.604649231125
Modelio regresijos kvadratų suma	50.394533319842

Tikriname hipotezes apie gauto modelio tinkamumą.

1) Hipotezės, apie bent vieno koeficiento nelygumą nuliui, tikrinimas (29 pav):

29 pav. Hipotezės apie visų koeficientų lygumą nuliui tikrinimas

3	K-lygties koeficientų skaičius				
31	N – imties dydis				
50.39	SSR				
2.6	SSE	***0.01***	alfa(reikšmingumo lygmuo)		
ANOVA	kvad.suma	laisv.laipsniai	Disp. Įverčiai	F	F(alfa)
regresijos	50.39	3	16.8	174.13	4.51
paklaidų	2.6	27	0.1		
viso	53	30			

Matome, kad nulinė hipotezė, apie visų koeficientų lygumą nuliui, yra atmetama.

2) Hipotezių, apie kiekvieno koeficiento nelygumą nuliui, tikrinimas (30,31,32 pav.):

30 pav. Hipotezės apie veiksnį A14 apibūdinančio koeficiento lygybę nuliui, tikrinimas

0.01			Bj įvertis	-0.23		
2.77	<?>	9.98	SSE	***2.6***	0.1	MSE
Jei < tai H0 atmetama			SSEj	187.3		
t.y. regr. Koeficientas tinka			k	3		
			n	31		

31 pav. Hipotezės apie veiksnio B14 lygybę nuliui apibūdinančio koeficiento tikrinimas

0.02			Bj įvertis	0.01		
2.47	<?>	2.6	SSE	***2.6***	0.1	MSE
Jei < tai H0 atmetama			SSEj	3694.86		
t.y. regr. Koeficientas tinka			k	3		
			n	31		

32 pav. Hipotezės apie konstantos lygumą nuliui tikrinimas

0.01			Bj įvertis	4.21		
2.77	<?>	4.98	SSE	***2.6***	0.1	MSE
Jei < tai H0 atmetama			SSEj	0.14		
t.y. regr. Koeficientas tinka			k	3		
			n	31		

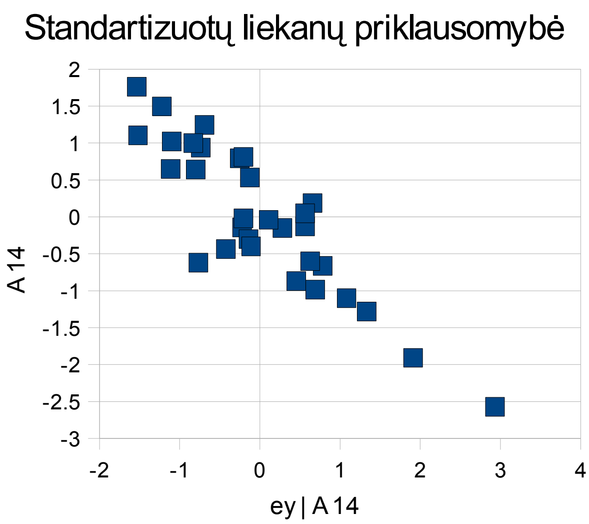
3) Multikolinearumo sąlyga (patikrinama modelio parinkimo metu);

4) Hipotezė apie modelio liekanų pasiskirstymą pagal normalųjį skirstinį:

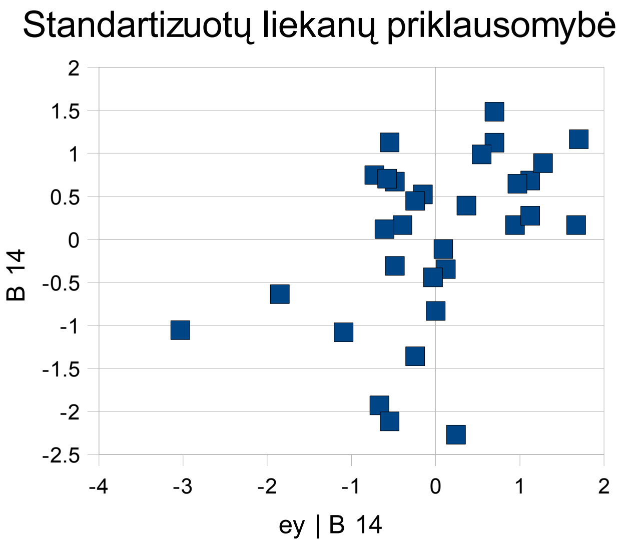
Tikriname pagal Kolmogorovo kriterijų. Pagal šį kriterijų liekanos pasiskirsčiusios pagal normalųjį skirstinį su tikimybe 0.99.

Tiesiškos prieklausomybės tikrinimas, braižant standartizuotų liekanų sklaidos diagramas (33 ir 34 pav.). Standartizuotų liekanų priklausomybė, parodanti priklausomo kintamojo tiesinę priklausomybę nuo veiksnio A14.

33 pav. Priklausomojo kintamojo tiesinė priklausomybė nuo veiksnio A14. Standartizuotųjų liekanų priklausomybė.



34 pav. Priklausomojo kintamojo tiesinė priklausomybė nuo veiksnio B14. Standartizuotųjų liekanų priklausomybė



. Galime pastebėti, kad hipotezės tenkinamos pakankamai tiksliai, tačiau tarp standartizuotų liekanų nėra labai tikslios tiesinės priklausomybės.

Aprašytą modelį apibendrinanti regresijos funkcijos forma yra tokia:

$$\hat{y}\left(a_{14,}a_{14}\right)=4.24264-0.1888\cdot a_{14}+0.08589\cdot b_{14} \qquad \textbf{(27)}$$

čia a_{14} ir b_{14} yra reikšmės, atitinkamai veiksnių: A14ir B14 prognozuojamu periodu.

Apibendrindamas pabandysiu įvertinti vieną aktyvą pagal (25) funkciją. Yra žinoma (statistikos departamento duomenimis), kad 2009 metų pirmą ketvirtį materialinės investicijos Lietuvoje sudarė 2.4227 mlrd litų. Taip pat, sakykime, kad nedarbas tuomet buvo 9proc. Jei spėsime, kad Lietuvoje tik 100 žmonių teigė norintys didesnio negu 1500 atlyginimo (2009 m, I ketvirčio ekonominėmis sąlygomis tas yra tikėtina). Tuomet pagal modelį vidutinio buto Kauno miesto centre vieno kvadratinio metro vertė 2009 metų pirmą ketvirtį buvo lygi maždaug 2510LT. Tai yra, vidutinis 45 kvadratinių metrų ploto dviejų kambarių butas Kauno miesto centre turėjo kainuoti apie 112 800LTL. Kiek ši kaina atrodo reali – tegul sprendžia skaitytojas.

4.2. PROGRAMINĖ REALIZACIJA IR INSTRUKCIJA VARTOTOJUI

Pirmas svarbus programos aspektas reikalingas dar duomenų surinkimo metu yra glodinančio ar/ir interpoliacinio splainio naudojimas. Kai turimi skirtingų periodų ir galbūt skirtingo patikimumo duomenys, tuomet duomenų papildymui ir susisteminimui naudojama procedūra:

```
file1 = uigetfile(["*"],"/home", "Pasirinkite duomenų failą", %f);
direktorija = dirname(file1);
file2 = uigetfile(["*"],"/home", "Pasirinkite svorių failą", %f);
BtPk=fscanfMat(file1)';
svoriai = fscanfMat(file2)';
dydis = size(BtPk);
kk = 0;
kkk = 0;
tt = 5; // koks svoris jau yra laikomas svarbiu
for j = 1:dydis(2)
  if svoriai(j) >= tt then kkk = kkk+1;
  end;
end;
xx = zeros(kkk);
for i = 1:dydis(2)
  if svoriai(i) >= tt then
    kk = kk+1;
    xx(kk) = i; // nurodomi numeriai, kuriuose yra svarbesni taškai
  end;
end;
x = linspace(1,dydis(2), dydis(2)); // graduojama nuo 1 iki paskutinės x koordinatės
[y, d] = lsq_splin(x, BtPk, svoriai, xx); // y - mažiausių kvadratų metodu sudarytas splainas
PPzingsn = 1/3;
nai = smooth([xx';y'], PPzingsn); // 0.33 - žingsnis, kurio norima. Galima padaryti mažesnį
nai1 = nai(1,:); // gautų naujų duomenų x koordinatė
nai2 = nai(2,:); // gautų naujų duomenų y koordinatė

file3 = direktorija + '/smoothingx.txt';
file4 = direktorija + '/smoothingy.txt';
fprintfMat(file3,(nai1'),'%8.5f');
fprintfMat(file4,(nai2'),'%8.5f');
```

Šiai procedūrai įvykdyti reikalingos dvi bylos, kuriose yra vienodas skaičius duomenų. Pirmoji (duomenų) byla aprašo turimus duomenis pastoviais periodais. Tais pastoviais periodais, kuriuose duomenų neturime, reikia palikti nulius. Vėliau, kuriant svorių vektorių kitoje byloje, prie atitinkamų pastovių periodų, kuriose informacija yra visiškai nepatikima arba nėra žinoma, reikia įrašyti nulius (arba kitus koeficientus, priklausomai nuo duomenų patikimumo). Programoje taip pat galima pakeisti kintamąjį „tt“, kuris aprašo tuos taškus, kuriuose informacija yra laikoma patikima. Realiai tai reiškia, koks skaičius duomenų užtikrina patikimą informaciją. Ši procedūra rezultatus palieka tame pačiame kataloge, kuriame yra ir parinkta duomenų byla. Rezultatų failas „smoothingy.txt“, kuriame nurodytos naujos gautos reikšmės. Duomenys šios programos metu bus suglodinti mažiausių kvadratų metodu.

Iš 17 lentelės (1 priedas) matome, kad suglodintu splainu prognozuotos reikšmės nedaug skiriasi nuo tikrųjų reikšmių (prognozuotų euro ir JAV dolerio valiutų santykio vidutinis reikšmių skirtumas 1.42proc).

Vėliau, jau turint visus reikiamus duomenis (veiksniai ir priklausomi kintamieji), veiksnių vertės atitinkamais periodais surašomos į stulpelius. Vienas stulpelis turi būti sudarytas vien tik iš vienetų. Šitas faktas nurodo, kad modelyje gali būti parinkta konstanta.

Jei programa taikoma įvertinti arbitražo įkainojimo teorijos modelį, tuomet duomenų sutvarkymui iškyla kai kurie papildomi reikalavimai:

visi veiksniai turi būti centruoti (jų vidurkis turi būti lygus nuliui);

priklausomas kintamasis turi būti nurodytas ne realia jo verte, tačiau suskaičiuotos grąžos per atitinkamus periodus, ir po to turėtų būti sukuriamas fiktyvus kintamasis, kuris būtų lygus priklausomojo kintamojo ir nerizikingos palūkanų normos tame periode skirtumui. Esant tokioms sąlygoms modelis skaičiuos arbitražo įkainojimo teorijos modeliui reikalingus aprašyti parametrus. Suprantama, kad kitos sąlygos, aprašytos darbe, turi būti patikrintos ir tokiam modeliui.

Jau parinkus veiksnius, ir surašius juos į matricą, kurioje vieno veiksnio istorinėm reikšmėm yra skiriama m eilučių (i-toji eilutė – i-tąjam periodui), tokią matricą reikia įrašyti į bylą „x.txt“, ir palikti tame pačiame kataloge, kaip ir bylą „y1.txt“, kurioje būtų priklausomojo kintamojo reikšmės, atitinkančios tuos pačius m periodų. Programos naudojimas yra glaudžiai susijęs su matematinėmis žiinomis, todėl teisingai parenkant ir suskirstant duomenis galima išlošti nemažai kompiuterinio skaičiavimo laiko. Šių dviejų bylų, esančių viename kataloge, jau užtenka darbo programai. Programos tekstas „nekoreliuotiems.sci“ arba „koreliuotiems.sci“ pirmiausiai turi būti atidarytas scilab aplinkoje, ir įvykdytas klavišais [Ctrl] ir [L], juos paspaudžiant vienu metu. Tuomet pasirodo dialogo langas, kuriame vartotojas turėtų nurodyti, kuriame kataloge yra darbui paruoštos bylos „x.txt“ ir „y1.txt“. Vykdant programinį kodą „koreliuotiems.sci“ - po dialogo lango dar kompiuteris per vartotojo sąsają paklausia vartotojo, kiek pirmų veiksnių yra koreliuoti. Tokiu atveju, kai vartotojui žinoma, kad dėl per didelio multikolinearumo (dispersijos mažėjimo koeficientas ne didesnis kaip 4) keli veiksniai tikrai nebus įtraukti į vieną modelį, šiuos veiksnius vartotojas gali įdėti bylos „x.txt“ pradiniuose stulpeliuose, ir vykdydamas programą, sutaupyti daug kompiuterinio laiko, nes šie veiksniai jau nebus perrinkinėjami vienas su kitu. Jei veiksnių skaičius yra nedidelis, tuomet paprastumo ir aiškumo dėlei geriau naudoti programos kodą „nekoreliuotiems.sci“. Programai baigus savo darbą, vartotojui nurodytame darbiniame kataloge yra pateikiamos kelios rezultatų bylos. Byloje „issirink.txt“ yra nurodoma, kurie veiksniai parinkti, koks yra gautas modelio tikslumas, taip pat čia nurodomos vidutinės liekanų kvadratų sumos, kai priklausomu kintamuoju yra nustatomas vienas iš tame modelyje parinktų veiksnių. Byloje „Bmatrix.txt“į i-tąją eilutę yra iš eilės surašomi kiekvienam koeficientui priskiriami regresijos lygties koeficientai. Jei koeficientas nėra parinktas, aiškumo dėlei jis yra prilygintas nuliui. Taip yra surašomos visų parinktų geriausių modelių eilutės. Byloje „SeSrR2sdj.txt“ yra iš eilės, kiekvienam modeliui, surašyti tokie koeficientai: pirmame stulpelyje įrašytas koreguotasis determinacijos koeficientas, antrame – modelio liekamųjų narių kvadratų suma, trečiame stulpelyje – regresijos kvadratų suma. Paskutiniai du koeficientai yra reikalingi hipotezėms tikrinti. Taip pat pateikiama byla „liekanos.txt“. Joje nurodomi kiekvieno parinkto modelio regresijos lygties liekamieji nariai. Pirmame stulpelyje pateikiami nariai to modelio, kuris turi vieną koeficientą, antrame to, kuris turi du koeficientus, ir taip toliau, iki to, kuris pateikiamas paskutinis.

Šią programą taip pat galima naudoti ir tikrinant penktąją modelio patikimumo sąlygą. Tuomet ypač naudingas yra bylos „liekanos.txt“ paskutinis stulpelis.

Taip pat prie programos pridedamos ir kelios skaičiuoklių „Open Office Spreadsheet“ arba „Microsoft Excel“ bylos, kuriose yra paprasta tikrinti hipotezes. Tačiau jei yra galimybė – galima hipotezes tikrinti ir kitais paketais. Šios programos esmė yra parinkti geriausio koreguoto determinacijos koeficiento modelį, turint labai daug veiksnių.

Reikia pastebėti, kad Scilab programa savo veikimo principu yra labai panaši į programą Matlab, dėl to šias procedūras galima nesunkiai pakeisti į matlab programinį kodą. Programos „nekoreliuoti.sci“ ir programos „koreliuoti.sci“ tekstai pateikti 4 priede.

5. REZULTATAI

Pagal modelį, kurį aprašo (25) lygtis, vidutinio buto Kauno miesto centre vieno kvadratinio metro vertė 2009 metų pirmą ketvirtį buvo lygi maždaug 2510LT. Tai yra, vidutinis 45 kvadratinių metrų ploto dviejų kambarių butas Kauno miesto centre turėjo kainuoti apie 112 800LTL. Todėl tikėtina, kad investuotojui turint vidutinį butą, už kurį siūloma daug didesnė suma – jį apsimoka parduoti (arbitražo galimybė). Ir atvirkščiai – jei investuotojas turi galimybę nusipirkti daug pigesnį butą – jis pasinaudos arbitražo galimybe, jei šiuos veiksmus atliks.

Pastebėta, kad daugelis veiksnių, esančių tiesinėj priklausomybėj su kitais veiksniais, stipriai koreliuoja tarpusavyje, taigi ir su vienu nežinomu ekonominiu veiksniu.

6. DISKUSIJA

Atliekant darbą buvo pastebėta, kad vengiant multikolinearumo dažnai pavykdavo gauti koeficientus prie veiksnių tais pačiais ženklais, kaip ir veiksnių koreliacijos su priklausomu kintamuoju koeficiento ženklas. Darbo metu buvo pastebėta, kad daug viešai pateikiamų ekonominių veiksnių gerai koreliuoja tarpusavyje, vadinasi – seka vieną ir tą patį nežinomą ekonominį veiksnį. Aišku tai, kad dažniausiai daugelis aktyvų su tam tikru vėlavimu taip pat seka tą patį veiksnį. Tačiau nagrinėjant daugialypės regresijos modelius yra naudinga rasti kitus veiksnius, kurie nekoreliuotų su pirmaisiais. Darbo metu buvo bandyta susieti ekonominius pokyčius su socialiniais pokyčiais. Atvirose rinkose perkant aktyvus labai stipriai veikia socialiniai veiksniai. Taip pat, tose atvirose rinkose yra daug didesnis šansas užimti patogią padėtį arbitražo pozicijoje. Mano nuomone – jei pavyktų gerai įvertinti kelis socialinius veiksnius, kurio priklausomybę nuo priklausomojo kintamojo pavyktų paversti į tiesinę, tuomet galima būtų statistiškai žymiai pagerinti daugialypės regresijos modeliais vertinamus įkainojimo metodus. Tai yra, kaip vertinti žmogiškąjį veiksnį įkainojime?

Anksčiau ekonomistai rinkdavosi daugiausia ekonominius veiksnius, ir pasitikėdavo pačia rinka. Tačiau dabar aktyvai nėra padengti dažnai niekuo, tik pasitikėjimu. Be to, paprastai aktyvų palūkanos keičiais dėl atsiradusio ar dingusio patikimumo. Tačiau įdomu, kad JAV obligacijų vertę gali pakeisti ir Federalinio rezervų banko spausdinami doleriai. Ir nors nuo 1971m. Šie doleriai nėra padengti auksu ar kuo kitu, tačiau obligacijų patikimumas lieka panašus su panašią grąža. Nuo kokių veiksnių tai priklauso? Kokiais veiksniais galima įvertinti patikimumą?

Radus veiksnį, apibūdinantį žmonių lūkesčius, tačiau nesunkiai nustatomą iš kitų rodiklių, labai tikėtina, būtų galima patikslinti būstų kainų prognozavimą. Tas pats pasakytina apie atlyginimus. Yra labai sunku nustatyti momentinį vidutinį atlyginimą, ir pagal jį įkainoti aktyvus, dėl to geriau naudoti veiksnius, gerai koreliuojančius su ateities aktyvo kainomis.

Mano manymu - tyrimui gerai parinkti tuos ekonominius ar socialinius veiksnius, kuriuos lengva nuspėti iš anksto. Kalbant apie Chen, Roll ir Ross siūlytus veiksnius reikia suprasti, jog nuo tada, kai 1971m. USD valiuta nėra padengta auksu, o JAV valstybės skola nuolatos auga – šios valstybės skolos popieriai nebėra tokie patikimi pasaulio rinkose [4], ir dėl to galima kritiškiau žiūrėti į veiksnius, kurių parinkimas buvo siūlytas prieš 20 metų. Be to [4] nurodoma, kad infliacijos skaičiavimas ES šalyse ir JAV skiriasi.

7. IŠVADOS

Pritaikius kompiuterinį veiksnių perrinkimą tapo įmanoma daug greičiau surasti gerus modelius;

Kuo didesnė koreguotojo determinacijos koeficiento reikšmė, tuo didesnis gali būti multikolinearumas;

Butų kainas gerai apibūdina vienas ekonominis veiksnys, su kuriuo gerai koreliuoja daugelis darbe nagrinėtų veiksnių (priklausantys darbe nagrinėtai grupei A);

Butų kainas gerai apibūdina žmonių lūkesčiai. Darbe jie buvo vertinti pagal žmonių skaičių, kurie tikisi gauti didelį atlyginimą. Tačiau tikslesnis socialinis veiksnys vis dar lieka nežinomas.

Butų kainas gerai apibūdina populiarios valiutos palūkanos. Tačiau nagrinėti veiksniai, gerai koreliuojantys su butų kainomis, neturėjo tiesinės priklausomybės, dėl to negalėjo būti taikomi tiesinėje prognozėje (žr. Veiksnio B3 ir butų kainų porinių stebėjimų grafiką, 2 priedas)

Modelis, gerai apibūdinantis butų kainų pokyčius (koreguotasis determinacijos koeficientas 0.96), gautas iš veiksnių: bedarbystė miestuose (proc), gyventojų lūkesčiai,matuojami bedarbių, siekiančių gauti daugiau, negu 1500LTL per mėnesį, skaičiumi (tūkstančiais), taip pat materialinėmis investicijomis į šalį (milijardais litų).

Jei aktyvo grąžos stipriai koreliuoja su kokiais nors veiksniais, ypač, jei ši priklausomybė tiesinė, tuomet naudingiau yra naudoti arbitražo įkainojimo teorijos modelį. Kitu atveju – paprastos daugialypės regresijos lygties prognozavimo modelį.

8. REKOMENDACIJOS

- Šis darbas yra labai gera pradžia platesnei ir gilesnei daugialypių regresijos lygčių analizei, ir arbitražo modelių sudarymui. Esminė dalis, ant kurios pastatyta ši programa, yra visų galimų rinkinių perrinkimas, ir kai kurių sąlygų patikrinimas skaitmeniniais metodais.
- Tęsiant darbą būtų galima automatizuoti kitus patikrinimus apie modelio koeficientus ir apie patį modelį, tokiu būdu dar skaičiavimo metu atmetant netinkamus modelius.
- Taip pat galima plėsti duomenų paiešką. Kad duomenys ne tik būtų dažnai atnaujinami, tačiau ir lengvai prieinami, nebrangūs, ir gerai koreliuotų ne su vienu pagrindiniu rinkos veiksniu, tačiau su kitu (tokiu, kaip žmonių lūkesčiai).
- Būtų įdomu pabandyti įkainoti apvaliąją medieną. Medienos kainos koreliuoja su daugelio veiksnių reikšmėmis, buvusiomis prieš kelis mėnesius. Naudojant tokią analizę galima pasinaudoti arbitražo situacija, jau iš anksto įvertinant būsimas medienos kainas.

9. PADĖKOS

1. Dėkoju darbo vadovui, už literatūrą ir patarimus;
2. Dėkoju šeimai už pagalbą ieškant medžiagos ir atliekant šį darbą;
3. Dėkoju universitetui, kad suteikė galimybes giliau žiūrėti į mokslą.

10. NUORODOS (ŠALTINIAI IR LITERATŪRA)

[1] – Investment science, David G. Luenberger, p. 197-223,Oxford University Press,USA 1997;

[2] – Statistika ir jos taikymai II, V. Čekanavičius, G. Murauskas, p122-196, TEV, Vilnius, 2008

[3] - http://en.wikipedia.org/wiki/Arbitrage_pricing_theory

[4] - http://www.marketnews.lt/naujiena/apie_jav_kinija_ir_rinku_sugriuvima_2009-01-30

[5] –Analysis of financial data, Gary Koop, p. 91-135, John Wiley & Sons, 2006

[6] - www.newyorkfed.org/research/staff_reports/sr216.pdf

[7] - http://en.wikipedia.org/wiki/Arbitrage

[8] - http://en.wikipedia.org/wiki/Least-squares_estimation_of_linear_regression_coefficients

[9] – Modulio „Matematinė statistika P160B101“ paskaitų konspektai. Dėstytojas doc. A. Jokimaitis.

[10] – http://www.scilab.org/product/man/ - išsamus kai kurių funkcijų aprašymas

[11] - Modern portfolio theory and investment analysis, Edwin J. Elton, Martin J. Gruber, Stephen J. Brown, William N. Goetzmann, p. 159-177, John Wiley & Sons, 7 edition, 2006;

1 PRIEDAS. PRADINIAI VEIKSNIAI. PRADINIAI DUOMENYS.

Parinkti pradiniai veiksniai, jų reikšmės. Parinktų butų kainų reikšmės.

12. lentelė: Stipriai koreliuojantys tarpusavyje veiksniai (pirmoji lentelė)

	A1	A2	A3	A4	A5	A6	A7	A8
2000								
	11.57	102.34	0.65	3.27	4.93	5.07	17.70	23.98
	11.71	92.33	0.67	3.74	4.57	5.19	17.65	28.03
	12.06	92.26	0.67	3.70	4.92	5.85	17.60	29.05
2001	10.98	92.70	0.63	4.06	4.77	5.48	17.95	21.15
	12.26	87.39	0.65	4.41	5.07	5.98	18.30	23.10
	12.50	82.13	0.67	4.27	5.19	5.76	18.60	24.18
	12.90	69.80	0.65	4.38	5.85	7.03	18.90	20.88
2002	11.41	75.56	0.63	4.07	5.48	5.84	18.70	18.90
	13.13	89.78	0.67	5.12	5.98	7.25	15.10	25.20
	13.68	83.49	0.70	4.93	5.76	6.64	13.60	23.38
	13.86	88.17	0.70	5.00	7.03	7.74	14.30	27.47
2003	12.68	84.78	0.67	5.05	5.84	6.53	14.80	27.85
	14.13	96.77	0.71	4.92	7.25	7.12	14.30	24.35
	14.82	134.95	0.73	5.50	6.64	7.35	12.70	25.32
	15.33	173.94	0.73	5.78	7.74	8.45	12.40	27.43
2004	13.52	174.48	0.74	5.52	6.53	7.29	14.40	28.67
	15.48	222.91	0.81	6.11	7.12	8.55	11.60	29.62
	16.37	206.25	0.85	6.73	7.35	8.89	11.40	31.02
	17.33	224.19	0.87	7.45	8.45	9.64	10.90	41.17
2005	15.06	293.97	0.86	6.82	7.29	8.43	10.30	35.73
	17.74	342.38	0.89	7.80	8.55	10.70	8.80	49.20
	19.15	411.78	0.95	8.56	8.89	11.26	7.70	54.68
	20.12	527.05	1.00	9.59	9.64	12.76	7.40	56.07
2006	17.34	451.47	1.00	9.07	8.43	11.86	6.10	56.07
	20.25	436.67	1.06	10.13	10.70	13.35	5.70	62.78
	22.21	387.66	1.12	10.03	11.26	13.93	5.60	68.03
	23.00	420.92	1.17	9.65	12.76	14.13	4.60	55.05
2007	20.22	493.48	1.18	9.65	11.86	13.72	5.00	53.77
	24.38	492.01	1.29	10.94	13.35	15.87	4.10	63.03
	26.20	529.42	1.36	11.56	13.93	15.97	4.00	70.41
	27.34	569.79	1.42	11.03	14.13	15.95	3.80	76.23
2008	12.97	73.15	0.66	4.40	5.71	6.78	18.76	22.30

13. lentelė: Veiksniai, stipriia koreliuojantys tarpusavyje (antroji lentelė)

	A9	A10	A11	A12	A13	A14	A15
2000							
	546	149.56	26.25	17.88	18.35	15.66	8.30
	471	151.81	25.02	19.01	17.51	17.88	8.34
	497	151.96	24.07	18.35	17.43	19.01	8.39
2001	324	148.18	24.52	17.51	17.75	18.35	8.42
	596	149.52	23.80	17.43	18.10	17.51	8.45
	519	149.20	23.45	17.75	18.37	17.43	8.46
	503	147.69	24.07	18.10	18.61	17.75	8.47
2002	309	144.46	23.75	18.37	19.11	18.10	8.49
	721	147.51	23.92	18.61	17.68	18.37	8.52
	584	149.28	25.77	19.11	14.23	18.61	8.56
	678	149.86	25.83	17.68	13.72	19.11	8.60
2003	424	146.12	26.81	14.23	14.55	17.68	8.61
	835	149.35	26.11	13.72	14.77	14.23	8.61
	820	151.45	26.95	14.55	13.82	13.72	8.62
	780	150.36	27.79	14.77	12.26	14.55	8.66
2004	589	152.30	30.19	13.82	13.22	14.77	8.78
	1200	157.05	29.93	12.26	13.76	13.82	8.99
	1085	164.04	30.54	13.22	11.21	12.26	9.25
	1127	167.67	29.95	13.76	11.34	13.22	9.50
2005	781	167.50	31.15	11.21	10.73	13.76	9.69
	1592	169.17	31.42	11.34	9.88	11.21	9.77
	1484	176.20	29.03	10.73	8.31	11.34	9.77
	1631	181.71	29.25	9.88	7.62	10.73	9.73
2006	1151	183.17	29.29	8.31	7.02	9.88	9.70
	1923	188.70	29.29	7.62	5.83	8.31	9.69
	1971	195.18	30.59	7.02	5.75	7.62	9.72
	2441	199.89	31.10	5.83	5.24	7.02	9.77
2007	1760	203.27	32.49	5.75	4.68	5.83	9.83
	2268	217.99	32.97	5.24	4.79	5.75	9.90
	2459	227.49	33.72	4.68	3.99	5.24	9.95
	2313	235.28	33.89	4.79	3.90	4.68	10.00
2008	530	150.24	24.26	3.99	4.03	4.79	10.02

14. lentelė: Įvairūs veiksniai (pirmoji lentelė)

	B1	B2	B3	B4	B5	B6	B7
2000							
	1.796	1.427	2.942	2.507	0.700	1.047	3.917
	1.457	1.739	2.655	3.014	1.799	1.049	3.709
	2.148	2.232	3.126	2.799	5.500	1.138	2.099
2001	1.427	1.855	3.338	2.704	7.279	1.062	3.916
	1.563	1.427	3.833	2.507	8.017	1.144	2.359
	1.489	1.739	4.551	3.014	8.592	1.177	1.895
	2.646	2.232	4.991	2.799	9.337	1.099	0.664
2002	1.778	1.549	4.844	2.395	8.402	1.122	1.851
	2.133	1.661	4.571	2.452	4.841	1.147	0.084
	1.711	2.004	4.435	2.550	2.989	1.009	0.312
	2.739	1.872	3.656	3.182	4.503	1.013	0.251
2003	1.475	1.544	3.279	2.665	4.199	0.953	0.619
	2.194	2.879	3.446	2.518	3.300	0.915	0.466
	1.843	3.683	3.438	2.454	6.500	0.869	0.555
	2.663	2.926	3.279	2.962	5.400	0.857	0.459
2004	1.771	3.023	2.861	2.507	5.800	0.795	0.610
	2.438	3.469	2.514	2.790	9.600	0.813	0.710
	2.161	4.667	2.145	2.358	6.700	0.820	0.584
	2.194	4.006	2.121	3.342	8.300	0.804	0.532
2005	1.612	3.722	2.120	2.689	6.400	0.737	0.490
	2.903	4.137	1.960	2.849	6.000	0.772	0.293
	2.695	4.740	2.116	2.447	6.200	0.826	0.243
	3.174	2.972	2.148	3.112	8.600	0.832	0.132
2006	2.786	2.653	2.150	2.750	9.100	0.845	0.051
	3.218	3.074	2.147	2.733	8.800	0.824	0.002
	3.902	3.749	2.107	2.414	9.200	0.786	0.035
	4.480	3.126	2.178	3.305	16.700	0.789	0.046
2007	4.061	3.011	2.489	2.827	12.500	0.758	0.065
	4.926	3.545	2.818	2.876	20.500	0.749	0.233
	4.405	4.311	3.055	2.387	20.000	0.739	0.724
	4.920	2.986	3.424	3.246	20.300	0.701	0.869
2008	5.382	2.183	4.906	2.951	22.800	0.680	0.871

15. lentelė: Įvairūs veiksniai (antroji lentelė)

	B8	B9	B10	B11	B12	B13	B14
2000							
	1.518	3.833	3.897	1.037	0.967	1.012	88
	1.761	4.551	6.398	1.067	0.933	1.022	65
	2.319	4.991	3.189	1.126	0.993	1.040	87
2001	1.076	4.844	2.280	1.147	1.047	1.047	84
	1.560	4.571	3.897	1.083	1.049	1.027	84
	1.985	4.435	6.398	1.183	1.138	1.057	99
	3.203	3.656	3.189	1.098	1.062	1.032	104
2002	1.315	3.279	2.244	1.116	1.144	1.037	106
	2.009	3.446	3.992	1.151	1.177	1.048	101
	2.089	3.438	6.564	1.071	1.099	1.023	104
	2.712	3.279	3.351	1.020	1.122	1.007	93
2003	1.362	2.861	2.399	1.006	1.147	1.002	87
	2.071	2.514	3.821	0.926	1.009	0.975	99
	2.231	2.145	7.206	0.850	1.013	0.947	101
	3.013	2.121	3.549	0.910	0.953	0.969	100
2004	1.508	2.120	2.903	0.835	0.915	0.942	128
	2.236	1.960	4.445	0.801	0.869	0.929	108
	2.711	2.116	7.977	0.821	0.857	0.936	89
	3.642	2.148	3.805	0.821	0.795	0.937	123
2005	1.807	2.150	2.551	0.752	0.813	0.909	110
	2.545	2.147	4.470	0.755	0.820	0.911	121
	3.137	2.107	8.864	0.813	0.804	0.933	100
	4.697	2.178	4.053	0.811	0.737	0.932	132
2006	2.303	2.489	2.863	0.846	0.772	0.946	119
	3.365	2.818	4.698	0.842	0.826	0.944	123
	3.880	3.055	9.576	0.778	0.832	0.920	97
	5.621	3.424	4.109	0.779	0.845	0.920	118
2007	3.554	3.725	2.976	0.758	0.824	0.912	153
	4.627	3.927	5.268	0.757	0.786	0.911	145
	5.209	4.176	10.462	0.744	0.789	0.906	159
	6.923	4.791	4.359	0.732	0.758	0.901	159
2008	3.067	3.934	4.446	0.679	0.749	0.879	180

16. lentelė: Vidutinės būstų kainos Kauno miesto centre (LTL už kvadratinį metrą)

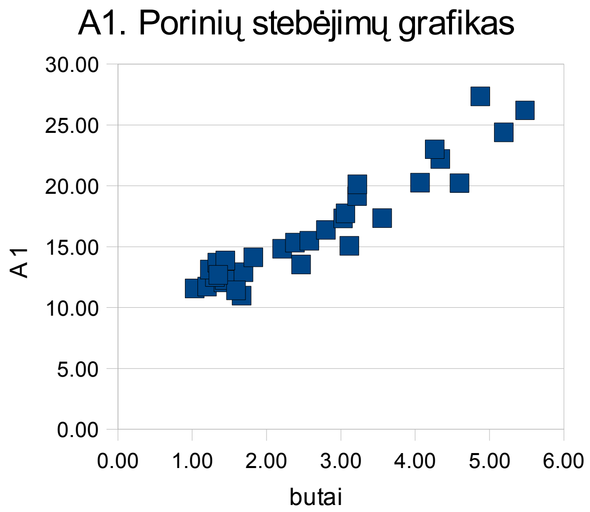
Periodas	Svoriai	Vidutinės vertės (LTL už kvadratinį metrą tą ketvirtį)	Prognozuotos reikšmės (tt=5)	Prognozuotos reikšmės (t =3)
1998 I	8	1345	1345	1345
1998 II	11	1224.91	1224.91	1224.91
1998 III	11	1292.27	1292.27	1292.27
1998 IV	8	1318	1318	1318
1999 I	5	1158.4	1090.33	1158.4
1999 II	1	1016	898.12	1023.44
1999 III	3	1288	815.44	1288
1999 IV	4	1838	823.31	1838
2000 I	3	1378.67	902.8	1378.67
2000 II	4	1060	1034.95	1060
2000 III	5	1108.4	1200.8	1108.4
2000 IV	7	1501.43	1420.25	1501.43
2001 I	10	1635.7	1664.42	1635.7
2001 II	4	1498.67	1440.52	1498.67
2001 III	5	1275.4	1307.74	1275.41
2001 IV	18	1692.11	1688.43	1692.1
2002 I	6	1588.17	1591.18	1588.29
2002 II	12	1238.75	1238.25	1238.53
2002 III	7	1337.14	1337.98	1338.52
2002 IV	10	1448.8	1447.05	1445.21
2003 I	12	1345.5	1350.85	1356.68
2003 II	15	1828.71	1822.08	1814.04
2003 III	0	2060.61	2210.41	2264.77
2003 IV	2	2292.5	2382.9	2492
2004 I	4	2592.5	2463.62	2477.06
2004 II	9	2458.11	2576.62	2483.05
2004 III	3	3137	2800.2	3116.95
2004 IV	4	3008	3029.55	3012.03
2005 I	9	3056.71	3114.1	3056.23
2005 II	7	3132.86	3062.82	3133.03
2005 III	5	3060.2	3218.98	3060.14
2005 IV	8	3437.88	3220.76	3437.89
2006 I	4	3001	3555.75	3000.99
2006 II	2	4994	4064.5	3513.52
2006 III	7	4229.75	4338.17	4229.75
2006 IV	11	4280.5	4262.05	4280.5
2007 I	4	4982.75	4596.92	4982.75
2007 II	15	5193.71	5193.34	5193.71
2007 III	5	5476.4	5476.4	5476.4
2007 IV	5	4877.2	4877.2	4877.2
2008 I	8	4716.88	4716.88	4716.88

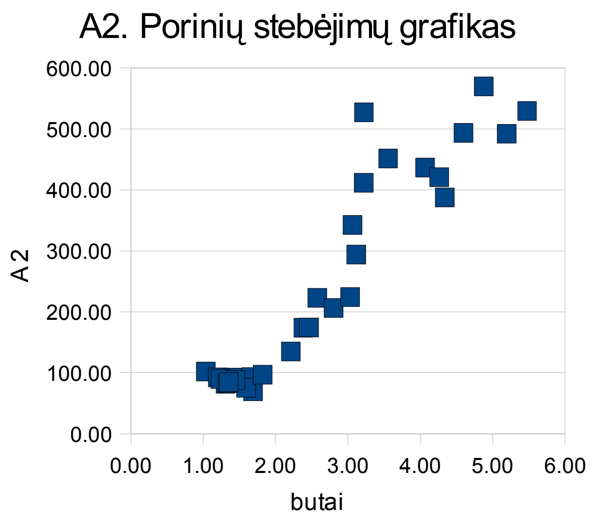
17. lentelė: Suglodintų ir realių verčių palyginimas valiurų EUR (euro) ir USD (JAV dolerio) kursų santykiui, 1991.01 - 2007.01 laikotarpiu, mėnesio pirmą darbo dieną

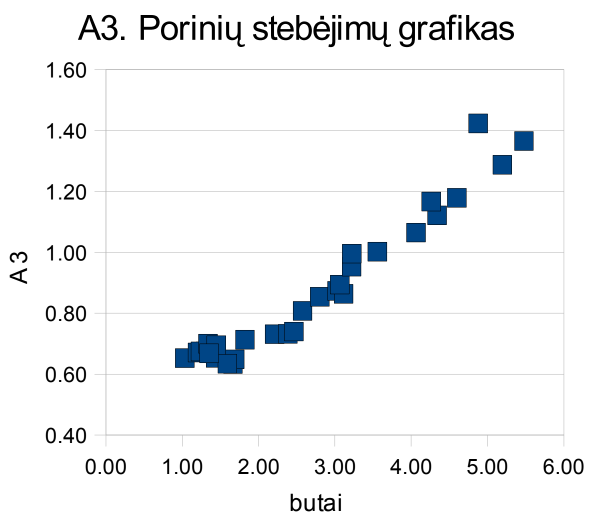
Tikros istorinės reikšmės	Reikšmės, prognozuotos glodinančiu ir interpoliaciniu splainu
0.8570 1.0670 1.0994 0.9290 0.8207 0.8320 0.8814 1.0491 1.1113 0.9261 0.8201 0.8290 0.9105 1.0787 1.1162 0.9153 0.8315 0.8462 0.9287 1.1264 1.1218 0.8952 0.8214 0.8446 0.9453 1.1375 1.1655 0.8495 0.8041 0.8258 0.9588 1.1786 1.1510 0.8693 0.7809 0.8422 0.9674 1.1468 1.1474 0.8904 0.7522 0.8240 0.9334 1.0620 1.1111 0.9103 0.7372 0.7921 0.9427 1.0679 1.0709 0.8573 0.7674 0.7777 0.9326 1.0832 1.0088 0.8635 0.7552 0.7862 0.9483 1.1439 1.0234 0.8349 0.7715 0.7837 0.9908 1.1274 1.0201 0.7952 0.7768 0.7794 0.9930 1.1826 1.0129 0.8026 0.8127 0.7893 1.0308 1.1772 1.0104 0.8007 0.8264 0.7862 1.0365 1.1423 1.0062 0.8130 0.8249 0.7582 1.0472 1.0976 0.9526 0.8347 0.8106 0.7580 1.0975	0.8570 1.0390 1.0994 0.9370 0.8201 0.8320 0.8775 1.0491 1.0932 0.9266 0.8201 0.8360 0.9028 1.0821 1.1024 0.9153 0.8202 0.8414 0.9287 1.1199 1.1218 0.8989 0.8165 0.8446 0.9510 1.1375 1.1446 0.8815 0.8041 0.8426 0.9653 1.1192 1.1580 0.8693 0.7806 0.8355 0.9674 1.0845 1.1474 0.8660 0.7545 0.8240 0.9558 1.0620 1.1051 0.8651 0.7372 0.8095 0.9401 1.0725 1.0503 0.8573 0.7371 0.7956 0.9326 1.1058 1.0088 0.8370 0.7507 0.7862 0.9423 1.1439 0.9988 0.8124 0.7715 0.7841 0.9650 1.1714 1.0071 0.7952 0.7937 0.7865 0.9930 1.1833 1.0129 0.7936 0.8131 0.7893 1.0192 1.1772 1.0007 0.8024 0.8264 0.7882 1.0388 1.1535 0.9767 0.8130 0.8314 0.7792 1.0472 1.1230 0.9526 0.8186 0.8316 0.7580 1.0433

2 PRIEDAS. PORINIŲ STEBĖJIMŲ GRAFIKAI

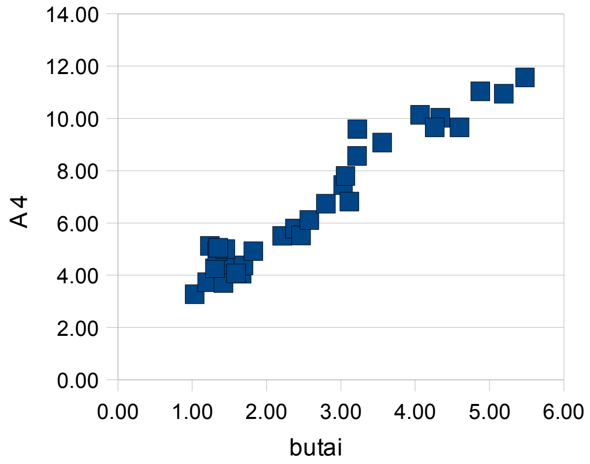
Veiksnių ir butų kainų porinių stebėjimų grafikai. Pagal veiksnio pavadinimą.



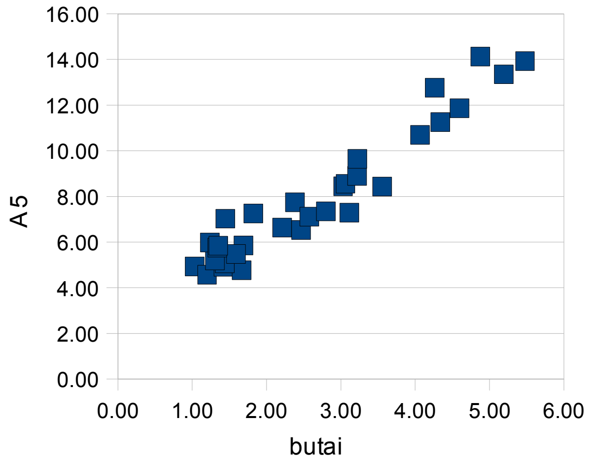




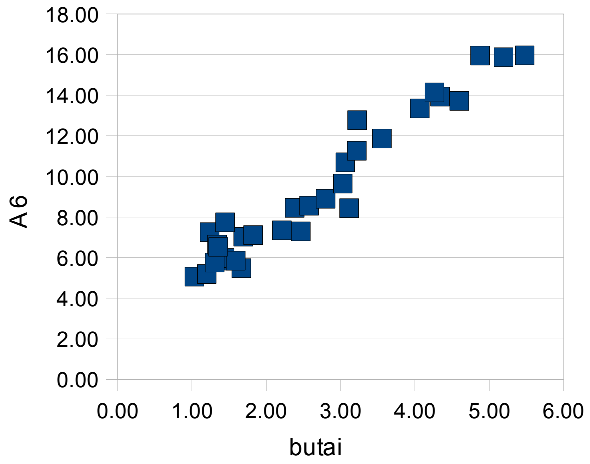
A4. Porinių stebėjimų grafikas



A5. Porinių stebėjimų grafikas

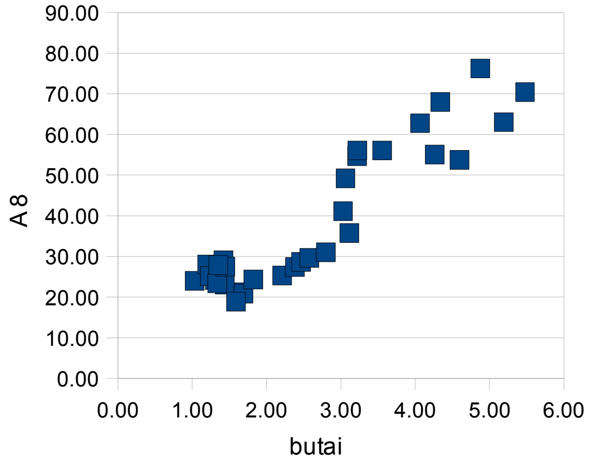


A6. Porinių stebėjimų grafikas

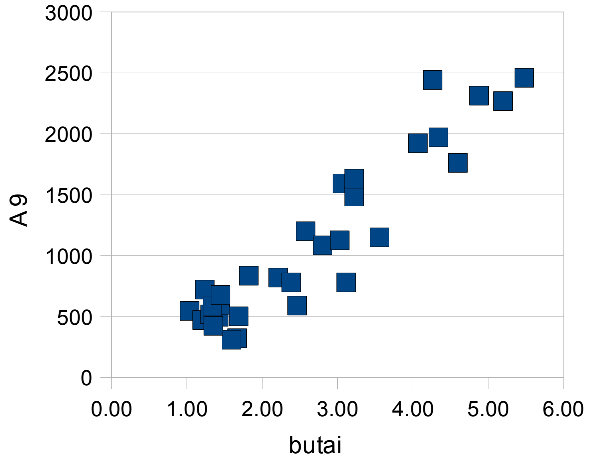


A7. Porinių stebėjimų grafikas

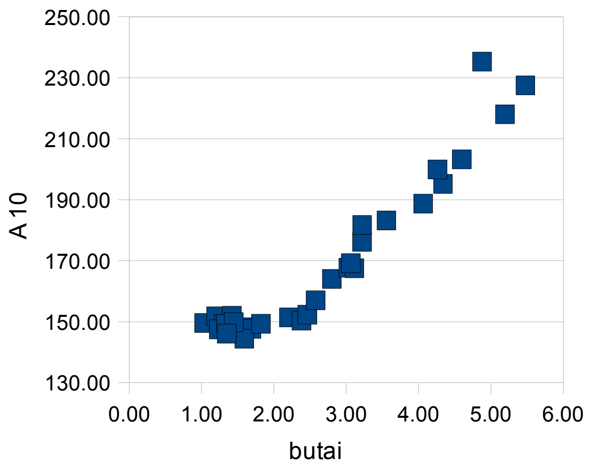
A8. Porinių stebėjimų grafikas

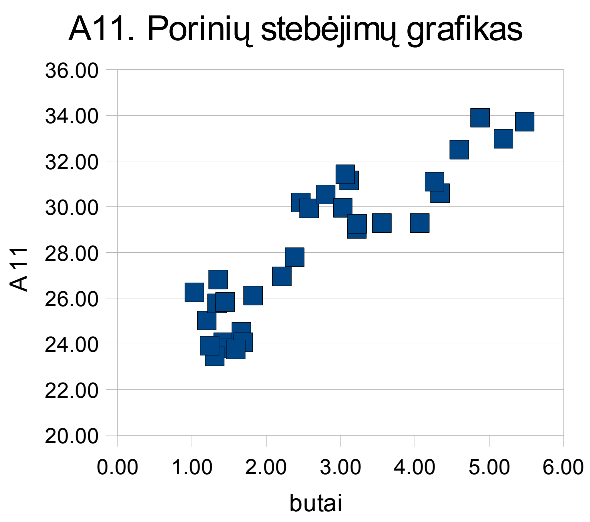


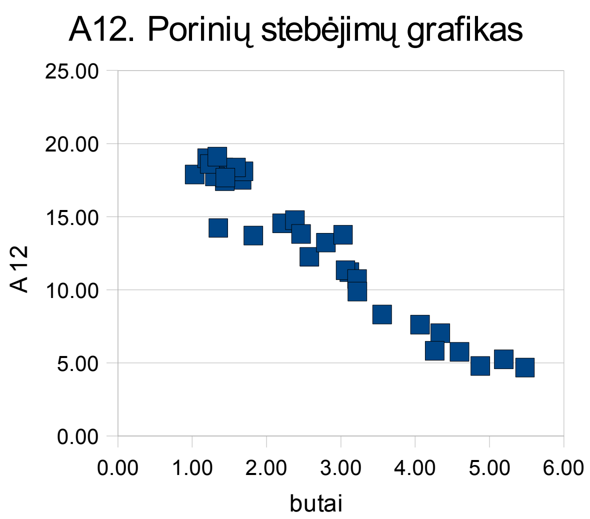
A9. Porinių stebėjimų grafikas

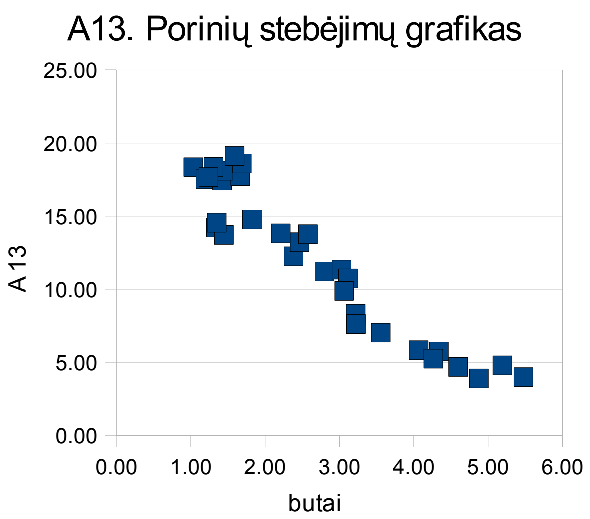


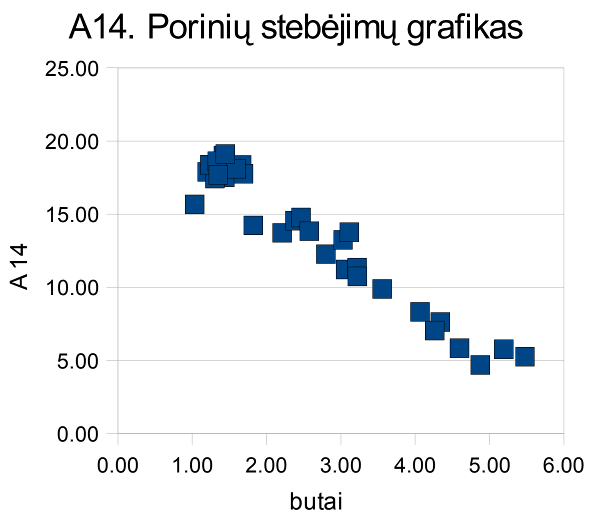
A10. Porinių stebėjimų grafikas

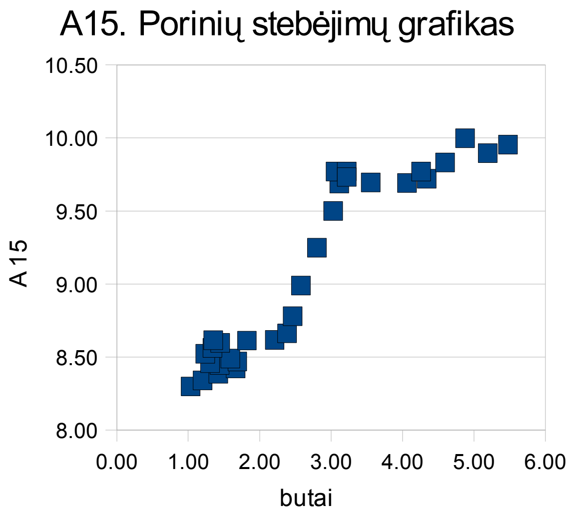


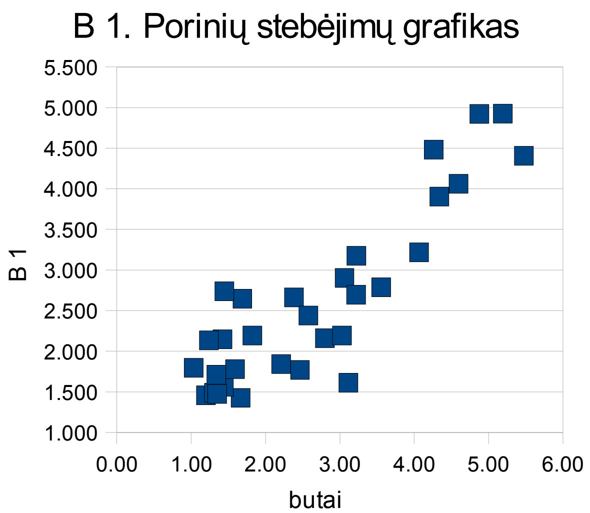

A11. Porinių stebėjimų grafikas
36.00
34.00
32.00
30.00
28.00
26.00
24.00
22.00
20.00
A 11
0.00
1.00
2.00
3.00
4.00
5.00
6.00
butai


A12. Porinių stebėjimų grafikas
25.00
20.00
15.00
10.00
5.00
0.00
A 12
0.00
1.00
2.00
3.00
4.00
5.00
6.00
butai

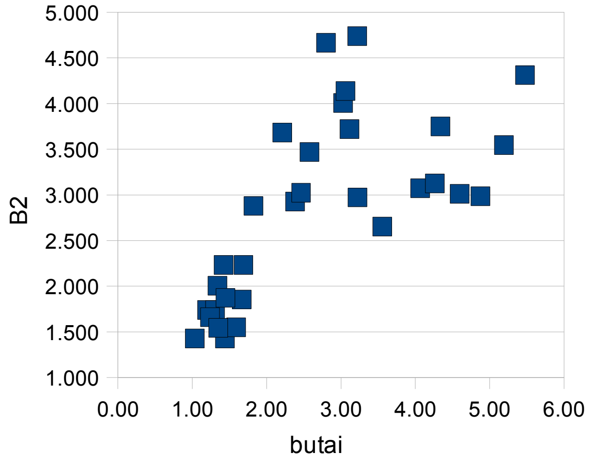

A13. Porinių stebėjimų grafikas
25.00
20.00
15.00
10.00
5.00
0.00
A 13
0.00
1.00
2.00
3.00
4.00
5.00
6.00
butai


A14. Porinių stebėjimų grafikas
25.00
20.00
15.00
10.00
5.00
0.00
A 14
0.00
1.00
2.00
3.00
4.00
5.00
6.00
butai

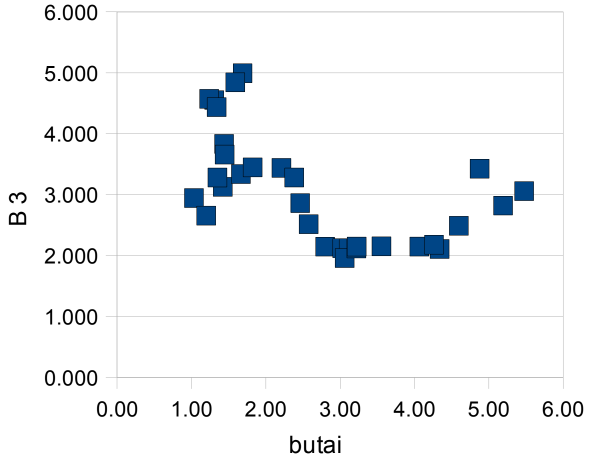

A15. Porinių stebėjimų grafikas
10.50
10.00
9.50
9.00
8.50
8.00
A 15
0.00
1.00
2.00
3.00
4.00
5.00
6.00
butai


B 1. Porinių stebėjimų grafikas
5.500
5.000
4.500
4.000
3.500
3.000
2.500
2.000
1.500
1.000
B 1
0.00
1.00
2.00
3.00
4.00
5.00
6.00
butai

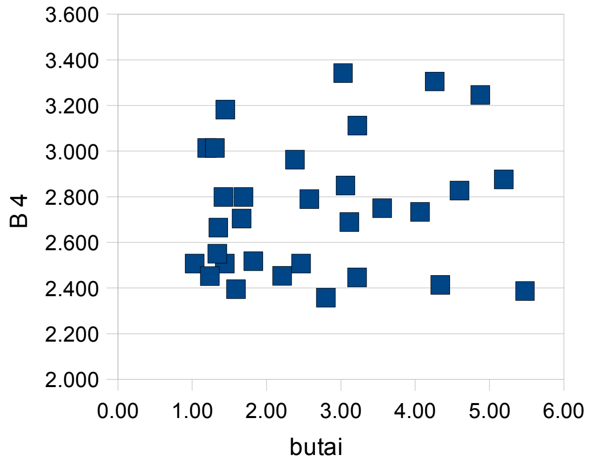
B 2. Porinių stebėjimų grafikas



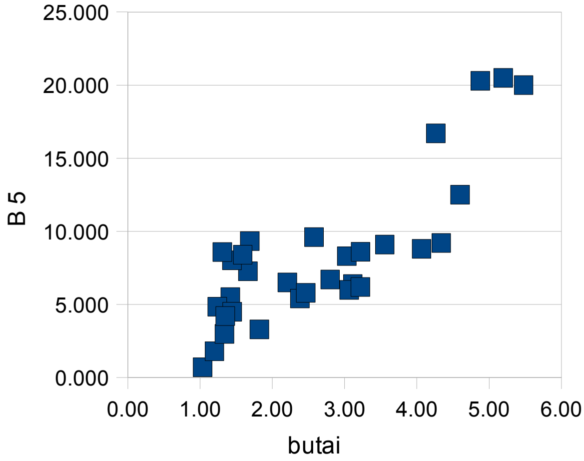
B 3. Porinių stebėjimų grafikas



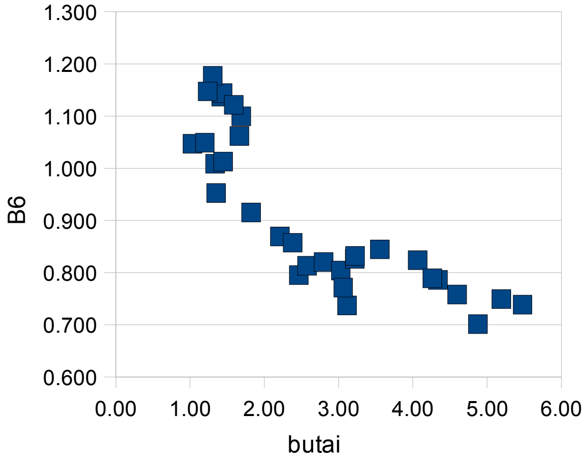
B 4. Porinių stebėjimų grafikas



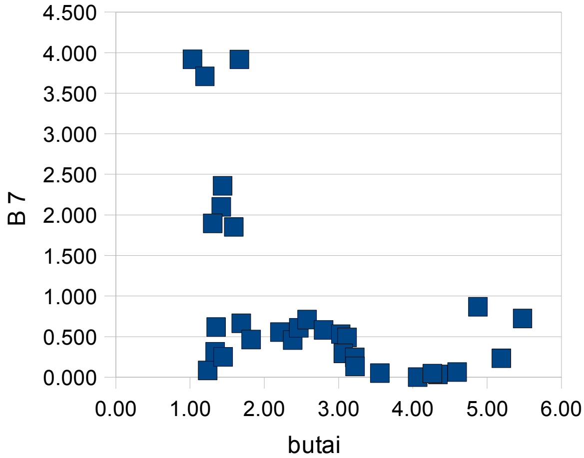
B 5. Porinių stebėjimų grafikas



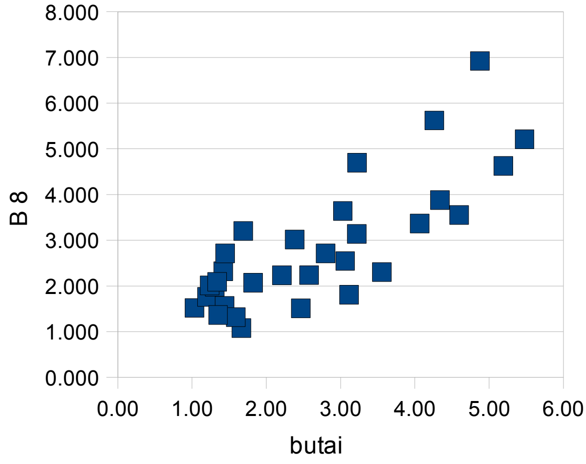
B 6. Porinių stebėjimų grafikas



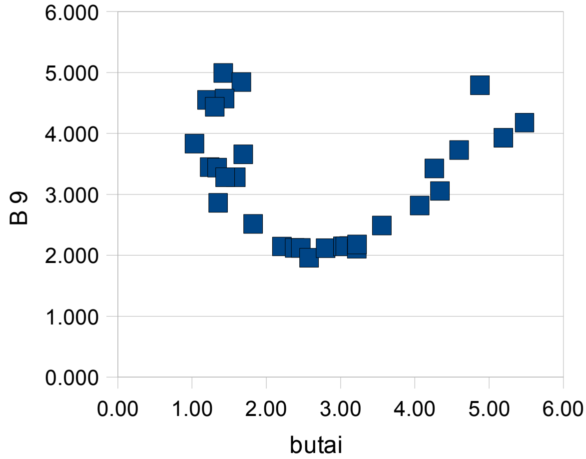
B 7. porinių stebėjimų grafikas



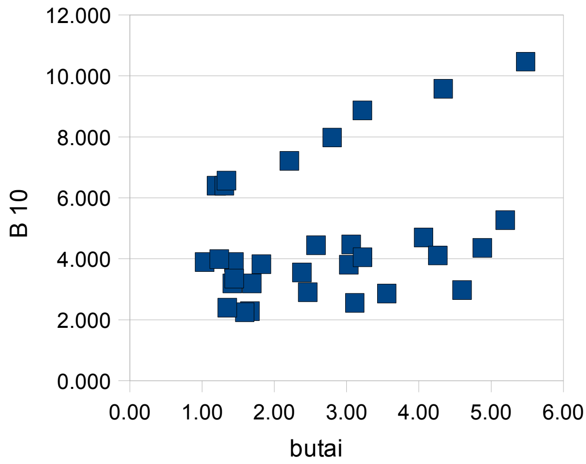
B 8. Porinių stebėjimų grafikas

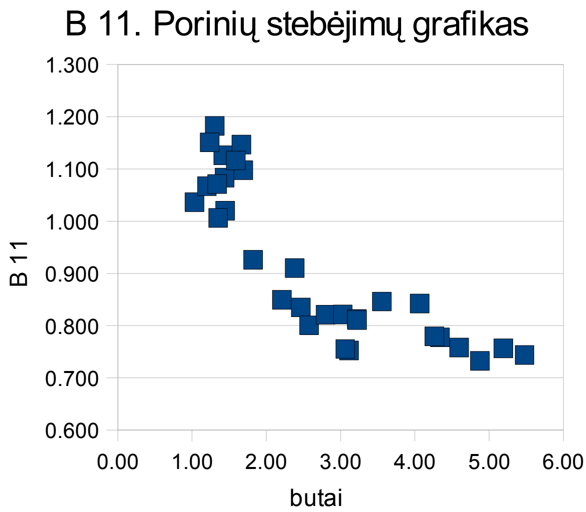


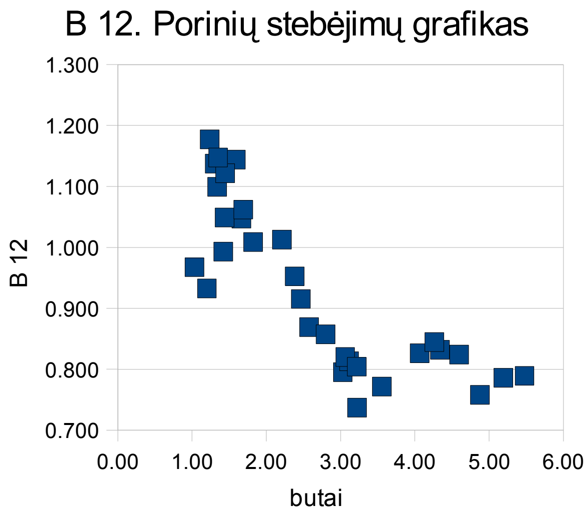
B 9. Porinių stebėjimų grafikas

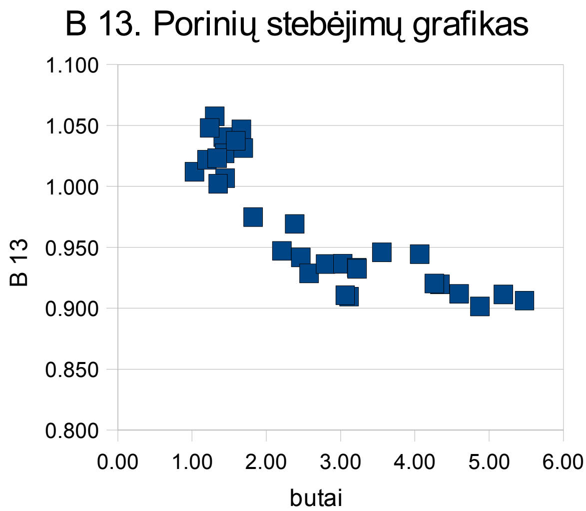


B 10. Porinių stebėjimų grafikas

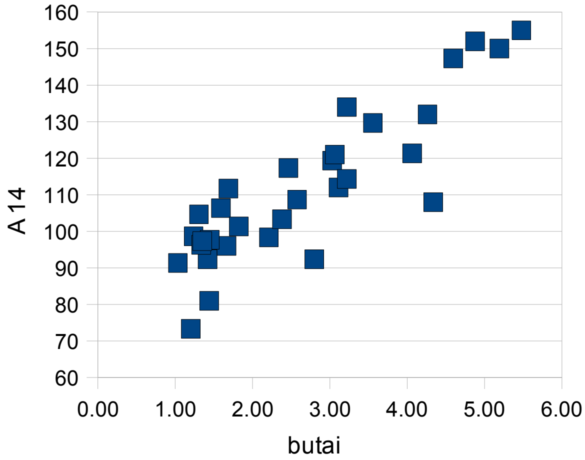



B 11. Porinių stebėjimų grafikas
1.300
1.200
1.100
1.000
0.900
0.800
0.700
0.600
B 11
0.00
1.00
2.00
3.00
4.00
5.00
6.00
butai


B 12. Porinių stebėjimų grafikas
1.300
1.200
1.100
1.000
0.900
0.800
0.700
B 12
0.00
1.00
2.00
3.00
4.00
5.00
6.00
butai


B 13. Porinių stebėjimų grafikas
1.100
1.050
1.000
0.950
0.900
0.850
0.800
B 13
0.00
1.00
2.00
3.00
4.00
5.00
6.00
butai

A14. Porinių stebėjimų grafikas



3 PRIEDAS. VEIKSNIUS APIBŪDINANČIOS KORELIACINĖS MATRICOS

18. lentelė: Veiksnius apibūdinanti darbinė koreliacinė matrica (pirmoji lentelė)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	1	0.92	0.98	0.96	0.98	0.98	-0.94	0.95	0.98	0.97
A2	0.92	1	0.95	0.97	0.91	0.95	-0.95	0.96	0.91	0.93
A3	0.98	0.95	1	0.96	0.98	0.98	-0.94	0.95	0.96	0.99
A4	0.96	0.97	0.96	1	0.96	0.99	-0.97	0.96	0.95	0.94
A5	0.98	0.91	0.98	0.96	1	0.98	-0.93	0.92	0.97	0.96
A6	0.98	0.95	0.98	0.99	0.98	1	-0.96	0.96	0.97	0.96
A7	-0.94	-0.95	-0.94	-0.97	-0.93	-0.96	1	-0.93	-0.93	-0.9
A8	0.95	0.96	0.95	0.96	0.92	0.96	-0.93	1	0.93	0.95
A9	0.98	0.91	0.96	0.95	0.97	0.97	-0.93	0.93	1	0.94
A10	0.97	0.93	0.99	0.94	0.96	0.96	-0.9	0.95	0.94	1
A11	0.86	0.86	0.87	0.87	0.86	0.85	-0.9	0.81	0.84	0.84
A12	-0.93	-0.94	-0.95	-0.96	-0.94	-0.95	0.96	-0.92	-0.93	-0.92
A13	-0.94	-0.95	-0.94	-0.97	-0.94	-0.96	0.99	-0.94	-0.92	-0.91
A14	-0.95	-0.94	-0.96	-0.95	-0.95	-0.95	0.94	-0.92	-0.95	-0.94
A15	0.9	0.95	0.91	0.96	0.89	0.93	-0.95	0.92	0.89	0.87
B1	0.93	0.84	0.92	0.87	0.95	0.93	-0.84	0.86	0.93	0.91
B2	0.63	0.61	0.58	0.65	0.59	0.61	-0.69	0.57	0.63	0.52
B3	-0.47	-0.61	-0.49	-0.57	-0.45	-0.51	0.64	-0.58	-0.53	-0.44
B4	0.25	0.23	0.23	0.2	0.26	0.24	-0.2	0.23	0.24	0.23
B5	0.82	0.72	0.84	0.76	0.83	0.81	-0.67	0.72	0.79	0.86
B6	-0.78	-0.8	-0.78	-0.81	-0.78	-0.78	0.87	-0.74	-0.76	-0.72
B7	-0.53	-0.49	-0.46	-0.6	-0.55	-0.58	0.64	-0.46	-0.51	-0.38
B8	0.92	0.79	0.87	0.82	0.9	0.87	-0.78	0.82	0.88	0.87
B9	-0.04	-0.12	0.03	-0.16	-0.04	-0.08	0.25	-0.02	-0.05	0.14
B10	0.42	0.28	0.36	0.36	0.33	0.34	-0.34	0.38	0.42	0.35
B11	-0.8	-0.82	-0.79	-0.83	-0.79	-0.8	0.88	-0.76	-0.79	-0.73
B12	-0.76	-0.87	-0.78	-0.81	-0.73	-0.78	0.82	-0.81	-0.76	-0.75
B13	-0.81	-0.83	-0.8	-0.84	-0.8	-0.81	0.89	-0.77	-0.8	-0.74
B14	0.79	0.8	0.82	0.8	0.82	0.81	-0.75	0.73	0.74	0.81

19. lentelė: Veiksnius apibūdinanti darbinė koreliacinė matrica (antroji lentelė)

	A11	A12	A13	A14	A15	B1	B2	B3	B4
A1	0.86	-0.93	-0.94	-0.95	0.9	0.93	0.63	-0.47	0.25
A2	0.86	-0.94	-0.95	-0.94	0.95	0.84	0.61	-0.61	0.23
A3	0.87	-0.95	-0.94	-0.96	0.91	0.92	0.58	-0.49	0.23
A4	0.87	-0.96	-0.97	-0.95	0.96	0.87	0.65	-0.57	0.2
A5	0.86	-0.94	-0.94	-0.95	0.89	0.95	0.59	-0.45	0.26
A6	0.85	-0.95	-0.96	-0.95	0.93	0.93	0.61	-0.51	0.24
A7	-0.9	0.96	0.99	0.94	-0.95	-0.84	-0.69	0.64	-0.2
A8	0.81	-0.92	-0.94	-0.92	0.92	0.86	0.57	-0.58	0.23
A9	0.84	-0.93	-0.92	-0.95	0.89	0.93	0.63	-0.53	0.24
A10	0.84	-0.92	-0.91	-0.94	0.87	0.91	0.52	-0.44	0.23
A11	1	-0.9	-0.9	-0.9	0.9	0.74	0.76	-0.68	0.17
A12	-0.9	1	0.96	0.97	-0.93	-0.85	-0.64	0.62	-0.17
A13	-0.9	0.96	1	0.94	-0.95	-0.85	-0.68	0.64	-0.22
A14	-0.9	0.97	0.94	1	-0.91	-0.87	-0.66	0.59	-0.15
A15	0.9	-0.93	-0.95	-0.91	1	0.77	0.74	-0.68	0.22
B1	0.74	-0.85	-0.85	-0.87	0.77	1	0.45	-0.32	0.31
B2	0.76	-0.64	-0.68	-0.66	0.74	0.45	1	-0.66	-0.06
B3	-0.68	0.62	0.64	0.59	-0.68	-0.32	-0.66	1	-0.17
B4	0.17	-0.17	-0.22	-0.15	0.22	0.31	-0.06	-0.17	1
B5	0.65	-0.76	-0.68	-0.76	0.67	0.84	0.37	-0.12	0.27
B6	-0.96	0.86	0.88	0.84	-0.86	-0.64	-0.81	0.73	-0.16
B7	-0.52	0.55	0.62	0.48	-0.59	-0.49	-0.54	0.24	-0.07
B8	0.69	-0.76	-0.79	-0.8	0.74	0.91	0.47	-0.29	0.47
B9	-0.3	0.17	0.22	0.12	-0.27	0.1	-0.55	0.43	0.12
B10	0.28	-0.27	-0.33	-0.35	0.29	0.26	0.53	-0.18	-0.33
B11	-0.95	0.87	0.88	0.86	-0.88	-0.66	-0.84	0.75	-0.13
B12	-0.86	0.81	0.82	0.83	-0.87	-0.63	-0.75	0.83	-0.25
B13	-0.96	0.88	0.88	0.87	-0.89	-0.67	-0.84	0.75	-0.14
B14	0.77	-0.77	-0.76	-0.78	0.75	0.76	0.45	-0.22	0.23

20. lentelė: Veiksnius apibūdinanti darbinė koreliacinė matrica (trečioji lentelė)

	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14
A1	0.82	-0.78	-0.53	0.92	-0.04	0.42	-0.8	-0.76	-0.81	0.85
A2	0.72	-0.8	-0.49	0.79	-0.12	0.28	-0.82	-0.87	-0.83	0.87
A3	0.84	-0.78	-0.46	0.87	0.03	0.36	-0.79	-0.78	-0.8	0.88
A4	0.76	-0.81	-0.6	0.82	-0.16	0.36	-0.83	-0.81	-0.84	0.86
A5	0.83	-0.78	-0.55	0.9	-0.04	0.33	-0.79	-0.73	-0.8	0.88
A6	0.81	-0.78	-0.58	0.87	-0.08	0.34	-0.8	-0.78	-0.81	0.88
A7	-0.67	0.87	0.64	-0.78	0.25	-0.34	0.88	0.82	0.89	-0.82
A8	0.72	-0.74	-0.46	0.82	-0.02	0.38	-0.76	-0.81	-0.77	0.81
A9	0.79	-0.76	-0.51	0.88	-0.05	0.42	-0.79	-0.76	-0.8	0.81
A10	0.86	-0.72	-0.38	0.87	0.14	0.35	-0.73	-0.75	-0.74	0.87
A11	0.65	-0.96	-0.52	0.69	-0.3	0.28	-0.95	-0.86	-0.96	0.79
A12	-0.76	0.86	0.55	-0.76	0.17	-0.27	0.87	0.81	0.88	-0.85
A13	-0.68	0.88	0.62	-0.79	0.22	-0.33	0.88	0.82	0.88	-0.82
A14	-0.76	0.84	0.48	-0.8	0.12	-0.35	0.86	0.83	0.87	-0.84
A15	0.67	-0.86	-0.59	0.74	-0.27	0.29	-0.88	-0.87	-0.89	0.83
B1	0.84	-0.64	-0.49	0.91	0.1	0.26	-0.66	-0.63	-0.67	0.84
B2	0.37	-0.81	-0.54	0.47	-0.55	0.53	-0.84	-0.75	-0.84	0.48
B3	-0.12	0.73	0.24	-0.29	0.43	-0.18	0.75	0.83	0.75	-0.31
B4	0.27	-0.16	-0.07	0.47	0.12	-0.33	-0.13	-0.25	-0.14	0.29
B5	1	-0.51	-0.31	0.8	0.26	0.2	-0.52	-0.51	-0.54	0.83
B6	-0.51	1	0.58	-0.6	0.49	-0.25	0.98	0.85	0.98	-0.7
B7	-0.31	0.58	1	-0.45	0.6	-0.14	0.59	0.33	0.59	-0.53
B8	0.8	-0.6	-0.45	1	0.1	0.32	-0.62	-0.63	-0.63	0.77
B9	0.26	0.49	0.6	0.1	1	-0.02	0.5	0.29	0.48	-0.05
B10	0.2	-0.25	-0.14	0.32	-0.02	1	-0.28	-0.26	-0.29	0.08
B11	-0.52	0.98	0.59	-0.62	0.5	-0.28	1	0.87	1	-0.7
B12	-0.51	0.85	0.33	-0.63	0.29	-0.26	0.87	1	0.88	-0.66
B13	-0.54	0.98	0.59	-0.63	0.48	-0.29	1	0.88	1	-0.71
B14	0.83	-0.7	-0.53	0.77	-0.05	0.08	-0.7	-0.66	-0.71	1

4 PRIEDAS. PROGRAMŲ TEKSTAI

Programos tekstas „nekoreliuotiems.sci“

```
//answ=messagebox("Ar su vienu katalogu dirbsite?", "Klausimas", "question",
["Taip" "Ne"], "modal");
answ = 1
//pasirenkame darbinę direktoriją, su failais "x.txt" ir "y.txt" bei "10proc"

if (answ == 1) then
    printf('\n Pasirinkite direktorija su: \n matrica X (stulp. sk. nurodo
parametrų sk.) [n x m]; \n Vektoriumi Y [n x t];')

directory = uigetdir('/home', "pasirink kataloga")
filename = directory + "/x.txt"
X = fscanfMat(filename);

filename = directory + "/y"+ string(1) +".txt"
Y = fscanfMat(filename);

// šie taikomi, jei norima X ir Y pasirinkti kitus
else
  printf('pasirinkite matricą X (stulp. sk. nurodo parametrų sk.) [n x m]')
 X =  fscanfMat(uigetfile());
  printf('pasirinkite vektorių stulpelį Y (eil. sk. nurodo duomenų sk.) (n x t)')
 Y = fscanfMat(uigetfile());

 directory = uigetdir()
end

// --- AA - modelio išmatuotos reikšmės,
// --- BB - turėtos reikšmės
// funkcijoms paduodamas vektorius-stulpelis
function [yse] = sserr(AA, BB)
  tt = size(AA);
  yse = zeros(tt(2))
 for i = 1:tt(2),
  ss = 0;
  for j = 1:tt(1) ,
    ss = ss + (AA(j, i) - BB(j, i))^2 ,
  end;
 yse(i)=ss;
 end;
```

```
endfunction;

// --- AA - Y teorinių arba praktinių reikšmių vektorius
// --- pilnoji kvadratų arba regresijos modelio suma
function [yse] = ssTR(AA, k)
  tt = size(AA);
 for i = 1:tt(2),
  ss = 0;
  for j = 1:tt(1) ,
    ss = ss + (AA(j, i) - k(i))^2 ,
  end;
 yse(i) = ss;
 end;
endfunction;

// -- skaičiuojamas vektoriaus AA vidurkis
function [ym] = ymean(AA)
 tt = size(AA);
 for i = 1:tt(2),
 yysum = 0;
  for j = 1:tt(1), yysum = yysum + AA(j, i);
  end;
 ym(i) = yysum/tt(1);
 end;
endfunction;

function [nai] = lyginti(failoVardas, rsqr, XX)
  pada = fscanfMat(failoVardas);
    param = size(XX);
  n = param(1); // duomenų skaičius
  m = param(2); // parametrų skaičius
  ff = (rsqr*(n-m) )/((1-rsqr)*(m-1));
  if pada(n, m) <= ff then nai = 1, else nai = 0,
    end;
endfunction;

// vektorius A nurodo iš n po k. t.y. iš viso ilgio, po tiek, kiek yra vienetų.
// A = [0 0 1 1] ir pan

// funkcija grąžina 0, jei eilutės nelygios ir 1, jei lygios
function [ga] = arlygios(A, t)
 ga = 1;
 d = size(A);
```

```
 for j = 1:1:d(2) ,
   if  ( A(t, j) <> A(t+1, j) ) then
     ga = 0
     break
   end;
 end;
endfunction;

// šaliname pasikartojančias eilutes
function [B] = mazinti(A)
 dyd=size(A)
 kk = 1;
  for i = 1:1:(dyd(1)-1),
   if ( arlygios(A, i) == 1 ) then
   continue
   end;
   for j = 1:1:dyd(2),
    B(kk, j) = A(i, j)
   end;
    kk = kk + 1 // pastumiame B masyvo eilutę
  end;
//if ( arlygios(A, dyd(1)-1) == 0 ) then
  for j = 1:dyd(2), B(kk, j) = A(dyd(1), j)
  end;
//end;
endfunction;
// gautoje B matricoje yra tiek gauti kėlinių vektoriai, eilutėmis

// k - prieš tai buvusio stulpelio ilgis
// t = { 0 ; 1 } prieš tai buvusio stulpelio skaičiai
// d - braizomo stulpelio vardiklis
// d1 - koeficiento skaitiklis braizomame stulpelyje

function A = comb( k, t, d, d1)
  if d1 < 0 | d < 1 then
     A = [];
     return
  end;
 if t == 1 then
   A =  [ [ ones((k*d1)/d, 1), comb((k*d1)/d, 1, d-1, d1-1)];[zeros((k*(d-d1))/d,
1), comb((k*(d-d1))/d, 0, d-1, d-d1-1) ] ]
  end; // if t 1 pabaiga
 if t == 0 then
```

```
   A =  [ [ ones((k*(d-d1))/d, 1),  comb( (k*(d-d1))/d , 1, d-1, d-d1-1)];
[ zeros((k*d1)/d, 1), comb((d1*k)/d, 0, d-1, d1-1) ]  ]
 end; // if t 0 pabaiga
endfunction;

// grąžina koeficientų, su kuriais modelio Rsq didžiausias, numerius
// KfNrMas - masyvas, nurodantis, kurie koeficientai buvo parinkti didžiausio Rsq modeliui
// Rsqur - grąžinamas gauto modelio maksimalus Rsq dydis
// KfSk    - funkcijai pateikiamas dydis, nurodantis, kiek koeficientų įtraukti į modelį
// Tai yra, gaunamame rezultate pateiktus KfSk koeficientus,
// su kuriais Rsq didžiausias KfSk koeficientų modelyje
// XXX - X nepriklausomų kintamųjų(faktorių) matrica (n x m)
// YYY - Y priklausomo kintamojo matrica (n x 1)
// funkcija ymean(Y) grąžina vektoriaus stulpelio "Y" vidurkį
// ssTr(Y, ymid) - pilnoji kvadratų arba regresijos modelio suma, ymid - vidurkis, Y - vektorius
function [KfNrMas, Rsqur, ZbB, RsqrAdj, SE1, SR1, aaa, liekanE, SEk, a] = KfRink(KfSk, XXX, YYY)
Ysize = size(YYY); //[m x t]
Xsize = size(XXX)  // [m x n]
Rsqur = zeros(Ysize(2), 1);  //[t x 1]
RsqrAdj = zeros(Ysize(2), 1) // [t x 1]
SE1 = zeros(Ysize(2), 1)
SR1 = zeros(Ysize(2), 1)
ZbB = zeros(Xsize(2), Ysize(2));   // [n x t]
KfNrMas = zeros(Xsize(2), Ysize(2))
zz = factorial(Xsize(2))/(factorial(Xsize(2)-KfSk)*factorial(KfSk))
TTT = comb( zz, 1, Xsize(2), KfSk)

Tsize = size(TTT)
aaa = []
a = 1;
SEk = []
for tt = 1:1:Tsize(1), // ciklas per visus galimus TTT variantus
Zx = zeros(Xsize(1), KfSk)
jj = 0;
 for kkk = 1:1:Xsize(2), // ciklas per X matricos stulpelius ir TTT matricos eilutes
   KfNrMasD(kkk) = TTT(tt, kkk)
   if( TTT(tt, kkk) == 1 ) then // jei tt-ame vektoriuje yra  kkk-tas vienetas
    jj = jj + 1
```

```
    for i = 1:1:Xsize(1), // ciklas per matricos X eilutes
      Zx(i, jj) = XXX(i, kkk);  // [m x p]; p = KfSk
    end; // for i pabaiga
  end;// if pabaiga
 end; // for kkk pabaiga
ZZb = inv(Zx'*Zx)*Zx'*YYY; // [p x t] = [[p x m][m x p]] [p x m][m x t]
YYYstg = Zx * ZZb;      // [m x p][p x t]  ZZb seniau buvo 1 x t
[aaaa, SEkk] = Multikolinearus(Zx)
SEk = [SEk, SEkk]
aaa = [aaa, aaaa]
//////////////////////////////////////////////////////////////////////
/////-- sia vieta isskiriu ja pritaikydamas papildomiems aktyvams --//
/////-- tai yra - didesnio (t) matavimo Y ir B masyvui -------------//
//////////////////////////////////////////////////////////////////////
  YYYaverage = ymean(YYY);
 liekanE = zeros(YYY)
 liekanE = liekanos(YYY, YYYstg)
  SSST = ssTR(YYY, YYYaverage);  // YYYstg
  SSSE = sserr(YYY, YYYstg);
// aukščiau esantys dydžiai ir taip yra [m x t] formato
// dėl to jiems žemiau esantis 'for it' ciklas netaikomas
 for it = 1:Ysize(2),
  // šis ciklas pereina per t aktyvų
  RRRsq(it) = 1 - SSSE(it)/SSST(it),// else RRRsq(it) = 1
  SSSR(it) = SSST(it) - SSSE(it)
  if ( RRRsq(it) > Rsqur(it) ) & ( nedidesni(aaa, tt, 4) == 1) then
   Rsqur(it) = RRRsq(it);
   RsqrAdj(it) = 1 - ((Ysize(1)-1)/(Ysize(1)-KfSk-1))*(1-Rsqur(it))
   SE1(it) = SSSE(it)
   SR1(it) = SSSR(it)
   purp = 1;
   a = tt;
    for iii = 1:Xsize(2), // nuo 1 iki n
      KfNrMas(iii, it) = KfNrMasD(iii),
      if( KfNrMas(iii, it) <> 0 ) then
        ZbB(iii, it) = ZZb(purp, it);
        purp = purp+1;
        else ZbB(iii, it) = 0,
      end; // if KfNrMas pabaiga
    end;  // for iii pabaiga
end; // if RRRsq pabaiga
clc()
end; // for it pabaiga
```

```
////////////////////////////////////////////////////////////////////////
////////////////////////////////////////////////////////////////////////
end;  // tt pabaiga
endfunction;

function [liekanE] = liekanos(Y, Ystg)
  sizeY = size(Y);
  for i = 1:sizeY(1)
    liekanE(i) = Y(i)-Ystg(i);
  end;
endfunction;
// tikrinama, ar matricos X, k-tojo stulpelio skaičiai ne didesni už a
// jei nedidesni - tuomet b = 1, jei didesni, tuomet 0

function b = nedidesni(X, k, a)
  b = 1;
  Xsize = size(X)
  for i = 1:Xsize(1),
     if X(i, k) > a then, b = 0, return;
     end;
  end;
endfunction



// X - nepriklausomu duomenu mmatrica (stulpeliais),
// k - priklausomos reiksmes stulpelis
function [rsqk, rsqkAd, SSEk] = FindRsqRsqadj(X, k)

 Xsize = size(X)
 Xk = zeros(Xsize(1))
 Xnew = zeros(Xsize(1), Xsize(2)-1)

// kopijuojame visus stulpelius iki k-tojo
 for t = 1:(k-1)
  for c = 1:Xsize(1)
    Xnew(c, t) = X(c, t)
  end; // for c pabaiga
 end; // for t pabaiga

// kopijuojame priklausomo kintamojo stulpeli
  for c = 1:Xsize(1)
     Xk(c) = X(c, k)
  end;
```

```
// kopijuojame visus stulpelius uz k-tojo
 for t = (k+1):Xsize(2)
  for c = 1:Xsize(1)
    Xnew(c, t-1) = X(c, t)
  end; // for c pabaiga
 end; // for t pabaiga

  if YraVienet(x) == 0 then // jei matrica neturi konstantos stulpelio
    Xnew = [ones(Xsize(1), 1), Xnew]; // del modelio su konstanta
  end;

//if abs(det(Xnew'*Xnew)) < 0.01 then
//  Xnew = zeros(Xnew)
//  rsqk = 1;
//  rsqkAd = 1;
//  SSEk = 0;
//  return;
//end;

 Bstog = inv(Xnew'*Xnew)*Xnew'*Xk
 Xstog = Xnew*Bstog
 //SSEk = zeros(Xk)
SSEk = sserr(Xk, Xstog);
SSTk = ssTR(Xk, ymean(Xk))

if abs(SSTk) < 0.000001 then
 SSTk = SSEk
end;

rsqk = 1 - SSEk/SSTk;
rsqkAd = 1 - ( 1 - rsqk )*(Xsize(1)-1)/(Xsize(1) - (Xsize(2) - 1) - 1)
endfunction;

function taip = YraVienet(X)
 taip = 0;
 ssx = size(X);
 for i = 1:ssx(2)
  for j = 2:ssx(1)
   if abs(X(j, i) - X(j-1, i)) > 0.00001 then
     break;
   end;
   if j == ssx(1) then
```

```
    taip = 1;
    return;
   end;
  end;
 end; // for i pabaiga
endfunction;

// jei VIF > 4 - galima itarti multikolinearuma
function [h, SEk] = Multikolinearus(x)

 Xsize = size(x);
 h = zeros(Xsize(2))
  rsq = 0.0;
  rsqadj = 0.0;
 for i = 1:Xsize(2)

  // skaičiuojame VIF - dispersijos mazejimo daugikli (variance inflation factor)
  [rsq, rsqadj, g] = FindRsqRsqadj(x, i)
  SEk(i) = g;
  h(i) = 1/(1-rsqadj);
 end; // for i pabaiga
endfunction;

filena = directory + "/issirink.txt"; // vartotojo pasirinkimui
failov = directory + "/matrica.txt"; // matricos nuskaitymui
fileBnm = directory + "/Bmatrix.txt" // b koeficientų matricos
fileSS = directory + "/SeSrR2adj.txt" // SSE SSR ir pataisyti determinacijos koef
failas
fileSe = directory + "/liekanos.txt" // liekanų masyvas

function [bTest, SEk] = vykdyti()
XXsize = size(X)
YYsize = size(Y)
  Zb = zeros(XXsize(2), YYsize(2))
// Knm - Koeficientų numerių masyvas
Knm = zeros(XXsize(2), XXsize(2))
 fd = mopen(filena, 'w');
 fdd = mopen(failov, 'w');
 fb = mopen(fileBnm, 'w');
 fssr = mopen(fileSS, 'w');
 bTest = zeros(X);
 mfprintf(fd, 'Nr.  Rsq   |  b1,   b2,   b3,   b4,   b5,   b6,   b7 \n    RsqAdj \
n');
```

```
 liekan = [];
 baigti = 0;
for KiekKoef = 1:(XXsize(2)), [Knm, RRssqq, Zb, RsqAdj, Se, Sr, bTest, SeE, SEk,
bb] = KfRink(KiekKoef, X, Y),
// rašome į failą
 liekan = [liekan, SeE];

  mfprintf(fd, '%25i', KiekKoef);
   mfprintf(fd, '\n');

 for jj = 1:YYsize(2)
  mfprintf(fd, '%i', KiekKoef)
     mfprintf(fd, 'rsq/VIF | ');
  mfprintf(fd, '%8.5f', RRssqq(jj));
    // rašome SSE SSR ir R2adj reikšmes
    mfprintf(fssr, '%15.12f ', RsqAdj(jj))
    mfprintf(fssr, '%15.12f ', Se(jj))
    mfprintf(fssr, '%15.12f ', Sr(jj))
    mfprintf(fssr, '\n')
   mfprintf(fd, ' ');

   thm = 0;
   for i = 1:XXsize(2),
    if Knm(i, jj) > 0.5 then
      thm = thm + 1;
       if bb <> 0 then, mfprintf(fd, '%5.3f ', bTest(thm, bb));
        else mfprintf(fd, ' bb=0'); end; // if bb pabaiga
     else  mfprintf(fd, '  0   ');
  end; // if Knm pabaiga
   end;
    mfprintf(fd, '\n SEE(k) ');
    a = size(SEk);

  if bb <> 0 then
   for i = 1:a(1);//KiekKoef,//(XXsize(2)-1),
    mfprintf(fd, '%5.3f\t ', SEk(i, bb));
   end; // for i pabaiga
  end; // if bb  pabaiga
    mfprintf(fd, '\n rsqadj/ArParinktas');

        mfprintf(fd, '  %8.5f', RsqAdj(jj));
   // spausdiname spausdinti b masyvus
 //  end; // for ti pabaiga
```

```
   // b koeficientų spausdinimo pabaiga
  for i = 1:XXsize(2),

      mfprintf(fb, '%18.15f ', Zb(i, jj)),

    mfprintf(fd, '%5.2f ', Knm(i, jj));
    mfprintf(fdd, '%18.15f ', Knm(i, jj));
    mfprintf(fdd, ' ');
  end;  // for i pabaiga
   mfprintf(fd, '\n');
   mfprintf(fdd, '\n');
   mfprintf(fb, '\n')

 end;  // for jj pabaiga
   mfprintf(fd, '\n');

if abs(RRssqq) < 0.0001 then
  baigti = baigti +1;
end;

if baigti > 1 then
   break;
end;

end;  // for KiekKoef pabaiga
fprintfMat(fileSe,liekan,'%5.2f');

  mclose(fd);
  mclose(fdd);
  mclose(fb);
  mclose(fssr);

endfunction;
tic()
[gagaa, sek] = vykdyti();
toc()
//M = fscanfMat(failov)
fscanfMat(filena)
function [a] = kuris()
  a = input("bu?");
endfunction;

//gdt = kuris()
```

```
//kuris = int(read(%io(1),1,1, '(a)'));

YYsize = size(Y)
bubu = string(1);
for i=1:(YYsize(2)-1),
  bubu = strcat([bubu; "; "; string(1)]),
end;
Matrixx = fscanfMat(fileBnm);
messagebox("Programos darbas baigtas", "Informacija apie darbo pabaigą", "info",
["Gerai"]);
```

Programos tekstas „koreliuotiems.sci“

```
answ = 1;
if (answ == 1) then
    printf('\n Pasirinkite direktorija su: \n matrica X (stulp. sk. nurodo
parametrų sk.) [n x m]; \n Vektoriumi Y [n x t];\n Bei statistikos reikšmių
failais (1proc 5proc ir/ar 10proc) ')

directory = uigetdir('/home/', "pasirink kataloga")
filename = directory + "/x.txt"
X = fscanfMat(filename);

filename = directory + "/y"+ string(1) +".txt"
Y = fscanfMat(filename);

// šie taikomi, jei norima X ir Y pasirinkti kitus
else
  printf('pasirinkite matricą X (stulp. sk. nurodo parametrų sk.) [n x m]')
 X =  fscanfMat(uigetfile());
  printf('pasirinkite vektorių stulpelį Y (eil. sk. nurodo duomenų sk.) (n x t)')
 Y = fscanfMat(uigetfile());

 directory = uigetdir()
 end

// --- AA - modelio išmatuotos reikšmės,
// --- BB - turėtos reikšmės
// funkcijoms paduodamas vektorius-stulpelis
function [yse] = sserr(AA, BB)
  tt = size(AA);
  yse = zeros(tt(2))
```

```
 for i = 1:tt(2),
  ss = 0;
  for j = 1:tt(1) ,
    ss = ss + (AA(j, i) - BB(j, i))^2 ,
  end;
 yse(i)=ss;
 end;
endfunction;

// --- AA - Y teorinių arba praktinių reikšmių vektorius
// --- pilnoji kvadratų arba regresijos modelio suma
function [yse] = ssTR(AA, k)
  tt = size(AA);
 for i = 1:tt(2),
  ss = 0;
  for j = 1:tt(1) ,
    ss = ss + (AA(j, i) - k(i))^2 ,
  end;
 yse(i) = ss;
 end;
endfunction;

// -- skaičiuojamas vektoriaus AA vidurkis

function [ym] = ymean(AA)
 tt = size(AA);
 for i = 1:tt(2),
 yysum = 0;
  for j = 1:tt(1), yysum = yysum + AA(j, i);
  end;
 ym(i) = yysum/tt(1);
 end;
endfunction;

function [nai] = lyginti(failoVardas, rsqr, XX)
  pada = fscanfMat(failoVardas);
    param = size(XX);
  n = param(1); // duomenų skaičius
  m = param(2); // parametrų skaičius
  ff = (rsqr*(n-m) )/((1-rsqr)*(m-1));
  if pada(n, m) <= ff then nai = 1, else nai = 0,
    end;
endfunction;
```

```
// vektorius A nurodo iš n po k. t.y. iš viso ilgio, po tiek, kiek yra vienetų.
//A = [0 0 1 1]

// funkcija grąžina 0, jei eilutės nelygios ir 1, jei lygios
function [ga] = arlygios(A, t)
 ga = 1;
 d = size(A);
 for j = 1:1:d(2) ,
   if  ( A(t, j) <> A(t+1, j) ) then
     ga = 0
     break
   end;
 end;
endfunction;



// šaliname pasikartojančias eilutes
function [B] = mazinti(A)
 dyd=size(A)
 kk = 1;
  for i = 1:1:(dyd(1)-1),
   if ( arlygios(A, i) == 1 ) then
   continue
   end;
   for j = 1:1:dyd(2),
    B(kk, j) = A(i, j)
   end;
    kk = kk + 1 // pastumiame B masyvo eilutę
  end;
//if ( arlygios(A, dyd(1)-1) == 0 ) then
  for j = 1:dyd(2), B(kk, j) = A(dyd(1), j)
  end;
//end;
endfunction;
// gautoje B matricoje yra tiek gauti kėlinių vektoriai, eilutėmis
// kk - kiek faktorių yra tarpusavyje gerai koreliuoti
function A = combi(k, t, d, d1, kk)

 A = [];
 kz = factorial(d)/(factorial(d-d1)*factorial(d1))
B = comb(kz, t, d, d1)
 for i = 1:kk,
```

```
 C = []
    for j = 1:(i-1),
      C = [C, zeros(kz, 1)]
    end;
    C = [C, ones(kz, 1)]
    for j = (i+1):kk
      C = [C, zeros(kz, 1)]
    end;
  A = [A;  C, B]
  end;
 //C = zeros(kz, kk)
// A = [A; C, B]
endfunction;

// k - prieš tai buvusio stulpelio ilgis
// t = { 0 ; 1 } prieš tai buvusio stulpelio skaičiai
// d - braizomo stulpelio vardiklis
// d1 - koeficiento skaitiklis braizomame stulpelyje

function A = comb( k, t, d, d1)
  if d1 < 0 | d < 1 then
     A = [];
     return
  end;
 if t == 1 then
   A =  [ [ ones((k*d1)/d, 1), comb((k*d1)/d, 1, d-1, d1-1)];[zeros((k*(d-d1))/d,
1), comb((k*(d-d1))/d, 0, d-1, d-d1-1) ] ]
  end; // if t 1 pabaiga
 if t == 0 then
   A =  [ [ ones((k*(d-d1))/d, 1),  comb( (k*(d-d1))/d , 1, d-1, d-d1-1)];
[ zeros((k*d1)/d, 1), comb((d1*k)/d, 0, d-1, d1-1) ]  ]
 end; // if t 0 pabaiga
endfunction;



// grąžina koeficientų, su kuriais modelio Rsq didžiausias, numerius
// KfNrMas - masyvas, nurodantis, kurie koeficientai buvo parinkti didžiausio Rsq
modeliui
// Rsqur - grąžinamas gauto modelio maksimalus Rsq dydis
// KfSk    - funkcijai pateikiamas dydis, nurodantis, kiek koeficientų įtraukti į
modelį
// Tai yra, gaunamame rezultate pateiktus KfSk koeficientus,
// su kuriais Rsq didžiausias KfSk koeficientų modelyje
```

```
// XXX - X nepriklausomų kintamųjų(faktorių) matrica (n x m)
// YYY - Y priklausomo kintamojo matrica (n x 1)
// BBB - B faktorių koeficientų matrica (m x 1)
// funkcija ymean(Y) grąžina vektoriaus stulpelio "Y" vidurkį
// ssTr(Y, ymid) - pilnoji kvadratų arba regresijos modelio suma, ymid - vidurkis,
Y - vektorius
// gh - parodo, kiek pirmų stulpelių stipriai koreliuoti tarpusavyje
function [KfNrMas, Rsqur, ZbB, RsqrAdj, SE1, SR1, aaa, liekanE, SEk, a] =
KfRink(KfSk, XXX, YYY, gh)
Ysize = size(YYY); //[m x t]
Xsize = size(XXX)  // [m x n]
Rsqur = zeros(Ysize(2), 1);  //[t x 1]
RsqrAdj = zeros(Ysize(2), 1) // [t x 1]
SE1 = zeros(Ysize(2), 1)
SR1 = zeros(Ysize(2), 1)
ZbB = zeros(Xsize(2), Ysize(2));   // [n x t]

KfNrMas = zeros(Xsize(2), Ysize(2))

zz = factorial(Xsize(2))/(factorial(Xsize(2)-KfSk)*factorial(KfSk))
//TTT = comb( zz, 1, Xsize(2), KfSk)
TTT = combi( zz, 1, Xsize(2)-gh, KfSk, gh)

Tsize = size(TTT)
aaa = []
a = 1;
SEk = []
for tt = 1:1:Tsize(1), // ciklas per visus galimus TTT variantus
Zx = zeros(Xsize(1), KfSk)
jj = 0;
 for kkk = 1:1:Xsize(2), // ciklas per X matricos stulpelius ir TTT matricos
eilutes
   KfNrMasD(kkk) = TTT(tt, kkk)
   if( TTT(tt, kkk) == 1 ) then // jei tt-ame vektoriuje yra  kkk-tas vienetas
    jj = jj + 1
    for i = 1:1:Xsize(1), // ciklas per matricos X eilutes
      Zx(i, jj) = XXX(i, kkk);  // [m x p]; p = KfSk
    end; // for i pabaiga
  end;// if pabaiga
 end; // for kkk pabaiga

ZZb = inv(Zx'*Zx)*Zx'*YYY; // [p x t] = [[p x m][m x p]] [p x m][m x t]
```

```
YYYstg = Zx * ZZb;      // [m x p][p x t]  ZZb seniau buvo 1 x t

[aaaa, SEkk] = Multikolinearus(Zx)
SEk = [SEk, SEkk]
aaa = [aaa, aaaa]

//////////////////////////////////////////////////////////////////////
/////-- sia vieta isskiriu ja pritaikydamas papildomiems aktyvams --//
/////-- tai yra - didesnio (t) matavimo Y ir B masyvui -------------//
//////////////////////////////////////////////////////////////////////
  YYYaverage = ymean(YYY);
 liekanE = zeros(YYY)
 liekanE = liekanos(YYY, YYYstg)
  SSST = ssTR(YYY, YYYaverage);  // YYYstg
  SSSE = sserr(YYY, YYYstg);
// aukščiau esantys dydžiai ir taip yra [m x t] formato
// dėl to jiems žemiau esantis 'for it' ciklas netaikomas
 for it = 1:Ysize(2),
  // šis ciklas pereina per t aktyvų
  RRRsq(it) = 1 - SSSE(it)/SSST(it),// else RRRsq(it) = 1
  SSSR(it) = SSST(it) - SSSE(it)
  if ( RRRsq(it) > Rsqur(it) ) & ( nedidesni(aaa, tt, 4) == 1) then
   Rsqur(it) = RRRsq(it);
   RsqrAdj(it) = 1 - ((Ysize(1)-1)/(Ysize(1)-(KfSk+1)-1))*(1-Rsqur(it))
   SE1(it) = SSSE(it)
   SR1(it) = SSSR(it)
   purp = 1;
   a = tt;
    for iii = 1:Xsize(2), // nuo 1 iki n
      KfNrMas(iii, it) = KfNrMasD(iii),
      if( KfNrMas(iii, it) <> 0 ) then
        ZbB(iii, it) = ZZb(purp, it);
        purp = purp+1;
        else ZbB(iii, it) = 0,
      end; // if KfNrMas pabaiga
    end;  // for iii pabaiga
  end; // if RRRsq pabaiga
 //end; // if aaa pabaiga
end; // for it pabaiga
//////////////////////////////////////////////////////////////////////
//////////////////////////////////////////////////////////////////////
end;  // tt pabaiga
endfunction;
```

```
function [liekanE] = liekanos(Y, Ystg)
  sizeY = size(Y);
  for i = 1:sizeY(1)
    liekanE(i) = Y(i)-Ystg(i);
  end;

endfunction;

// tikrinama, ar matricos X, k-tojo stulpelio skaičiai ne didesni už a
// jei nedidesni - tuomet b = 1, jei didesni, tuomet 0

function b = nedidesni(X, k, a)
  b = 1;
  Xsize = size(X)
  for i = 1:Xsize(1),
     if X(i, k) > a then, b = 0, return;
     end;
  end;
endfunction



// X - nepriklausomu duomenu mmatrica (stulpeliais),
// k - priklausomos reiksmes stulpelis
function [rsqk, rsqkAd, SSEk] = FindRsqRsqadj(X, k)

 Xsize = size(X)
 Xk = zeros(Xsize(1))
 Xnew = zeros(Xsize(1), Xsize(2)-1)

// kopijuojame visus stulpelius iki k-tojo
 for t = 1:(k-1)
  for c = 1:Xsize(1)
    Xnew(c, t) = X(c, t)
  end; // for c pabaiga
 end; // for t pabaiga

// kopijuojame priklausomo kintamojo stulpeli
  for c = 1:Xsize(1)
     Xk(c) = X(c, k)
  end;

// kopijuojame visus stulpelius uz k-tojo
```

```
 for t = (k+1):Xsize(2)
  for c = 1:Xsize(1)
    Xnew(c, t-1) = X(c, t)
  end; // for c pabaiga
 end; // for t pabaiga

  if YraVienet(x) == 0 then // jei matrica neturi konstantos stulpelio
    Xnew = [ones(Xsize(1), 1), Xnew]; // kodas: peleda (: jei rimtai, tai del
modelio su konstanta
  end;

 Bstog = inv(Xnew'*Xnew)*Xnew'*Xk
 Xstog = Xnew*Bstog
 //SSEk = zeros(Xk)
SSEk = sserr(Xk, Xstog);
SSTk = ssTR(Xk, ymean(Xk))

if abs(SSTk) < 0.000001 then
 SSTk = SSEk
end;

rsqk = 1 - SSEk/SSTk;
rsqkAd = 1 - ( 1 - rsqk )*(Xsize(1)-1)/(Xsize(1) - (Xsize(2) - 1) - 1)
endfunction;

function taip = YraVienet(X)
 taip = 0;
 ssx = size(X);
 for i = 1:ssx(2)
  for j = 2:ssx(1)
   if abs(X(j, i) - X(j-1, i)) > 0.00001 then
     break;
   end;
   if j == ssx(1) then
    taip = 1;
    return;
   end;
  end;
 end; // for i pabaiga
endfunction;

// jei VIF > 4- galima itarti multikolinearuma
function [h, SEk] = Multikolinearus(x)
```

```
 Xsize = size(x);
 h = zeros(Xsize(2))
  rsq = 0.0;
  rsqadj = 0.0;
 for i = 1:Xsize(2)
  // skaičiuojame VIF - dispersijos mazejimo daugikli (variance inflation factor)
  [rsq, rsqadj, g] = FindRsqRsqadj(x, i)
  SEk(i) = g;
  h(i) = 1/(1-rsqadj);
 end; // for i pabaiga
endfunction;



filena = directory + "/issirink.txt"; // vartotojo pasirinkimui
failov = directory + "/matrica.txt"; // matricos nuskaitymui
fileBnm = directory + "/Bmatrix.txt" // b koeficientų matricos
fileSS = directory + "/SeSrR2adj.txt" // SSE SSR ir pataisyti determinacijos koef
failas
fileSe = directory + "/liekanos.txt" // liekanų masyvas

bubu = string(1)
[ok, ghh]=getvalue("Faktoriu skaičiaus pasirinkimas",["kiek pirmų faktorių yra
stipriai koreliuoti tarpusavyje?"],list("vec",1),[bubu])



function [bTest, SEk] = vykdyti()
XXsize = size(X)
YYsize = size(Y)
  Zb = zeros(XXsize(2), YYsize(2))
// Knm - Koeficientų numerių masyvas
Knm = zeros(XXsize(2), XXsize(2))
 fd = mopen(filena, 'w');
 fdd = mopen(failov, 'w');
 fb = mopen(fileBnm, 'w');
 fssr = mopen(fileSS, 'w');
// fse = mopen(fileSe, 'w')
 bTest = zeros(X);
baigti = 0;
 mfprintf(fd, 'Nr.  Rsq   |  b1,   b2,   b3,   b4,   b5,   b6,   b7 \n    RsqAdj \
n');
for KiekKoef = 1:(XXsize(2)-ghh), [Knm, RRssqq, Zb, RsqAdj, Se, Sr, bTest, SeE,
SEk, bb] = KfRink(KiekKoef, X, Y, ghh),
// rašome į failą
```

```
clc();
fprintfMat(fileSe,SeE,'%5.2f');

  mfprintf(fd, '%25i', KiekKoef);
   mfprintf(fd, '\n');

 for jj = 1:YYsize(2)
  mfprintf(fd, '%i', KiekKoef)
     mfprintf(fd, 'rsq/VIF | ');
  mfprintf(fd, '%8.5f', RRssqq(jj));
    // rašome SSE SSR ir R2adj reikšmes
    mfprintf(fssr, '%15.12f ', RsqAdj(jj))
    mfprintf(fssr, '%15.12f ', Se(jj))
    mfprintf(fssr, '%15.12f ', Sr(jj))
    mfprintf(fssr, '\n')
   mfprintf(fd, ' ');

   thm = 0;
   for i = 1:XXsize(2),
    if Knm(i, jj) > 0.5 then
      thm = thm + 1;
       if bb <> 0 then, mfprintf(fd, '%5.3f ', bTest(thm, bb));
        else mfprintf(fd, ' bb=0'); end; // if bb pabaiga
     else  mfprintf(fd, '  0   ');
  end; // if Knm pabaiga
   end;
    mfprintf(fd, '\n SEE(k) ');
    a = size(SEk);

  if bb <> 0 then
   for i = 1:a(1);//KiekKoef,//(XXsize(2)-1),
    mfprintf(fd, '%5.3f\t', SEk(i, bb));
   end; // for i pabaiga
  end; // if bb  pabaiga
    mfprintf(fd, '\n rsqadj/ArParinktas');


        mfprintf(fd, '  %8.5f', RsqAdj(jj));

  for i = 1:XXsize(2),

      mfprintf(fb, '%18.15f ', Zb(i, jj)),
```

```
    mfprintf(fd, '%5.2f ', Knm(i, jj));
    mfprintf(fdd, '%18.15f', Knm(i, jj));
    mfprintf(fdd, ' ');
  end;  // for i pabaiga
   mfprintf(fd, '\n');
   mfprintf(fdd, '\n');
   mfprintf(fb, '\n')

 end;  // for jj pabaiga
   mfprintf(fd, '\n');

if abs(RRssqq) < 0.0001 then
  baigti = baigti +1;
end;

if baigti > 1 then
   break;
end;

end;  // for KiekKoef pabaiga
  mclose(fd);
  mclose(fdd);
  mclose(fb);
  mclose(fssr);
// čia yra baigti skaičiuoti geriausi koeficientai.
endfunction;
tic()
[gagaa, sek] = vykdyti();
toc()
fscanfMat(filena)
function [a] = kuris()
  a = input("bu?");
endfunction;
YYsize = size(Y)
bubu = string(1);
for i=1:(YYsize(2)-1),
  bubu = strcat([bubu; "; "; string(1)]),
end;
Matrixx = fscanfMat(fileBnm);
messagebox("Programos darbas baigas.", "Informacija", "question", ["Taip"],
"modal");
```

5 PRIEDAS. ALGORITMŲ ILIUSTRACIJOS

35 pav. Algoritmo dalis