

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MATEMATINĖS STATISTIKOS KATEDRA

AGNĖ IKVILDAITĖ

MUITINĖS IMPORTO PROCEDŪRŲ TAIKYMO ANALIZĖ

MAGISTRO DIPLOMINIS DARBAS

DARBO VADOVAS:

DOC. VYTAUTAS KAZAKEVIČIUS

VILNIUS

2006

TURINYS

1.	Įvadas	3
2.	Apie duomenis	4
2.1.	Duomenų struktūra	4
2.2.	Duomenų sąryšiai	5
3.	Aproksimacija	9
3.1.	Tinkamiausios regresijos funkcijos suradimas	9
3.2.	Regresijos funkcijos analizė ištraukiant dalį duomenų	11
4.	Laiko eilučių indeksai	13
4.1.	Indeksai siejantys gretimus laiko eilutės narius	13
4.2.	Indeksai siejantys laiko eilutės narius nutolusius vienas nuo kito per 12 mėnesių 16	
4.3.	Prognozavimas pritaikant indeksus	18
5.	Apibendrinimas	20
5.1.	Prognozių palyginimas	20
5.2.	Išvados	21
6.	Literatūra	22
7.	Summary	23
8.	Priedai	24

1. ĮVADAS

Svarbus muitinės veiklos rodiklis – muitinės importo procedūrų skaičius. Šio darbo tikslas išanalizuoti muitinės importo procedūrų taikymą. Analizei naudojami kas mėnesį renkami duomenys apie Lietuvos muitinėje įformintų importo deklaracijų skaičių. Importo deklaracijos gali būti pateikiamos įforminti vienai iš šešių muitinės importo procedūrų: *išleidimas į laisvą apyvartą, laikinasis įvežimas perdirbti, laikinasis įvežimas, muitinis sandėliavimas, perdirbimas muitinei prižiūrint, reimportas*. Importo deklaracijos yra įforminamos penkiose teritorinėse muitinėse Vilniaus, Kauno, Klaipėdos, Šiaulių ir Panevėžio (toliau – VTM, KTM, LTM, STM ir PTM). Kiekvienoje teritorinėje muitinėje gali būti įforminama bet kuri iš anksčiau išvardintų importo procedūrų.

Nagrinėjami duomenys nuo 2004 m. birželio mėn. iki 2006 kovo mėn. Ankstesni duomenys nenagrinėjami, nes yra nepalyginami su duomenimis renkamais nuo Lietuvos įstojimo į ES (2004 m. gegužės mėn. duomenys į analizę neįtraukiami dėl pereinamojo laikotarpio įtakos duomenų tikslumui ir palyginamumui).

Importo deklaracijų analizės tikslas nustatyti – įvairius sąryšius, kuriais remiantis būtų galima kuo tiksliau prognozuoti importo procedūrų skaičių iš turimų, kol kas negausių, statistinių duomenų ir į ką atsižvelgiant tas prognozes koreguoti priklausomai nuo stebėjimų, kurie bus gaunami laikui bėgant.

2. APIE DUOMENIS

2.1. DUOMENŲ STRUKTŪRA

Bendras Muitinės importo deklaracijų skaičius yra pagrindinis analizės objektas. Importo deklaracijų skaičius randamas remiantis formule:

$$D = \sum_{i=1}^5 \sum_{j=1}^6 D_{ij},$$

čia D – importo deklaracijų skaičius,

i – teritorinė muitinė: 1 – Vilniaus,

2 – Kauno,

3 – Klaipėdos,

4 – Šiaulių,

5 – Panevėžio;

j – procedūros tipas: 1 – išleidimas į laisvą apyvertą,

2 – laikinasis įvežimas perdirbti,

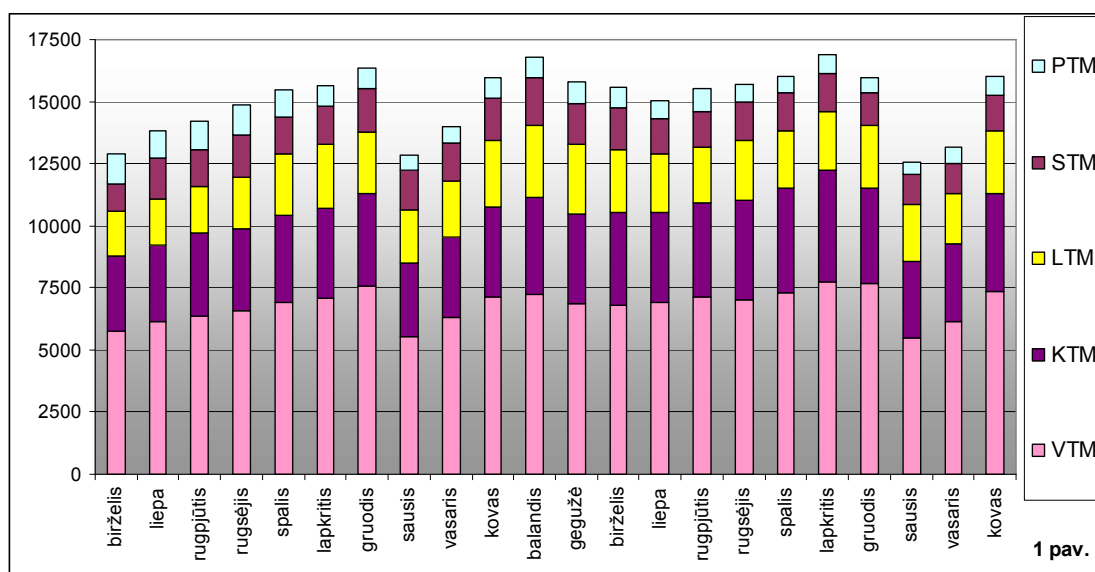
3 – laikinasis įvežimas,

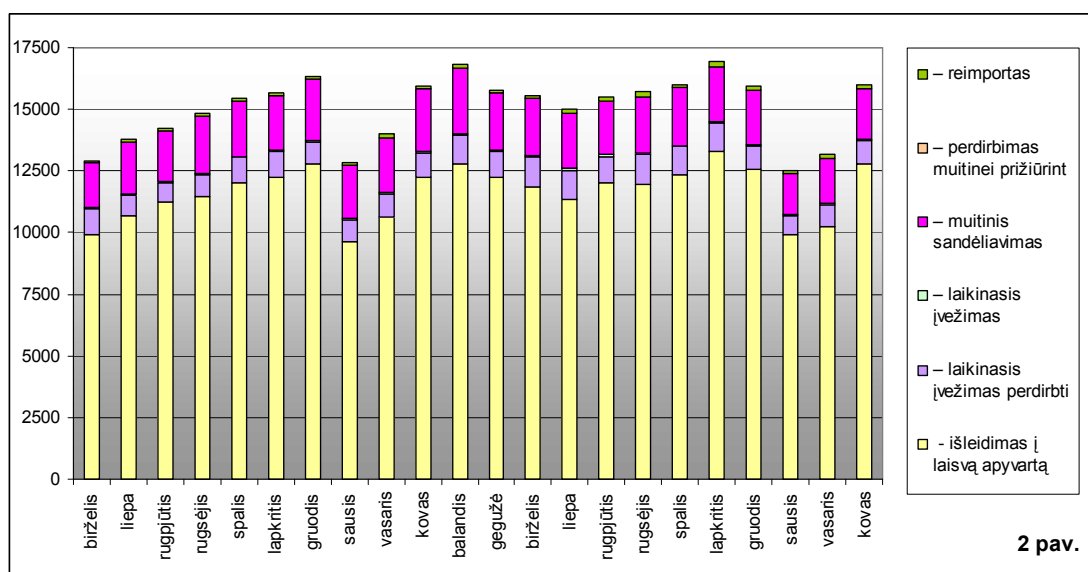
4 – muitinis sandėliavimas,

5 – perdirbimas muitinei prižiūrint,

6 – reimportas.

Žemiau pateikti du grafikai iliustruoja koks, nagrinėjamoju laikotarpiu, yra pasiskirstymas pagal teritorinę muitinę (*1 pav.*) ir pritaikytos procedūros tipą (*2 pav.*).





2 pav.

Iš 1 pav. matome, kad daugiausiai importo procedūrų yra įforminama VTM vidutiniškai apie 45 proc. visų procedūrų kiekvieną mėnesį, kitos teritorinės muitinės sudaro tokias procentines dalis: KTM – 24 proc., LTM – 15 proc., STM – 10 proc., PTM – 6 proc.

Iš 2 pav. matome, kad daugiausiai yra įforminama išleidimo į laisvą apyvartą procedūrų vidutiniškai tai sudaro 77 proc. visų procedūrų kiekvieną mėnesį, kitos procedūros pasiskirsto taip: muitinis sandėliavimas 15 proc., laikinas įvežimas perdirbti 7 proc., reimportas 1 proc., laikinas įvežimas perdirbti ir perdirbimas muitinei prižiūrint net kartu paėmus nesudaro vieno procento.

Kiekvieną mėnesį, išleidimo į laisvą apyvartą procedūros VTM sudaro vidutiniškai 36 proc. muitinėje per mėnesį įforminamų importo procedūrų.

2.2. DUOMENŲ SĄRYŠIAI

Norint nustatyti bendro įformintų deklaracijų skaičiaus priklausomybę nuo atskiroje teritorinėje muitinėje įformintų deklaracijų skaičiaus atlikta regresinė analizė. 1 lentelėje pateikiami determinacijos koeficientai, kurie parodo tiesinės priklausomybės stiprumą tarp bendro įformintų importo procedūrų skaičiaus ir kiekvienos teritorinės muitinės įformintų importo procedūrų skaičiaus atskirai.

1 lentelė

Teritorinė muitinė	Determinacijos koeficientas*
VTM	0,93
KTM	0,79
LTM	0,53
STM	0,32
PTM	0,00

* – determinacijos koeficientas apskaičiuojamas pagal formulę:

$$r^2 = 1 - \frac{SSE}{SST}, \text{ kur } SSE = \sum (Y_j - \hat{Y}_j)^2, \text{ SST} = (\sum Y_j^2) - \frac{(\sum Y_j)^2}{n}.$$

Kaip ir buvo galima tikėtis, paaiškėjo, kad stipriausia tiesinė priklausomybė yra nuo VTM duomenų. *Priede Nr.1* pateikiami grafikai vaizduojantys realų ir regresijos lygtimi aproksimuojant gautą įformintų procedūrų skaičių.

Iš šių rezultatų seka išvada, kad žinodami VTM importo procedūrų skaičių galime gana tiksliai įvertinti bendrą importo procedūrų skaičių.

Jeigu iš turimų duomenų eilutės atmetame 4-ių, 3-jų, 2-jų arba tik 1-o paskutinio mėnesio duomenis ir iš turimos trumpesnės duomenų eilutės įvertiname tiesinės regresijos lygties $y=a+bx+\varepsilon$ koeficientus (a – pastovioji regresijos dalis (angl. *intercept*), b – koeficientas prie priklausomojo kintamojo, ε – klaida daroma pakeičiant tikrą priklausomybės formą tiesine) gauname rezultatus pateiktus 2 lentelėje.

2 lentelė

Stebėjimų skaičius	Determinacijos koeficientas	Intercept \hat{a}	VTM \hat{b}	Regresijos lygties sudarymui nenaudotų stebėjimų absoliutinių nuokrypių nuo tikrosios reišmės vidurkis	Regresijos lygties sudarymui naudotų stebėjimų absoliutinių nuokrypių nuo tikrosios reišmės vidurkis	Visų absoliutinių nuokrypių nuo tikrosios reišmės vidurkis
22	0,927	1883,05	1,94	-	249	249
21	0,926	1800,21	1,96	414	249	247
20	0,931	2168,51	1,91	447	237	256
19	0,913	2348,09	1,88	358	244	260
18	0,944	1537,07	2,01	503	215	267

Įvertis \hat{a} didėja mažėjant įverčiui \hat{b} . Įverčio \hat{a} reikšmių išsibarstymas simetriškas taškui 1943, o įverčio \hat{b} reikšmių išsibarstymas simetriškas taškui 1,94. Kai stebėjimų skaičius 22, \hat{a} ir \hat{b} reikšmės artimiausios įverčių išsibarstymo simetrijos taškams.

Determinacijos koeficientai visais atvejais yra labai geri, taigi tiesinė priklausomybė parinkta pagrįstai. Nors tuomet, kai stebėjimų skaičius mažiausias (18), turime didžiausią determinacijos koeficientą, bet tuo atveju gautos lygties patikimumas prognozavimui nėra labai geras, nes pagal šią lygtį prognozuojant gautų reikšmių absoliutinių nuokrypių nuo tikrųjų reikšmių vidurkis yra didžiausias (503), nors didelę determinacijos koeficiento reikšmę paaiškina tai, jog regresijos lygties sudarymui naudotų stebėjimų absoliutinių nuokrypių nuo tikrosios reišmės vidurkis yra mažiausias (215). Mažiausias prognozuojamų reikšmių absoliutinių nuokrypių nuo tikrųjų reikšmių vidurkis (358) gaunamas kai stebėjimų skaičius 19, bet šia lygtimi naudojantis prognozės greičiausiai bus blogesnės, nei naudojantis lygtimis gautomis, kai įtraukti visi 22, 21 arba 20 turimų stebėjimų, nes šios lygtys geriau atitinka stebėtus duomenis. Vidutiniškai geriausiai visus duomenis, tiek prognozuojamus, tiek ir stebėtus, aprašo lygtis kurios sudarymui buvo naudojamas 21 stebėjimas, bet jos nuokrypių vidurkiai labai panašūs į gautus įtraukiant visus 22 stebėjimus. Todėl prisiminus, kad įverčiai \hat{a} ir \hat{b} yra artimi įverčių išsibarstymo simetrijos taškams, kai įtraukti visi 22 stebėjimai, galime daryti išvadą, kad prognozavimui reiktų naudotis lygtimi: $y(t) = 1883,05 + 1,94x(t)$, čia $y(t)$ – bendras importo deklaracijų skaičius momentu t , $x(t)$ – importo deklaracijų skaičius VTM momentu t .

2 priede esančiuose grafikuose pateikiamos stebėtos reikšmės ir kiekviena iš penkių f-jų pasinaudojus gautos aproksimacijos, bei prognozės (prognozuojamos reikšmės išskirtos ryškesniais taškais, atitinkamoje aproksimuojančioje kreivėje).

Pateikiame funkcijas, kuriomis remiantis nubraižyti grafikai

$$\hat{y}^{22}(t) = 1883,05 + 1,94x(t),$$

$$\hat{y}^{21}(t) = 1800,21 + 1,96x(t),$$

$$\hat{y}^{20}(t) = 2168,51 + 1,91x(t),$$

$$\hat{y}^{19}(t) = 2348,09 + 1,88x(t),$$

$$\hat{y}^{18}(t) = 1537,07 + 2,01x(t).$$

Padarėme išvadą, kad prognozavimui geriausiai tinka visus stebėjimus naudojant gauta f-ja. Tais pačiais kriterijais remiantis, tik šiek tiek blogesnė yra f-ja

gauta naudojant vienu mažiau stebėjimų. Taigi galime daryti prielaidą, kad į prognozavimo f-ją įtraukus dar daugiau narių, ji būtų dar tikslesnė.

Ateityje norint prognozuoti šiuo būdu, reikia pakoreguoti gautą f-ją atsižvelgiant ir į vėlesnius stebėjimus.

Šiame skyrelyje nustatyta sąryšį tarp bendro importo procedūrų skaičiaus ir Vilniaus teritorinėje muitinėje įformintų importo procedūrų skaičiaus, praktikoje galima pritaikyti, pvz. jeigu per trumpą laiką reikalingas bendras importo procedūrų skaičius, o duomenų iš visų teritorinių muitinių greitai neįmanoma gauti, užtektų gauti tik VTM duomenis tam, kad gan gerai įvertinti bendrą deklaracijų skaičių.

3. APROKSIMACIJA

3.1. TINKAMIAUSIOS REGRESIJOS FUNKCIJOS SURADIMAS

Nagrinėjami duomenys sudaro 22 stebėjimų laiko eilutę. Ši laiko eilutė per trumpa, kad naudojantis statistiniais paketais būtų galima joje išskirti tendą, sezoniškumą, cikliškumą ir atsitiktinumą (kai kuriems statistiniams paketams reikia pateikti penkių metų duomenis). Kadangi duomenys mėnesiniai, jie apima nepilnus dvejus metus. Remiantis grafiniu duomenų vaizdu, žinant duomenų prigimtį, bei intuityviai numanant nuo ko jie priklauso, galima daryti išvada, kad nagrinėjamiems duomenims turi įtakos sezonas (ketvirtis, mėnuo). Atsižvelgiant į šį faktą ir panaudojant regresinės analizės metodus bus ieškoma polinominės regresinės kreivės geriausiai atspindinčios duomenis.

Į kreivės lygtį įtrauksime trendo kintamąjį l (laiko momentas įgyjantis sveikąsias reikšmes atitinkančias mėnesio numerį, pradedant skaičiuoti nuo pirmojo į regresinę analizę įtraukto mėnesio) ir kintamuosius atspindinčius mėnesius – t_i , ($i=1, \dots, 12$, t_i įgyja reikšmę 1, kai nagrinėjamų duomenų mėnesio numeris pradedant skaičiuoti nuo pirmojo į regresinę analizę įtraukto mėnesio lygus $i+12n$, $n=0, 1, \dots$) bei ketvirčius – k_j , ($j=1, \dots, 4$, k_j įgyja reikšmę 1, kai nagrinėjamų duomenų ketvirčio numeris skaičiuojant nuo metų pradžios lygus j). Tokie kintamieji vadinami *dummy* (angl.) kintamaisiais. Regresinė kreivės lygtis užrašoma taip:

$$d = a + b \cdot l + \sum_{i=1}^{12} b_i \cdot t_i + \sum_{j=1}^4 c_j \cdot k_j, \quad (1)$$

čia a – pastovioji dalis (*intercept*), b – koeficientas prie laiko momento, b_i – koeficientai prie dummy kintamųjų t_i , atspindintys i -tojo mėnesio įtaką, c_j koeficientai prie dummy kintamųjų k_j , atspindintys j -tojo ketvirčio įtaką. Šiuos koeficientus įvertinsime iš turimų duomenų mažiausiųjų kvadratų metodu, t. y. minimizuodami skirtumų tarp realių ir aproksimuojamų reikšmių kvadratų sumą. Taip pat išnagrinėsime dar kelis kitus regresinės kreivės lygties variantus:

$$d = a + b \cdot l + \sum_{i=1}^{12} b_i \cdot t_i, \quad (2)$$

$$d = a + b \cdot l + \sum_{j=1}^4 c_j \cdot k_j. \quad (3)$$

(2) – tai lygtis į kurią, įtraukti dummy kintamieji atspindintys mėnesių įtaką,

(3) – tai lygtis į kurią, įtraukti dummy kintamieji atspindintys tik metų ketvirčio įtaką.

3 lentelėje pateikiami determinacijos ir koreliacijos koeficientai bei skirtumų tarp realių ir prognozuojamų reikšmių kvadratų sumos (Final value).

3 lentelė

Lygtis	Determinacijos koeficientas	Koreliacijos koeficientas	Final Value
(1)	0,8647	0,92989	4 939 273,80
(2)	0,8647	0,92989	4 939 273,80
(3)	0,4510	0,67156	20 041 892,67

Iš šios lentelės matome, kad trečiosios lygties visi koeficientai mažiausi, o kitų dviejų visi koeficientai vienodi.

Turimoje laiko eilutėje susumavę ketvirčių duomenis ir įvertinę koeficientus regresinės kreivės lygties, kurios pavidalas

$$\hat{d}_Q = a + b \cdot l + c_I \cdot k_I + c_{II} \cdot k_{II} + c_{III} \cdot k_{III} + c_{IV} \cdot k_{IV} \quad (4)$$

gauname tokius duomenis:

determinacijos koeficientas – 0,902818,

koreliacijos koeficientas – 0,950167,

skirtumų tarp realių ir prognozuojamų reikšmių kvadratų suma (Final value) – 5 057 167.

Matome, kad determinacijos ir koreliacijos koeficientai yra geresni nei anksčiau nagrinėtais atvejais, o skirtumų tarp realių ir prognozuojamų reikšmių kvadratų suma yra šiek tiek didesnė nei geriausia iš anksčiau nagrinėtų atvejų

3 priede pateikiami grafikai vaizduojantys realius duomenis, realių duomenų aproksimaciją gautosiomis regresijos kreivėmis bei prognozę artimiausiems 9 mėnesiams. Funkcijos, pagal kurias gauti rezultatai yra šios:

$$d^{(1)} = 10904 + 54 l - 165 t_1 + 482t_2 + 882t_3 + 1238t_4 + 866t_5 + 1351t_6 + 1173t_7 - 736t_8 + 90t_9 + 2428t_{10} + 2186t_{11} + 1110t_{12} + 1782k_1 + 3130k_2 + 2601k_3 + 3390k_4, \text{ čia}$$

$$l = 1, 2, \dots, \quad t_1 = \begin{cases} 1, & \text{jei birželis,} \\ 0, & \text{jei kitas mėnuo.} \end{cases}, \dots, \quad t_{12} = \begin{cases} 1, & \text{jei gegužė,} \\ 0, & \text{jei kitas mėnuo.} \end{cases}$$

$$k_1 = \begin{cases} 1, & \text{jei I ketvirtis,} \\ 0, & \text{jei kitas ketvirtis.} \end{cases}, \dots, \quad k_4 = \begin{cases} 1, & \text{jei IV ketvirtis,} \\ 0, & \text{jei kitas ketvirtis.} \end{cases}$$

$$d^{(2)} = 13420 + 54l + 449t_1 + 567t_2 + 967t_3 + 1323t_4 + 1740t_5 + 2225t_6 + 2047t_7 - 1470t_8 - 644t_9 + 1693t_{10} + 2799t_{11} + 1724t_{12}, \text{ čia } l = 1, 2, \dots,$$

$$t_l = \begin{cases} 1, & \text{jei birželis,} \\ 0, & \text{jei kitas mėnuo.} \end{cases}, \dots, t_{12} = \begin{cases} 1, & \text{jei kovas,} \\ 0, & \text{jei kitas mėnuo.} \end{cases}$$

$$d^{(3)} = 11377 + 75l + 1583k_1 + 3201k_2 + 2803k_3 + 3791k_4, \text{ čia } l = 1, 2, \dots$$

$$k_l = \begin{cases} 1, & \text{jei I ketvirtis,} \\ 0, & \text{jei kitas ketvirtis.} \end{cases}, \dots, k_4 = \begin{cases} 1, & \text{jei IV ketvirtis,} \\ 0, & \text{jei kitas ketvirtis.} \end{cases}$$

Pavyzdžiui, jei prognozuojame 2006 m. balandžio mėnesiui, turime tokius rezultatus:

$$d^{(1)} = 10904 + 54 \cdot 23 + 2186 \cdot 1 + 3130 \cdot 1 = 17462$$

$$d^{(2)} = 13420 + 54 \cdot 23 + 2799 \cdot 1 = 17461$$

$$d^{(3)} = 11377 + 75 \cdot 23 + 3201 \cdot 1 = 16303,$$

nes $t = 23$ (23 mėnuo, skaičiuojant nuo pirmo mėnesio įtraukto į analizę), $t_{11} = 1$, (nes balandis pradant skaičiuoti nuo birželio yra 11 mėn.), $k_2 = 1$, (nes balandis priklauso II metų ketvirčiui).

3 lentelėje pateikti duomenys ir grafikų informacija leidžia daryti išvadą, kad iš visų išnagrinėtųjų kreivių labiausiai tinkama mėnesinėms prognozėms (2) lygtimi aprašoma kreivė. Nors (1) lygtimi aprašoma kreivė duoda tokius pačius rezultatus kaip (2) lygtimi aprašoma kreivė, bet jai reikalingi 4 papildomi kintamieji.

Ketvirtiniams duomenims prognozuoti tinkama (4) lygtimi aprašoma kreivė. Naudojantis ja gauti rezultatai nedaug (apie 1 proc.) skiriasi nuo susumuotų pagal ketvirčius rezultatų gautų naudojantis (2) lygtimi.

3.2. REGRESIJOS FUNCIJOS ANALIZĖ ĮTRAUKIANT DALĮ DUOMENŲ

Remiantis laiko eilučių teorija visos prognozavimui skirtos funkcijos yra gaunamos išanalizavus ne viso žinomo laikotarpio duomenis, o kažkokią duomenų dalį, neįtraukiant paskutiniųjų stebėjimų. Dažniausiai funkcijos sudarymui naudojama laiko eilutė, kurios ilgis 12 mėnesių arba 4 ketvirčių stebėjimais trumpesnis (jei duomenys mėnesiniai atmetama 12 paskutinių stebėjimų, jei ketvirtiniai 4 paskutiniai stebėjimai). Atmestieji stebėjimai naudojami gautos prognozavimo funkcijos patikrinimui, jos koregavimui ir sprendimo apie jos pritaikomumą padarymui.

Laiko eilutę, kurios ilgis tik 22 stebėjimai nuodugniai ištirti ir padaryti tiksliai išvadas sudėtinga. Nėra galimybės atmesti duomenis ir vėliau jais tikrinti gautuosius rezultatus. Turimų 22 stebėjimų yra per mažai, kad gautais rezultatais būtų galima drąsiai pasikliauti. Norint ateityje naudotis regresinės kreivės lygtimis nuolat būtina perskaičiuoti koeficientus įtraukiant vis naujų stebėjimų duomenis.

4 lentelėje pateikti rezultatai gauti iš turimos laiko eilutės pabaigos atmetus 1, 2, ir 3 stebėjimus ir įvertinus skrelyje 3.1. gautos (2) regresinės lygties koeficientus.

4 lentelė

Įtrauktų stebėjimų skaičius	(1)	22	21	20	19
Determinacijos koeficientas	(2)	0,8647	0,8663	0,8933	0,8980
Koreliacijos koeficientas	(3)	0,92989	0,93074	0,94512	0,94764
Final Value	(4)	4 939 273,80	4 754 185,00	3 414 378,75	2 559 327,00
Skirtumų kvadrato, tarp realios ir aproksimuojančios gautos reikšmės, dalis tenkanti vienam stebėjimui	(5)= (4)/(1)	224 512,45	237 709,25	179 704,14	142 184,83
Absoliutinių skirtumų, tarp realiųjų ir prognozuojamųjų reikšmių, vidurkis	(6)	-	641	1285	1439

Matome, kad nors ir nežymiai, bet anksčiau aptarti koeficientai gerėja mažėjant stebėjimų skaičiui. Taip pat mažėja ir kvadratinis nuokrypis, aproksimuojančios reikšmės nuo realiosios, bet prognozuojamiems rezultatams šis stebėjimų mažinimas, kaip ir galima buvo numanyti, atsiliepia neigiamai. 4 lentelės (6) eilutėje matome, kad absoliutinis vidutinis skirtumas didėja, todėl prognozėms geriau naudoti kreivės lygtis, kurių koeficientai gauti naudojant kuo daugiau stebėjimų.

4 priede pateikiami grafikai, kuriuose pavaizduoti realūs duomenys, jų aproksimacijos kreivėmis gautomis naudojant 22-u, 21-a, 20-t ir 19-ka laiko eilutės stebėjimų, ir pagal regresijos kreivės funkcijas suprognozuotos tolimesnės laiko eilutės reikšmės. 4 priedo paskutiniame grafike pavaizduoti visi 4 aproksimacijų variantai, realioji laiko eilutė ir visos prognozės (prognozės pažymėtos kvadratukais, atitinkamoje aproksimuojančioje kreivėje). Matome, kad kuo mažiau duomenų naudojama aproksimacijai, tuo labiau nutolusios prognozuojamosios reikšmės nuo realiųjų reikšmių.

Taigi norint taikyti regresijos lygtimis aprašomas kreives prognozavimui, būtina jų suradimui naudoti kuo ilgesnę laiko eilutę.

4. LAIKO EILUČIŲ INDEKSAI

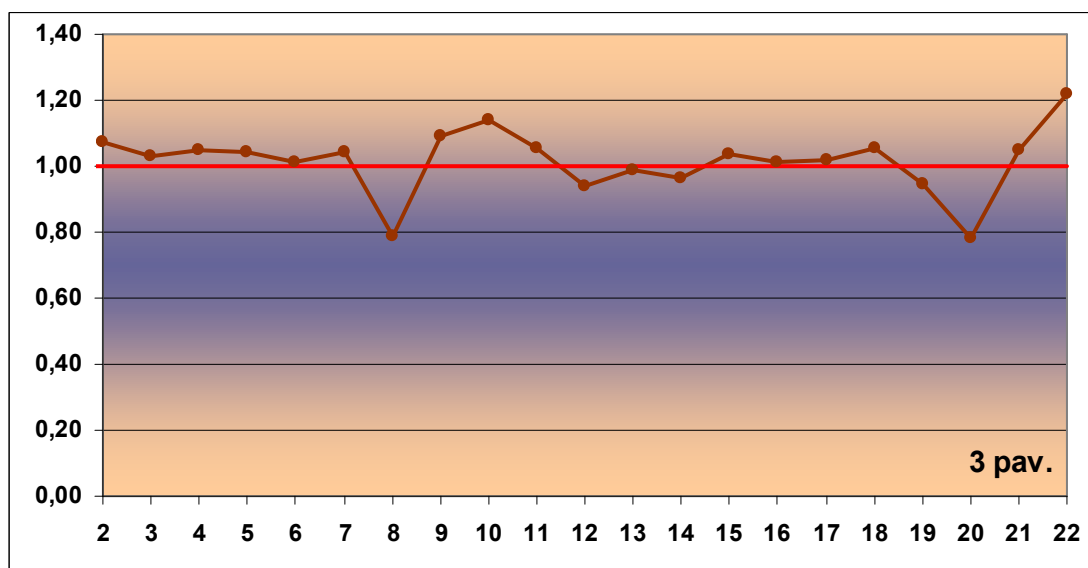
4.1. INDEKSAI SIEJANTYS GRETIMUS LAIKO EILUTĖS NARIUS

Laiko eilutės x_1, \dots, x_n indeksais vadinsime kintamuosius i_k , kurie apibrėžiami taip:

$$i_k = \frac{x_k}{x_{k-1}}, k = \overline{2, n}. \quad (5)$$

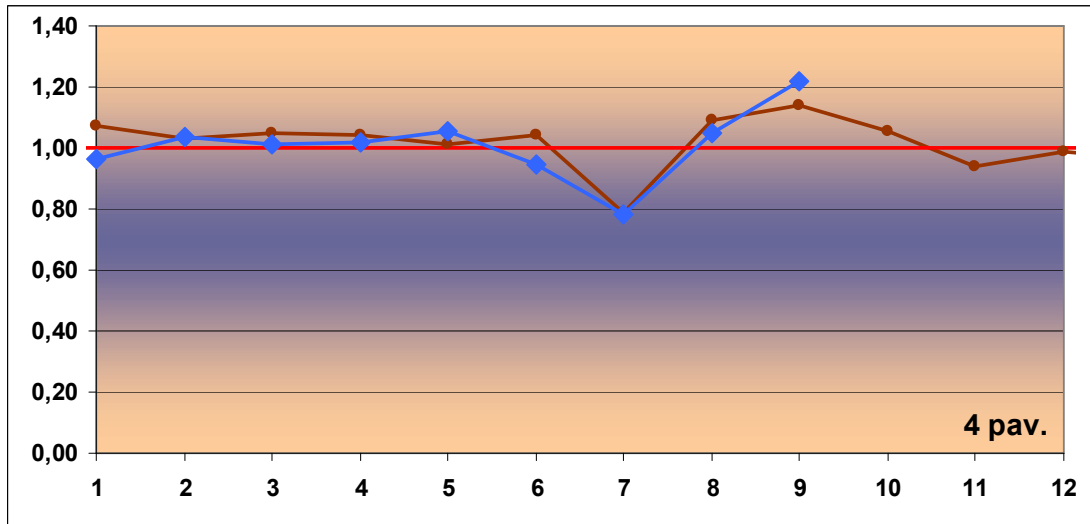
Jeigu laiko eilutė neturi didelių nuosmukių ir pakilimų, tuomet indeksų i_k įgyjamos reikšmės yra netoli 1. Indeksas i_k įgyja reikšmę didesnę už 1, kai k-tasis narys yra didesnis už prieš jį buvusį ir įgyja reikšmę mažesnę už 1, kai k-tasis narys yra mažesnis už prieš jį buvusį.

Laiko eilutę x_1, \dots, x_n pavertę į indeksų eilutę i_2, \dots, i_n galime stebėti jos pokyčius laike, bei daryti išvadas apie laiko eilutės tendencijas. Indeksų eilutės ilgis visuomet vienu vienetu mažesnis nei laiko eilutės ilgis. Nubrėšime indeksų eilutę, kuri atspindini įformintų importo procedūrų skaičiaus laiko eilutę (3 pav.).

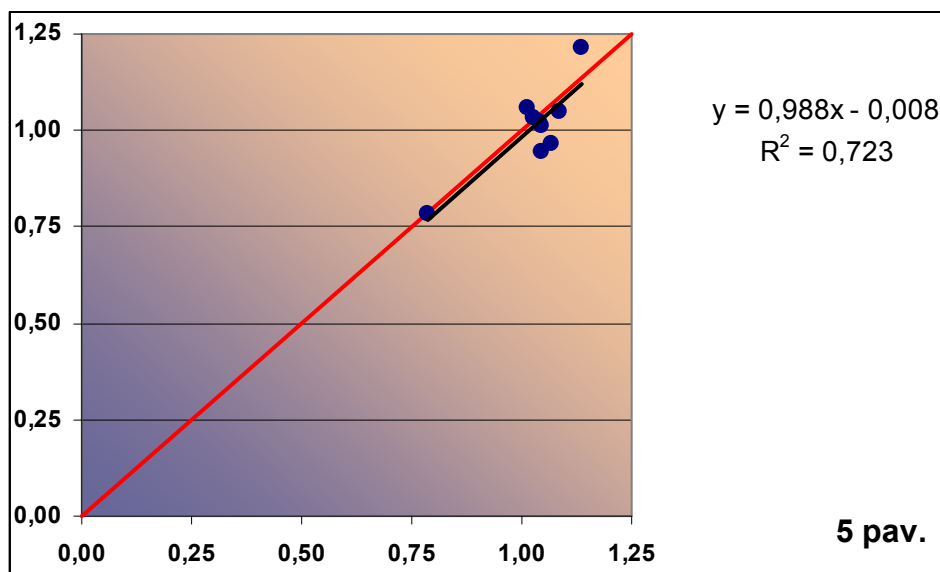


Iš šio grafiko matome, kad dauguma indeksų netoli 1. Visi indeksai pasiskirstę intervale $[0,785; 1,216]$, šio intervalo ilgis 0,431, o vidurio taškas 1, taigi išsibarstymo intervalas yra simetriškas tiesei $y=1$. Pastebime kelias išsiskiriančias reikšmes t. y. lokalius f-jos minimumus taškuose 8 ir 20, bei lokalius maksimumus taškuose 10 ir 22. Kadangi $20-8=12$ ir $22-10=12$ tai akivaizdu, kad turime laiko eilutę, kurioje pasireiškia sezoniškumas, t. y. pakilimai ir nuosmukiai kartojasi kas 12 mėnesių.

Nubrėškime dar vieną grafiką, kurio x ašyje žymėsime skaičius nuo 1 iki 12 atitinkančius 12 metų mėnesių, o indeksus koordinacių plokštumoje atidėsime dviem sekomis ($21/12=1,75$; $[1,75]=2$), pirmojoje indeksai nuo 2 iki 13 antrojoje nuo 14 iki 22. Gauti rezultatai pavaizduoti 4 pav.



Iš šio grafiko matome, kad indeksai i_k ir i_{k+12} beveik sutampa, t. y. panašu, jog kas 12-tas indeksas beveik atsikartoja. Nubrėškime dar vieną grafiką, t. y. grafiką atspindinti i_{k+12} priklausomybę nuo i_k , $k=2, \dots, 10$ (5 pav.). Koordinacių plokštumoje papildomai nubrėžtas lygties $y=x$ grafikas, matome, kad jis beveik sutampa su trendo tiese, kurios lygtis pateikta 5 pav. Šios lygties koeficientas prie kintamojo x beveik lygus 1, o laisvasis narys artimas 0, todėl ji beveik sutampa su lygtimi $y=x$. Taigi iš to seka išvada, kad kas 12 indeksas beveik nesiskiria.



Sugrįžkime prie nagrinėjamos laiko eilutės ir formulės (5) pagal kurią iš laiko eilutės gaunami indeksai. Išreiškime iš jos x_k :

$$x_k = x_{k-1} \cdot i_k, k = 2, 3, \dots \quad (6)$$

Taigi matome, kad kiekvieną laiko eilutės narį galima išreikšti per prieš tai buvusį narį ir atitinkamą indeksą. Momentu k , indeksas i_k nėra žinomas, bet žinomi indeksai i_{k-12n} , $n=1, 2, \dots$, todėl (6) formulę perrašykime vietoje i_k įrašydami i_{k-12} , arba $\frac{1}{n} \sum_{z=1}^n i_{k-12z}$, priklausomai nuo to ar manome, jog laiko eilutės tolimesnės reikšmės priklauso, tik nuo 12 paskutinių stebėjimų ar nuo $12n$ paskutinių stebėjimų.

Žinoma, laiko eilutė yra per trumpa daryti griežtoms išvadoms, ypačingai dėl to, kad 5 pav. pavaizduotas grafikas turi tik 9 taškus. Jeigu šiame grafike pavaizdavus nemažiau, kaip 12 taškų trendo lygties koeficientas išliktų lygus 1, o R^2 pakankamai arti 1, tuomet indeksus būtų galima taikyti prognozavimui.

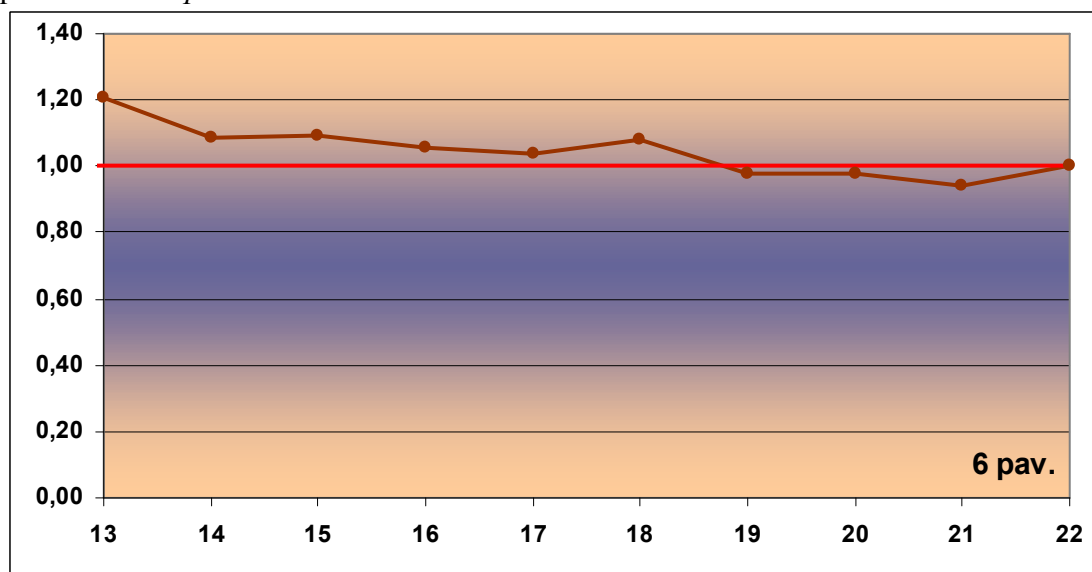
Panagrinėkime importo procedūrų skaičiaus logaritmų skirtumus $\ln(x_k) - \ln(x_{k-1})$, $k \geq 2$, arba $\ln(i_k)$, nes tai tas pats. Patikrinkime ar šie skirtumai nėra „baltasis triukšmas“. Tai galime patikrinti atlikę autokoreliacinę analizę, jos rezultatai pateikti 5 priede. Su tikimybe 0,90 priimame, kad tai „baltasis triukšmas“, nors su tikimybe 0,95 to daryti negalime. Bet įtraukus daugiau duomenų tikėtina, kad hipotezė apie „baltąjį triukšmą“ bus atmetama ir su tikimybe 0,95.

4.2. INDEKSAI SIEJANTYS LAIKO EILUTĖS NARIUS NUTOLUSIUS VIENAS NUO KITO PER 12 MĖNESIŲ

Panagrinėkime indeksus apibrėžiamus šiek tiek kitaip nei 4.1. skyrelyje:

$$i_k^{12} = \frac{x_k}{x_{k-12}}, k = \overline{13, n}.$$

Šios indeksų eilutės ilgis lygus $n-12$. Nagrinėjamoju muitinės importo procedūrų skaičiaus atveju turime indeksų eilutės ilgį 10 ($22-12=10$). Ši indeksų eilutė pavaizduota 6 pav.



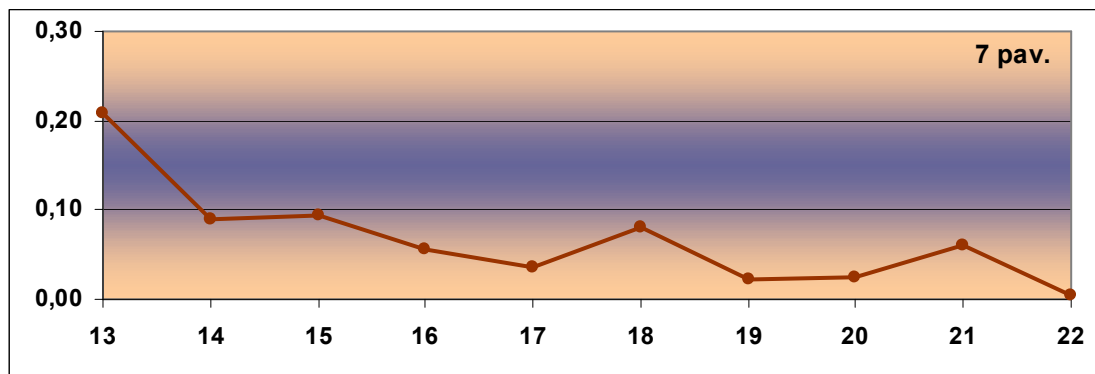
Indeksai i_k^{12} pasiskirstę intervale $[0,94; 1,21]$, šio intervalo ilgis 0,27, o vidurio taškas 1,07, šio išsibarstymo intervalas nėra simetriškas tiesei $y=1$.

Pastebime kelis dalykus:

1) 13-to mėnesio deklaracijų skaičius labiausiai skiriasi nuo 1-mo mėnesio deklaracijų skaičiaus nors abu atitinka birželio mėnesį (13-tas – 2005 metų, o 1-mas 2004 metų), lyginant su kitų atitinkamų mėnesių skirtumais. Taip pat matome, kad kovo mėnesio duomenys 2005 m. sutampa su 2006 m., t. y. $i_{22}^{12}=1,00$.

2) 7-ni iš 10-ties indeksų yra didesni už 1.

3) Didėjant k , skirtumo $|i_k^{12} - 1|$ reikšmės kinta, bet artėja prie 0. (žr. 7 pav.).



Ši indeksų eilutė, per trumpa daryti apibendrinančioms laiko eilutę išvadoms, reikėtų mažiausiai trijų metų stebėjimų. Bet atsižvelgus į laiko eilutės prigimtį, išsišokančią indekso i_{13}^{12} reikšmę galima paaiškinti: šio indekso apskaičiavimui naudojamas pirmas laiko eilutės elementas, kuris gali neatitikti eilutės prigimties, nes veiksniai įtakojantys šį rodiklį dar galėjo būti nenusistovėję po „šoko“, t. y. Lietuvos įstojimo į ES 2004 m. gegužės 1 d. Taip pat galėtume paaiškinti ir palyginus labai didelį skirtumą tarp indeksų $i_2=1,07$ ir $i_{14}=i_{2+12}=0,96$ nagrinėtų 5.1. skyrelyje.

Išvados naudingos laiko eilutės prognozavimui:

1) Atitinkamo mėnesio reikšmė vidutiniškai yra 5 proc. didesnė, nei prieš tai buvusiais metais (arba 3 proc. didesnė, jei nuspręsimė, nesiremti 2004 m. birželio mėn. duomenimis).

Jei skaičiuosime, kiek vidutiniškai procentų padidėja mėnesio reikšmė nuo prieš tai buvusių metų atitinkamo mėnesio reikšmės naudodamiesi slenkančiu vidurkiu (periodas lygus 6), gausime, kad vidutinis didėjimo procentas mažėja nuo 9 iki 0 (turint pakankamai duomenų logiška būtų skaičiuoti 12 periodų slenkantį vidurkį, tuomet jis atsispindėtų visų metų pakilimų ir nuosmukių vidurkį).

2) Pastebimas vidutiniškai 7 proc. pokytis nuo stebėtos reikšmės praėjusių metų atitinkamą mėnesį (arba 5 proc. pokytis, jei nuspręsimė, nesiremti 2004 m. birželio mėn. duomenimis).

Jei naudosimės slenkančiu vidurkiu (periodas lygus 6), tai vidutinis pokytis lyginant su pernai metų duomenimis mažėja nuo 9 iki 4 proc. (slenkančių vidurkių eilutė atrodo taip: 9 proc., 6 proc., 5 proc., 5 proc., 4 proc.).

Prognozei gauti reikia naudotis formule:

$$x_k = x_{k-12} \cdot i_k^{12}, k = 13, 14, \dots \quad (7),$$

Joje, kaip ir 4.1. skyrelio analogiškoje formulėje indeksą i_k^{12} pakeiskime išraiška $\frac{1}{n} \sum_{z=1}^n i_{k-12z}^{12}$, ir parinkime $n=1$, jei norime prognozuoti tik remdamiesi paskutiniu atitinkamu indeksu arba $n>1$, jei norime prognozuoti pagal n paskutinių indeksų reikšmes.

4.3. PROGNOZAVIMAS PRITAIKANT INDEKSUS

4.1. ir 4.2. skyreliuose nagrinėjome du indeksus susietus su laiko eilute aprašomu muitinės importo procedūrų skaičiumi. Norint įsitikinti, kad samprotavimai teisingi reikėtų patikrinti ar pagal (6) ir (7) formules gaunamos reikšmės sutampa su stebimomis reikšmėmis.

4.1. skyrelyje gavome 21 indeksą ($i_{2,...,i_{22}}$), kas dvyliktas indeksas atspindi tokį pat mėnesį (tik skirtingų metų), taigi 9 iš jų atspindi tą patį metų mėnesį, todėl į (6)

formulę įrašysime $\frac{i_{k-12} + i_{k-24}}{2}$, kai prognozuojama reikšmė x_k , $k>25$, ir

$$i_{k-12}, \quad \text{kai } k \leq 24.$$

6 priede pavaizduotame grafike matome realų importo procedūrų skaičių, ir importo procedūrų skaičių gautą remiantis (6) formule tiek buvusiems 9 laiko momentams, tiek ir prognozę būsimiems 12 laiko momentų. Prognozę būsimiems laiko momentams bus galima patikrinti vėliau, o buvusių 9 mėnesių prognozuojamų reikšmių nuo tikrųjų absoliutinis skirtumas sudaro vidutiniškai apie 2,3 proc., daugiausiai (5,5 proc.) skiriasi liepos mėnesio tikroji reikšmė nuo prognozuojamosios. Ši didžiausią skirtumą galima paaiškinti remiantis dviem faktais: 1) prognozei buvo naudotas tik vienas indeksas, o ne atitinkamų indeksų aritmetinis vidurkis, 2) indeksas naudotas prognozei gautas iš duomenų, kurie galėjo būti netipiniai t. y. labiau įtakoti „šoko“ negu kiti.

Kad prognozės būtų patikimesnės, būtina įtraukti kuo daugiau mėnesių duomenų. Turint galimybę įtraukti bet kokį kiekį duomenų, taip pat reikės papildomos analizės, kad nuspresti ar visi duomenys naudingi, ir ar tikėtina, kad jie atspindės tai ko tikimės ateityje. Gali paaiškėti, kad remtis visais turimais duomenimis nenaudinga, o tinkamiausi prognozei tik kelių paskutinių metų duomenys.

7 priede pavaizduotas kitas grafikas. Jame matome realų importo procedūrų skaičių, ir remiantis (7) formule gautą prognozę būsimiems 12 laiko momentų. Tam kad pasinaudoti (7) formule reikėjo nuspręsti kokius indeksus naudoti. Kad gauti 12 indeksų i_k^{12} reikia turėti 24 stebėjimus, o nagrinėjamoju atveju turime tik 22. Iš stebėjimų eilutės galime gauti 10 indeksų (bet skyrelyje 4.2. pastebėjome, kad vienas iš indeksų atrodo labai nepatikimas prognozavimui, todėl jo neskaičiuojame įprastu būdu), mums dar reikalingi 2 indeksai (arba 3, nes vieno iš galimų apskaičiuoti nenaudosime), kad galėtume prognozuoti metams. Taigi, tuos tris indeksus gauname taip: suskaičiuojame aritmetinį turimų 9 indeksų vidurkį. Bet į formulę (7) įrašinėjame, ne šiuos indeksus, o jų slenkančio vidurkio išraiškas (periodas lygus 4), pažymėkime šias išraiškas \hat{i}_k^{12} , taigi prognozavimui naudosimės tokia išraiška:

$$x_k = \hat{i}_k^{12} \cdot x_{k-12} = \frac{i_{k-11}^{12} + i_{k-12}^{12} + i_{k-13}^{12} + i_{k-14}^{12}}{4} \cdot x_{k-12}, \quad (8)$$

matome, kad

$$\hat{i}_k^{12} = \frac{i_{k-11}^{12} + i_{k-12}^{12} + i_{k-13}^{12} + i_{k-14}^{12}}{4},$$

tai slenkantis vidurkis indeksų atitinkančių k-tąjį, k-1-ąjį, k-2-ąjį ir k+1-ąjį mėnesį. (Pastaba: $k-12n=k$, pvz. $k=1$ = birželis, tai $k=27=3+12\cdot 2=3$ =rugpjūtis).

Šios prognozės kol kas niekaip negalime patikrinti. Ateityje ji bus tikrinama ir koreguojama parenkant tinkamiausius indeksus arba jų išraiškas.

5. APIBENDRINIMAS

5.1. PROGNOZIŲ PALYGINIMAS

Ankstesniuose skyreliuose išskyrėme tris f-jas skirtas prognozavimui ((2), (6) ir (8)). Tai trim nepriklausomais būdais gautos f-jos. Kol kas sudėtinga įvertinti, kuri iš jų tinkamiausia, nes duomenų, kuriais remiantis būtų galima patikrinti yra per mažai. Todėl 8 *priede* tiesiog pateiktas grafikas, kuriame pavaizduotos visomis trimis f-jomis naudojantis gaunamos prognozės. Šie grafikai atrodo gana panašiai, bet kuris iš jų tiksliausias parodys laikas.

Bet jau dabar galima išskirti silpnąsias kiekvienos prognozės vietas:

(2) f-ja gauta turint tik po vieną balandžio ir gegužės mėnesių stebėjimą, todėl šiuos mėnesius geriausiai aproksimuoja gauta f-ja, bet prognozei gali būti visiškai netinkama;

Pritaikant (6) f-ją prognozei nebuvo galimybės naudotis visų atitinkamų indeksų vidurkiais, nes indeksų tik 21, be to ši prognozė remiasi tik viena t. y. paskutiniąja stebėtos eilutės reikšme, o toliau visi skaičiavimai atliekami naudojantis jau suprognozuotomis reikšmėmis, todėl kuo ilgesnis prognozavimo periodas, tuo labiau prognozė gali būti nutolusi nuo tikrosios reikšmės. Bent vienas blogai parinktas indeksas gali padaryti didelės įtakos tolimesnei prognozei. Taip pat prognozės tikslumui gali turėti įtakos ir momentas, nuo kurio pradedama prognozuoti. Jei pradinė reikšmė yra netipinė laiko eilutės reikšmė, tuomet gauta prognozė gali turėti teigiamą arba neigiamą poslinkį. Tam, kad sušvelninti tokius netikslumus ir įvairias išsišonančias reikšmes reikėtų taikyti įvairaus periodiškumo slenkančius vidurkius;

Remiantis (8) f-ją nubrėžtas prognozės grafikas bendromis tendencijomis nelabai išsiskiria iš kitų dviejų, bet (8) f-jos pritaikomumas šių duomenų prognozavimui mažiausiai ištirtas, todėl prognozė gali būti pati netiksčiausia.

Kai laiko eilutė tokia trumpa, kiekvienas naujas stebėjimas yra labai svarbus ir gali turėti didelės įtakos bet kuriai prognozės funkcijai.

5.2. IŠVADOS

Prognozuoti reiškinius besiremiant tik tam tikra praeities stebėjimų aibe, sudėtingas dalykas. Kai praeities stebėjimų aibė nedidelė prognozavimas tampa dar sudėtingesnis. Tuomet nebeįveda ir įvairūs statistiniai paketai. Tokiais atvejais jie pasiūlo elementariausią prognozavimo būdą, t. y. kiekvieną būsimą reikšmę prognozuoja prieš tai buvusių stebėjimų vidurkiu. Bet tokios prognozės dažniausiai netenkina, taigi šiuo darbu buvo bandoma rasti būdą, kaip turint mažai duomenų galima juos analizuoti ir pritaikyti prognozavimui. Tikėtina, kad analizės metu gautomis funkcijomis sudarytos prognozės gali būti gana tolimos tikrosioms laiko eilutės reikšmėms, bet jų koeficientus ateityje atitinkamai pakoregavus bus galima prognozuoti nebegaištant daug laiko analizei.

Prognozės gali būti tolimos nuo realių reikšmių ir dėl tokių priežasčių kurios sunkiai įvertinamos prognozavimo momentu. Kelios iš tokių priežasčių gali būti šios: naujų šalių priėmimas į Europos Sąjungą, Euro įvedimas Lietuvoje, Šv. Velykų data, naujų teisės aktų įsigaliojimas ir įvairūs kiti su prekyba susijusių reiškinių pokyčiai.

Ieškant funkcijų skirtų prognozavimui buvo atlikta ir daugiau įvairių skaičiavimų, kurie nepasirodė naudingi, vis dėl tos pačios per trumpos duomenų aibės. Duomenų stygius atrodo dar didesnis, kai analizuoti bandoma remiantis ne mėnesiniais, o ketvirtiniais duomenimis, o laiko eilutės atspindinčios metinius duomenis apskritai neįmanoma sudaryti.

Turint daugiau duomenų būtų galima analizuoti ir lyginti fiksuoto ilgio sumines duomenų eilutes (pvz. atitinkančias vienus metus, t. y. 12 mėn.), tokios eilutės 1-asis narys sausio mėn. duomuo, 2-asis – sausio ir vasario mėn. duomenų suma, ..., 12-asis – sausio, vasario, ..., gruodžio mėn. duomenų suma. Panauginėjus kelias to paties reiškinio fiksuoto ilgio laiko eilutes galima rasti būdą, kaip rezultatus panaudoti prognozavimui.

Importo procedūrų skaičius tai duomuo, kuris kasmet yra prognozuojamas Muitinės departamente. Šis darbas turėtų palengvinti prognozavimo procesą. O prognozės bus gaunamos remiantis matematiniais samprotavimais ir jais pagrįstais skaičiavimais, todėl, tikėkimės, bus tikslesnės nei iki šiol.

Ateityje nustačius, kuris prognozavimo būdas yra patikimiausias, bus bandoma jį automatizuoti, kad būtų dar patogiau juo naudotis.

Šį darbą rašant įgyta patirtis, taip pat galės padėti analizuoti ir kitus duomenis bei geitai rasti tinkamiausią būdą jų prognozavimui.

6. LITERATŪRA

1. Čekanavičius V., Murauskas G. *Statistika ir jos taikymai I*. Vilnius: TEV, 2001.
2. Čekanavičius V., Murauskas G. *Statistika ir jos taikymai II*. Vilnius: TEV, 2002.
3. Keller G., Warrack B. *Statistics for management and economics*. Thomson Learning, Inc. 2003.
4. Deibold F.X., *Elements of Forecasting*. Thomson South-Western, 2004.
5. Sakalauskas V., *Duomenų analizė su STATISTICA*. Vilnius: Margi raštai, 2003.

7. SUMMARY

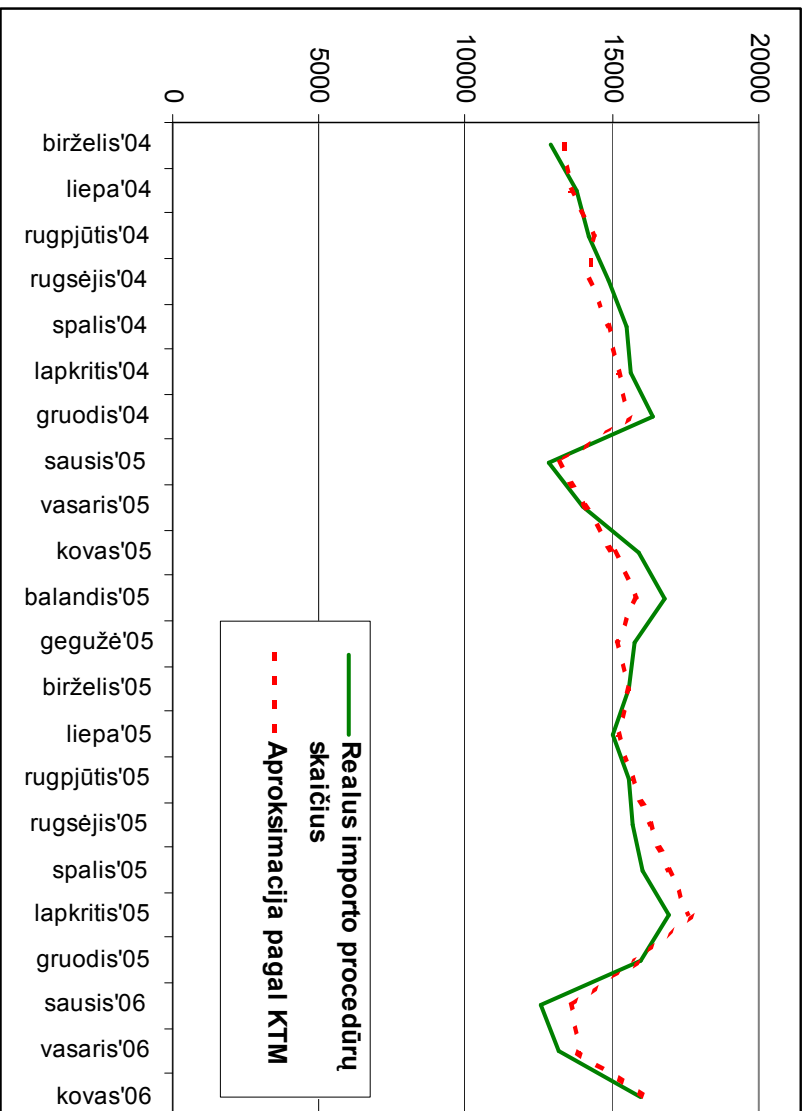
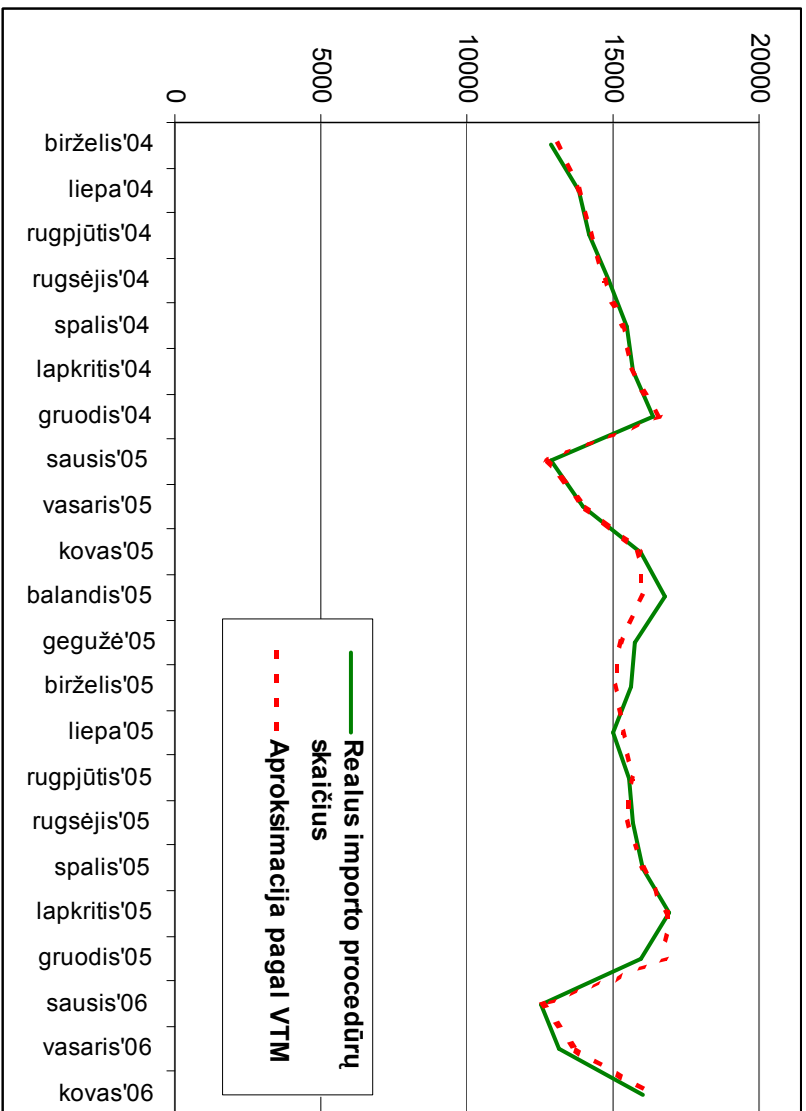
Analysis of Customs Procedures at Import

Monthly data of customs procedures at import are collected in Customs department. From the 1st of May 2004 the data became incomparable with the data collected before; differences appeared because Lithuania became EU state.

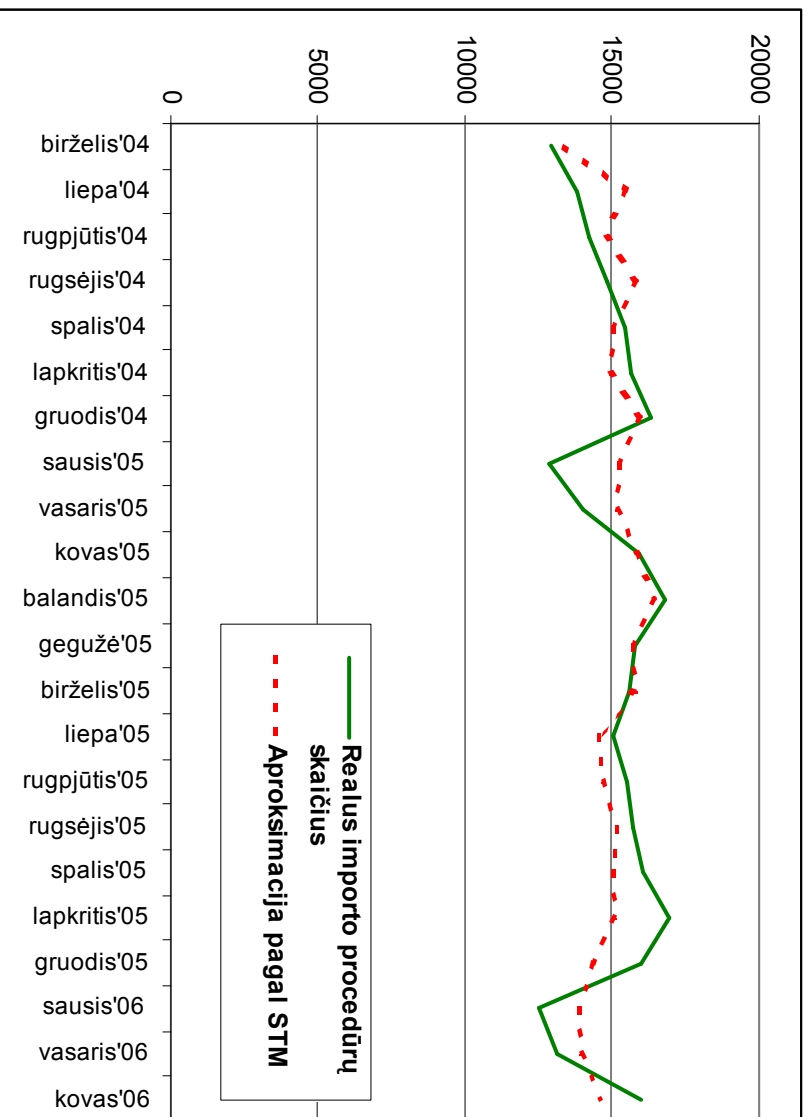
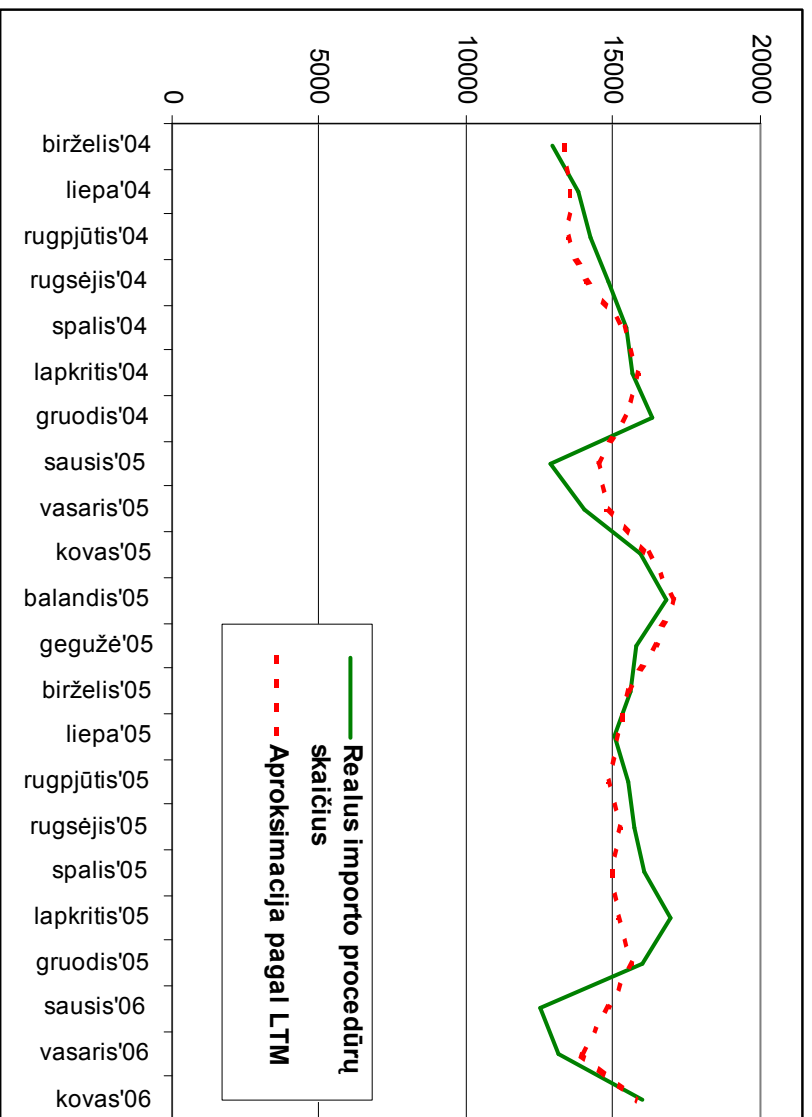
In this work time series expressing monthly number of procedures at import were analysed and some functions for forecasting were found. The functions were found using regression analysis, dummy variables and two types of indexes linking different time series elements. Available time series were quite short for drawing strict inferences therefore the best function for forecasting was not distinguished. It was concluded to compare future observations with predictions and to adjust coefficients of forecasting functions due to new observations.

8. PRIEDAI

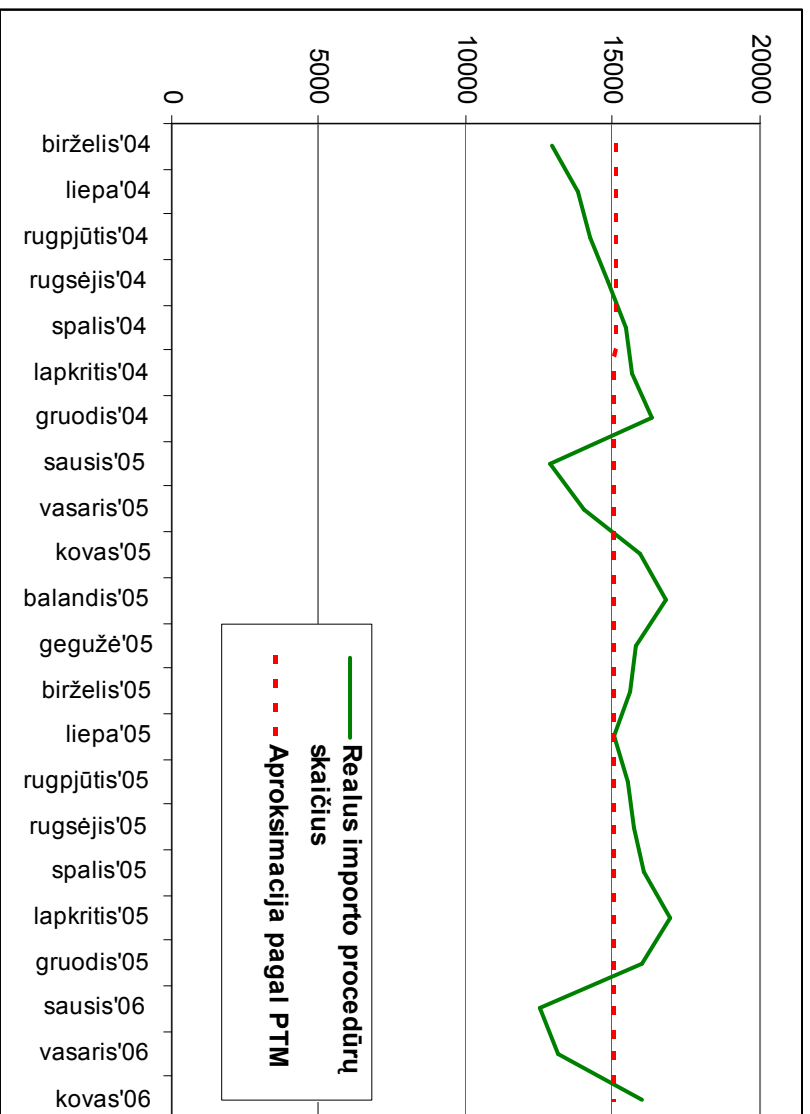
Priedas Nr.1

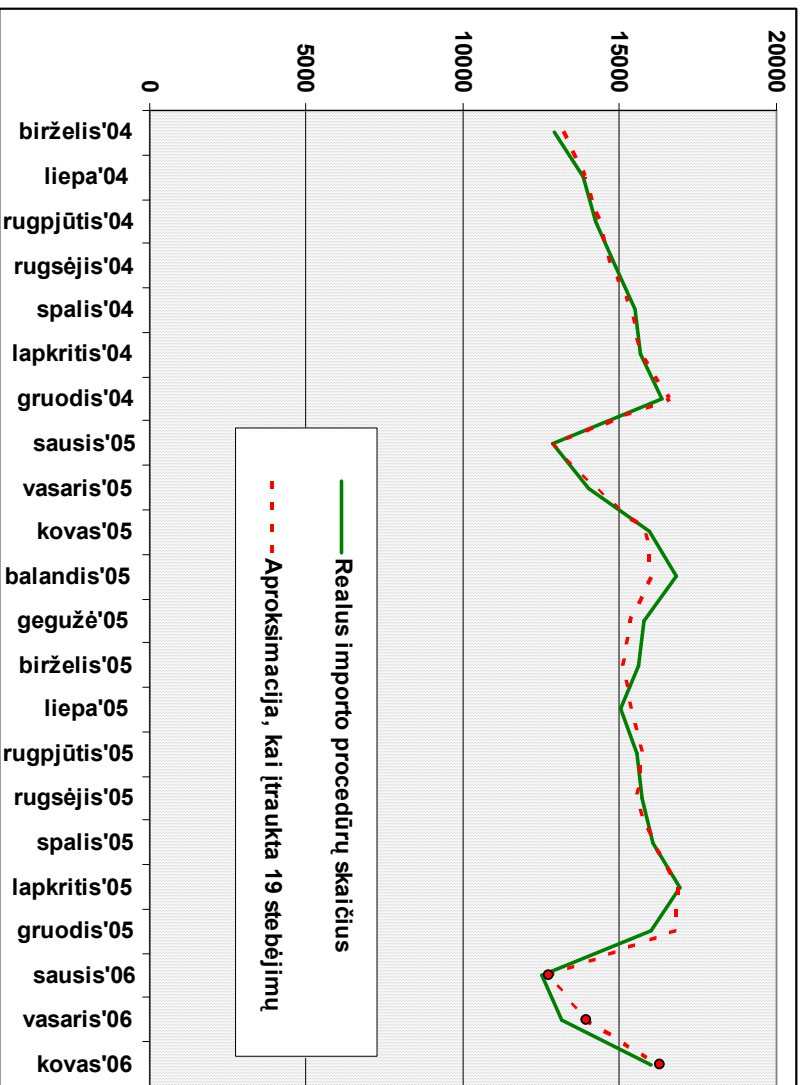
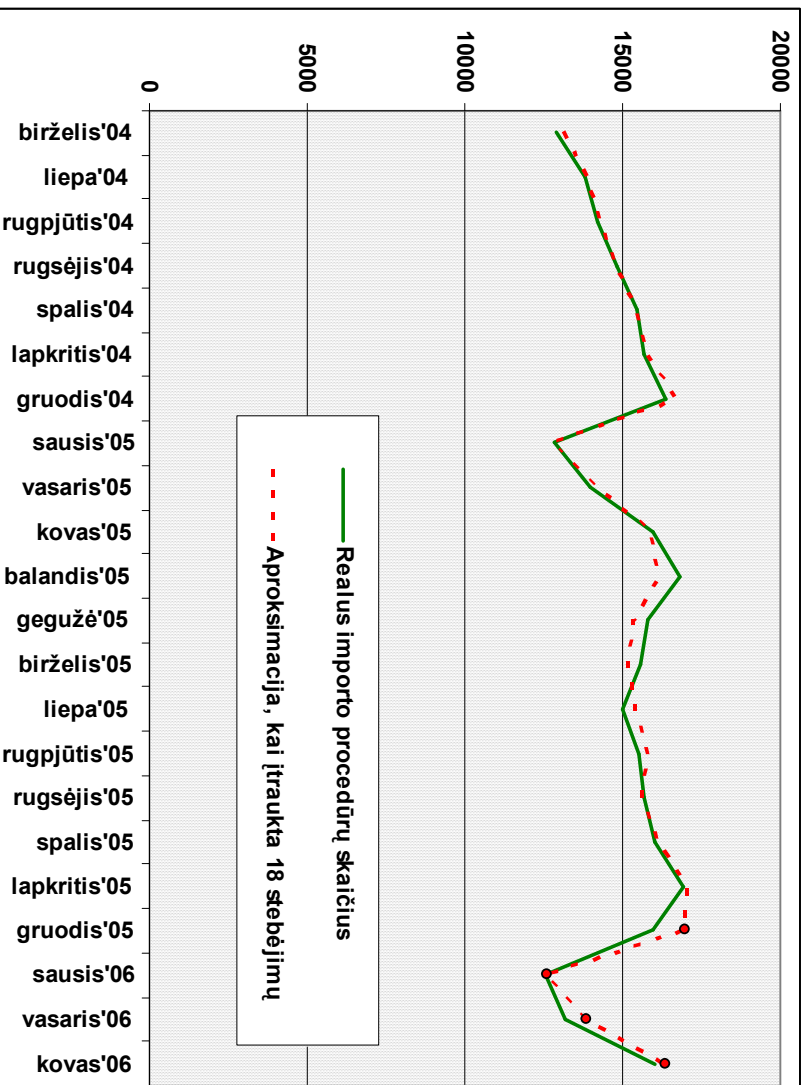


Priedas Nr.1 (tēsinys)

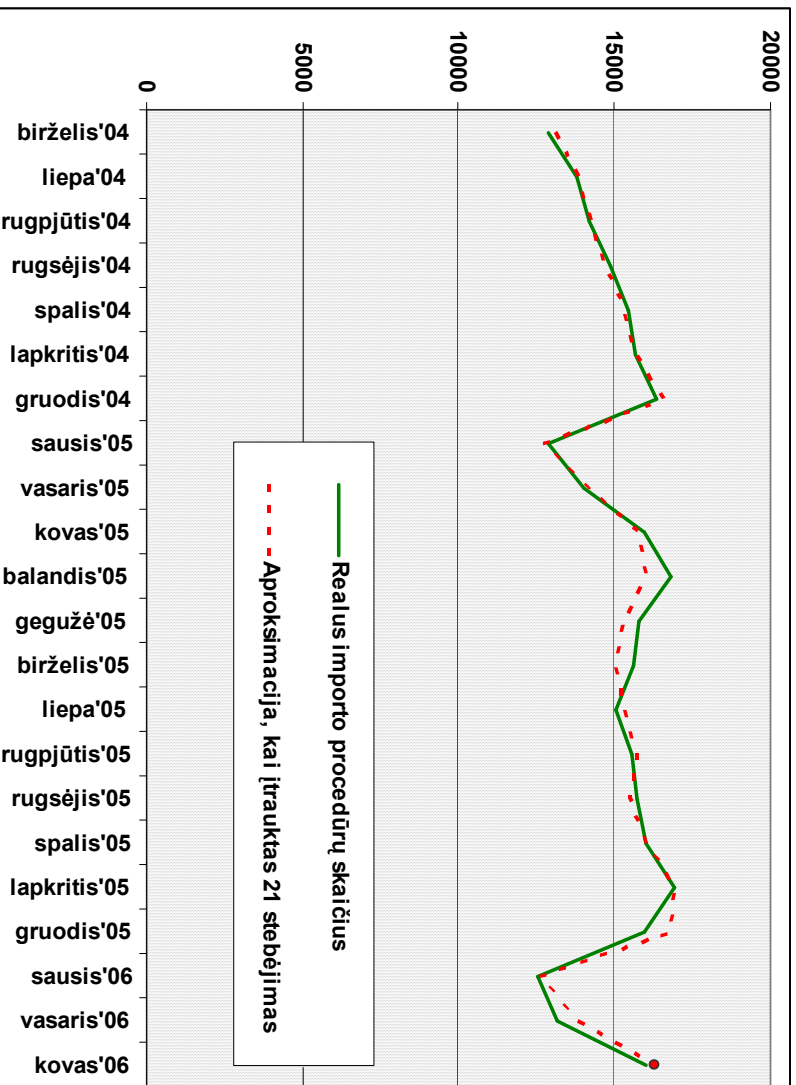
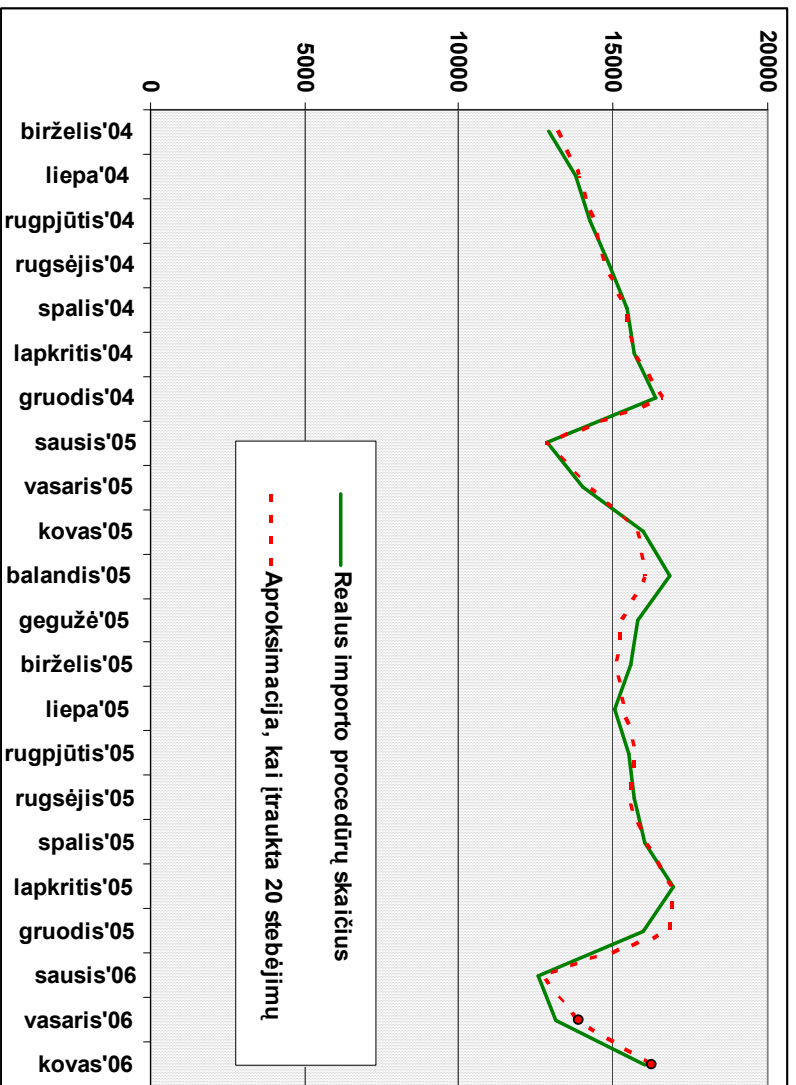


Priedas Nr.1 (tēsinys)

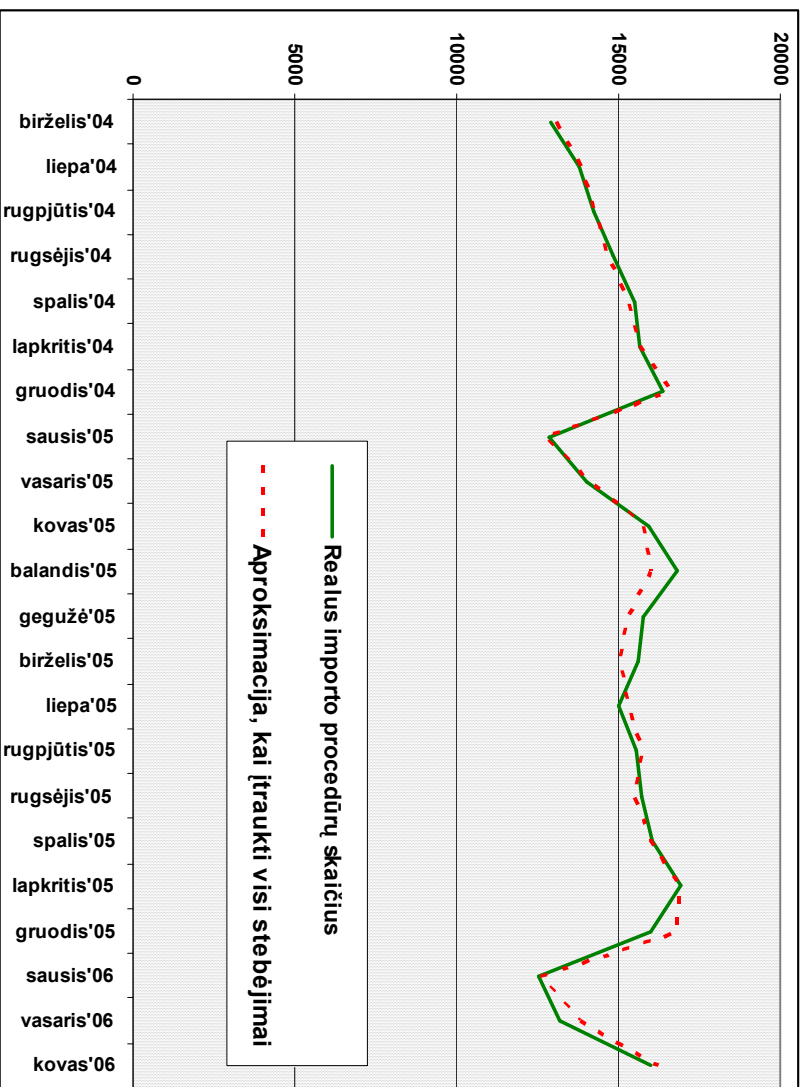


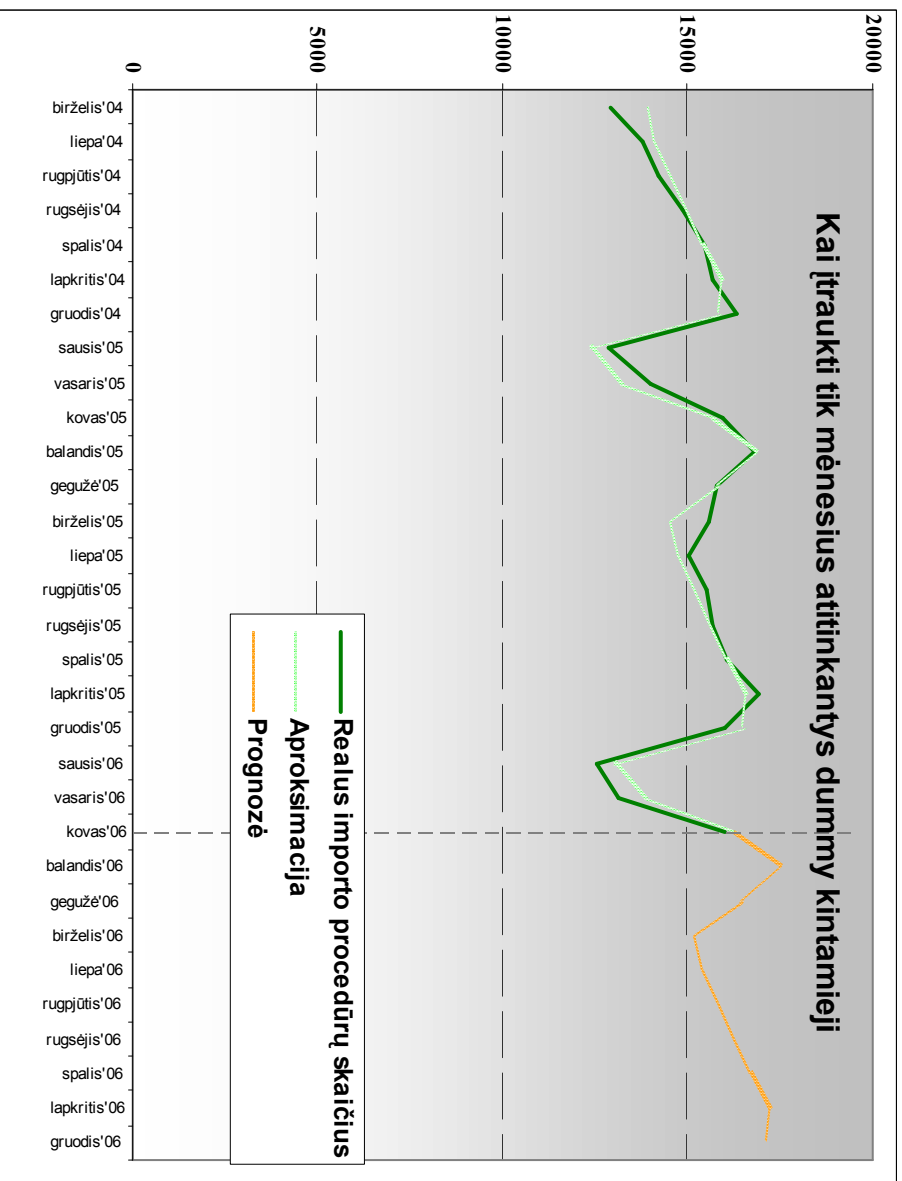
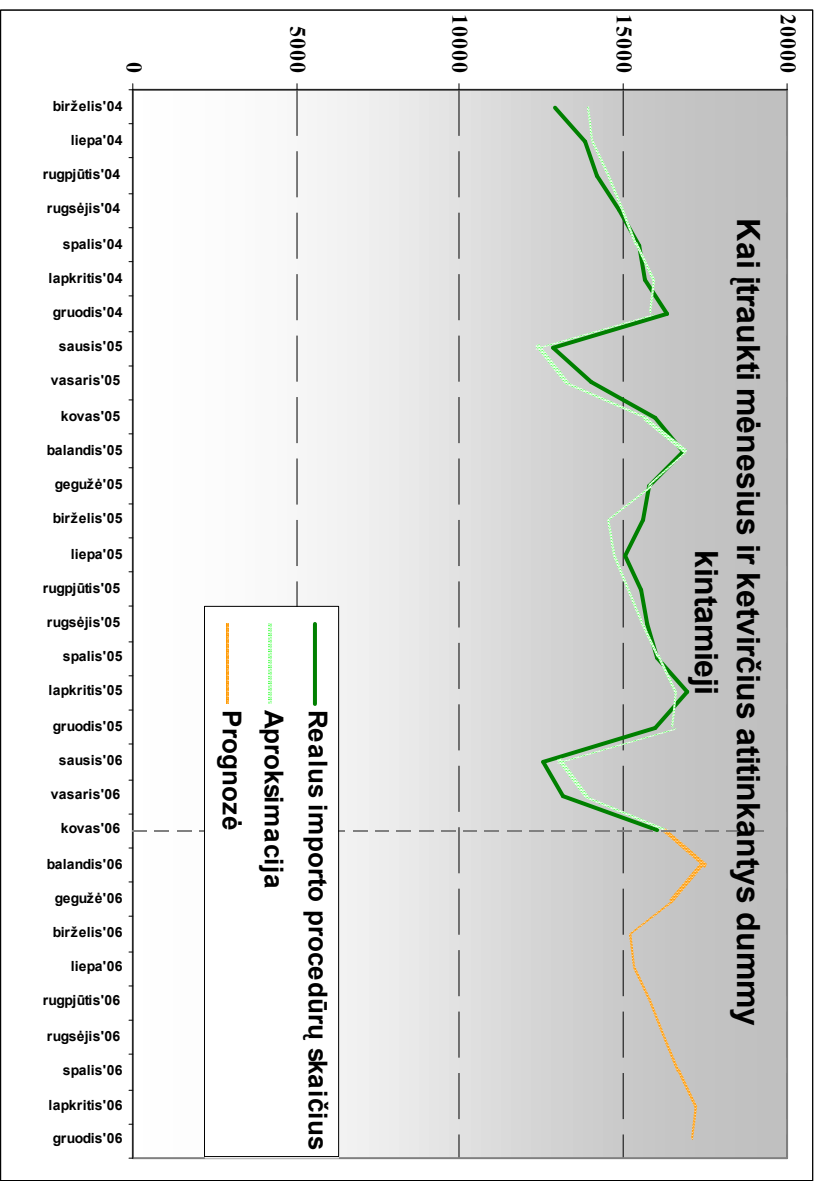


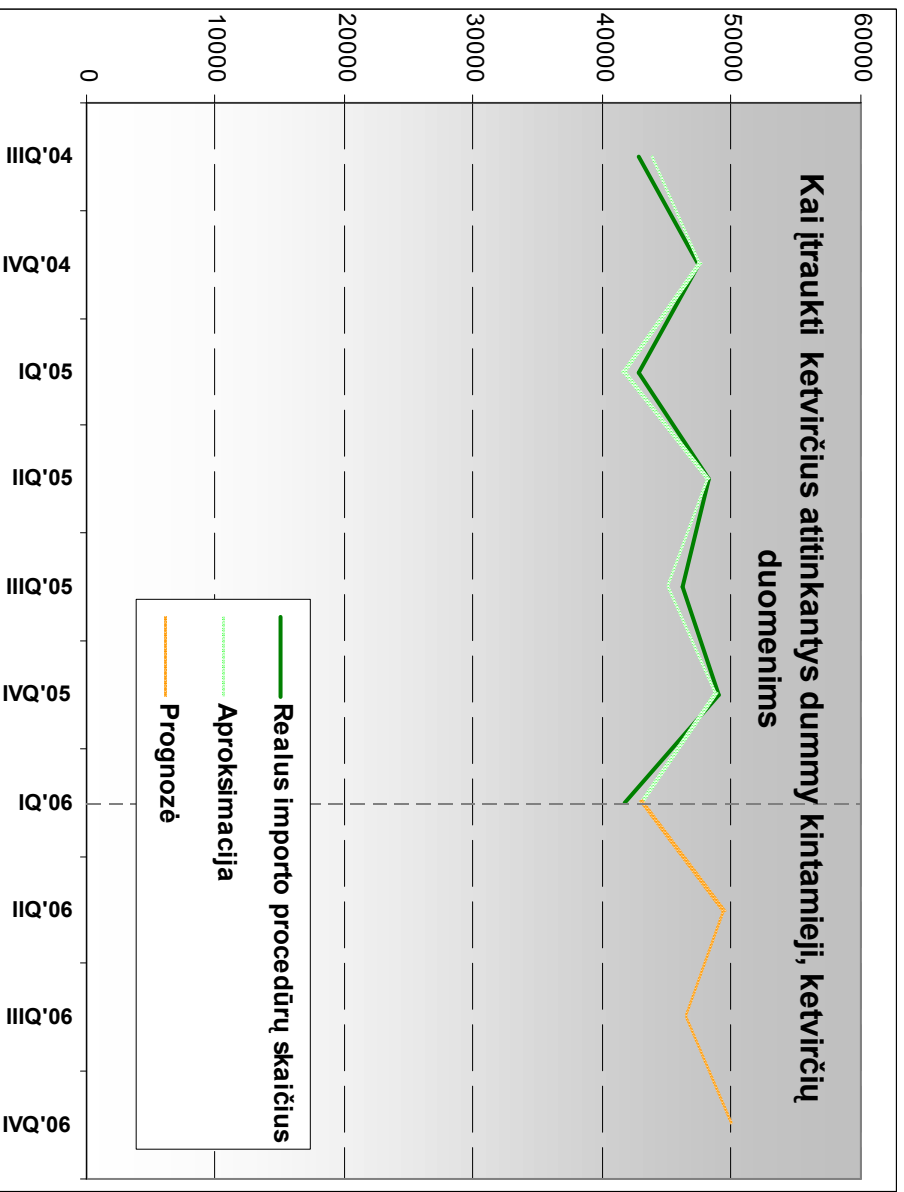
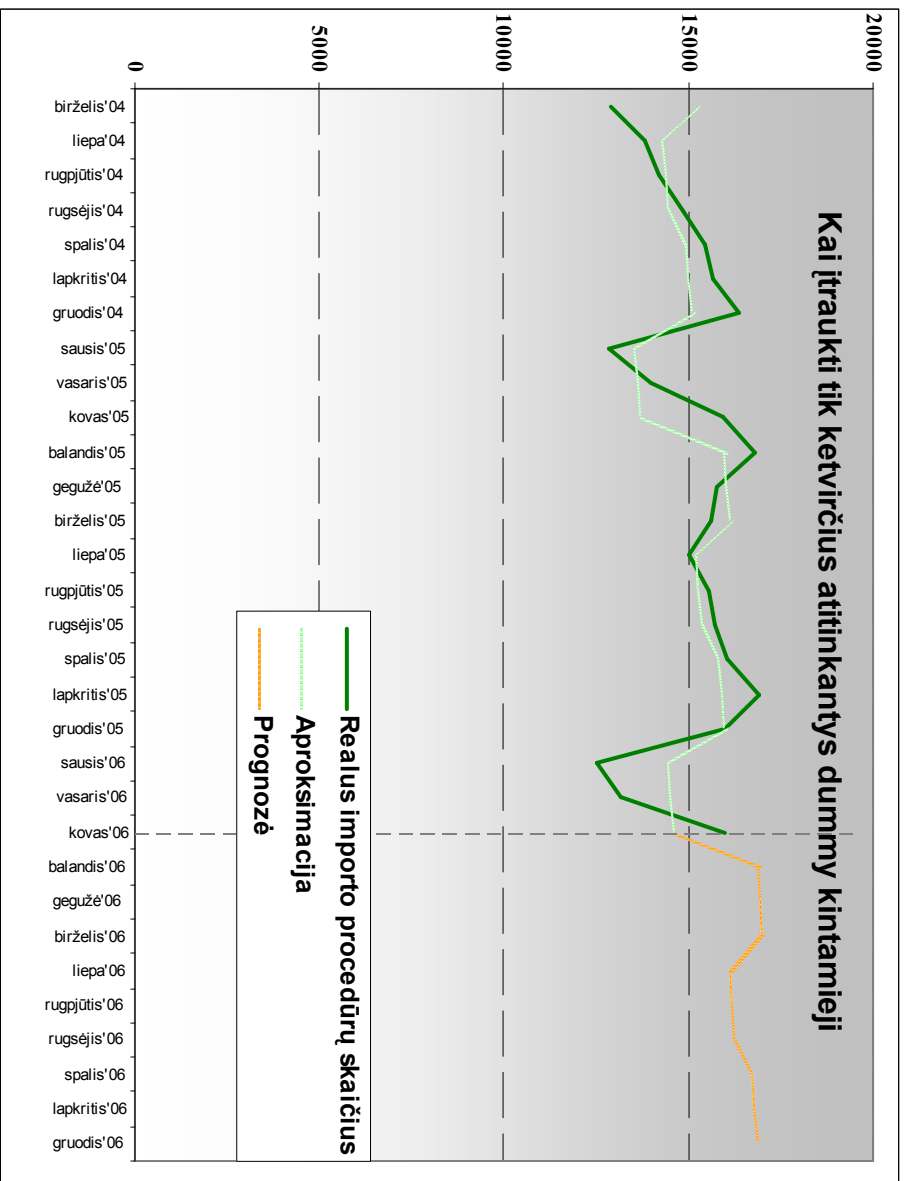
Priedas Nr.2 (tęsinys)

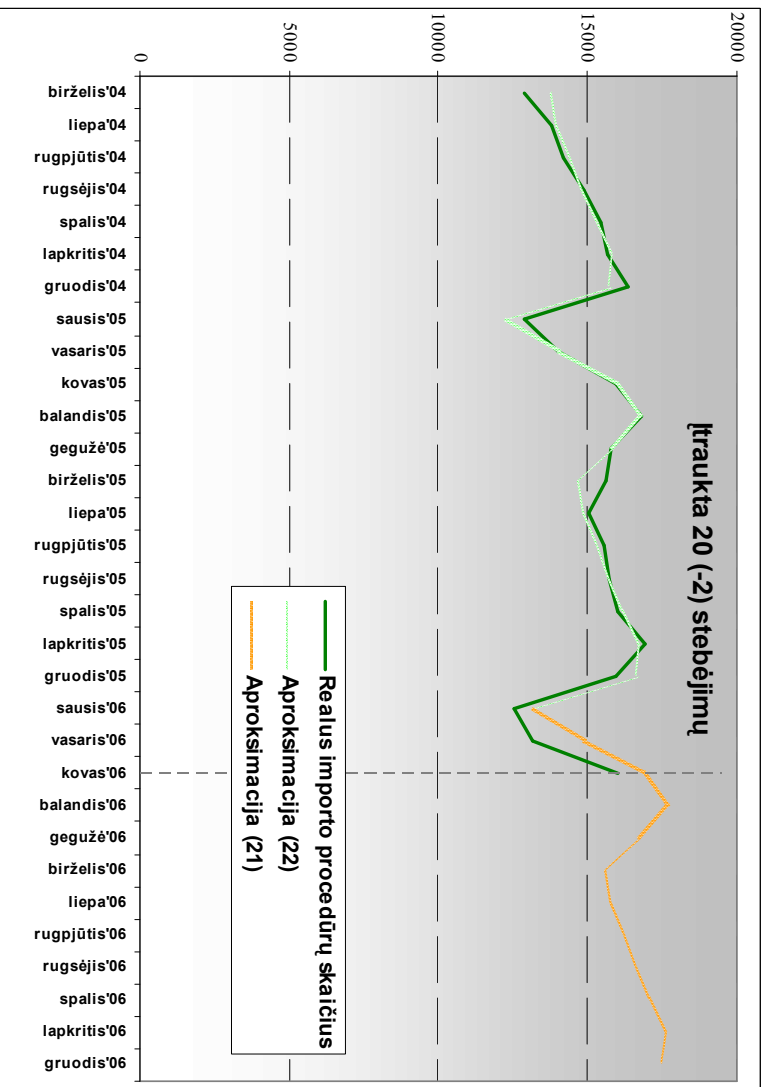
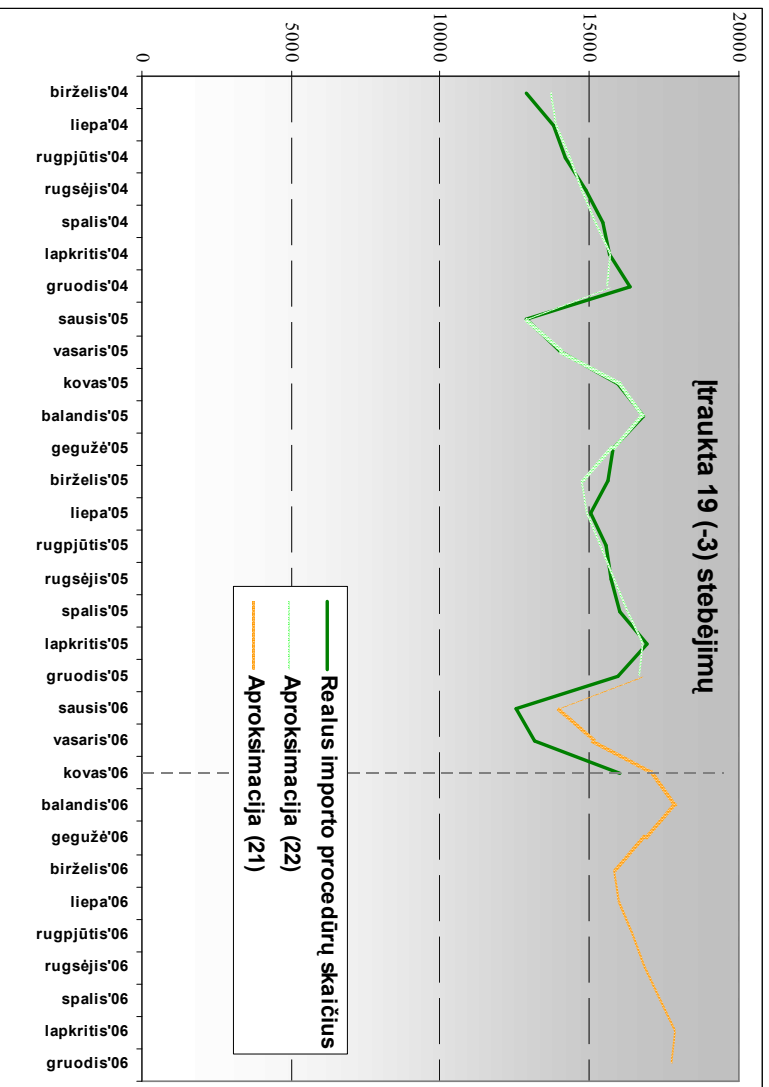


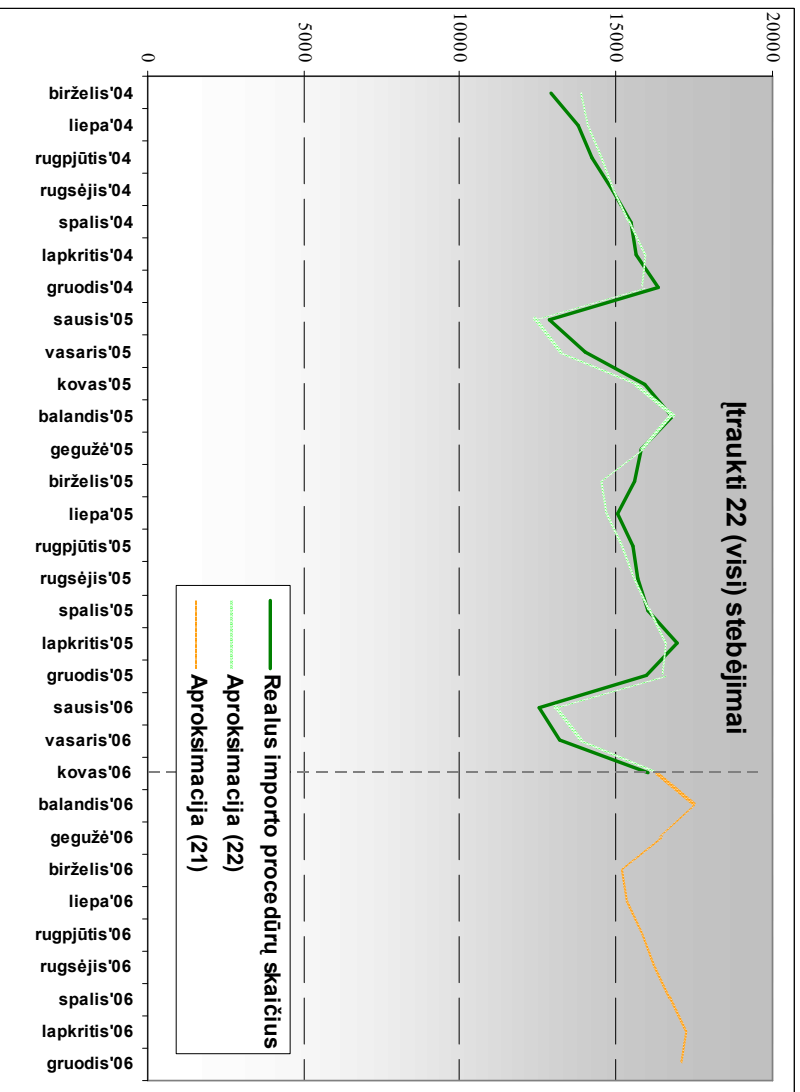
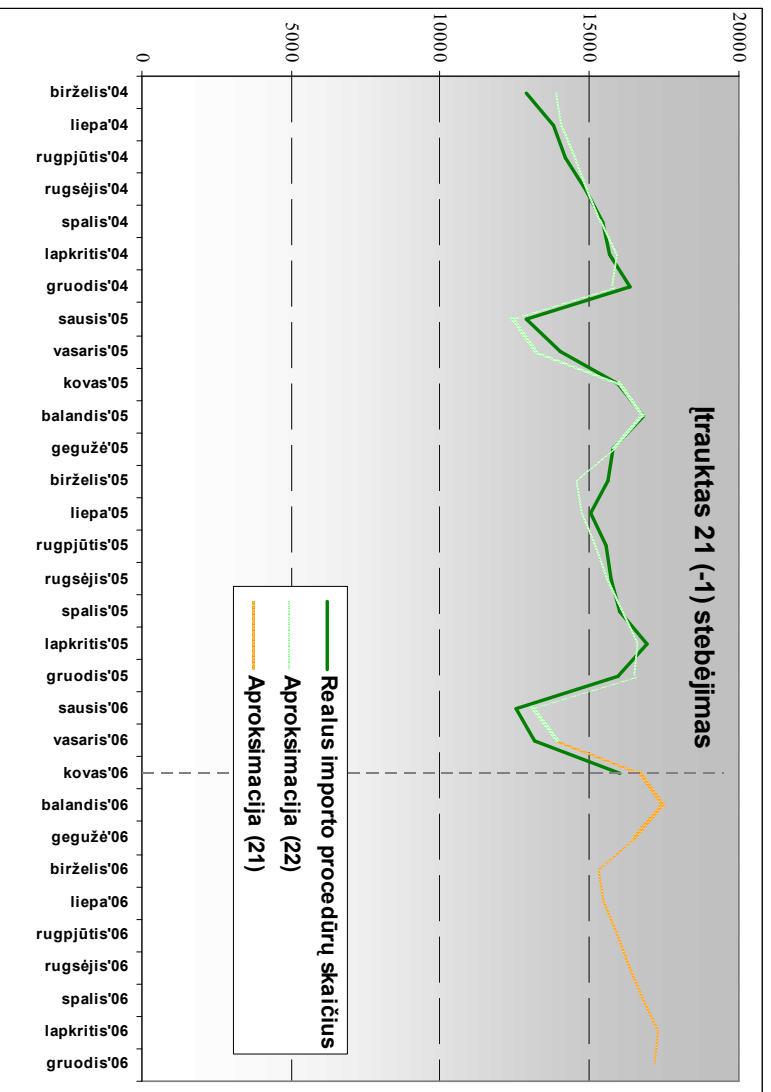
Priedas Nr.2 (tęsinys)



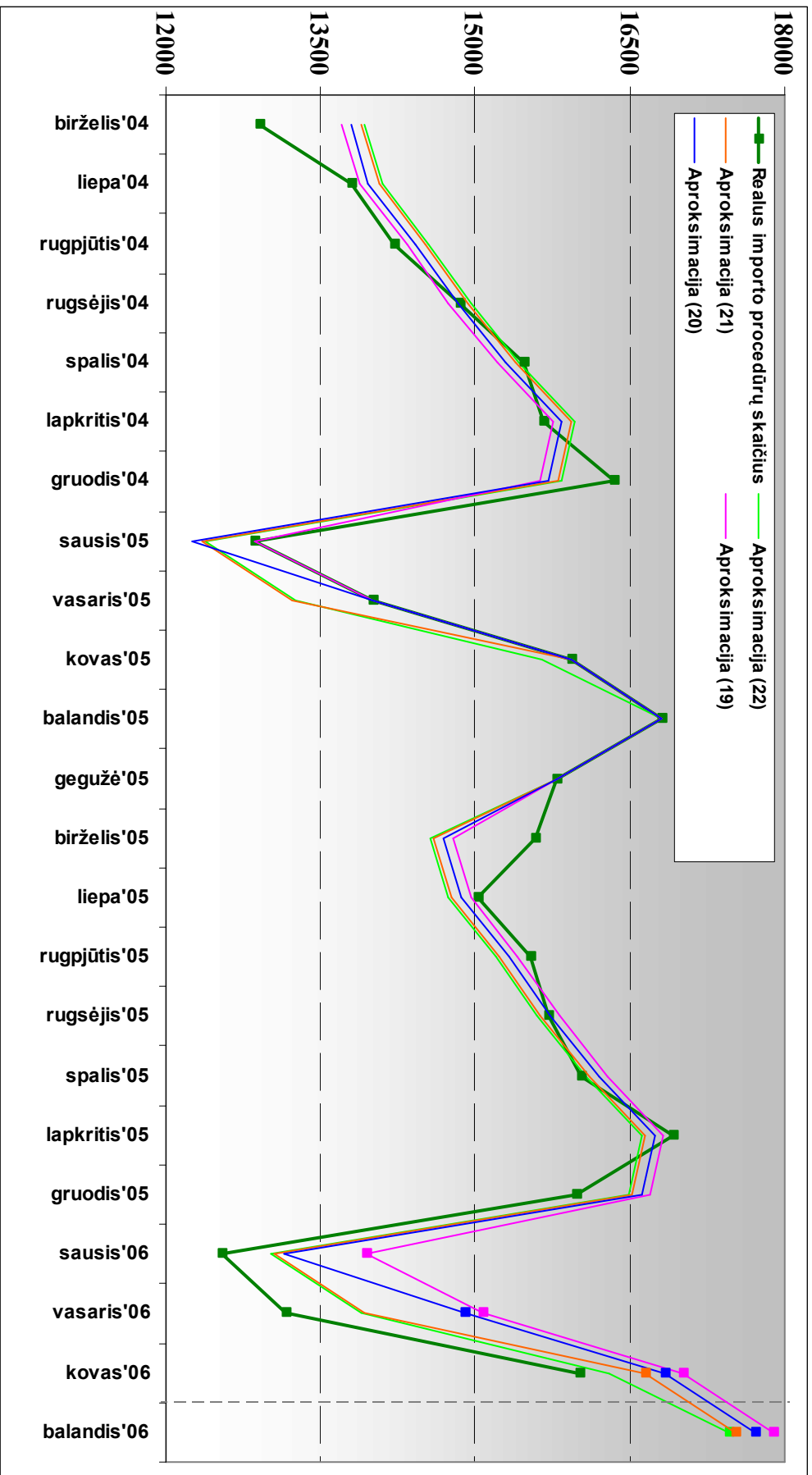




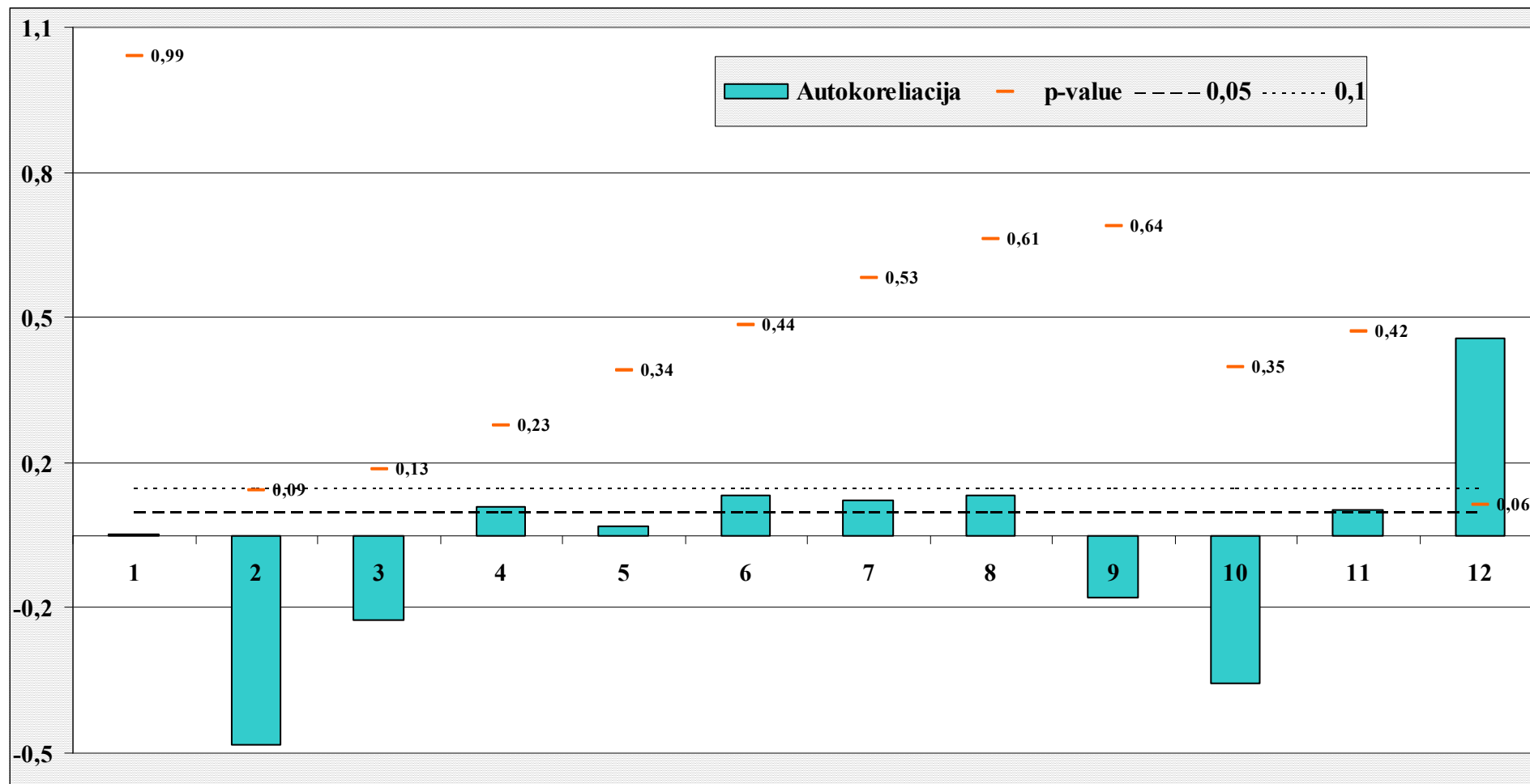




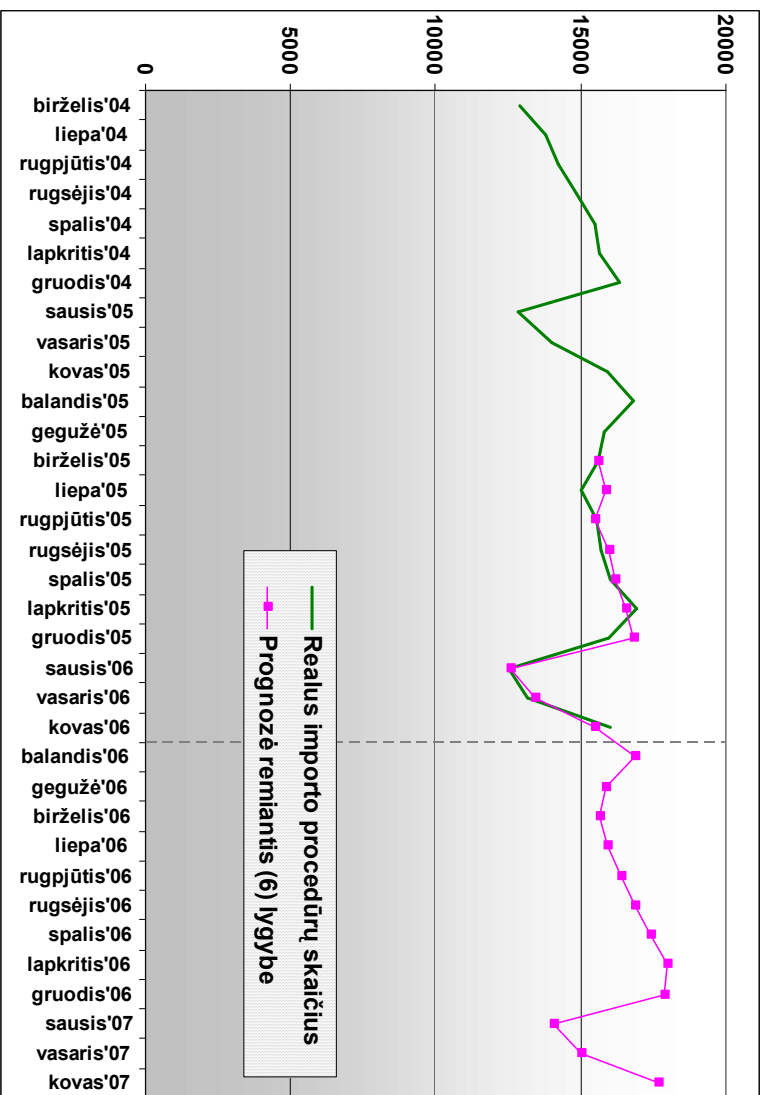
Priedas Nr.4 (tęsinys)



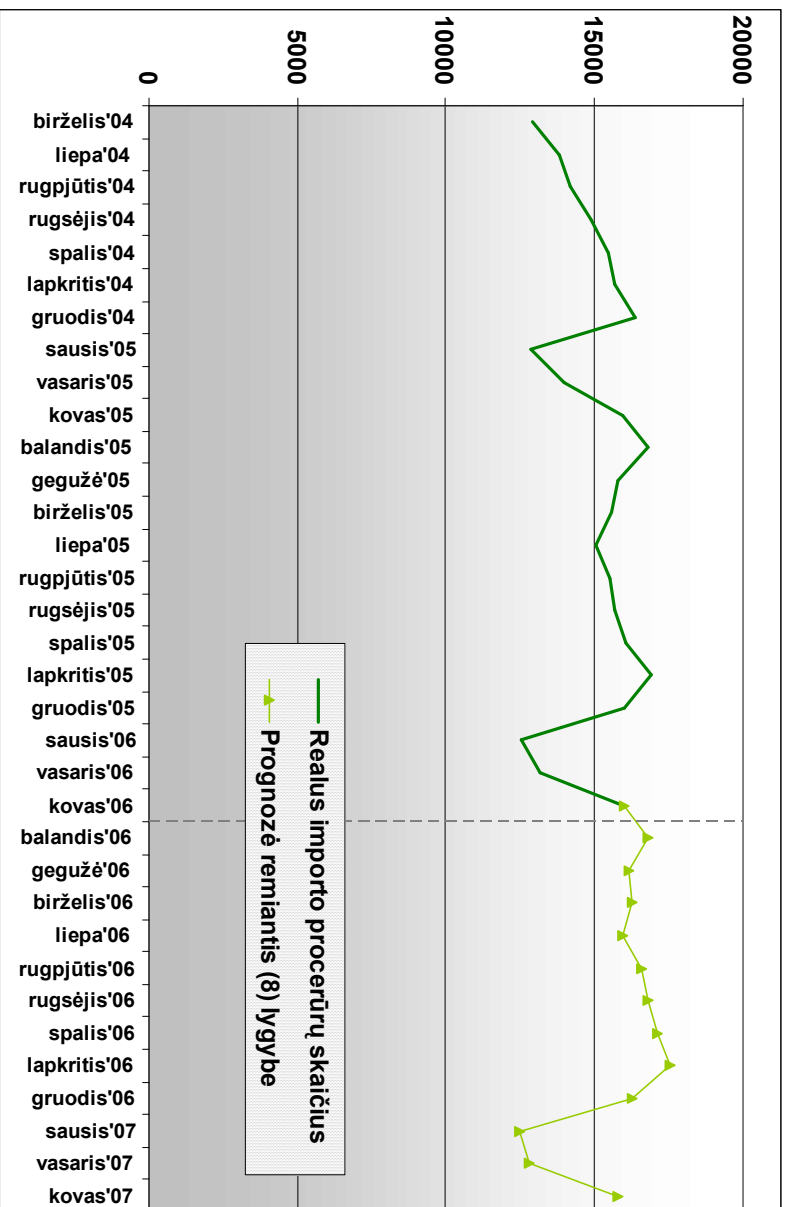
Priedas Nr.5

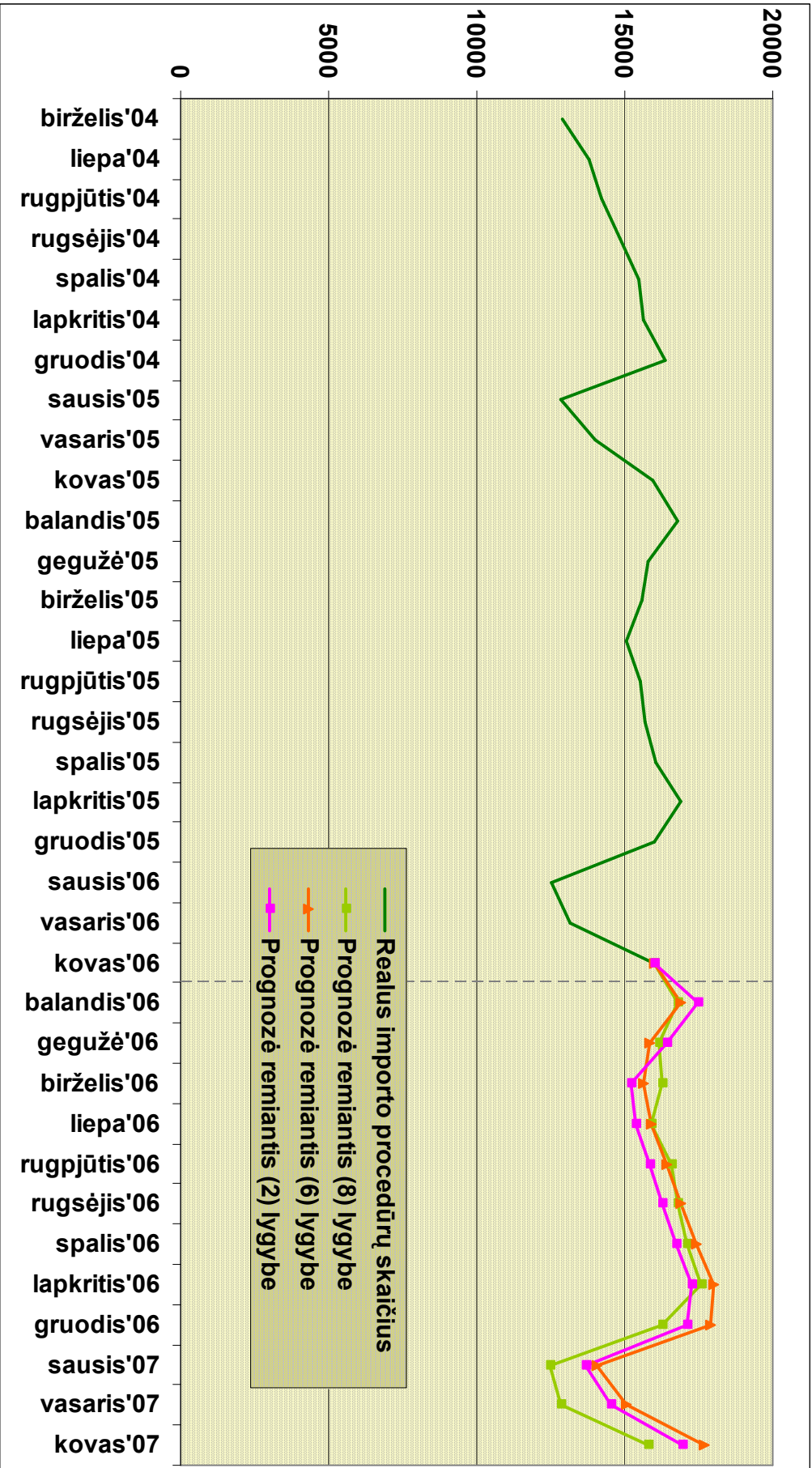


Priedas Nr.6



Priedas Nr.7





Priedas Nr.8

