Vilniaus Universitetas

Matematikos ir informatikos fakultetas

Matematinės statistikos katedra

Valerij Jarmulkovič

Neapskaičiuotų cigarečių ir alkoholio vartojimo įverčiai

Estimating consumption of unrecorded cigarettes and alcohol

Bakalauro baigiamasis darbas

Vilnius, 2016

Darbo vadovas:

doc. Rimantas Eidukevičius

Recenzentas:

prof. Vytautas Kazakevičius

Registracijos Nr.:.....

Darbo gynimo data:.....

Turinys

Įvadas **2**

1 Metodų ir panašių tyrimų apžvalga **3**

1.1 Pagrindinės sąvokos . . . 3

1.2 Taikomi tyrimo metodai . . . 3

1.2.1 Tiesioginiai vertinimo metodai . . . 4

1.2.2 Netiesioginiai vertinimo metodai . . . 4

1.3 Kitų panašių tyrimų aptarimas . . . 5

1.4 ARIMA modelis . . . 6

2 Neapskaičiuoto alkoholio vertinimas **7**

2.1 Turimi pradiniai duomenys neapskaičiuotam alkoholiui vertinti . . . 7

2.2 Duomenų paruošimas skaičiavimams . . . 8

2.3 Netiesioginio vertinimo metodo - ARIMA taikymas . . . 10

2.3.1 Vertinimo etapai ir taikomas modelis . . . 10

2.3.2 Įverčių μ ir β suradimas bei neapskaičiuoto alkoholio įvertinimas . . . 11

2.4 Laiko eilučių simuliacijos modelio tinkamumui vertinti . . . 14

3 Neapskaičiuotų cigarečių vertinimas **17**

3.1 Pradiniai duomenys . . . 17

3.2 Duomenų paruošimas skaičiavimams . . . 18

3.3 ARIMA modelio parinkimas ir neapskaičiuoto tabako vertinimas . . . 19

Išvados **22**

Santrauka **23**

Summary **24**

Literatūros ir šaltinių sąrašas **25**

1 priedas. SAS programos kodai **26**

neapskaičiuoto alkoholio vertinimas . . . 26

neapskaičiuoto tabako vertinimas . . . 31

2 priedas. Output'ai **34**

neapskaičiuoto alkoholio vertinimo output'ai 34

neapskaičiuoto tabako vertinimo output'ai 48

Įvadas

Pasaulio sveikatos organizacijos tyrimų duomenimis Europos regionas yra daugiausiai alkoholio suvartojantis žemynas, o Lietuva viena iš daugiausiai Europoje alkoholio suvartojančių šalių [3]. Sulig tuo auga ir alkoholio vartojimo sąlygojamų ligų bei mirčių skaičius. Lietuvoje mirtingumas nuo kepenų cirozės, atsitiktinio apsinuodijimo alkoholiu ir kai kurių kitų su alkoholio vartojimu susijusių priežasčių yra vienas aukščiausių Europos Sąjungos valstybėse. Tuo tarpu nuo rūkymo sukeltų ligų kasmet pasaulyje miršta daugiau nei 3 mln. gyventojų [2]. Ir nors statistiniai duomenys rodo, kad cigarečių suvartojimas Lietuvoje mažėja, tiek alkoholio, tiek cigarečių sukeliama žala kelia būtinybę nustatyti šių produktų vartojimo mastą ir įvertinti galimą riziką.

Laikomasi nuomonės, kad būtent neapskaičiuotas alkoholis ir cigaretės sukelia didžiausią žalą žmogaus organizmui, todėl, siekiant įvertinti tokio suvartojimo mastą, baigiamajam darbui tikslingai pasirinkta tema – neapskaičiuoto alkoholio ir cigarečių suvartojimo vertinimas.

Darbe naudojami 2001 - 2014 metų legalaus alkoholio suvartojimo ir susirgimų, kuriuos sukelia alkoholis, metiniai, taip pat, 2011 m. sausio - 2016 m. vasario mėnesių duomenys apie mažmeninės tabako gamybos pridėtinės vertės mokestį ir tabako sukeliamą ligotumą. Taikant netiesioginį vertinimo metodą - ARIMA, siekiama parengti neapskaičiuoto alkoholio ir cigarečių apyvartos Lietuvoje įvertinimą.

1 Metodų ir panašių tyrimų apžvalga

1.1 Pagrindinės sąvokos

Apskaičiuotas alkoholis – alkoholiniai gėrimai, parduoti mažmeninėje prekyboje ir maitinimo įmonėse.

Apskaičiuoto alkoholio suvartojimas – mažmeninėje prekyboje parduotų alkoholinių gėrimų kiekis litrais, paverstas į absoliutų (100%) gryną alkoholį, tenkantį vienam 15 metų ir vyresniam gyventojui.

Neapskaičiuotas alkoholis – į neapskaičiuoto alkoholio sampratą įeina namų gamybos alkoholiniai gėrimai, kontrabandiniai alkoholio produktai, nelegalios gamybos alkoholis, turistinis importas, bei surogatinis alkoholis, kuris nėra skirtas gerti (pavyzdžiui, kvepalai, odekolonai, burnos skalavimo skysčiai ir pan.).

Neapskaičiuoto alkoholio suvartojimas – absoliutaus 100% alkoholio kiekis, neįeinantis į apskaičiuoto alkoholio suvartojimą, tenkantis vienam 15 metų ir vyresniam gyventojui.

Sergamumas – naujai nustatytų ligų atvejų skaičius, tenkantis 100 tūkst. gyventojų.

Ligotumas – visų žinomų ligos atvejų bendras skaičius, tenkantis 100 tūkst. gyventojų, ataskaitinių metų pabaigoje [1].

Mokėtinas PVM – pridėtinės vertės mokestis įskaičiuotas į prekės kainą, kurį pirkėjas turi sumokėti pardavėjui (standartinis tarifas - 21%).

1.2 Taikomi tyrimo metodai

Gardino valstybinio medicinos universiteto profesorius dr. Yury Razvodovsky išsamiai aprašė neapskaičiuoto alkoholio suvartojimo tyrimo metodus. Profesorius pažymi, kad egzistuoja daug vertinimo metodų, tačiau kiekvienas jų turi savo privalumų ir trūkumų. Apytikslį neapskaičiuotą vartojimą galima nustatyti vadovaujantis policijos ir muitinės duomenimis apie konfiskuotą namų gamybos alkoholį, suklastotą alkoholį, alkoholio kontrabandą, keliautojų importą, taip pat duomenimis apie pramonės ir medicinos tikslais naudojamą alkoholį ir kt. Kaip nurodo Y. Razvodovsky, neapskaičiuoto alkoholio vertinimo metodai gali būti išskiriami į tiesioginius ir netiesioginius [7].

1.2.1 Tiesioginiai vertinimo metodai

Tiesioginiai metodai apima gyventojų apklausas, kurios reprezentuoja visą tiriamą regioną. Tokiu atveju rengiami įvairiausi klausimynai ir tyrimo technikos, apklausiami skirtingo amžiaus ir lyties regiono gyventojai. Kaip nurodo profesorius, pats paprasčiausias ir dažniausiai naudojamas metodas įvertinti apskaičiuotą ir neapskaičiuotą suvartojimą, įvertinus skirtumą tarp bendro alkoholio suvartojimo lygio ir apskaičiuoto alkoholio pardavimų – tai gyventojų apklausa dėl suvartojamo alkoholio kiekio ir dažnumo. Profesorius savo straipsnyje mini tyrimo būdą, kurio metu respondentams yra siūlomas alkoholis didelėmis ir mažėjančiomis dozėmis. Lyginant su alkoholio pardavimų statistika, šio metodo pranašumas tas, kad jis atskleidžia alkoholio vartojimo modelį, t. y. alkoholio vartojimo dažnumą, suvartotą kiekį ir perdėto alkoholio vartojimo dažnumą. Kita vertus, respondentai linkę nuslėpti savo alkoholio suvartojimo kiekį. Dar vienas profesoriaus minimas metodo trūkumas – gauti rezultatai nėra pakankamai reprezentatyvūs, kadangi panašiuose tyrimuose retai dalyvauja iš tiesų daug geriantys žmonės, kurie linkę vartoti neapskaičiuotą alkoholį. Nepaisant metodo trūkumų, tiesioginiai metodai (tokie kaip gyventojų apklausos), profesoriaus nuomone, yra vienas tiksliausių būdų nustatyti neapskaičiuoto suvartojimo mąstą šalyje. Pažymėtina, kad tiesioginiu metodu atliktų tyrimų rezultatai gali būti puikus būdas patikrinti gautus neapskaičiuoto alkoholio suvartojimo duomenis netiesioginiais metodais.

1.2.2 Netiesioginiai vertinimo metodai

Skandinavijos tyrėjas, Stokholmo universiteto Socialinių tyrimų instituto profesorius Thor Norström pristatė tyrimą, kurio tikslas buvo įvertinti neapskaičiuoto alkoholio daromos žalos vienam gyventojui pokyčius. Neapskaičiuoto alkoholio vartojimas buvo vertinamas per neatitikimus tarp stebėtų alkoholio daromos žalos rodiklių pokyčių ir pokyčių, kurių galima tikėtis iš apskaičiuoto alkoholio suvartojimo. Profesorius pažymi, kad didelė panašių tyrimų dalis remiasi alkoholio pardavimų ir įvairių alkoholio daromos žalos rodiklių santykiu. Įvertinti šį santyki dažniausiai naudojamas toliau pateikiamas netiesioginis vertinimo modelis [5]:

$$I_t = \beta A_t + N_t;$$

Kur I_t – dažnai logoritmuojamas žalos indikatorius - mirtingumas nuo kepenų cirozės laiko momentu t,A_t – apskaičiuoto absoliutaus (100%) alkoholio pardavimai, tenkantys vienam 15 metų ir vyresniam gyventojui litrais, laiko momentu t, β – vertinamas alkoholio efekto dydžio parametras žalos indikatoriui I,N_t – triukšmo terminas į kurį įeina visi kiti

faktoriai, išskyrus apskaičiuoto alkoholio pardavimus, kurie turi įtakos mirtingumui nuo kepenų cirozės.

Kitaip tariant, apskaičiuojamas skirtumas tarp stebėtų žalos rodiklio pokyčių ir tikėtinų alkoholio pardavimų pokyčių turėtų išreikšti neapskaičiuoto vartojimo pokyčius. Tačiau profesorius pažymi, kad egzistuoja nuomonių, jog nėra jokio ryšio tarp apskaičiuoto ir neapskaičiuoto suvartojimo, ir tokiu būdu egzistuoja tam tikras poslinkis apskaičiuojant alkoholio poveikio ir triukšmo kriterijus. Profesorius pripažįsta, kad sunku įvertinti kiek patikima yra pastaroji prielaida ir kaip galimi nukrypimai gali paveikti tyrimo rezultatus.

Žalos rodikliai gali varijuoti priklausomai nuo bendro alkoholio suvartojimo. Kaip nurodo profesorius, kai kurie rodikliai, kaip pavyzdžiui apsinuodijimas alkoholiu, praktiškai priklauso vien nuo neapskaičiuoto vartojimo, tačiau didžiajai daliai rodiklių įtakos turi ir kiti veiksniai. Todėl tampa svarbu tinkamai pasirinkti žalos rodiklius. Profesoriaus atliktame tyrime rodiklių pasirinkimą nulėmė dvi aplinkybės. Visų pirma, šis rodiklis turi būti glaudžiai susijęs su vieno vidutinio gyventojo suvartojimu, antra, šis rodiklis neturi turėti jokios įtakos kitiems, aiškias tendencijas turintiems, veiksniams.

Pažymėtina, kad T. Norström savo skaičiavimuose remiasi ARIMA (angl. autoregressive integrated moving average) prognozavimo metodu, kuris leidžia, lyginant apskaičiuoto alkoholio suvartojimą ir alkoholio vartojimo pasekmes, apytiksliai nustatyti, kokios būtų pasekmės, jeigu tikrasis alkoholio suvartojimas būtų toks, koks yra oficialiai apskaičiuoto alkoholio suvartojimas. T. Norström nuomone, geriausiai neapskaičiuoto alkoholio suvartojimą atskleidžia bendras alkoholio vartojimo sąlygojamo mirtingumo rodiklis.

Kaip galima matyti, neapskaičiuoto suvartojimo paplitimo vertinimas atrodo pakankamai opi problema. Atsižvelgiant į pateiktą informaciją, galima daryti išvadą, kad vertinant neapskaičiuotą suvartojimą, skirtingų tyrėjų, skirtingose šalyse naudojami šaltiniai ir tyrimo metodai gali skirtis ir vieno teisingo metodo nėra [5].

1.3 Kitų panašių tyrimų aptarimas

2007 metais tyrėjai Svetlana Popova (Kanada), Jurgen Rehm (Kanada, Šveicarija), Jayadeep Patra (Kanada) ir Witold Zatonski (Lenkija) pristatė tyrimą, kurio metu lygintas Centrinės ir Rytų Europos šalių alkoholio suvartojimas su kitų Europos šalių suvartojimu. Kaip pažymėjo tyrėjai, Centrinė ir Rytų Europa yra daugiausiai alkoholio suvartojantis regionas. Svarbu atkreipti dėmesį, kad vadovaujantis tyrimo duomenimis, Lietuva buvo trečia daugiausiai apskaičiuoto alkoholio suvartojanti šalis šiame regione [6]. Tyrimas parodė, kad

Europos šalyse aukštesnis ir neapskaičiuoto alkoholio suvartojimas. Įvertinus turimus duomenis buvo nustatyta, kad Lietuva patenka į šalių grupę, kurioje neapskaičiuoto alkoholio suvartojimas yra aukštas. Kaip nurodo tyrėjai, aukštas neapskaičiuoto alkoholio suvartojimas atsispindi atsižvelgus į augantį su alkoholio vartojimu siejamų ligų dažnį. Kaip galima matyti, tyrimo metu vadovautasi prielaida, kad neapskaičiuotas alkoholis yra prastesnės kokybės, todėl sukelia didesnį pavojų sveikatai.

1.4 ARIMA modelis

Apibrėžimas. Tarkime, kad turime tam tikro kintamojo Y reikšmių stebėjimus, kuriuos gauname laiko momentais t=1,2,...T. Tada eilutę - $Y_1, Y_2, ..., Y_T$, vadiname laiko eilute.

Sakysime, kad laiko eilutė $Y_1, Y_2, ..., Y_T$ yra ARMA(p,q), jei Y_t yra stacionari ir:

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_n Y_{t-n} + \theta_t + \theta_1 \varepsilon_{t-1} + ... + \theta_m \varepsilon_{t-m},$$

čia $\phi_n \neq 0$, $\theta_m \neq 0$ ir $\sigma^2 > 0$. Parametrus p ir q vadinsime atitinkamai autoregresijos ir slenkančio vidurkio, o baltuoju triukšmu vadinsime stacionarų, kvadratu integruojamą procesą $(\varepsilon_t, t \in Z)$, kurio vidurkis bus lygus nuliui ir dispersija lygi σ^2.

Jei slenkančio vidurkio parametras $q = 0$, tai tuomet turimas toks autoregresijos procesas AR(p):

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_n Y_{t-n} + \varepsilon_t = \sum_{i=1}^{n} \phi_i Y_{t-i} + \varepsilon_t$$

Jei autoregresijos parametras $p = 0$, tai tuomet turimas slenkančio vidurkio procesas MA(q):

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_m \varepsilon_{t-m} = \sum_{i=1}^{m} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

kur $\phi_1, ...\phi_n$ ir $\theta_1, ..., \theta_m$ yra konstantos ir ε_t - anksčiau apibrėžtas baltasis triukšmas.

Įtraukus į ARMA modelį skirtumų panaudojimą, gauname ARIMA(p,d,q) procesą, kur d yra integruojamumo parametras. Pažymime $\triangle Y_t = Y_t - Y_{t-1}$, $\triangle^2 Y_t = \triangle(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2}, ..., \triangle^d Y_t = \triangle(\triangle^{d-1} Y_t)$. Gautas Y_t procesas ir yra ARIMA(p,d,q) [9].

2 Neapskaičiuoto alkoholio vertinimas

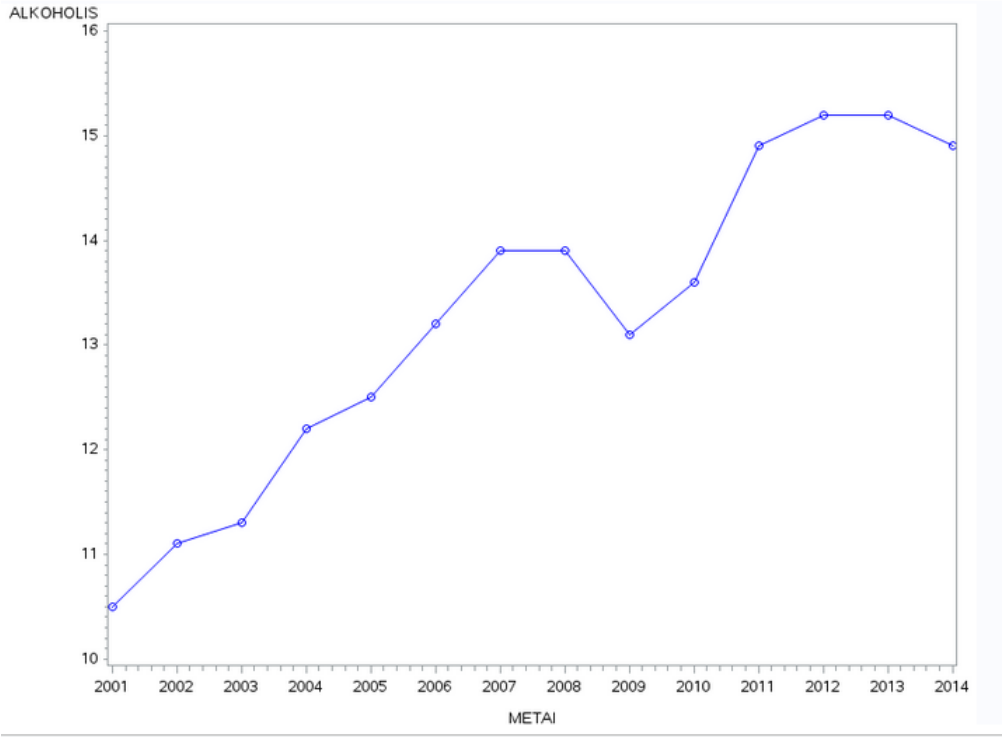
Tolimesni skaičiavimai, naudojantis programa SAS, bus atliekami šiais žingsniais:

- duomenų paruošimas skaičiavimams;
- laiko eilutės vertinimas ir parametrų μ ir β suradimas;
- neapskaičiuoto alkoholio kiekio įvertinimas;
- laiko eilučių simuliavimas, modelio tinkamumui įvertinti;
- neapskaičiuoto alkoholio vertinimas iš simuliuotų duomenų, laikant žinomą tik apskaičiuotą alkoholį.

2.1 Turimi pradiniai duomenys neapskaičiuotam alkoholiui vertinti

Šiame skyriuje trumpai aprašysime turimus duomenis, naudojamus tolimesniuose skaičiavimuose. Turima apskaičiuoto absoliutaus (100 %) alkoholio, tenkančio vienam, 15 bei vyresniam amžiaus gyventojui, metinė laiko eilutė. Apskaičiuoto alkoholio duomenys buvo gauti iš oficialios statistikos portalo. Lietuvos statistikos departamento pateikiami legaliai suvartojamo alkoholio duomenys yra skelbiami tik nuo 2000 metų. Dėl šios priežasties, turėsime gan trumpą laiko eilutę.

1 pav. *Apskaičiuotas alkoholis, tenkantis vienam 15 metų ir vyresniam asmeniui, litrais.*



Taip pat neapskaičiuoto alkoholio vertinime naudosime alkoholio sukeliamų ligų duomenis, gautus iš Higienos instituto. Atrinktos ligos, kurioms didžiausią įtaką daro suvartojamas alkoholis. Norint apibendrinti turimus duomenis, jiems bus pritaikytas pagrindinių komponenčių metodas ir gautas naujas kintamasis, tolimesniuose skaičiavimuose bus naudojamas kaip bendras visų susirgimų kintamasis. Stebima susirgimų laiko eilutė prasideda nuo 2000 metų, tačiau, norint pateikti konkretesnius rezultatus, duomenys yra pertvarkomi ir pasinaudojus Lietuvos statistikos departamento duomenimis apie Lietuvos gyventojų skaičių, suskaičiuojame kiek susirgimų tenka 100 tūkst. gyventojų. Nuolatinių gyventojų skaičius Lietuvoje skelbiamas tik nuo 2001 metų, todėl mūsų turimi duomenys sumažėjo dar vieneriais metais.

Pradiniai susirgimų duomenys buvo suskirstyti į tokias grupes:

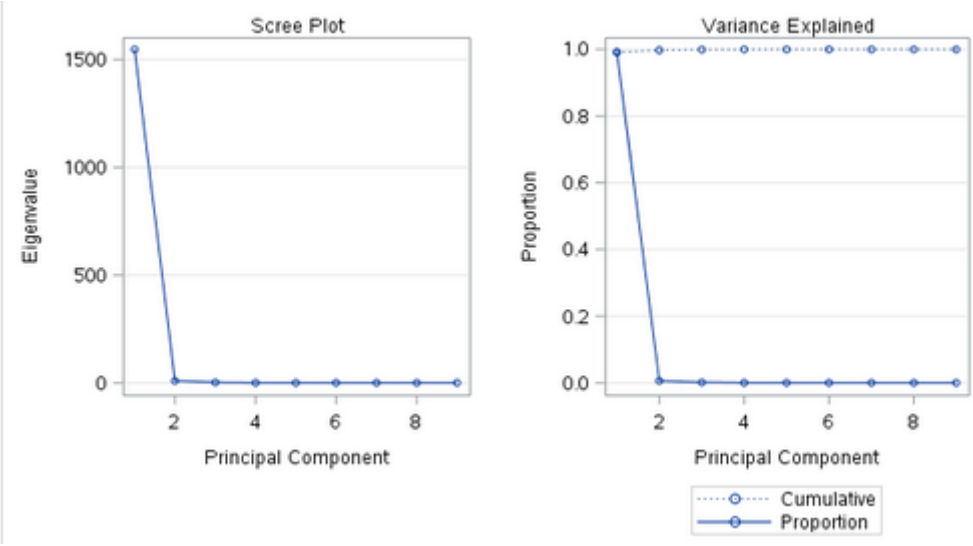
- X45 - Atsitiktinis apsinuodijimas alkoholiu bei jo poveikis;
- X65 - Tyčinis apsinuodijimas ir apnuodijimas alkoholiu;
- Y15 - Apsinuodijimas alkoholiu bei jo poveikis, kai ketinimas nežinomas;
- F10 - Psichikos ir elgesio sutrikimai dėl alkoholio vartojimo;
- G31.2 - Nervų sistemos degeneracija, sukelta alkoholio;
- G62.1 - Alkoholio sukelta polineuropatija;
- I42.6 - Alkoholinė kardiomiopatija;
- K70 - Alkoholinė kepenų liga;
- K86.0 - Lėtinis alkoholinis pankreatitas.

2.2 Duomenų paruošimas skaičiavimams

Turimi susirgimų duomenys yra to pačio matavimo, todėl mums nereikės centruoti ir normuoti turimus duomenis. Pagrindinę komponentę skaičiuosime iš kovariacijų matricos, t.y. naudojant COV pasirinkimą, kintamieji su didelėmis dispersijom bus labiau susiję su komponentėm su didelėmis tikrinėmis reikšmėmis, o kintamieji su mažomis dispersijomis, bus labiau susiję su komponentėmis su mažomis tikrinėmis reikšmėmis.

Pritaikę procedūrą, gauname, kad pirmoji pagrindinė komponentė paaiškina 99% visos dispersijos. Todėl matome, kad pirmosios komponentės tikrinės reikšmės, suteikia gerą duomenų apibendrinimą.

2 pav. *Kovariacijų matricos tikrinės reikšmės (kairėje) ir dispersijų santykis (dešinėje)*



Iš tikrinių vektorių matricos (2 lent.) apskaičiuojame pirmąją pagrindinę komponentę:

3 pav. *Tikriniai vektoriai*

Eigenvectors									
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
F10	0.993553	-.101141	-.046815	-.010196	-.016662	0.006504	-.000891	0.002307	-.001430
G31	0.051817	0.074981	0.624982	0.196495	0.745080	-.020364	0.082910	-.003535	0.005884
G62	0.039974	0.161887	0.658831	0.325652	-.656592	-.018575	-.014594	0.006718	0.019262
I42	-.001014	0.011104	-.084915	0.065541	-.059161	-.201504	0.949587	0.107144	0.176265
K70	0.091753	0.976751	-.180491	-.022622	0.054602	-.018156	-.022111	-.024156	-.008174
K86	-.011781	-.051735	-.364990	0.921967	0.078687	0.007416	-.084810	-.018858	-.012214
X45	-.003295	0.023793	0.009122	0.016656	-.012832	0.965856	0.228871	-.066067	-.096296
X65	0.000640	0.010986	-.007534	-.003188	0.023293	0.153074	-.171513	0.285881	0.929889
Y15	-.000711	0.020297	-.001576	0.010911	0.009182	0.043583	-.041424	0.949434	-.307149

$$\begin{aligned}\text{Prin1} = {} & 0.993553 \times (\text{F10}) \\ & + 0.051817 \times (\text{G31}) \\ & + 0.039974 \times (\text{G62}) \\ & \dots \\ & \dots \\ & + (\text{-}0.00711) \times (\text{Y15})\end{aligned}$$

Apskaičiave pagrindinę komponentę ir duomenis įrašę į lentelę turime tokius sutvarkytus duomenis, kuriuos ir naudosime neapskaičiuoto alkoholio vertinimui:

4 pav. *Sutvarkyti duomenys, paruošti neapskaičiuoto alkoholio įverčiams skaičiuoti.*

METAI	ALKOHOLIS	susirg
2001	10.5	179.3
2002	11.1	164.09
2003	11.3	175.05
2004	12.2	191.63
2005	12.5	237.8
2006	13.2	239.33
2007	13.9	290.66
2008	13.9	263.01
2009	13.1	230.33
2010	13.6	242.73
2011	14.9	270.06
2012	15.2	256.22
2013	15.2	264.04
2014	14.9	245.81

2.3 Netiesioginio vertinimo metodo - ARIMA taikymas

2.3.1 Vertinimo etapai ir taikomas modelis

Laiko eilučių vertinimas naudojant netiesioginį vertinimo metodą ARIMA paprastai atliekamas tokiais etapais [10]:

1. **Atpažinimas** (angl. *Identification*). Šiame etape, remiantis duomenų grafikais, autokoreliacijomis, dalinėmis autokoreliacijomis bei kita informacija, parenkamas preliminarus ARIMA modelis.

2. **Įvertinimas** (angl. *Estimation*). Parinktiems ARIMA p , d ir q parametrams įvertinami θ ir ϕ, SAS'e pagal nutylėjimą naudojant sąlyginį mažiausių kvadratų vertinimo metodą (angl. Conditional least squares estimation).

3. Galutinis modelio **apibendrinimas** (angl. *Diagnostic cheking*). Iš liekanų autokoreliacijų grafiko tikrinama ar prieš tai parinktas modelis neturi trūkumų ir galutinai pataisomas modelis.

4. **Prognozavimas** (angl. *Forecast*). Prognozuojamas pasirinktas skaičius reikšmių į priekį (SAS programoje nurodžius lead = 0, prognozuojamos reikšmės tik turimiems stebėjimams).

Turimiems duomenims, skaičiavimuose bus naudojamas toks modelis:

$$S_t = (A_t N_t)^\beta \mu \varepsilon_t$$

kur S_t - susirgimai alkoholio sukeliamomis ligomis, tenkantys 100 tūkst. gyventojų, A_t - legalaus (apskaičiuoto) absoliutaus (100%) alkoholio pardavimai, tenkantys vienam 15 m. ir vyresniam gyventojui, litrais, μ - pastovus vidurkio parametras, β - parametras atspindintis alkoholio įtaka sukeltiems susirgimams ir ε_t - baltas triukšmas (angl. *White noise*).

Čia N_t yra ARMA procesas, kuris atspindi ne tik apskaičiuoto alkoholio pardavimų poveikį daromą susirgimams. Todėl, galime manyti, kad būtent šis parametras ir bus mūsų ieškomas neapskaičiuotas alkoholis, jei laikysime, kad alkoholis yra vienintėlis faktorius, darantis įtaką turimiems susirgimams.

Logoritmavus lygtį, turime tokį tiesinės regresijos su autoregresinėmis paklaidomis modelį [8]:

$$\ln S_t = \ln \mu + \beta \ln A_t + \beta \ln N_t + \ln \varepsilon_t$$

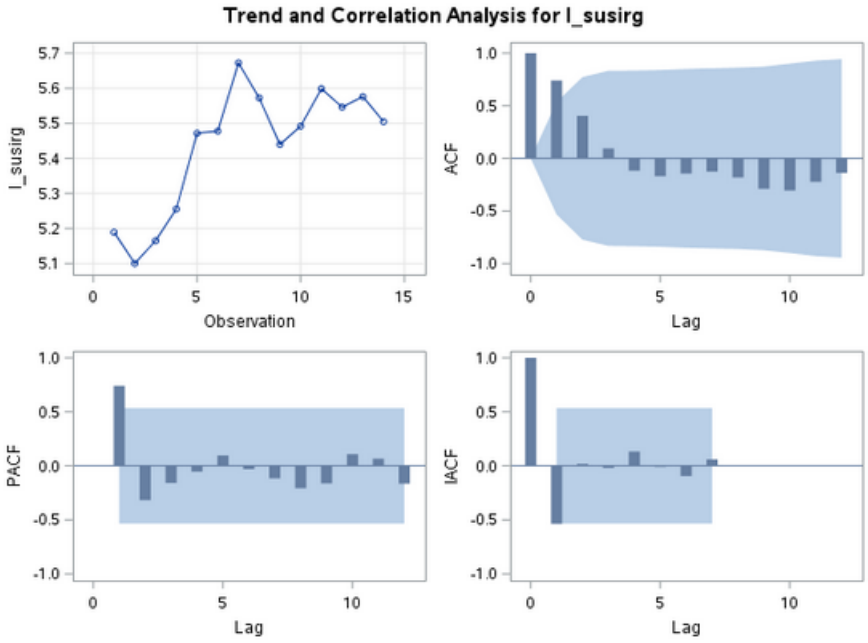
Išsireiškę ieškomą parametrą, gauname:

$$N_t = \exp\left(\tfrac{1}{\beta}(\ln S_t - \ln \mu - \beta \ln A_t - \ln \varepsilon_t)\right)$$

2.3.2 Įverčių μ ir β suradimas bei neapskaičiuoto alkoholio įvertinimas

Pirmiausiai, norint surasti μ ir β įverčius, reikia įvertinti ARIMA(p,d,q) modelio eilę.

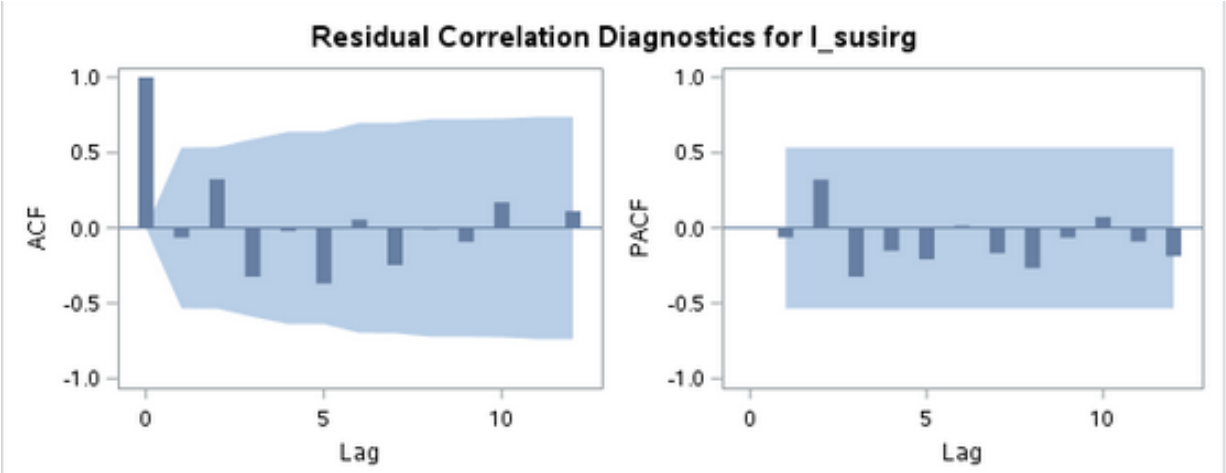
5 pav. *Logoritmuotų susirgimų laiko eilutės, autokoreliacijų, dalinių autokoreliacijų bei atvirkštinių autokoreliacijų įverčių grafikai.*



Iš logoritmuotų susirgimų grafiko nematyti ilgalaikio trendo ar sezoniškumo, todėl vertinimui naudosime ARIMA(p,0,q) modelį. Iš autokoreliacijų ir dalinių autokoreliacijų grafikų nustatysime apytiksles ARIMA modelio, autoregresijos parametro p ir slenkančio vidurkio q reikšmes. Iš autokoreliacijų grafiko matome, kad tiriama laiko eilutė yra stacionari, nes po trečios eilės koeficientai staigiai krenta ir svyruoja aplink nulį. Iš dalinių autokoreliacijų grafiko aiškiai matome, kad turimas pirmos eilės autoregresijos procesas: antros ir aukštesnių eilių koeficientai yra nereikšmingi, t.y. beveik lygūs nuliui. Taigi, vertinsime pagal ARIMA(1,0,0) modelį.

Iš liekanų autokoreliacijų ir dalinių autokoreliacijų grafikų nustatome modelio tinkamumą. Matome, kad visų eilių, pradedant pirmąja, reikšmės pasiskirsto apie nulį. Tai mums leidžia daryti išvadas, kad laiko eilutei parinktas tinkamas modelis. Pasinaudoję prognozavimo funkcija, gauname prognozuojamas reikšmes bei liekanas, kurias įrašome į lentelę.

6 pav. *Liekanų autokoreliacijų ir dalinių autokoreliacijų grafikai*



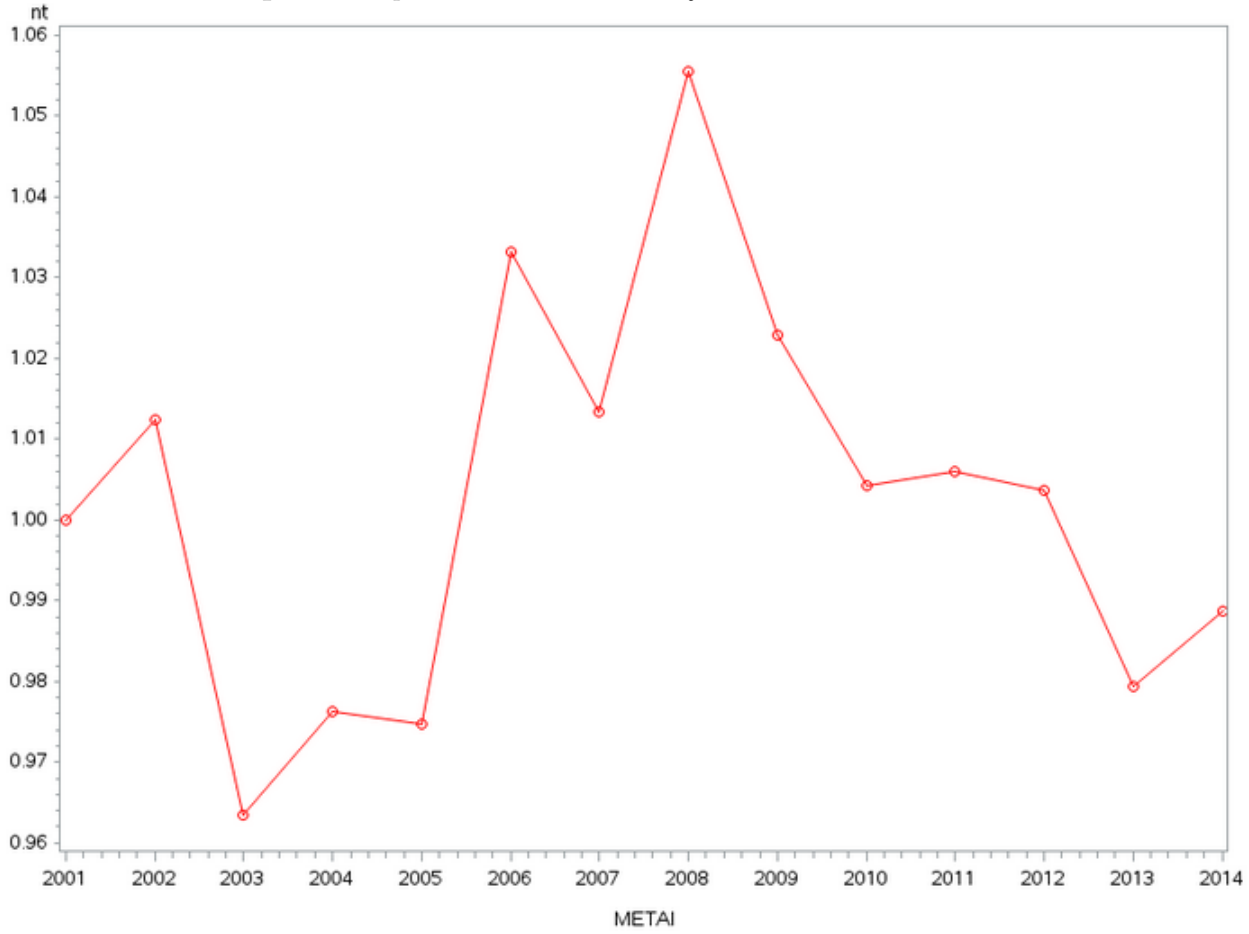
Taip pat, procedūra ARIMA atspausdina ieškomus μ ir β parametrus.

7 pav. *μ β ir pirmos eilės autoregresijos koeficiento įverčiai*

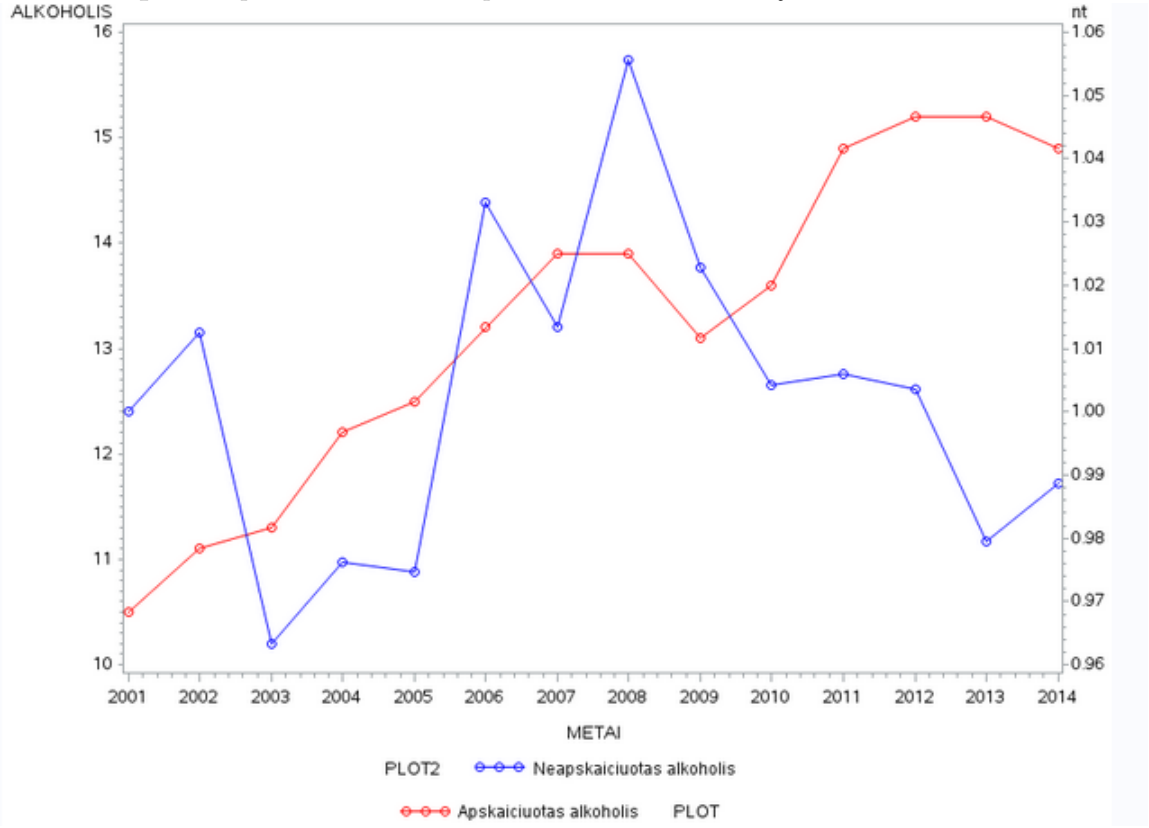
Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	2.20987	0.66657	3.32	0.0069	0	l_susirg	0
AR1,1	0.39314	0.29522	1.33	0.2099	1	l_susirg	0
NUM1	1.25023	0.25949	4.82	0.0005	0	l_alkoholis	0

Turimi visi parametrai, reikalingi neapskaičiuotam alkoholiui vertinti. Įsistatę į formulę turimas reikšmes, gauname rezultatus, kuriuos pavaizduojame grafikuose.

8 pav. *Neapskaičiuoto alkoholio įverčiai*



9 pav. *Apskaičiuoto ir neapskaičiuoto alkoholio įverčiai*



Gauti neapskaičiuoto alkoholio įverčiai pavaizduoti lentelėje. Kartu vartotojams pateikiami ir pradiniai rodikliai.

10 pav. *Apskaičiuoto bei neapskaičiuoto alkoholio kaita ir pagrindinių komponenčių metodu paskaičiuotas kintamasis - susirgimai sukelti alkoholio*

METAI	ALKOHOLIS	susirg	nt
2001	10.5	179.3	1.0000000126
2002	11.1	164.09	1.0124794835
2003	11.3	175.05	0.9633688653
2004	12.2	191.63	0.9762778947
2005	12.5	237.8	0.9746478593
2006	13.2	239.33	1.0331880997
2007	13.9	290.66	1.0133328012
2008	13.9	263.01	1.0555183103
2009	13.1	230.33	1.0228559663
2010	13.6	242.73	1.0041900759
2011	14.9	270.06	1.0059619027
2012	15.2	256.22	1.0036110142
2013	15.2	264.04	0.9794393925
2014	14.9	245.81	0.9887426454

2.4 Laiko eilučių simuliacijos modelio tinkamumui vertinti

Norint įvertinti modelio tinkamumą, simuliuosime apskaičiuoto ir neapskaičiuoto alkoholio laiko eilutes. Kaip minėjo profesorius Thor Norström, sunku įvertinti ryšį tarp neapskaičiuoto ir apskaičiuoto alkoholio. Viena vertus, galima laikytis nuomonės, kad bendras alkoholio poreikio padidėjimas nulems ir apskaičiuoto bei neapskaičiuoto alkoholio vartojimo padidėjimą, kas reiškia teigiamą koreliaciją. Kita vertus, taip pat įtikinama atrodo tai, kad apskaičiuoto alkoholio suvartojimo mažėjimas (pavyzdžiui, dėl išaugusių alkoholio kainų) galėtų būti kompensuojamas neapskaičiuoto alkoholio vartojimo padidėjimu, kas, profesoriaus nuomone, parodytų neigiamą koreliaciją [5]. Taigi, apibendrinant profesoriaus įžvalgas, galima daryti išvadą, kad nustatyti tikslų šių dviejų priešingų veiksnių poveikį yra labai sunku.

Todėl, buvo bandoma simuliuoti laiko eilutes su įvairiomis tarpusavio priklausomybėmis. Pateiksime simuliacijų pavyzdį su neigiama koreliacija tarp legalaus ir neapskaičiuoto alkoholio.

Simuliuojame tokias laiko eilutes:

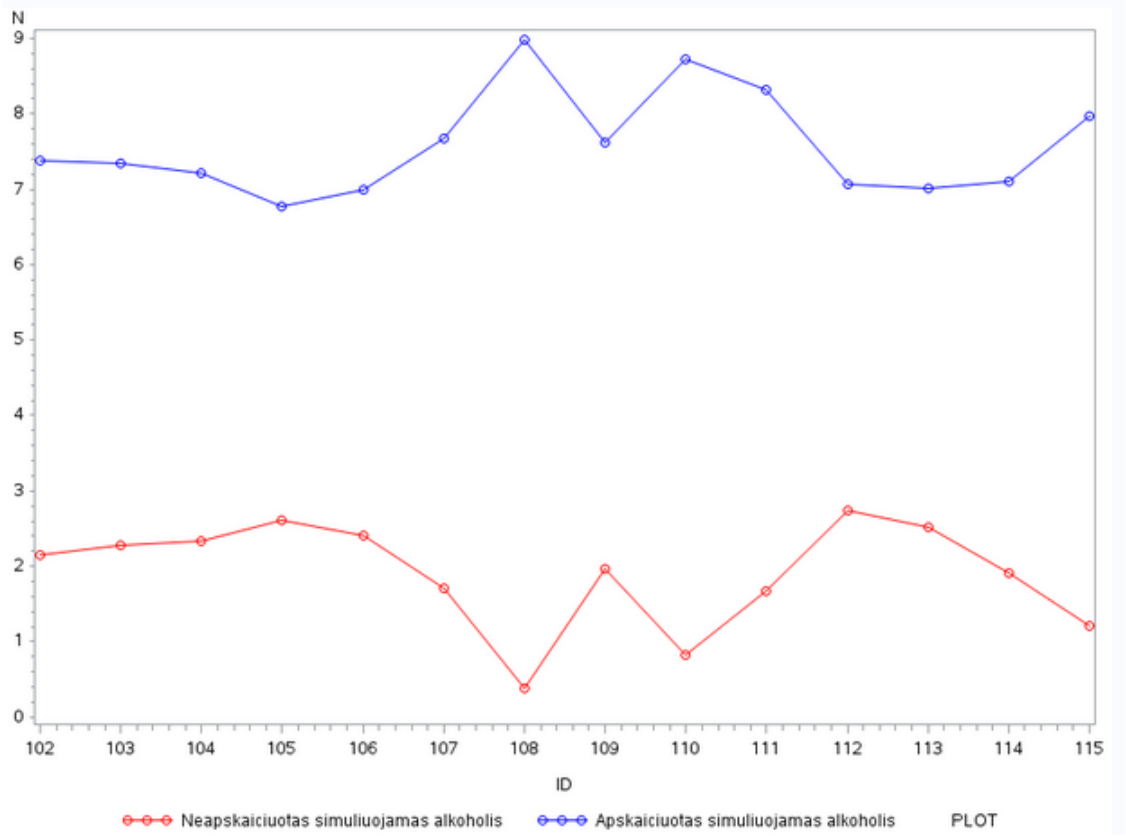
$$A_t = \mu_A + \alpha_t^A A_{t-1} - \gamma_t^A N_{t-1} A_{t-1}$$

$$N_t = \mu_N + \alpha_t^N N_{t-1} - \gamma_t^N A_{t-1} N_{t-1}$$

kur μ_A, μ_N, α_t^A,α_t^N ir γ_t^A, γ_t^N - yra pasirenkamos konstantos.

Generuojame tokias paklaidas, kad tarp jų būtų koreliacija su pasirinktu koreliacijos koeficientu ρ. Pasirinke neigiamą koeficientą netoli vieneto, gauname tokias simuliuojamas laiko eilutes:

11 pav. *Apskaičiuoto bei neapskaičiuoto simuliuojamų alkoholių grafikas*



Pirmiausiai, turimiems simuliuotiems duomenims pagal žemiau pateiktą formulę apskaičiuosime susirgimų kintamąjį:

$$\ln I_t = \ln \mu + \beta \ln A_t + \beta \ln N_t + \varepsilon_t$$

kur konstantas μ ir β parenkame panašias į tokias, kaip gavome vertinant nelegalų alkoholį iš realių duomenų. Tuomet laikydami kaip realiame gyvenime, kad turimi tik apskaičiuoto alkoholio duomenys, remdamiesi netiesioginio ARIMA vertinimo metodu, vertinsime neapskaičiuoto alkoholio suvartojimą, kuris turėtų būti panašus į simuliuojama neapskaičiuoto alkoholio laiko eilutę.

Gaunami tokie rezultatai:

12 pav. *Simuliuojamu ir neapskaiciuoto alkoholio gauto is simuliuojamu duomenu grafikas*



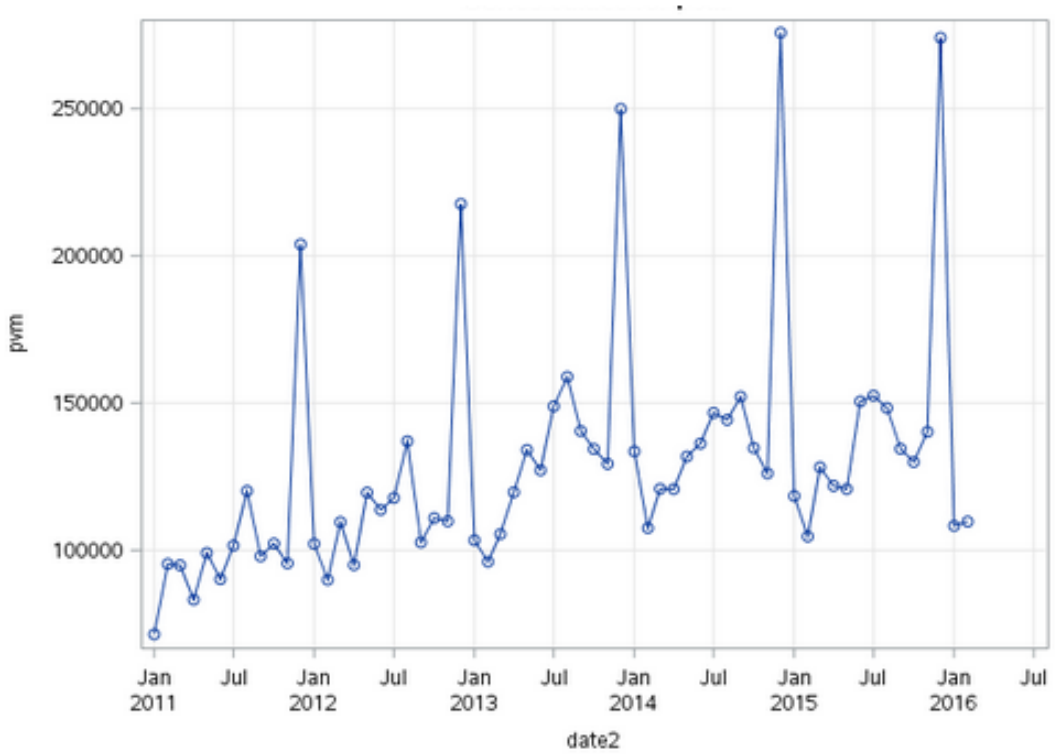
Matome, kad laikant turimus simuliuotus apskaičiuoto alkoholio duomenis, gauti neapskaičiuoto alkoholio duomenys yra pasiskirstę netoli vieneto. Tai reiškia, kad eksponentės laipsnis yra arti nulio. Apibendrindami, galėtume teigti, kad metodo tinkamumas ne visiškai pasiteisino.

3 Neapskaičiuotų cigarečių vertinimas

3.1 Pradiniai duomenys

Neapskaičiuotam tabakui vertinti naudosime mažmeninės tabako gamybos mokėtino PVM duomenis eurais, gautus iš Muitinės. Turima mėnesinių duomenų laiko eilutė nuo 2011 metų pradžios:

13 pav. *Mokėtinas pridėtinės vertės mokęstis (PVM) mažmeninei tabako gamybai*



Kita svarbi laiko eilutė, kurios pagalba vertinsime neapskaičiuoto tabako suvartojimą, yra ligotumas susijęs su tabako sukeliamomis ligomis. Turimos diagnozės buvo suskirstytos į 13 pagrindinių grupių:

- burnos navikai;
- kvėpavimo sistemos navikai;
- krūtinės angina;
- ūminis miokardo infarktas;
- lėtinė išeminė širdies liga;
- smegenų infarktas;

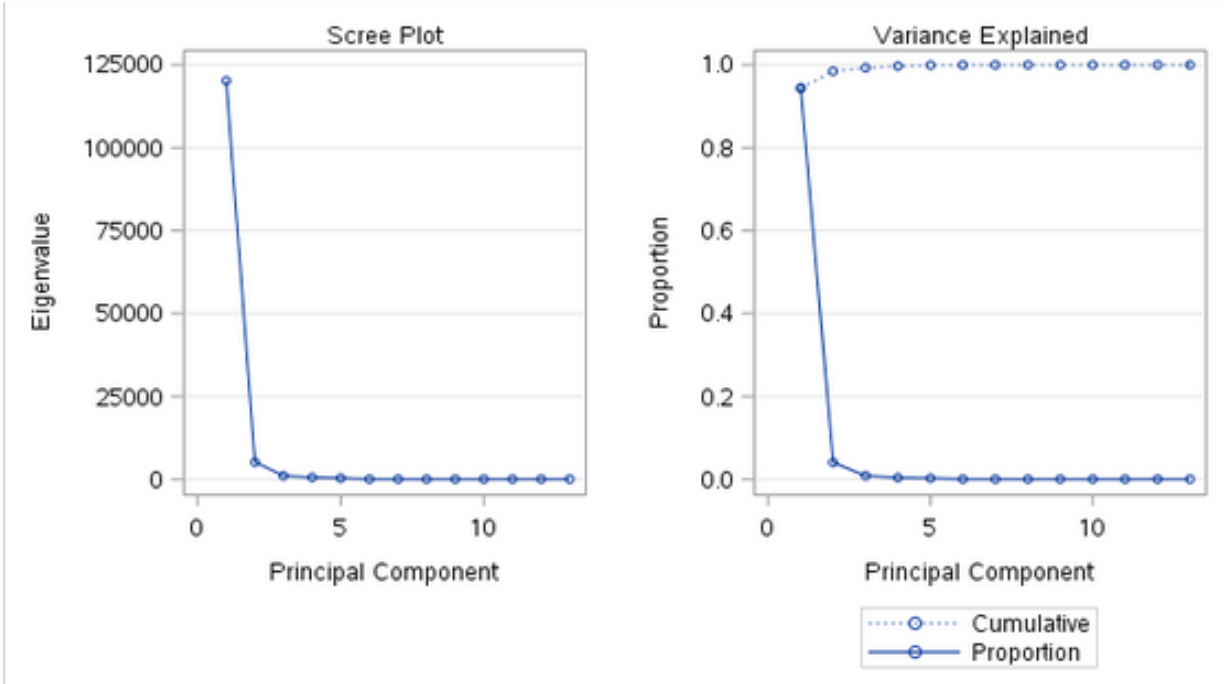
- kitos smegenų kraujagyslių ligos;
- smegenų kraujagyslių ligų pasekmės;
- aterosklerozė (išskyrus širdį ir smegenis);
- kitos kraujotakos ligos;
- bronchitai;
- emfizemos ir kitos lėtinės obstrukcinės plaučių ligos;
- astma.

3.2 Duomenų paruošimas skaičiavimams

Analogišku principu kaip ir su alkoholiu, tabako sukeltoms diagnozėms pritaikome pagrindinių komponenčių metodą ir apskaičiuojame naują kintamąjį, kurį naudosime tolimesniuose skaičiavimuose kaip apibendrintą rūkymo sukeliamų ligų kintamąjį.

Taigi, iš dešinėje pusėje esančio grafiko matome, kad pirmoji komponentė paaiškina apie 94% visos dispersijos. Apskaičiavę pirmąją pagrindinę komponentę, gauname naują tabako sukeliamų ligų kintamąjį, kurį naudosime tolimesniuose skaičiavimuose.

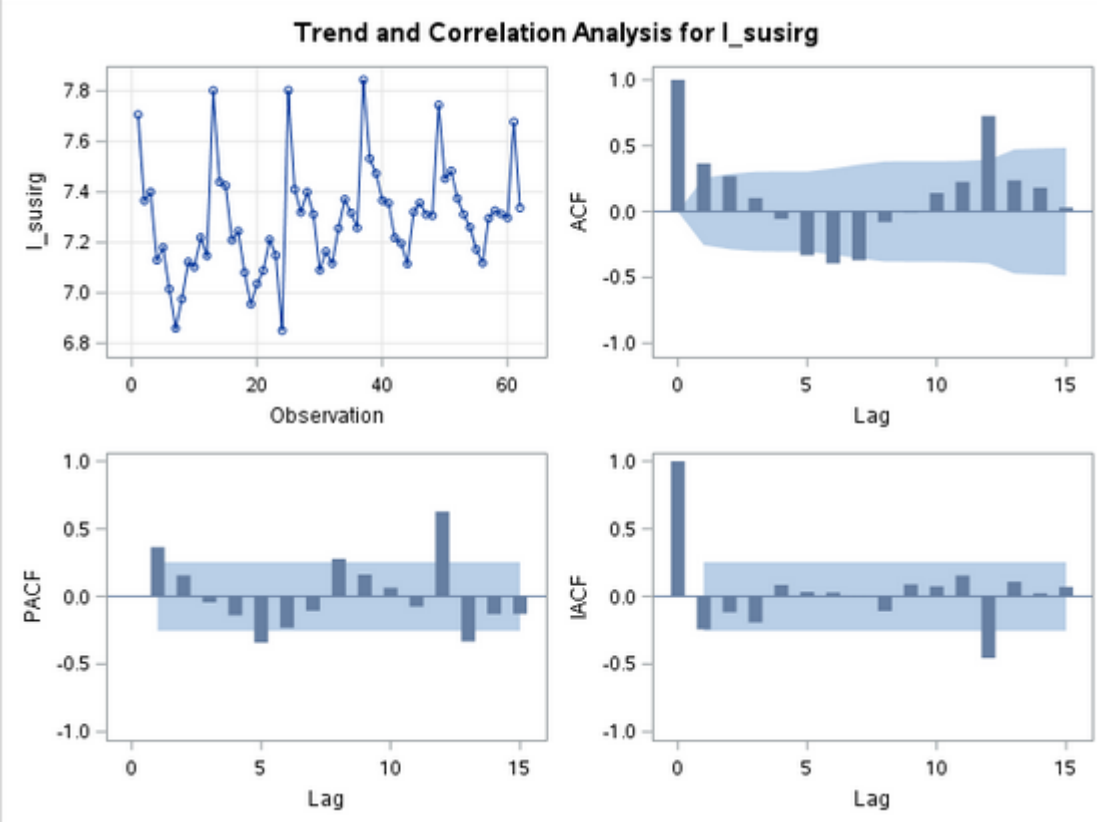
14 pav. *Tikrinių reikšmių (kairėje) ir dispersijų santykių (dešinėje) grafikai*



3.3 ARIMA modelio parinkimas ir neapskaičiuoto tabako vertinimas

Pritaikę atpažinimo (angl. Identification) ARIMA modelio etapą turimai ligotumų laiko eilutei, akivaizdžiai matome, kad laiko eilutei būdingas sezoniškumas, t.y. ligos, pagrinde kvėpavimo sistemos navikai, pasireiškia žiemos sezono metu.

15 pav. *Ligotumo, autokoreliacijų, dalinių autokoreliacijų bei atvirkštinių autokoreliacijų grafikai*



Norint eliminuoti sezoniškumą, naudosime fiktyvius kintamuosius (angl. dummy variables) ir laiko eilutę vertinsime pagal tokį modelį:

$$S_t = \mu + \beta PVM_t + \beta_1 Q_{1,t} + \beta_2 Q_{2,t} + \beta_3 Q_{3,t} + \beta NPVM_t + \varepsilon_t$$

čia S_t - rūkymo sukeltas ligotumas, tenkantis 100 tūkst. gyventojų,μ - vidurkio parametras, β - efekto parametras, atspindintis pridėtinės vertės mokęsčio poveikį rūkymo sukeltomis ligomis, PVM_t - apskaičiuotas pridėtinės vertės mokestis mažmeninei tabako gamybai, $\beta_1, ...\beta_3$ - fiktyvių parametrų įverčiai, NPVM - neapskaičiuotas pridėtinės vertės mokestis mažmeninei tabako gamybai, kurį laikome ieškomu parametru, ε_t - baltas triukšmas

$$Q_{1,t} = \begin{cases} 1, & \text{jei mėnuo} = 12, 1 \text{ arba } 2 \text{ (žiemos sezonas)} \\ 0, & \text{kitu atveju} \end{cases}$$

$$Q_{2,t} = \begin{cases} 1, & \text{jei mėnuo} = 3,\ 4 \text{ arba } 5 \text{ (pavasario sezonas)} \\ 0, & \text{kitu atveju} \end{cases}$$

$$Q_{3,t} = \begin{cases} 1, & \text{jei mėnuo} = 6,\ 7 \text{ arba } 8 \text{ (vasaros sezonas)} \\ 0, & \text{kitu atveju} \end{cases}$$

Ketvirtojo ketvirčio nėra būtina įtraukti, nes jis automatiškai įsitraukia, kai pirmieji trys fiktyvus kintamieji yra lygus nuliui [4].

Logoritmavus parametrus ir išsireiškus nežinomą kintamąjį, gauname:

$$NPVM_t = \exp\left(\frac{1}{\beta}(\ln S_t - \ln \mu - \beta \ln PVM_t - \beta_1 Q_{1,t} - \beta_2 Q_{2,t} - \beta_3 Q_{3,t} - \ln \varepsilon_t)\right)$$

Turimiems duomenims pritaikius procedūrą ARIMA, atsižvelgiant į dalinių autokoreliacijų ir autokoreliacijų grafikus, paklaidoms buvo parinktas ARIMA(2,0,2) modelis, kuris geriausiai paaiškino hipotezę, kad paklaidos yra baltas triukšmas.

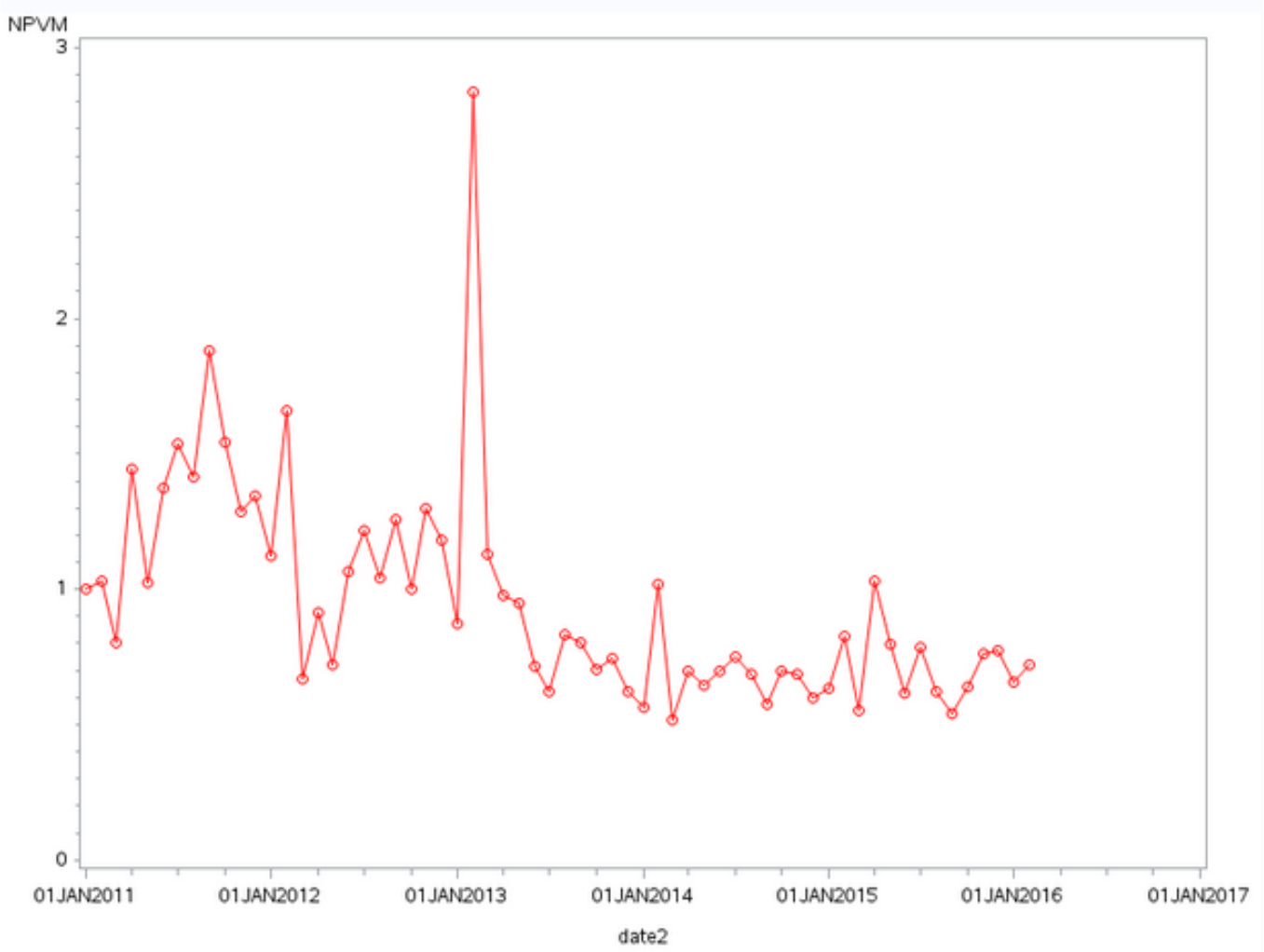
16 pav. *ARIMA(2,0,2) modelio gauti parametrų įverčiai*

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	11.28475	0.60229	18.74	<.0001	0	l_susirg	0
MA1,1	0.77507	0.18979	4.08	0.0002	1	l_susirg	0
MA1,2	-0.49915	0.17862	-2.79	0.0072	2	l_susirg	0
AR1,1	0.51730	0.19767	2.62	0.0115	1	l_susirg	0
AR1,2	0.35406	0.19664	1.80	0.0775	2	l_susirg	0
NUM1	-0.35010	0.05245	-6.67	<.0001	0	l_PVM	0
NUM2	0.29560	0.02952	10.01	<.0001	0	Q1	0
NUM3	0.04680	0.03653	1.28	0.2057	0	Q2	0
NUM4	-0.12519	0.02924	-4.28	<.0001	0	Q3	0

Iš įverčių lentelės p-reikšmių, matome, kad visi parametrai yra reikšmingi. Taip pat, iš žiemos sezono p-reikšmės akivaizdžiai matoma, kad žiemos sezonas prideda stiprų indėlį rezultatams.

Atlikdami analogiškus skaičiavimus kaip ir su neapskaičiuoto alkoholio vertinimu, pagal išsireikštą formulę, gauname neapskaičiuotą pridėtinės vertės mokestį mažmeninės tabako gamybos gaminiams ir rezultatus pavaizduojame grafiškai.

17 pav. *Neapskaičiuoto pridėtinės vertės mokesčio mažmeninei tabako gamybai eurais, grafikas*



Kaip matome, rezultatai yra pasiskirstę netoli vieneto, kas tikrai negali sudaryti neapskaičiuoto PVM, kai tikrasis, žinomas PVM siekia iki 200 tūkst. eurų. Apibendrindami, galime teigti, kad gauti neapskaičiuoti PVM mažmeninei tabako gamybai rezultatai, taip pat kaip ir neapskaičiuoto alkoholio yra ne visai tikslūs.

Išvados

Apibendrinant baigiamąjį darbą, svarbu paminėti, kad Lietuvoje nėra nusistovėjusios tradicijos reguliariai vykdyti neapskaičiuoto alkoholio ir cigarečių vartojimo vertinimą. Papildomai pažymėtina, kad alkoholio ir tabako suvartojimas Lietuvoje yra skaičiuojamas pakankamai neseniai. Dėl šios priežasties tiksliems ir išsamiems rezultatams pasiekti vis dar trūksta oficialių ir didesnį laikotarpį apimančių duomenų, kurie ypač vertingi vertinant neapskaičiuoto alkoholio ir tabako daromą žalą ir su šiuo vartojimu susijusių pažeidimų mastą.

Atlikus neapskaičiuoto vartojimo vertinimo metodų apžvalgą, buvo nustatyta, kad mokslinėje literatūroje išskiriami du pagrindiniai neapskaičiuoto vartojimo vertinimo metodai – tiesioginis ir netiesioginis. Kiekvienas iš šių modelių turi savo privalumų ir trūkumų, todėl siekiant gauti tiksliausius rezultatus tikslinga šiuos du metodus taikyti kartu.

ARIMA statistinio prognozavimo metodas parodė, kokios pasekmės žmogaus sveikatai turėtų būti, jei tikrasis alkoholio suvartojimas būtų toks, koks yra apskaičiuojamo alkoholio suvartojimas. Rezultatai parodė, kad pasirinktas metodas ne visai atitiko išsikeltus tikslus, kadangi nebuvo pilnai tinkamas.

Apibendrinus baigiamąjį darbą, galima teigti, kad išsikeltas tikslas įgyvendintas dalinai, ką sąlygojo išsamių ir didesnį laiko tarpą apimančių duomenų trūkumas. Nustatyta, kad alkoholio suvartojimui taikytas metodas, taikant jį neapskaičiuoto cigarečių vartojimo atvejui, neatspindi neapskaičiuoto cigarečių suvartojimo realijų. Atsižvelgiant į pateiktą informaciją, šis baigiamasis darbas gali būti puikus atspirties taškas ir praktinis pavyzdys tolimesniems tyrimams, o gauti tyrimo rezultatai gali būti aktualūs siekiant nustatyti neapskaičiuoto alkoholio ir cigarečių žalą ir mastą. Ateityje atliekant panašius tyrimus rekomenduotina surinkti kaip įmanoma išsamesnius ir ilgesnį laikotarpį apimančius oficialius duomenis. Taip pat naudinga kompleksiškai pritaikyti vieną iš tiesioginių metodų. Darytina išvada, kad pasitelkus šio tyrimo praktinį pavyzdį ir reguliariai skaičiuojamus valstybės statistinius duomenis, ateityje tyrimų rezultatai bus daug tikslesni.

Santrauka

Lietuvoje nėra nusistovėjusios tradicijos reguliariai vykdyti neapskaičiuoto alkoholio ir cigarečių vartojimo vertinimą, kuris leistų nustatyti tokio vartojimo daromą socialinę ir ekonominę žalą visuomenei. Atsižvelgiant į tai, šio baigiamojo darbo tikslas – parengti neapskaičiuoto alkoholio ir cigarečių apyvartos Lietuvoje įvertinimą. Tyrimo metu taikytas ARIMA statistinio prognozavimo metodas, skirtas laiko eilučių analizei. Darbą sudaro trys pagrindinės struktūrinės dalys: pirmoje darbo dalyje apibrėžiamos pagrindinės darbe naudojamos sąvokos, apžvelgiami tiesioginis ir netiesioginis neapskaičiuoto vartojimo vertinimo metodai, aptariami panašūs tyrimai bei pristatomas tyrimo metu taikytas modelis; antroje darbo dalyje pateikiamas neapskaičiuoto alkoholio vertinimas; trečioje darbo dalyje aprašomas neapskaičiuotų cigarečių vertinimas. Atlikus tyrimą buvo nustatyta kad alkoholio suvartojimui taikytas metodas, taikant jį neapskaičiuoto cigarečių vartojimo atvejui, neatspindi neapskaičiuoto cigarečių suvartojimo realijų. Taip pat pažymėta, kad siekiant gauti tiksliausius rezultatus, kol kas nepakanka turimų valstybės statistinių duomenų. Prieita išvados, kad pasitelkus atlikto tyrimo praktinį pavyzdį ir reguliariai skaičiuojamus valstybės statistinius duomenis, ateityje tyrimų rezultatai bus daug tikslesni .

Summary

In Lithuania, there is not well-established tradition of regular alcohol and cigarettes consumption estimation, which could determine a consumer-related social and economic harm to the public. Taking that fact, the final goal was to to prepare unrecorded alcohol and cigarettes consumption turnover in Lithuania. The study applied statistical ARIMA forecasting method for time series analysis. The work consists of three main structural parts: the first part defines key terms used in the work, gives an overview of the direct and indirect methods for estimating consumption, used estimation methods are discussed and presented similar research study applied to the model; the second part consists of estimation of unrecorded alcohol consumption; The third part describes the estimation of unrecorded cigarettes. The research revealed that the method which was used to estimate unrecorded alcohol consumption did not fit for estimation of unrecorded cigarettes consumption either. It is also noted that in order to get the most accurate results, there is insufficient statistical data available. It was concluded that through the practical example and regulary calculated statistical data, research results will be more accurate in the future.

Literatūros ir šaltinių sąrašas

[1] V. Cicėnienė, R. Gaidelytė, M. Garbuvienė.*Pagrindinės sveikatos statistikos sąvokos, jų apibrėžimai ir skaičiavimas*, Higienos instituto Sveikatos informacijos centras, Vilnius, 2010, p. 11, http://sic.hi.lt/data/stat_leid.pdf

[2] E. Danila. Rūkymo paplitimas, mirtingumas nuo rūkymo sukeltų ligų, ekonominė žala, *Biomedicina*, 2001, Nr. 1, p. 70-74, http://www.elibrary.lt/resursai/Ziniasklaida/Medicine/Biomedicina/2001/BM_2001_01_1048.pdf

[3] *Global status report on alcohol and health 2014*, World Health Organization 2014, 29 p., 222 p., http://apps.who.int/iris/bitstream/10665/112736/1/9789240692763_eng.pdf?ua=1

[4] J. Hyndman, A. Kostenko. Minimum sample size requirements for seasonal forecasting models, *Foresight*, 2007, Issue 6 Spring, p. 12-15, http://www.bishophill.com/admin/sidebar_images/1741759940_test.pdf

[5] NORSTRÖM, Thor.*Estimating changes in unrecorded alcohol consumption in Norway using indicators of harm.* Volume 93, Issue 10, 1998, p. 1531-1538.

[6] POPOVA, Svetlana; REHM, Jurgen; PATRA, Jayadeep; ZATONSKI, Witold. Comparing alcohol consumtion in Central and Eastern Europe to other European countries.*Alcohol & Alcoholism.* Vol. 42, No. 5. 2007, p. 465-473. http://alcalc.oxfordjournals.org/content/alcalc/42/5/465.full.pdf

[7] RAZVODOVSKY, Yury. Unrecorded Alcohol Consumption: Quantitative Methods of Estimation. *Alcoholism 2010*; 45(1). 2010, p. 15-24. https://goo.gl/tvSdeH

[8] *Regression with ARIMA errors, Cross correlation functions, and Relationships between 2 Time Series*, The Pennsylvania State University, 2016, https://goo.gl/EQ8E5S

[9] L. Stabingienė. *Ekonometrika*, Klaipėda, 2014, http://www.ilab.lt/stabingiene/sk8_1.html

[10] *The Box-Jenkins Method*, NCSS Statistical Software, Chapter 470, p. 5-7, https://goo.gl/Id9ZXw

1 priedas. SAS programos kodai

neapskaičiuoto alkoholio vertinimas

```
/* Ikeliami duomenys */
data principal_components;
    infile '/home/valerkaa1/my_content/fsprojects/bakal/duom.txt' dlm='09'x;
    input metai alkoholis F10 G31 G62 I42 K70 K86 X45 X65 Y15;
run;

/* Taikomas pagrindiniu komponenciu metodas */
ods trace on;
ods graphics on;
proc princomp data=principal_components cov;
    var F10 G31 G62 I42 K70 K86 X45 X65 Y15;
    ods output Eigenvectors=komponenetes;
run;
ods graphics off;
ods trace off;
/* Skaiciuojama nauja komponente */
proc iml ;
    use principal_components;
    read all;
    close principal_components;
    use komponenetes;
    read all;
    close komponenetes;
    row=nrow(F10);
    col=ncol(F10);
    nauja_komp=J(row, col, 0);

    do i=1 to nrow(F10);
        nauja_komp[i, ]=prin1[1, ]*F10[i, ]+prin1[2, ]*G31[i, ]+prin1[3, ]*G62[i,
            ]+prin1[4, ]*I42[i, ]+prin1[5, ]*K70[i, ]+
prin1[6, ]*K86[i, ]+prin1[7, ]*X45[i, ]+prin1[8, ]*X65[i, ]+prin1[9, ]*Y15[i, ];
    end;
    nauja_komp=round(nauja_komp, 0.01);
    create nauja_komponente var {metai alkoholis nauja_komp};
    append;
    close nauja_komponente;
    quit;

/* logoritmuojami duomenys */
data duomenys_l_susirg (rename=(nauja_komp=susirg));
    set nauja_komponente;
    l_susirg=log(nauja_komp);
    l_alkoholis=log(alkoholis);
    t=_N_;
run;
```

```
/* logoritmuojami duomenys */
data duomenys_l_susirg (rename=(nauja_komp=susirg));
    set nauja_komponente;
    l_susirg=log(nauja_komp);
    l_alkoholis=log(alkoholis);
    t=_N_;
run;


/* laiko eilutes vertinimas */
ods trace on;

proc arima data=duomenys_l_susirg;
    identify var=l_susirg nlag=13 crosscorr=l_alkoholis;
    estimate input=l_alkoholis;
    run;
    estimate p=1 input=l_alkoholis MAXITER=200;
    forecast lead=0 out=prognoze;
    ODS output ParameterEstimates=iverciai;
    run;

ods trace off;


/* sukuriamas naujas kintamasis, kuris lygus eilutes numeriui */
data prognoze;
set prognoze;
t=_N_;
run;
/* jungiamos lenteles pagal t */
data sujungta;
    merge duomenys_l_susirg prognoze;
    by t;
run;
```

```
/* SQL pagalba, gauti iverciai irasomi i makrokintamuosius */
proc sql noprint;
select estimate as miu into: miu from iverciai where Parameter="MU";
quit;

proc sql noprint;
    select estimate as beta into: beta from iverciai where Parameter="NUM1";
quit;

/* Skaiciuojamas neapskaiciuotas alkoholis */
data neapskc_alk;
    set sujungta;
    nt=exp((l_susirg-&beta*l_alkoholis-residual-&miu)/&beta);
run;


/* Braizomas neapskaiciuoto alkoholio grafikas */
GOPTIONS RESET=(axis, legend, pattern, symbol, title, footnote) INTERPOL=JOIN;
SYMBOL1 height=1 cv=red value=circle;
PROC GPLOT DATA=neapskc_alk;
    PLOT nt*metai /overlay FRAME LVREF=1 CVREF=BLACK;
    RUN;
QUIT;

/* Braizomi apskaiciuoto ir neapskaiciuoto alkoholio grafikai */
GOPTIONS RESET=(axis, legend, pattern, symbol, title, footnote) INTERPOL=JOIN;
SYMBOL1 height=1 cv=red value=circle;
SYMBOL2 height=1 cv=blue value=circle;
legend1  value=('Apskaiciuotas alkoholis');
legend2  value=('Neapskaiciuotas alkoholis');
PROC GPLOT DATA=neapskc_alk;
    PLOT alkoholis*metai /legend=legend1;
    PLOT2 nt*metai /legend=legend2;
    RUN;
QUIT;
/* Makro programa simuliuojanti laiko eilutes */
%MACRO SIMULIACIJA(N=,alpha1=, alpha2=,miuAt=,miuNt=,gama1=,gama2=,ro=);

DATA GENERAVIMAS;

  ARRAY At At1 -At&N;
  ARRAY Nt Nt1-Nt&N;
  ARRAY e_At  e_At1 -e_At&N ;
  ARRAY e_Nt  e_Nt1 -e_Nt&N ;

  At(1)=0;
  Nt(1)=0;
  e_At(1) =0;
  e_Nt(1) =0;

    DO J=2 TO &N;

       e_At(J) =RANNOR(-1111);
       vid=&ro*e_At(J);
       std=sqrt(1-&ro*&ro);
       e_Nt(J) =vid+std*RANNOR(-111);

At(J)=&miuAt + At(J-1)*&alpha1- Nt(J-1)*&gama1*At(J-1)+ e_At(J);
Nt(J)=&miuNt + Nt(J-1)*&alpha2 - At(J-1)*&gama2*Nt(J-1) + e_Nt(J);

    END;

OUTPUT;
```

```
/* Transponuojama lentele. Is eiluciu pertvarkoma i stulpeli*/
DATA GENERAV1;SET GENERAVIMAS;
   KEEP At1-At&N ;PROC TRANSPOSE OUT=GENERAV1;
   DATA GENERAV1;SET GENERAV1;ID=_N_;At=COL1;OUTPUT;DROP COL1;

DATA GENERAV2;SET GENERAVIMAS;
   KEEP Nt1-Nt&N;PROC TRANSPOSE OUT=GENERAV2;
   DATA GENERAV2;SET GENERAV2;ID=_N_;N=COL1;OUTPUT;DROP COL1;

   DATA GENERAV3;SET GENERAVIMAS;
   KEEP e_Nt1-e_Nt&N;PROC TRANSPOSE OUT=GENERAV3;
   DATA GENERAV3;SET GENERAV3;ID=_N_;e_Nt=COL1;OUTPUT;DROP COL1;

   DATA GENERAV4;SET GENERAVIMAS;
   KEEP e_At1-e_At&N;PROC TRANSPOSE OUT=GENERAV4;
   DATA GENERAV4;SET GENERAV4;ID=_N_;e_At=COL1;OUTPUT;DROP COL1;

/*  Sujungiamos lenteles ir paliekami stebejimai kur ID > 101 */
DATA GENERAVIMAS;MERGE GENERAV1 GENERAV2 GENERAV3 GENERAV4;BY ID; where ID>101;
            KEEP ID At N e_At e_Nt;

/* Breziamas neapskaiciuoto ir apskaiciuoto simuliuojamu laiko eiluciu grafikai */
GOPTIONS RESET=(axis, legend, pattern, symbol, title, footnote) INTERPOL=JOIN;
SYMBOL1 height=1 cv=red value=circle;
SYMBOL2 height=1 cv=blue value=circle;
legend1 label=('') value=('Neapskaiciuotas simuliuojamas alkoholis'
                          'Apskaiciuotas simuliuojamas alkoholis');
PROC GPLOT DATA=generavimas;
    PLOT (N At)*ID /overlay FRAME LVREF=1 CVREF=BLACK legend=legend1;
    RUN;
QUIT;
/*Pridedamas stulpelis lygus eilutes skaiciui. (Pagal ji jungsime lenteles) */
data baba;
set generavimas;
t=_N_;
run;

/* Sujungiamos lenteles */
data sujj;
    merge neapskc_alk baba;
    by t;
run;

/* Logoritmuojami kintamieji ir skaiciuojami susirgimai */
data sujj;
set sujj;
A_log=log(At);
N_log=log(N);
ln_It= &miu+&beta*A_log+&beta*N_log+rannor(-111);
run;

/* ARIMA metodu vertinami susirgimai */
ods trace on;
proc arima data=sujj;
    identify var=ln_It nlag=13 crosscorr=A_log;
    estimate input=A_log;
    run;
    estimate q=2 input=A_log maxiter=200;
    forecast lead=0 out=pr;
    ODS output ParameterEstimates=iv;
    run;
    ods trace off;
```

```
data pr;
set pr;
t=_N_;
run;
/* Rusiuojamos ir jungiamos lenteles */

data su;
    merge sujj pr;
    by t;
run;

/* Gauti iverciai miu ir beta irasomi i makrokintamuosius */

proc sql noprint;
    select estimate as miu2 into: miu2 from iv where Parameter="MU";
quit;

proc sql noprint;
    select estimate as beta2 into: beta2 from iv where Parameter="NUM1";
quit;

/* Skaiciuojamas neapskaiciuotas alkoholis */
data ne_alk;
    set su;
    nt_sim=exp((ln_It-&beta2*A_log-&miu2-residual)/&beta2);
run;

/* Rusiuojama pagal metus */
proc sort data=ne_alk out=gal;
    by metai;
run;
/* Braizomas neapskaiciuoto alkoholio, gauto is apskaiciuoto simuliuoto
alkoholio, grafikas */
GOPTIONS RESET=(axis, legend, pattern, symbol, title, footnote) INTERPOL=JOIN;
SYMBOL1 height=1 cv=red value=circle;
PROC GPLOT DATA=gal;
    PLOT nt_sim*metai /overlay FRAME LVREF=1 CVREF=BLACK;
    RUN;
QUIT;


/* Braizomas apskaiciuoto, neapskaiciuoto simuliuotu ir neapskaiciuoto alkoholio,
gauto is simuliuojamo apskaiciuoto alkoholio duomenu, grafikai */
GOPTIONS RESET=(axis, legend, pattern, symbol, title, footnote) INTERPOL=JOIN;
SYMBOL1 height=1 cv=blue value=circle;
SYMBOL2 height=1 cv=red value=circle;
SYMBOL4 height=1 cv=orange value=circle;
legend1 label=(position=(bottom right)) value=('apsk. alk. is sim. duomenu'
                                               'Neapsk. alk. simuliuojamas'
                 'neapskaiciuotas alkoholis gautas is sim. apskc. alkoholio');
PROC GPLOT DATA=gal;
    PLOT (At N nt_sim)*metai /overlay FRAME legend=legend1;
    RUN;
QUIT;


%MEND SIMULIACIJA;
%SIMULIACIJA (N=115,alpha1=0.47, alpha2=0.4,miuAt=4.2,miuNt=4,gama1=0.07,gama2=0.05,ro=-0.99);
```

neapskaičiuoto tabako vertinimas

```
/* Isikeliami duomenys */
data principal_components;
infile '/home/valerkaa1/my_content/fsprojects/bakal/amb_duom_100_tukst.txt' dlm='09'x;
input metai menuo PVM burnos_navikai kvepavimo_sist_navikai krutines_angina
uminis_miokardo_infarktas letine_isemine_sirdies smegenu_infarktas kitos_smeg_kraujag_lig
smegen_kraujag_ligos ateroskleroz_isk_sird_ir_smeg kitos_kraujot_ligos bronchitai
emfizemos_ir_kt_obst_pl_ligos astma;
run;

/* Taikomas pagrindiniu komponenciu metodas */
ods trace on;
ods graphics on;
proc princomp data=principal_components cov;
   var burnos_navikai kvepavimo_sist_navikai krutines_angina uminis_miokardo_infarktas
letine_isemine_sirdies smegenu_infarktas kitos_smeg_kraujag_lig smegen_kraujag_ligos
ateroskleroz_isk_sird_ir_smeg kitos_kraujot_ligos bronchitai emfizemos_ir_kt_obst_pl_ligos
astma;
   ods output
   Eigenvectors=komponenetes;
run;
ods graphics off;
ods trace off;
/* Skaiciuojama nauja komponente */
proc iml;
use principal_components;
   read all;
close principal_components;
use komponenetes;
    read all;
close komponenetes;

row=nrow(burnos_navikai);
col=ncol(burnos_navikai);
nauja_komp=J(row,col,0);
do i=1 to nrow(burnos_navikai);
nauja_komp[i,]=prin1[1,]*burnos_navikai[i,]+prin1[2,]*kvepavimo_sist_navikai[i,]
+prin1[3,]*krutines_angina[i,]+prin1[4,]*uminis_miokardo_infarktas[i,]+prin1[5,]*
letine_isemine_sirdies[i,]+prin1[6,]*smegenu_infarktas[i,]+prin1[7,]*
kitos_smeg_kraujag_lig[i,]+prin1[8,]*smegen_kraujag_ligos[i,]+prin1[9,]*
ateroskleroz_isk_sird_ir_smeg[i,]+prin1[10,]*kitos_kraujot_ligos[i,]+
prin1[11,]*bronchitai[i,]+prin1[12,]*emfizemos_ir_kt_obst_pl_ligos[i,]+prin1[13,]*astma[i,];
end;
nauja_komp=round(nauja_komp,0.01);


create nauja_komponente var {metai menuo PVM nauja_komp};
append;
close nauja_komponente;

quit;
```

```
/* logoritmuojami duomenys bei isivedami fiktyvus kintamieji */
data tabakas (rename=(nauja_komp=ligotumai));
    set nauja_komponente;
    l_ligotumai=log(nauja_komp);
    l_PVM=log(PVM);
    t=_N_;
    if menuo = 12 or menuo= 1 or menuo = 2 then do;
        Q1 = 1;
        Q2 = 0;
        Q3 = 0;
     end;
     else if menuo= 3 or menuo=4 or menuo=5 then do;
        Q1 = 0;
        Q2 = 1;
        Q3 = 0;
     end;
     else if menuo = 6 or menuo = 7 or menuo = 8 then do;
        Q1 = 0;
        Q2 = 0;
        Q3 = 1;
     end;
     else do;
        Q1 = 0;
        Q2 = 0;
        Q3 = 0;
     end;
run;
/* Is metu ir menesio stulpeliu, padarome nauja datos stulpeli*/
data tabakas;
    set tabakas;
    date2=mdy(menuo, 1, metai);
    format date2 date9.;
run;
/* Isrusiuojame pagal nauja datos kintamaji */
proc sort data=tabakas;
    by date2;
run;
/* Breziame pvm ir ligotumo laiko eiluciu grafikus */
ods graphics on;
proc timeseries data=tabakas plot=series;
    id date2 interval=month;
    var ligotumai pvm;
run;
ods graphics off;

/* taikome atitinkama modeli ir vertiname laiko eilute */
ods graphics on;
ods trace on;
proc arima data=tabakas;
    identify var=l_ligotumai crosscorr=(l_PVM Q1 Q2 Q3);
    estimate p=2 q=2 input=(l_PVM Q1 Q2 Q3) maxiter=200;
    forecast lead=0 id=date2 interval=month out=prognozes;
    ODS OUTPUT ParameterEstimates=iverciai;
    run;
ods trace off;
```

```
/* Lenteleje sukuriamas naujas kintamasis, atitinkantis eilutes numeri */
data prognozes;
set prognozes;
t=_N_;
run;
/* sujungiamos lenteles pagal t */
data sujungta;
    merge tabakas prognozes;
    by t;
run;

/* Ieskomi iverciai irasomi i makrokintamuosius */
proc sql noprint;
    select estimate as miu into: miu from iverciai where Parameter="MU";
quit;

proc sql noprint;
    select estimate as beta1 into: beta1 from iverciai where Parameter="NUM1";
quit;
proc sql noprint;
    select estimate as beta2 into: beta2 from iverciai where Parameter="NUM2";
quit;
proc sql noprint;
    select estimate as beta3 into: beta3 from iverciai where Parameter="NUM3";
quit;
proc sql noprint;
    select estimate as beta4 into: beta4 from iverciai where Parameter="NUM4";
quit;
/* Neapskaiciuoto pvm skaiciavimai */
data sujungta;
    set sujungta;
    NPVM=exp((l_ligotumai-&beta1*l_PVM-&beta2*Q1-&beta3*Q2-&beta4*Q3-&miu-residual)/&beta1);
run;


/* Breziamas neapskaiciuoto pvm grafikas */
GOPTIONS RESET=(axis, legend, pattern, symbol, title, footnote) INTERPOL=JOIN;
SYMBOL1 height=1 cv=red value=circle;
SYMBOL2 height=1 cv=blue value=circle;

PROC GPLOT DATA=sujungta;
    PLOT NPVM*date2 /overlay FRAME LVREF=1 CVREF=BLACK;
    RUN;
QUIT;

/* Breziamas neapskaiciuoto ir apskaiciuoto PVM grafikai */
GOPTIONS RESET=(axis, legend, pattern, symbol, title, footnote) INTERPOL=JOIN;
SYMBOL1 height=1 cv=red value=circle;
SYMBOL2 height=1 cv=blue value=circle;

PROC GPLOT DATA=sujungta;
    PLOT NPVM*date2 /overlay FRAME LVREF=1 CVREF=BLACK;
    PLOT2 PVM*date2 /overlay FRAME LVREF=1 CVREF=BLACK;
    RUN;
QUIT;
```

2 priedas. Output'ai

neapskaičiuoto alkoholio vertinimo output'ai

Observations	14
Variables	9

Simple Statistics						
	F10	G31	G62	I42	K70	K86
Mean	230.6571429	9.142857143	7.374285714	0.7800000000	24.76142857	5.137142857
StD	39.0898957	2.325535269	2.004624324	0.1880261848	4.73265393	1.120511069

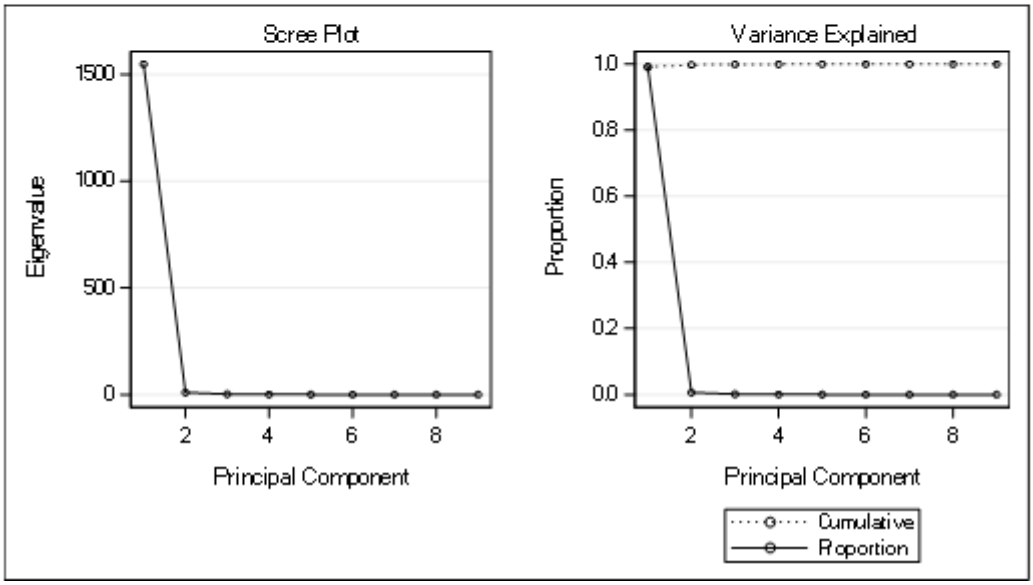
Simple Statistics			
	X45	X65	Y15
Mean	0.2300000000	0.0528571429	0.1192857143
StD	0.2160484276	0.0604379620	0.0935179362

Covariance Matrix									
	F10	G31	G62	I42	K70	K86	X45	X65	Y15
F10	1528.019945	79.537501	61.247021	-1.561800	140.158035	-18.036540	-5.091577	0.974740	-1.113487
G31	79.537501	5.408114	4.066056	-0.204100	7.832942	-1.319714	-0.236323	0.056668	-0.039290
G62	61.247021	4.066056	4.018519	-0.134908	6.923078	-1.131564	-0.144454	0.037102	-0.014404
I42	-1.561800	-0.204100	-0.134908	0.035354	-0.006646	0.130377	0.004754	-0.001608	0.003700
K70	140.158035	7.832942	6.923078	-0.006646	22.398013	-2.031049	-0.246508	0.199188	0.092686
K86	-18.036540	-1.319714	-1.131564	0.130377	-2.031049	1.255545	0.052823	-0.012345	0.012775
X45	-5.091577	-0.236323	-0.144454	0.004754	-0.246508	0.052823	0.046677	0.002092	0.009085
X65	0.974740	0.056668	0.037102	-0.001608	0.199188	-0.012345	0.002092	0.003653	0.002871
Y15	-1.113487	-0.039290	-0.014404	0.003700	0.092686	0.012775	0.009085	0.002871	0.008746

Total Variance	1561.1945654

Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	1547.80950	1538.07029	0.9914	0.9914
2	9.73920	7.47031	0.0062	0.9977
3	2.26890	1.43458	0.0015	0.9991
4	0.83432	0.33181	0.0005	0.9997
5	0.50251	0.47757	0.0003	1.0000
6	0.02494	0.01408	0.0000	1.0000
7	0.01086	0.00674	0.0000	1.0000
8	0.00412	0.00390	0.0000	1.0000
9	0.00022		0.0000	1.0000

Eigenvectors									
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
F10	0.993553	-.101141	-.046815	-.010196	-.016662	0.006504	-.000891	0.002307	-.001430
G31	0.051817	0.074981	0.624982	0.196495	0.745080	-.020364	0.082910	-.003535	0.005884
G62	0.039974	0.161887	0.658831	0.325652	-.656592	-.018575	-.014594	0.006718	0.019262
I42	-.001014	0.011104	-.084915	0.065541	-.059161	-.201504	0.949587	0.107144	0.176265
K70	0.091753	0.976751	-.180491	-.022622	0.054602	-.018156	-.022111	-.024156	-.008174
K86	-.011781	-.051735	-.364990	0.921967	0.078687	0.007416	-.084810	-.018858	-.012214
X45	-.003295	0.023793	0.009122	0.016656	-.012832	0.965856	0.228871	-.066067	-.096296
X65	0.000640	0.010986	-.007534	-.003188	0.023293	0.153074	-.171513	0.285881	0.929889
Y15	-.000711	0.020297	-.001576	0.010911	0.009182	0.043583	-.041424	0.949434	-.307149



Name of Variable = l_susirg	
Mean of Working Series	5.432895
Standard Deviation	0.173776
Number of Observations	14

Autocorrelations				
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	0.030198	1.00000	|**********|	0
1	0.022351	0.74013	|*******. |	0.267261
2	0.012206	0.40420	|********. |	0.386891
3	0.0028520	0.09444	|**. |	0.415962
4	-0.0035439	-.11735	**| . |	0.417490
5	-0.0051265	-.16976	***| . |	0.419840
6	-0.0043700	-.14471	***| . |	0.424715
7	-0.0038165	-.12638	***| . |	0.428222
8	-0.0054974	-.18204	****| . |	0.430878
9	-0.0087459	-.28962	******| . |	0.436337
10	-0.0092210	-.30535	******| . |	0.449859
11	-0.0067424	-.22327	****| . |	0.464427
12	-0.0041961	-.13895	***| . |	0.472032
13	-0.0012481	-.04133	*| . |	0.474944

"." marks two standard errors

Inverse Autocorrelations		
Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	-0.53753	| **********| . |
2	0.01764	| . | . |
3	-0.01981	| . | . |
4	0.13219	| . |***. |
5	-0.00904	| . | . |
6	-0.09455	| . **| . |
7	0.05992	| . |* . |

Partial Autocorrelations			
Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	
1	0.74013	| . |*************** |	
2	-0.31754	| . ******| . |	
3	-0.15909	| . ***| . |	
4	-0.05371	| . *| . |	
5	0.09589	| . |** . |	
6	-0.03100	| . *| . |	
7	-0.11613	| . **| . |	
8	-0.20731	| . ****| . |	
9	-0.16342	| . ***| . |	
10	0.10769	| . |** . |	
11	0.06595	| . |* . |	
12	-0.16556	| . ***| . |	
13	-0.03930	| . *| . |	

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
6	14.28	6	0.0266	0.740	0.404	0.094	-0.117	-0.170	-0.145
12	30.89	12	0.0020	-0.126	-0.182	-0.290	-0.305	-0.223	-0.139

Correlation of l_susirg and l_alkoholis	
Variance of input =	0.013646
Number of Observations	14

Crosscorrelations			
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
-13	-0.0021607	-.10644	| . **| . |
-12	-0.0054540	-.26867	| . *****| . |
-11	-0.0083006	-.40889	| . ********| . |
-10	-0.0099061	-.48798	| .**********| . |
-9	-0.0077537	-.38195	| . ********| . |
-8	-0.0040988	-.20191	| . ****| . |
-7	-0.0000586	-.00289	| . | . |
-6	0.0019267	0.09491	| . |** . |
-5	0.0021088	0.10388	| . |** . |
-4	0.0033300	0.16404	| . |*** . |

Crosscorrelations			
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
-3	0.0058728	0.28930	| . |****** . |
-2	0.0093523	0.46070	| . |********* . |
-1	0.014439	0.71128	| . |************** |
0	0.018079	0.89057	| . |****************** |
1	0.014381	0.70842	| . |************** |
2	0.0074354	0.36627	| . |******* . |
3	0.0019035	0.09377	| . |** . |
4	-0.0033350	-.16429	| . ***| . |
5	-0.0043977	-.21663	| . ****| . |
6	-0.0056977	-.28067	| . ******| . |
7	-0.0039776	-.19594	| . ****| . |
8	-0.0037567	-.18506	| . ****| . |
9	-0.0052425	-.25825	| . *****| . |
10	-0.0059999	-.29556	| . ******| . |
11	-0.0043491	-.21424	| . ****| . |
12	-0.0031831	-.15680	| . ***| . |
13	-0.0011565	-.05697	| . *| . |

"." marks two standard errors

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr> |t|	Lag	Variable	Shift
MU	2.01847	0.50392	4.01	0.0017	0	l_susirg	0
NUM1	1.32480	0.19532	6.78	<.0001	0	l_alkoholis	0

Constant Estimate	2.018472
Variance Estimate	0.007289
Std Error Estimate	0.085374
AIC	-27.3279
SBC	-26.0497
Number of Residuals	14

** AIC and SBC do not include log determinant.*

Correlations of Parameter Estimates		
Variable Parameter	l_susirg MU	l_alkoholis NUM1
l_susirg MU	1.000	-0.999
l_alkoholis NUM1	-0.999	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
6	11.28	6	0.0801	0.337	0.275	-0.259	-0.224	-0.418	-0.196
12	17.64	12	0.1270	-0.280	-0.045	-0.005	0.213	0.079	0.083

Model for variable l_susirg	
Estimated Intercept	2.018472

Input Number 1	
Input Variable	l_alkoholis
Overall Regression Factor	1.324804

Input Number 1	
Input Variable	l_alkoholis
Overall Regression Factor	1.324804

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr> |t|	Lag	Variable	Shift
MU	2.20987	0.66657	3.32	0.0069	0	l_susirg	0
AR1,1	0.39314	0.29522	1.33	0.2099	1	l_susirg	0
NUM1	1.25023	0.25949	4.82	0.0005	0	l_alkoholis	0

Constant Estimate	1.341089
Variance Estimate	0.006901
Std Error Estimate	0.083074
AIC	-27.3108
SBC	-25.3936
Number of Residuals	14

** AIC and SBC do not include log determinant.*

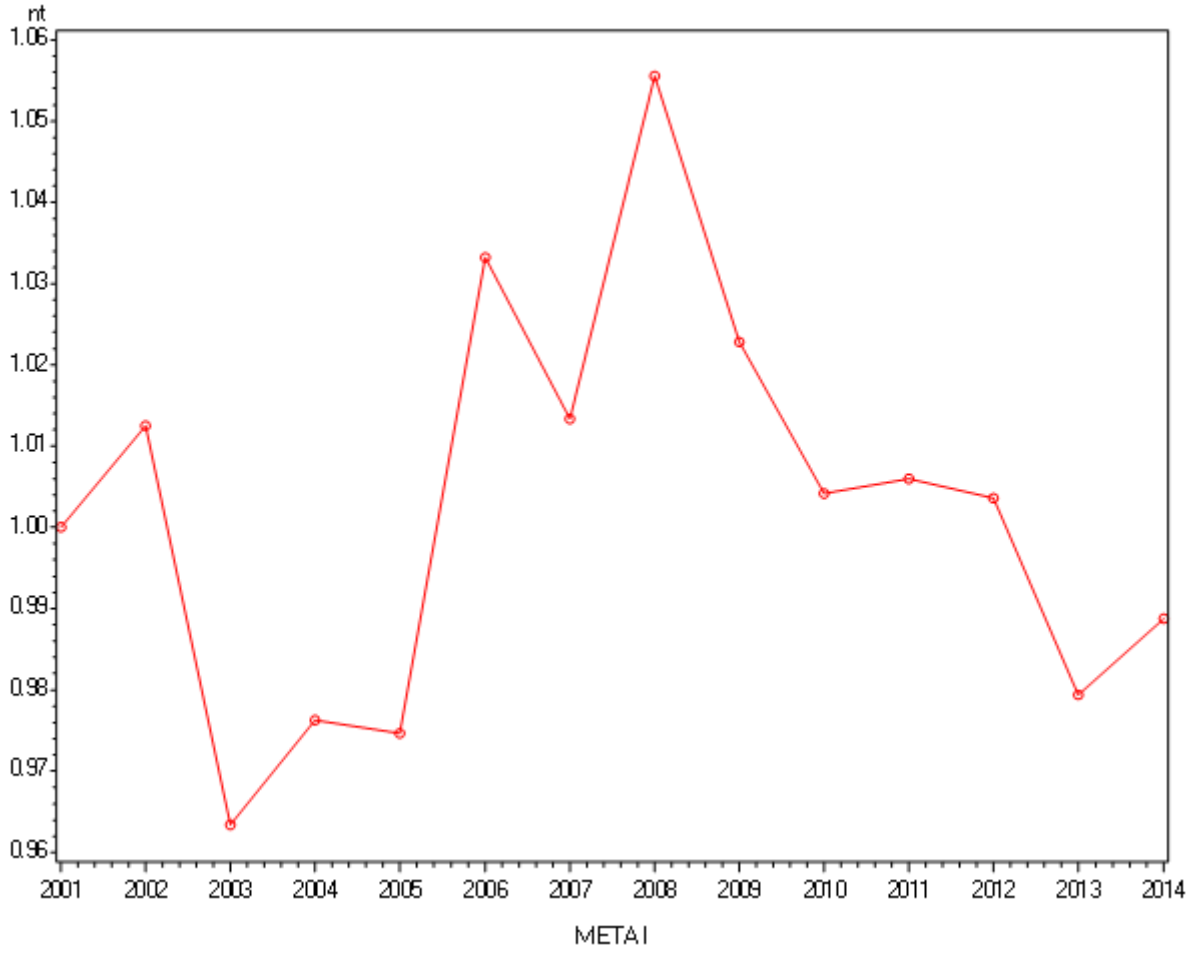
Correlations of Parameter Estimates			
Variable Parameter	l_susirg MU	l_susirg AR1,1	l_alkoholis NUM1
l_susirg MU	1.000	0.152	-0.999
l_susirg AR1,1	0.152	1.000	-0.156
l_alkoholis NUM1	-0.999	-0.156	1.000

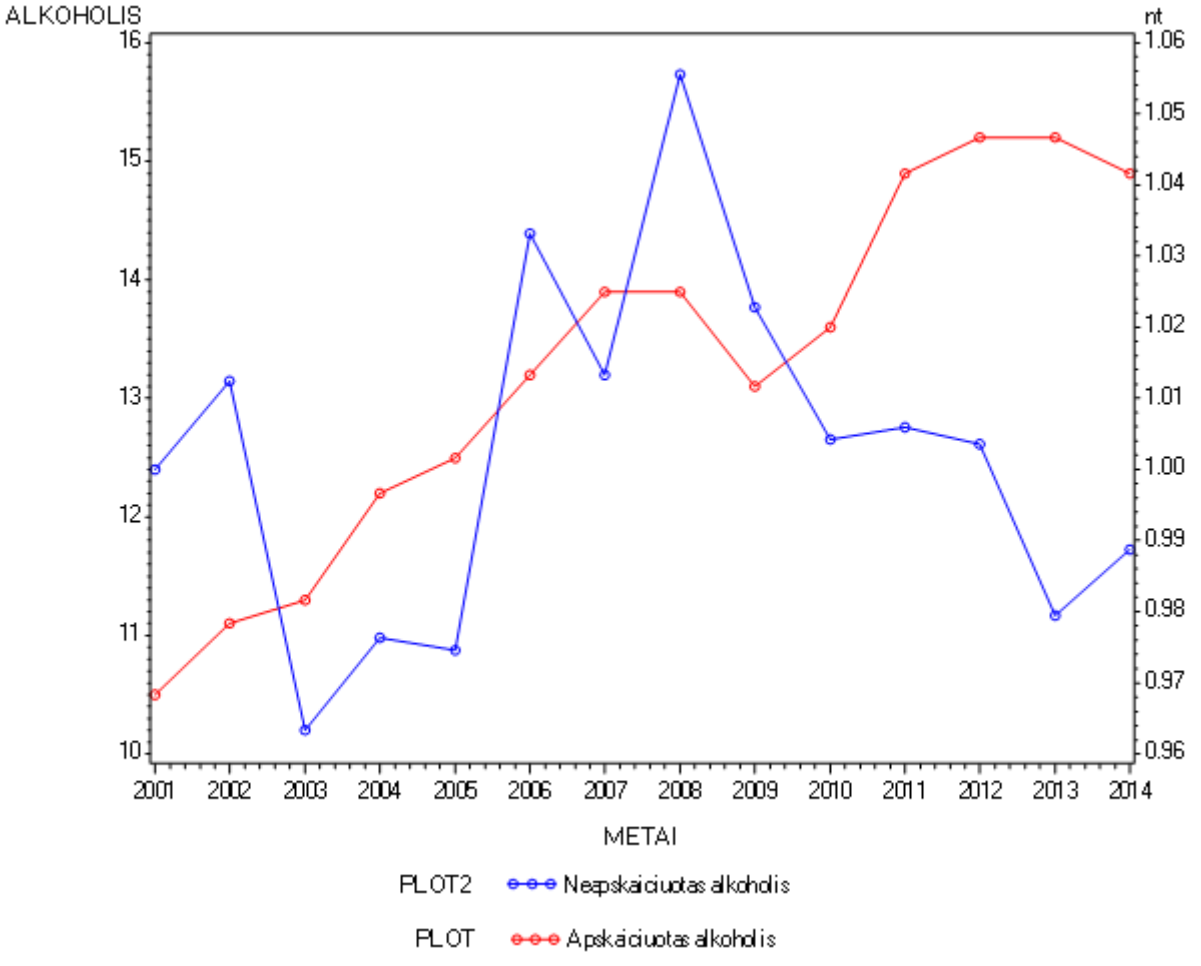
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
6	7.72	5	0.1722	-0.064	0.320	-0.329	-0.028	-0.372	0.051
12	13.28	11	0.2756	-0.246	-0.005	-0.088	0.177	0.008	0.116

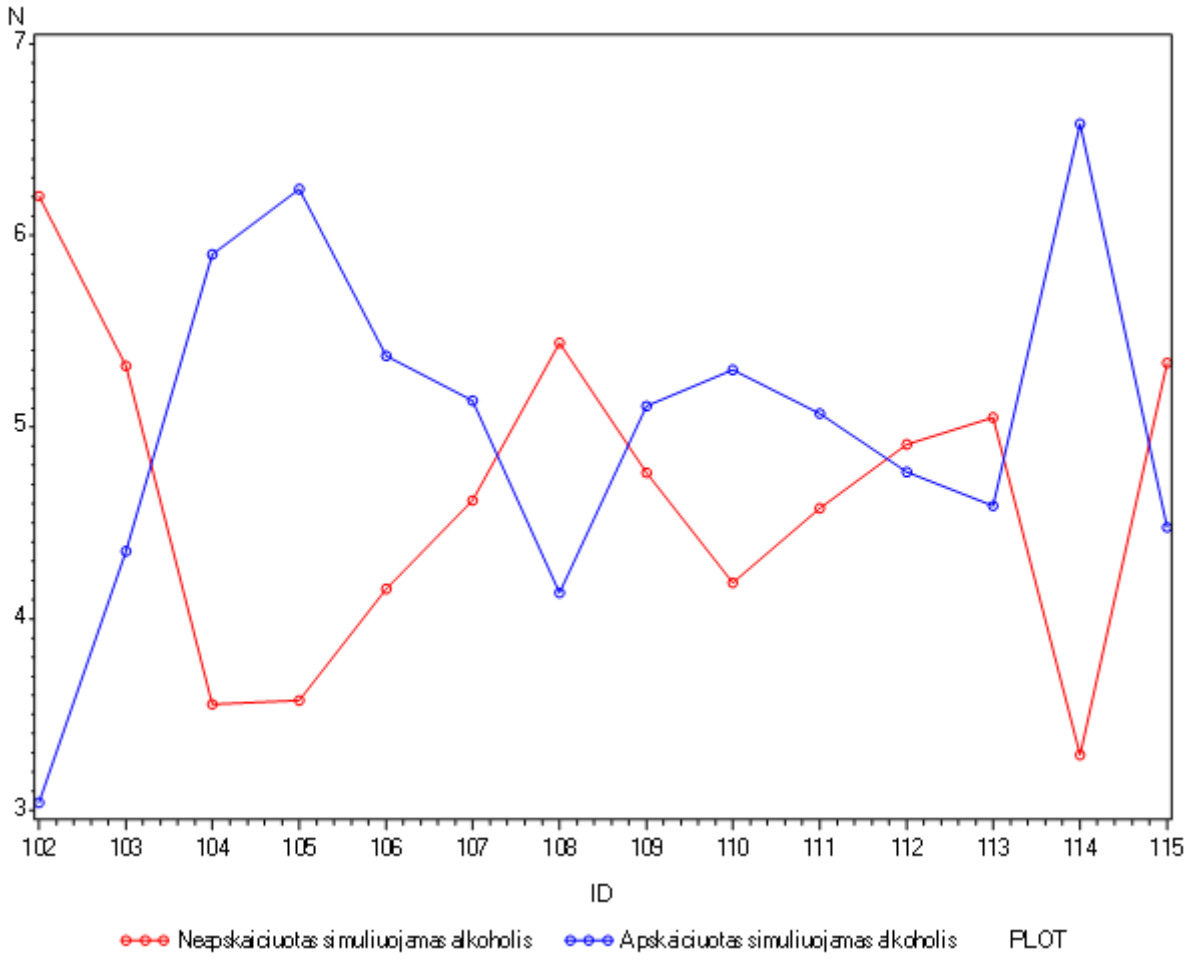
Model for variable l_susirg	
Estimated Intercept	2.209867

Autoregressive Factors	
Factor 1:	1 - 0.39314 B**(1)

Input Number 1	
Input Variable	l_alkoholis
Overall Regression Factor	1.250227


nt
1.06
1.05
1.04
1.03
1.02
1.01
1.00
0.99
0.98
0.97
0.96
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
METAI


ALKOHOLIS
nt
16
15
14
13
12
11
10
1.06
1.05
1.04
1.03
1.02
1.01
1.00
0.99
0.98
0.97
0.96
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
METAI
PLOT2
Neapskaiciuotas alkoholis
PLOT
Apskaiciuotas alkoholis



Name of Variable = ln_It	
Mean of Working Series	5.98998
Standard Deviation	1.123621
Number of Observations	14

Autocorrelations				
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	1.262524	1.00000	|********************|	0
1	0.437287	0.34636	|******* . |	0.267261
2	0.450567	0.35688	|******* . |	0.297601
3	-0.264673	-.20964	| .****| . |	0.326743
4	-0.240931	-.19083	| .****| . |	0.336213
5	-0.514561	-.40757	| .********| . |	0.343863
6	-0.097252	-.07703	| . **| . |	0.376791
7	-0.190160	-.15062	| . ***| . |	0.377914
8	-0.015818	-.01253	| . | . |	0.382178
9	-0.065181	-.05163	| . *| . |	0.382208
10	-0.044684	-.03539	| . *| . |	0.382705
11	-0.054198	-.04293	| . *| . |	0.382939
12	-0.019162	-.01518	| . | . |	0.383283
13	-0.012496	-.00990	| . | . |	0.383326

"." marks two standard errors

Inverse Autocorrelations		
Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	-0.19329	| . ****| . |
2	-0.40622	| . ********| . |
3	0.11618	| . |** . |
4	0.20100	| . |**** . |
5	0.08502	| . |** . |
6	-0.16724	| . ***| . |
7	0.07064	| . |* . |

Partial Autocorrelations		
Lag	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1	0.34636	| . |******* . |
2	0.26921	| . |***** . |
3	-0.48173	| .**********| . |
4	-0.13293	| . ***| . |
5	-0.08776	| . **| . |
6	0.18194	| . |**** . |
7	-0.10691	| . **| . |
8	-0.28999	| . ******| . |
9	0.03524	| . |* . |
10	-0.01954	| . | . |
11	-0.00325	| . | . |
12	-0.11685	| . **| . |
13	-0.11223	| . **| . |

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
6	10.46	6	0.1067	0.346	0.357	-0.210	-0.191	-0.408	-0.077
12	11.54	12	0.4833	-0.151	-0.013	-0.052	-0.035	-0.043	-0.015

Correlation of ln_It and A_log	
Variance of input =	0.034504
Number of Observations	14

Crosscorrelations			
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
-13	0.0050947	0.02441	| . | . |
-12	-0.012691	-.06081	| . *| . |
-11	0.012873	0.06168	| . |* . |
-10	-0.049003	-.23479	| . *****| . |
-9	0.016220	0.07771	| . |** . |
-8	-0.030115	-.14429	| . ***| . |
-7	0.0019106	0.00915	| . | . |
-6	0.0093588	0.04484	| . |* . |
-5	0.028212	0.13517	| . |*** . |
-4	0.0035091	0.01681	| . | . |
-3	-0.010720	-.05136	| . *| . |

Crosscorrelations			
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
-2	-0.0077196	-.03699	| . *| . |
-1	-0.043313	-.20752	| . ****| . |
0	-0.023613	-.11313	| . **| . |
1	-0.0010535	-.00505	| . | . |
2	0.102249	0.48990	| . |**********. |
3	0.062875	0.30125	| . |****** . |
4	0.102315	0.49021	| . |**********. |
5	-0.0032800	-.01572	| . | . |
6	-0.063640	-.30492	| . ******| . |
7	-0.073309	-.35124	| . *******| . |
8	-0.0046205	-.02214	| . | . |
9	-0.011574	-.05545	| . *| . |
10	0.0090548	0.04338	| . |* . |
11	-0.0011818	-.00566	| . | . |
12	-0.0098160	-.04703	| . *| . |
13	-0.0080218	-.03843	| . *| . |

"." marks two standard errors

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr> |t|	Lag	Variable	Shift
MU	7.08095	2.78458	2.54	0.0258	0	ln_It	0
NUM1	-0.68435	1.73500	-0.39	0.7002	0	A_log	0

Constant Estimate	7.080946
Variance Estimate	1.454093
Std Error Estimate	1.205858
AIC	46.81352
SBC	48.09163
Number of Residuals	14

** AIC and SBC do not include log determinant.*

Correlations of Parameter Estimates		
Variable Parameter	ln_It MU	A_log NUM1
ln_It MU	1.000	-0.993
A_log NUM1	-0.993	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
6	10.39	6	0.1093	0.328	0.410	-0.188	-0.136	-0.398	-0.107
12	12.11	12	0.4373	-0.192	-0.033	-0.049	-0.056	-0.038	-0.031

Model for variable ln_It	
Estimated Intercept	7.080946

Input Number 1	
Input Variable	A_log
Overall Regression Factor	-0.68435

Input Number 1	
Input Variable	A_log
Overall Regression Factor	-0.68435

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr> |t|	Lag	Variable	Shift
MU	7.98538	2.06513	3.87	0.0031	0	ln_It	0
MA1,1	-0.27587	0.25490	-1.08	0.3045	1	ln_It	0
MA1,2	-0.62897	0.27145	-2.32	0.0430	2	ln_It	0
NUM1	-1.30790	1.33965	-0.98	0.3519	0	A_log	0

Constant Estimate	7.985378
Variance Estimate	1.120553
Std Error Estimate	1.058562
AIC	44.61318
SBC	47.16941
Number of Residuals	14

** AIC and SBC do not include log determinant.*

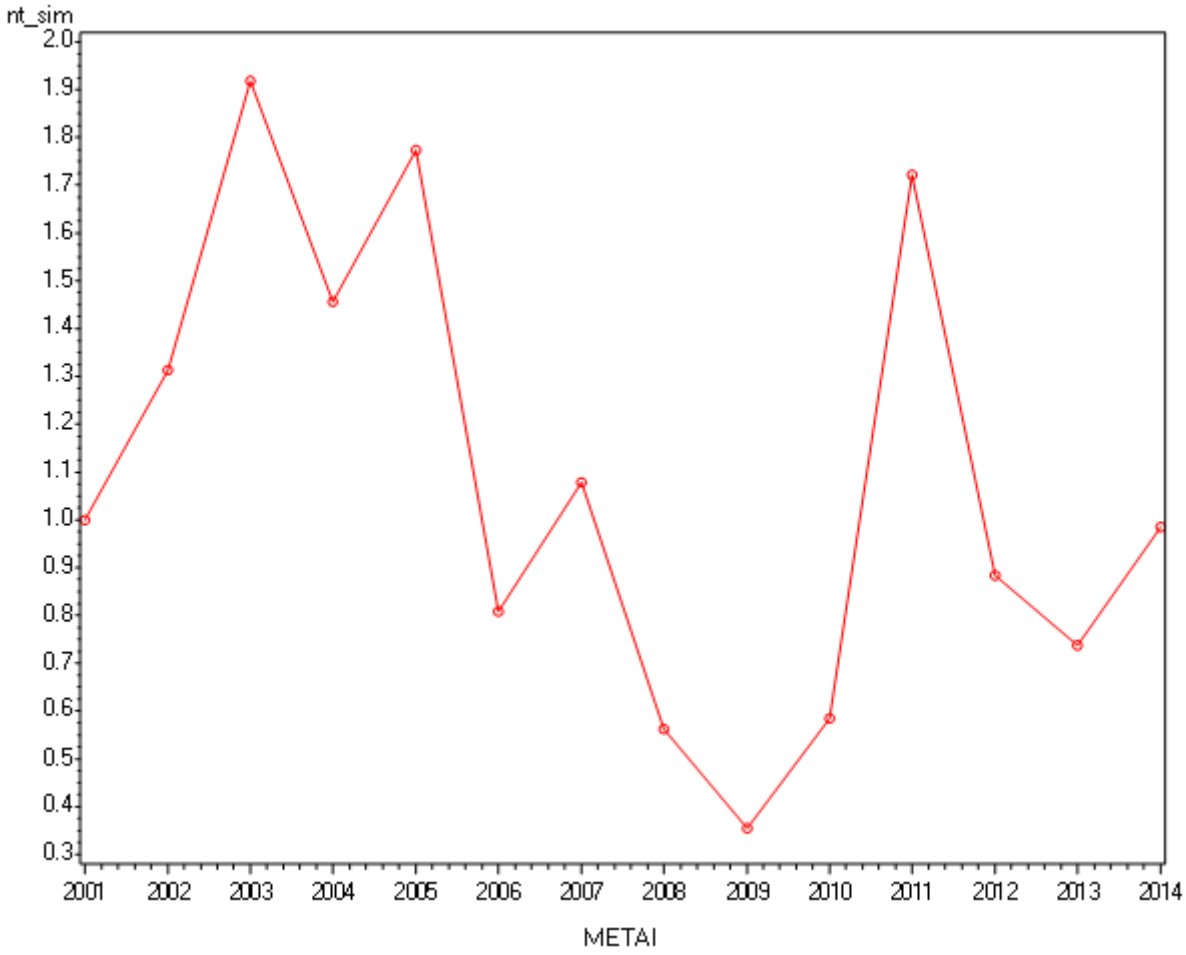
Correlations of Parameter Estimates				
Variable Parameter	ln_It MU	ln_It MA1,1	ln_It MA1,2	A_log NUM1
ln_It MU	1.000	0.018	-0.238	-0.972
ln_It MA1,1	0.018	1.000	0.138	-0.013
ln_It MA1,2	-0.238	0.138	1.000	0.264
A_log NUM1	-0.972	-0.013	0.264	1.000

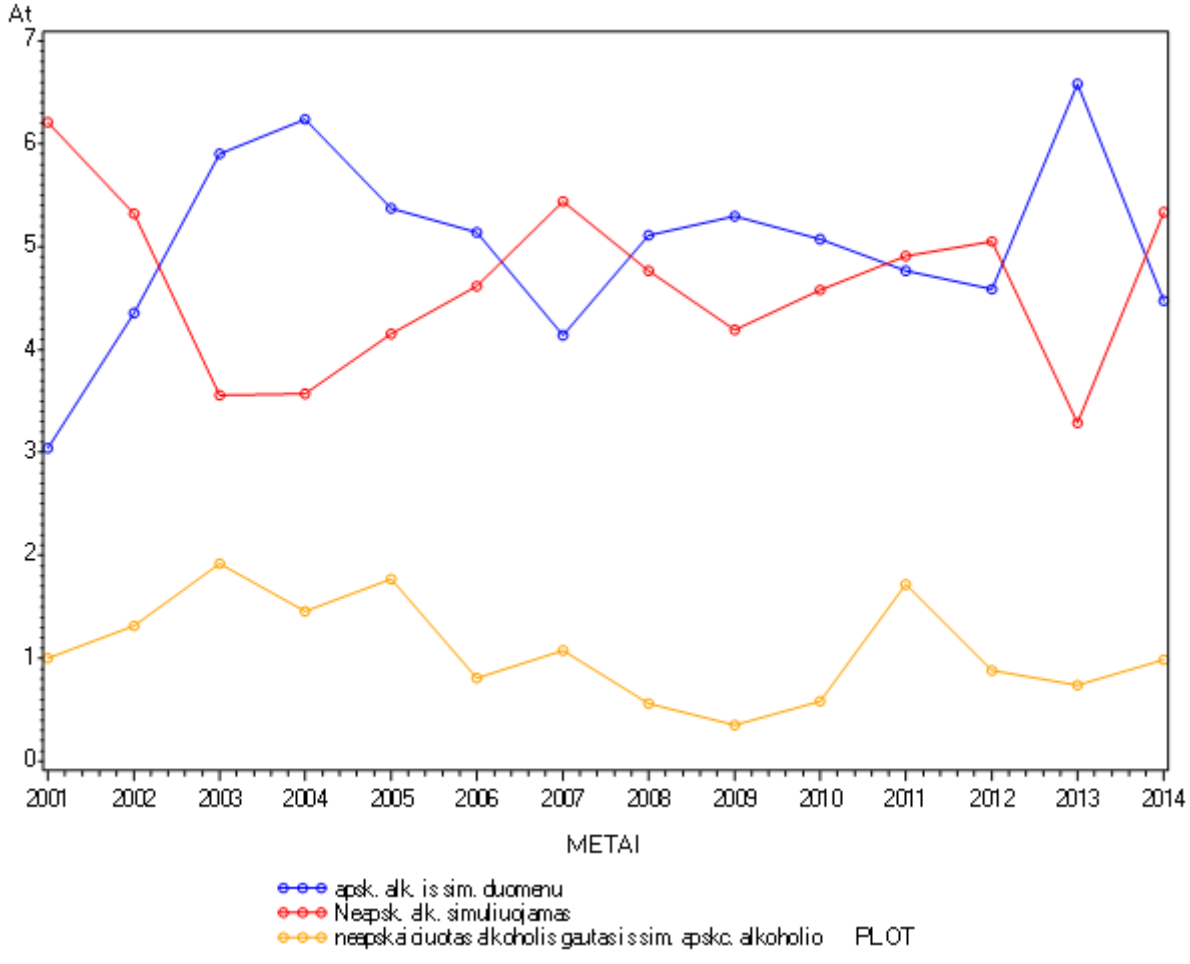
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
6	2.54	4	0.6372	0.031	0.072	-0.062	0.020	-0.307	-0.010
12	3.63	10	0.9624	-0.087	0.033	0.019	-0.106	-0.016	-0.036

Model for variable ln_It	
Estimated Intercept	7.985378

Moving Average Factors	
Factor 1:	1 + 0.27587 B**(1) + 0.62897 B**(2)

Input Number 1	
Input Variable	A_log
Overall Regression Factor	-1.3079


nt_sim
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
METAI


At
7
6
5
4
3
2
1
0
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
METAI
apsk. alk. is sim. duomenu
Neapsk. alk. simuliuojamas
neapskaiciuotas alkoholis gautas is sim. apskc. alkoholio
PLOT

neapskaičiuoto tabako vertinimo output'ai

Observations	62
Variables	13

Simple Statistics				
	burnos_navikai	kvepavimo_sist_navikai	krutines_angina	uminis_miokardo_infarktas
Mean	17.20451613	42.96774194	1000.488871	30.29048387
StD	2.90554342	5.04632683	250.061851	5.13969554

Simple Statistics			
	letine_isemine_sirdies	smegenu_infarktas	kitos_smeg_kraujag_lig
Mean	347.3038710	93.67451613	437.1366129
StD	126.1327889	11.73804878	106.8333412

Simple Statistics				
	smegen_kraujag_ligos	ateroskleroz_isk_sird_ir_smeg	kitos_kraujot_ligos	bronchitai
Mean	186.0472581	87.43629032	786.5743548	125.0122581
StD	37.5896222	14.73080460	158.0128167	44.4506545

Simple Statistics		
	emfizemos_ir_kt_obst_pl_ligos	astma
Mean	259.8758065	537.5341935
StD	52.3542473	77.4340319

Covariance Matrix			
	burnos_navikai	kvepavimo_sist_navikai	krutines_angina
burnos_navikai	8.44218	13.22938	381.12388
kvepavimo_sist_navikai	13.22938	25.46541	938.70185
krutines_angina	381.12388	938.70185	62530.92950
uminis_miokardo_infarktas	13.07583	24.06868	920.20387
letine_isemine_sirdies	57.93172	279.94006	28533.75467
smegenu_infarktas	6.98834	29.79795	2458.90758
kitos_smeg_kraujag_lig	220.62292	477.05461	25568.30695
smegen_kraujag_ligos	68.66427	154.64517	9237.30346
ateroskleroz_isk_sird_ir_smeg	25.52789	59.77488	3577.45177
kitos_kraujot_ligos	311.27255	690.50904	38328.50970
bronchitai	37.77847	116.26504	8991.13704
emfizemos_ir_kt_obst_pl_ligos	49.36360	158.79931	12233.38854
astma	139.91611	324.79569	17422.91573

Covariance Matrix			
	uminis_miokardo_infarktas	letine_isemine_sirdies	smegenu_infarktas
burnos_navikai	13.07583	57.93172	6.98834
kvepavimo_sist_navikai	24.06868	279.94006	29.79795
krutines_angina	920.20387	28533.75467	2458.90758
uminis_miokardo_infarktas	26.41647	282.01383	30.21154
letine_isemine_sirdies	282.01383	15909.48044	1303.97213
smegenu_infarktas	30.21154	1303.97213	137.78179
kitos_smeg_kraujag_lig	460.99663	10297.94637	943.92594

Covariance Matrix			
	uminis_miokardo_infarktas	letine_isemine_sirdies	smegenu_infarktas
smegen_kraujag_ligos	151.00031	4010.78516	353.18757
ateroskleroz_isk_sird_ir_smeg	56.87069	1576.54931	145.85531
kitos_kraujot_ligos	668.82905	15922.77317	1476.03002
bronchitai	99.80033	4396.84282	362.93330
emfizemos_ir_kt_obst_pl_ligos	150.70258	6210.06725	554.01566
astma	294.23747	7167.52931	736.32455

Covariance Matrix			
	kitos_smeg_kraujag_lig	smegen_kraujag_ligos	ateroskleroz_isk_sird_ir_smeg
burnos_navikai	220.62292	68.66427	25.52789
kvepavimo_sist_navikai	477.05461	154.64517	59.77488
krutines_angina	25568.30695	9237.30346	3577.45177
uminis_miokardo_infarktas	460.99663	151.00031	56.87069
letine_isemine_sirdies	10297.94637	4010.78516	1576.54931
smegenu_infarktas	943.92594	353.18757	145.85531
kitos_smeg_kraujag_lig	11413.36280	3929.05715	1509.61833
smegen_kraujag_ligos	3929.05715	1412.97970	542.06798
ateroskleroz_isk_sird_ir_smeg	1509.61833	542.06798	216.99660
kitos_kraujot_ligos	16827.19452	5866.75189	2271.37176
bronchitai	3552.75882	1273.38329	491.11058
emfizemos_ir_kt_obst_pl_ligos	4787.64652	1757.42151	694.00924
astma	7678.05991	2674.26131	1061.81856

Covariance Matrix				
	kitos_kraujot_ligos	bronchitai	emfizemos_ir_kt_obst_pl_ligos	astma
burnos_navikai	311.27255	37.77847	49.36360	139.91611
kvepavimo_sist_navikai	690.50904	116.26504	158.79931	324.79569
krutines_angina	38328.50970	8991.13704	12233.38854	17422.91573
uminis_miokardo_infarktas	668.82905	99.80033	150.70258	294.23747
letine_isemine_sirdies	15922.77317	4396.84282	6210.06725	7167.52931
smegenu_infarktas	1476.03002	362.93330	554.01566	736.32455
kitos_smeg_kraujag_lig	16827.19452	3552.75882	4787.64652	7678.05991
smegen_kraujag_ligos	5866.75189	1273.38329	1757.42151	2674.26131
ateroskleroz_isk_sird_ir_smeg	2271.37176	491.11058	694.00924	1061.81856
kitos_kraujot_ligos	24968.05024	5335.13622	7296.02409	11491.41086
bronchitai	5335.13622	1975.86068	2106.29825	2656.08567
emfizemos_ir_kt_obst_pl_ligos	7296.02409	2106.29825	2740.96721	3512.28254
astma	11491.41086	2656.08567	3512.28254	5996.02929

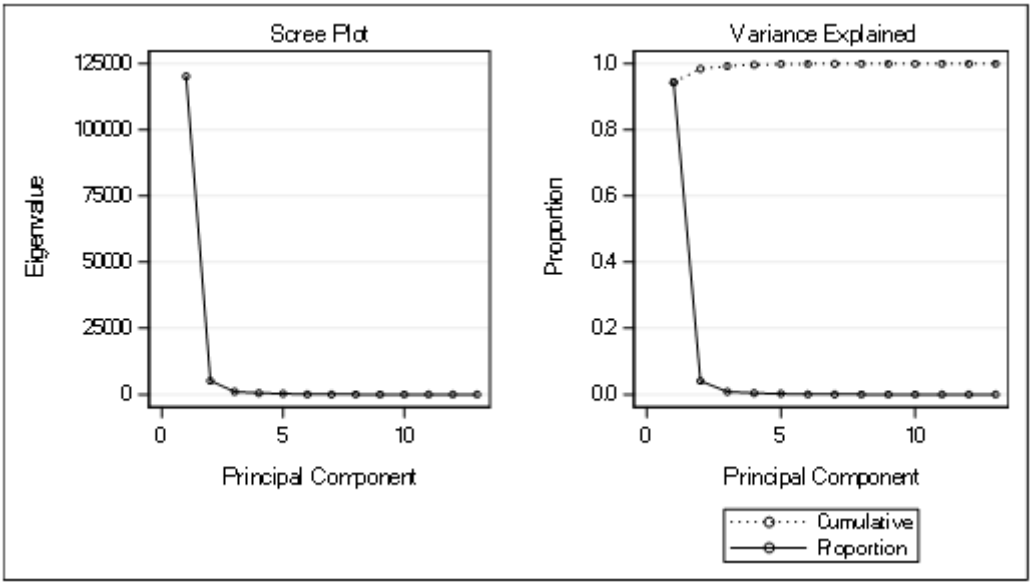
Total Variance	127362.76233

Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	120233.136	115058.176	0.9440	0.9440
2	5174.960	4125.167	0.0406	0.9847
3	1049.794	539.250	0.0082	0.9929
4	510.543	212.137	0.0040	0.9969
5	298.406	252.766	0.0023	0.9992
6	45.640	20.684	0.0004	0.9996
7	24.956	11.916	0.0002	0.9998
8	13.040	7.165	0.0001	0.9999
9	5.875	1.275	0.0000	0.9999
10	4.600	3.866	0.0000	1.0000
11	0.735	0.046	0.0000	1.0000
12	0.689	0.302	0.0000	1.0000
13	0.387		0.0000	1.0000

Eigenvectors							
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
burnos_navikai	0.004545	0.030339	-.011160	-.007254	0.004981	-.018448	0.066840
kvepavimo_sist_navikai	0.011138	0.040774	-.000791	0.007925	0.020974	0.001457	0.011428
krutines_angina	0.719655	-.098681	-.289589	-.252696	-.548399	0.150757	-.030924
uminis_miokardo_infarktas	0.010833	0.037015	-.024353	0.007339	0.031866	0.086283	0.053257
letine_isemine_sirdies	0.326496	-.761377	0.024694	0.333047	0.325395	-.282076	-.082823
smegenu_infarktas	0.028708	-.038017	0.064947	0.153735	0.046357	0.406038	0.127292
kitos_smeg_kraujag_lig	0.297946	0.361025	-.082783	-.085530	0.382859	-.217271	-.578713
smegen_kraujag_ligos	0.107006	0.050979	-.056093	0.045752	0.024194	-.322118	0.639196
ateroskleroz_isk_sird_ir_smeg	0.041531	0.016165	0.000540	0.065732	-.006090	0.019643	0.288566
kitos_kraujot_ligos	0.446414	0.423294	-.039514	0.138560	0.465990	0.075892	0.336769
bronchitai	0.104517	-.098070	0.578851	-.705591	0.095554	-.255424	0.121630
emfizemos_ir_kt_obst_pl_ligos	0.142068	-.157080	0.369131	-.054861	0.238767	0.687672	-.019774
astma	0.205114	0.241869	0.653992	0.517324	-.399070	-.172642	-.114885

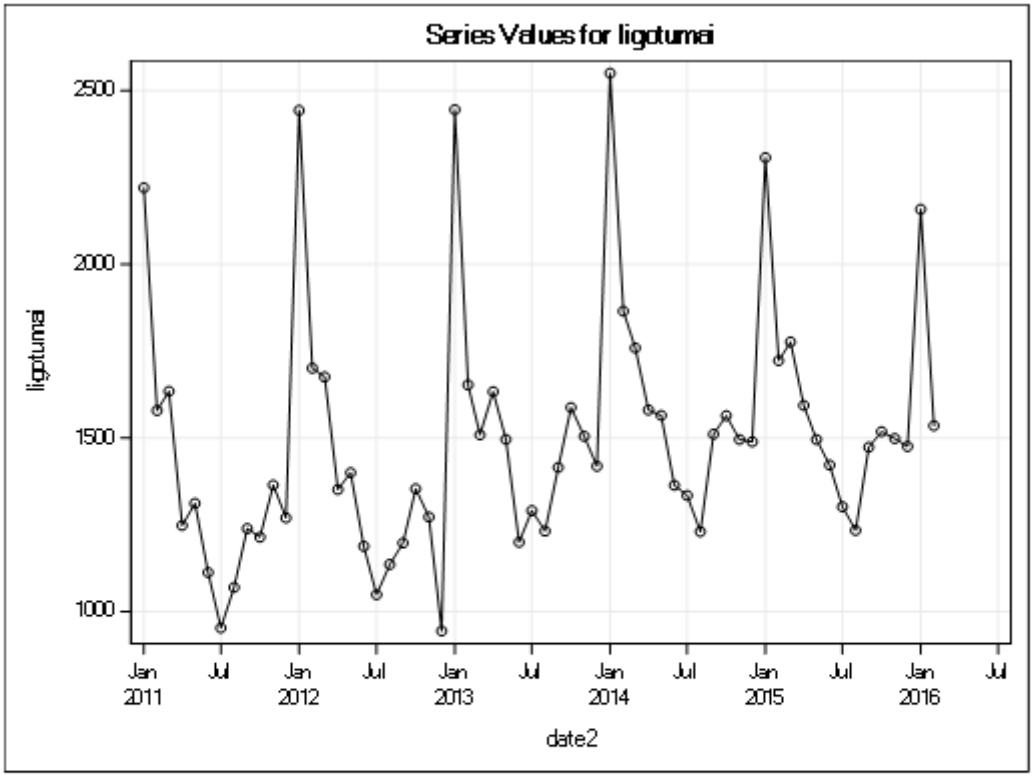
Eigenvectors						
	Prin8	Prin9	Prin10	Prin11	Prin12	Prin13
burnos_navikai	0.145601	0.114999	0.201072	-.066098	-.135731	0.946785

Eigenvectors						
	Prin8	Prin9	Prin10	Prin11	Prin12	Prin13
kvepavimo_sist_navikai	0.154559	0.232004	0.361718	-.368517	0.807180	-.040975
krutines_angina	0.005015	-.007010	0.001314	-.007029	0.009075	0.002975
uminis_miokardo_infarktas	0.179617	-.035199	0.842677	0.365823	-.249717	-.216284
letine_isemine_sirdies	-.056269	0.032760	0.069639	-.021025	-.014974	0.010582
smegenu_infarktas	-.619861	-.401913	0.088687	0.289939	0.333078	0.195004
kitos_smeg_kraujag_lig	0.104188	-.092618	-.089712	0.397419	0.210038	0.092113
smegen_kraujag_ligos	0.407977	-.410905	-.159427	0.248765	0.212101	0.014815
ateroskleroz_isk_sird_ir_smeg	-.126790	0.762771	-.159772	0.523964	0.109493	-.007106
kitos_kraujot_ligos	-.210671	0.098715	0.048532	-.380853	-.234180	-.089919
bronchitai	-.220892	0.000691	0.092101	-.001539	0.007655	0.005307
emfizemos_ir_kt_obst_pl_ligos	0.498039	0.003232	-.188013	0.044731	0.007530	-.011031
astma	0.025224	-.007564	0.027028	0.005250	-.019618	-.001363



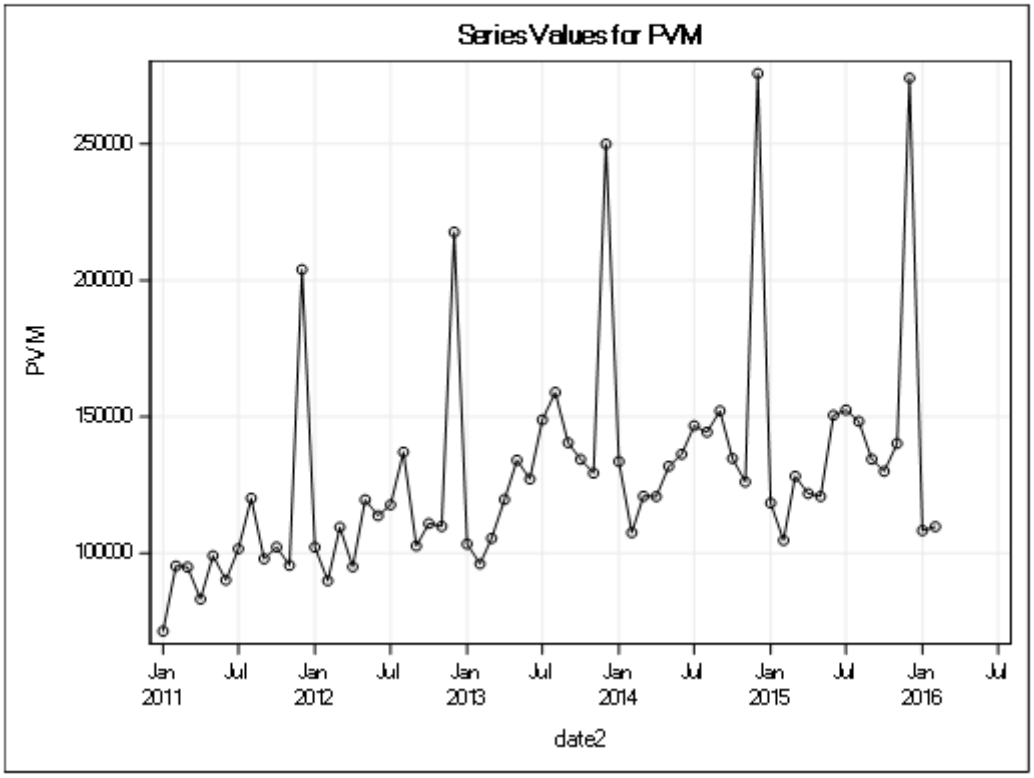
Input Data Set	
Name	WORK.TABAKAS
Label	
Time ID Variable	date2
Time Interval	MONTH
Length of Seasonal Cycle	12

Variable Information	
Name	ligotumai
Label	
First	JAN2011
Last	FEB2016
Number of Observations Read	62



Variable Information	
Name	PVM
Label	
First	JAN2011

Variable Information	
Last	FEB2016
Number of Observations Read	62



Name of Variable = l_ligotumai	
Mean of Working Series	7.291171
Standard Deviation	0.212036
Number of Observations	62

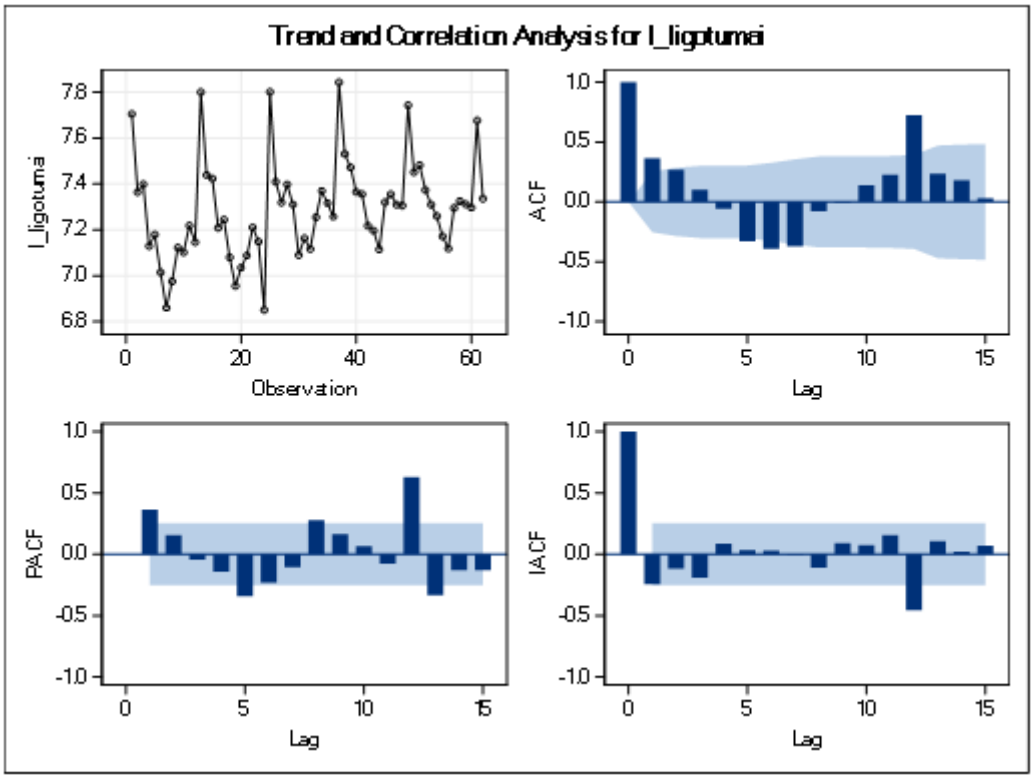
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
6	33.01	6	<.0001	0.365	0.268	0.102	-0.056	-0.332	-0.394
12	90.65	12	<.0001	-0.370	-0.080	-0.007	0.141	0.225	0.726

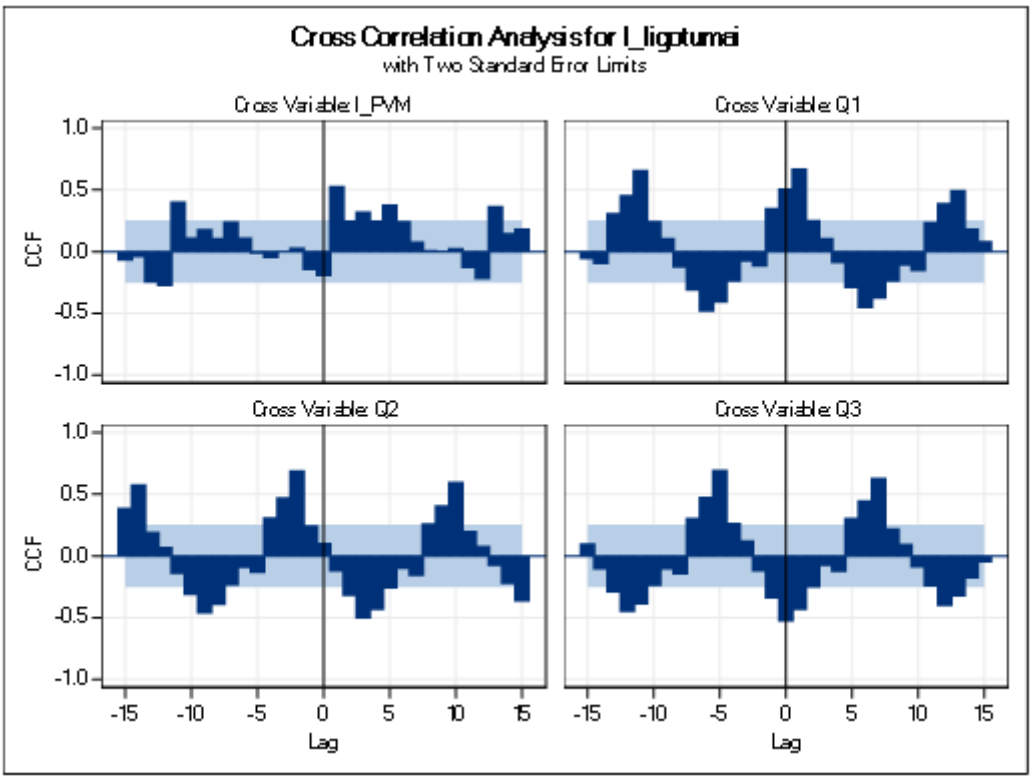
Correlation of l_ligotumai and l_PVM	
Variance of input =	0.067465
Number of Observations	62

Correlation of l_ligotumai and Q1	
Variance of input =	0.199011
Number of Observations	62

Correlation of l_ligotumai and Q2	
Variance of input =	0.183403
Number of Observations	62

Correlation of l_ligotumai and Q3	
Variance of input =	0.183403
Number of Observations	62





Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr> |t|	Lag	Variable	Shift
MU	11.28475	0.60229	18.74	<.0001	0	l_ligotumai	0
MA1,1	0.77507	0.18979	4.08	0.0002	1	l_ligotumai	0
MA1,2	-0.49915	0.17862	-2.79	0.0072	2	l_ligotumai	0
AR1,1	0.51730	0.19767	2.62	0.0115	1	l_ligotumai	0
AR1,2	0.35406	0.19664	1.80	0.0775	2	l_ligotumai	0
NUM1	-0.35010	0.05245	-6.67	<.0001	0	l_PVM	0
NUM2	0.29560	0.02952	10.01	<.0001	0	Q1	0
NUM3	0.04680	0.03653	1.28	0.2057	0	Q2	0
NUM4	-0.12519	0.02924	-4.28	<.0001	0	Q3	0

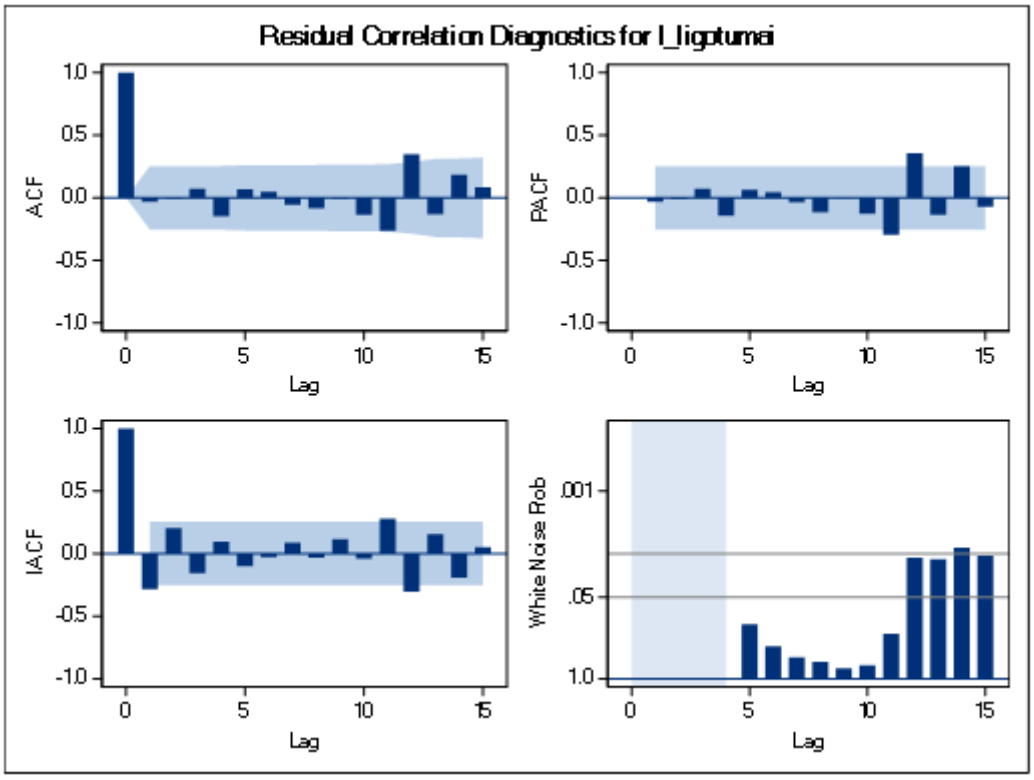
Constant Estimate	1.451572
Variance Estimate	0.014657
Std Error Estimate	0.121065
AIC	-77.5929

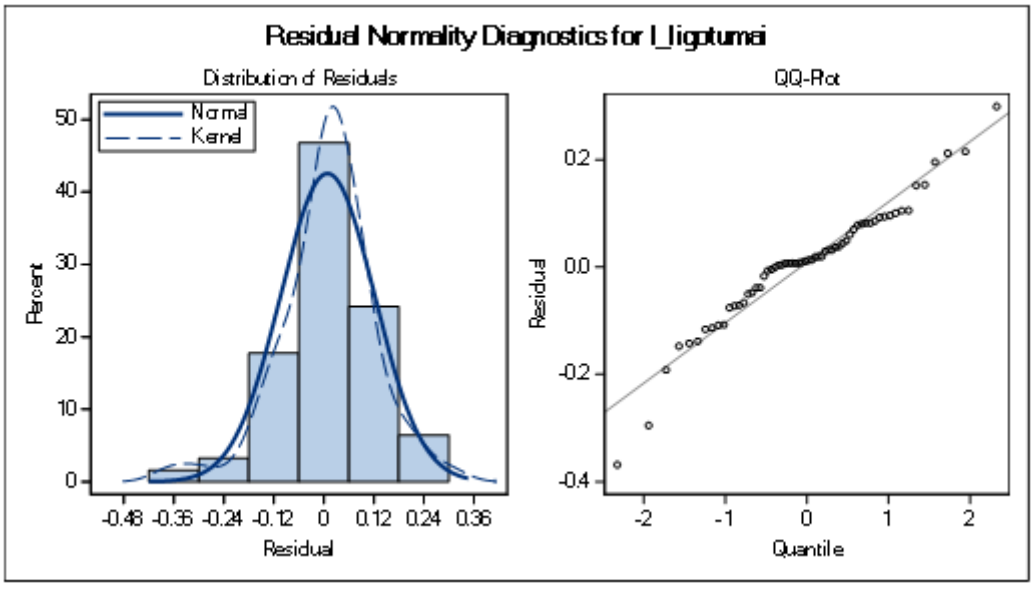
SBC	-58.4487
Number of Residuals	62

** AIC and SBC do not include log determinant.*

Correlations of Parameter Estimates									
Variable Parameter	l_ligotumai MU	l_ligotumai MA1,1	l_ligotumai MA1,2	l_ligotumai AR1,1	l_ligotumai AR1,2	l_PVM NUM1	Q1 NUM2	Q2 NUM3	Q3 NUM4
l_ligotumai MU	1.000	0.120	-0.223	-0.038	0.042	-0.995	0.140	-0.018	0.205
l_ligotumai MA1,1	0.120	1.000	-0.652	0.706	-0.628	-0.117	-0.073	0.054	0.006
l_ligotumai MA1,2	-0.223	-0.652	1.000	-0.509	0.626	0.223	-0.011	0.045	-0.034
l_ligotumai AR1,1	-0.038	0.706	-0.509	1.000	-0.879	0.033	-0.025	0.083	0.024
l_ligotumai AR1,2	0.042	-0.628	0.626	-0.879	1.000	-0.048	0.035	-0.061	-0.034
l_PVM NUM1	-0.995	-0.117	0.223	0.033	-0.048	1.000	-0.182	-0.012	-0.219
Q1 NUM2	0.140	-0.073	-0.011	-0.025	0.035	-0.182	1.000	0.606	0.210
Q2 NUM3	-0.018	0.054	0.045	0.083	-0.061	-0.012	0.606	1.000	0.620
Q3 NUM4	0.205	0.006	-0.034	0.024	-0.034	-0.219	0.210	0.620	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr> ChiSq	Autocorrelations					
6	2.47	2	0.2906	-0.021	0.002	0.084	-0.134	0.082	0.063
12	19.18	8	0.0139	-0.041	-0.067	0.007	-0.122	-0.252	0.358
18	27.40	14	0.0171	-0.121	0.191	0.093	-0.058	0.184	0.005
24	30.28	20	0.0655	-0.002	0.074	-0.071	-0.021	-0.124	0.051


Residual Correlation Diagnostics for l_ligptumai
ACF
PACF
IACF
White Noise Prob
Lag


Residual Normality Diagnostics for l_ligptumai
Distribution of Residuals
Normal
Kernel
Percent
Residual
QQ-Plot
Residual
Quantile

Model for variable l_ligotumai	
Estimated Intercept	11.28475

Autoregressive Factors	
Factor 1:	1 - 0.5173 B**(1) - 0.35406 B**(2)

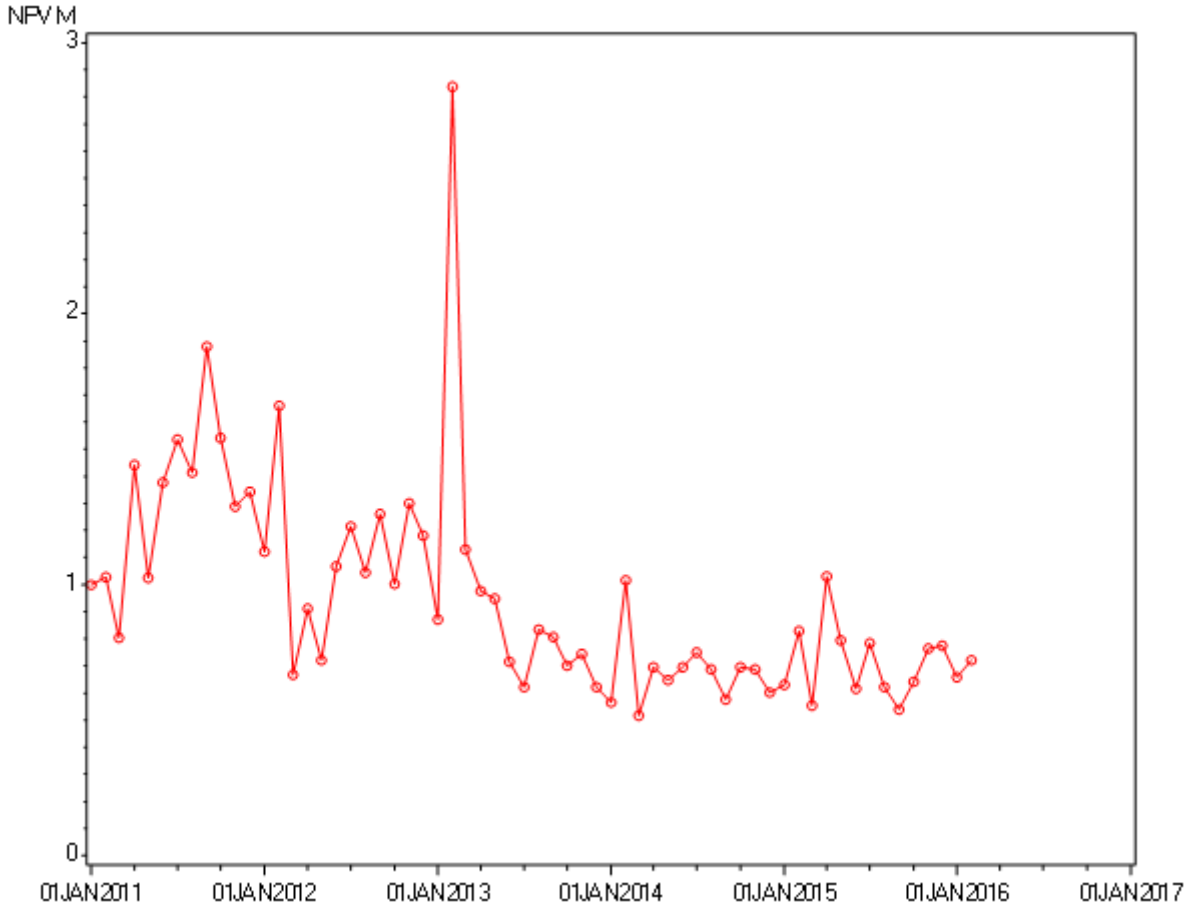
Moving Average Factors	
Factor 1:	1 - 0.77507 B**(1) + 0.49915 B**(2)

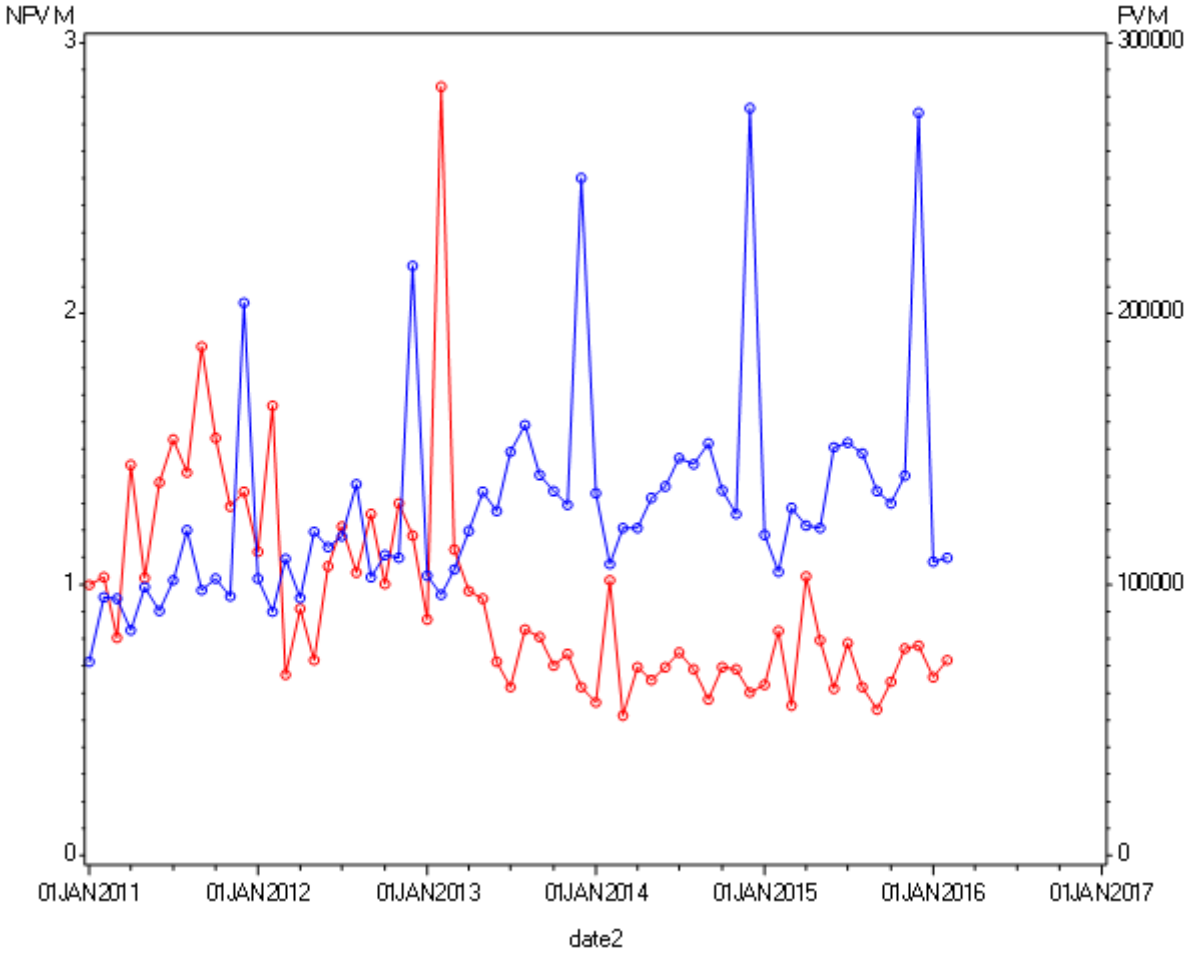
Input Number 1	
Input Variable	l_PVM
Overall Regression Factor	-0.3501

Input Number 2	
Input Variable	Q1
Overall Regression Factor	0.295595

Input Number 3	
Input Variable	Q2
Overall Regression Factor	0.046799

Input Number 4	
Input Variable	Q3
Overall Regression Factor	-0.12519


NPVM
3
2
1
0
01JAN2011
01JAN2012
01JAN2013
01JAN2014
01JAN2015
01JAN2016
01JAN2017


NPVM
PVM
3
2
1
0
300000
200000
100000
0
01JAN2011
01JAN2012
01JAN2013
01JAN2014
01JAN2015
01JAN2016
01JAN2017
date2