

Nelegalaus alkoholio suvartojimo įvertinimas: metodika ir pavyzdžiai

Tadas Telksnys

2014 m. lapkričio 13 d.

1 Nörstrom siūloma metodika ir jos apibendrinimas

Pažymėkime A_t – oficialų alkoholio suvartojimą t -taisiais metais, I_t – kokio nors indikatoriaus, priklausančio nuo bendro alkoholio suvartojimo, reikšmę t -taisiais metais. Nörstrom siūlo naudoti semilogaritminį ARIMAX modelį:

$$\ln I_t = \beta A_t + \ln N_t, \quad (1)$$

kur N_t yra, bendru atveju, ARIMA(p, d, q) procesas. Modelis (1) natūraliai apibendrinamas panaudojant bendriausią ARIMAX formą (sekdami ankstesne lygybe, konstantos (vidurkio) nenaudosime), t.y.:

$$\nabla^d \ln I_t = \omega_0 \nabla^d A_t - \omega_1 \nabla^d A_{t-1} - \dots - \omega_m \nabla^d A_{t-m} + \ln N_t, \quad (2)$$

čia $\ln N_t$ – ARIMA(p, d, q) procesas; $\nabla X_t = X_t - X_{t-1}$ – diferencijavimo operatorius. Toks modelis leidžia lanksčiau modeliuoti $\ln I_t$ priklausomybę nuo A_t , kadangi galima nagrinėti ne pačias laiko eilučių reikšmes, o jų skirtumus, be to, galima įtraukti daugiau A_t narių, t.y. leidžiama padaryti, kad $\ln I_t$ priklausytų nuo A_t nebūtinai tais pačiais metais, o keliais metais atgal.

Dar bendresnis modelis yra *transfer function* modelis, kuris leidžia naudoti racionaliąją ryšio funkciją:

$$\nabla^d \ln I_t = \frac{\omega_0 \nabla^d A_t - \omega_1 \nabla^d A_{t-1} - \dots - \omega_m \nabla^d A_{t-m}}{\nabla^d A_t - \delta_1 \nabla^d A_{t-1} - \delta_2 \nabla^d A_{t-2} - \dots - \delta_n \nabla^d A_{t-n}} + N_t, \quad (3)$$

kur galioja anksčiau minėtos prielaidos.

Kad pernelyg nenukryptume nuo literatūroje plačiai naudojamos metodikos, (3) modeliu nesinaudosime. Atsižvelgiant į žemiau pateiktą analizę ir turimos imties pobūdį neatrodo, kad šis modelis pagerintų rezultatus.

2 Nelegalaus alkoholio suvartojimo įvertinimas

Pateiktai analizei naudotasi duomenimis iš WHO HFADB offline versijos. Naudojami du indikatoriai – hom_t žymėsime kintamąjį „homicide and intentional injury, 0–64, per 100 000“, $psych_t$ – kintamąjį „Incidence of alcoholic psychosis per 100 000“. Šie kintamieji pasirinkti dėl to, kad turima daugiau duomenų palyginti su kitais indikatoriais – abi laiko eilutės registruotos nuo 1985 iki 2010 metų.

Įprasta į laiką neatsižvelgianti koreliacinė analizė pateikta 1 pav. Kaip matome, norimo didelio ryšio tarp kintamųjų nėra, kuomet į juos žiūrima bendrai. Tai gali pasikeisti kai atsižvelgiama į laiką, tačiau bendra tendencija vistiek išlieka, kad reikėtų stipriau (bet ne per stipriai) koreliuojančių kintamųjų.

Tolimesnė analizė parodo, kad visos trys tiriamos laiko eilutės yra nestacionarios. Kad jas stacionarizuoti, pakanka vieno diferencijavimo, taigi $d = 1$ ir toliau visuose skaičiavimuose dirbsime su skirtumais $\nabla \ln I_t = \ln \frac{I_t}{I_{t-1}}$, $\nabla A_t = A_t - A_{t-1}$. Be to, modeliuojant tiek $\ln I_t$, tiek A_t įprastiniais ARIMA modeliais gaunama jų eilė ARIMA(0, 1, 0). Tai šiek tiek komplikuoja analizės interpretavimą, kadangi tai reiškia, jog skirtumai tarp gretimų laiko eilučių reikšmių yra baltas triukšmas, kitaip tariant, atsitiktinis skaičius.

Koreliacijų $\text{cor}(I_{t+\tau}, A_t)$ grafikai pateikiami 2 pav.

Įvertinus turimus duomenis, parenkami geriausi (2) tipo modeliai abiemis indikatoriams. Gauname:

$$\ln \frac{\text{psych}_t}{\text{psych}_{t-1}} = 0.06924 (A_t - A_{t-1}) + \ln N_t, \quad (4)$$

koeficiento $\omega_0 = 0.06924$ reikšmingumo lygmuo $p = 0.0766$, taigi koeficientas nereikšmingas, tačiau balansuoja ant reikšmingumo ribos. Modelio su reikšmingesniu koeficientu rasti nepavyko. N_t išreiškiamas:

$$N_t = \frac{\text{psych}_t}{\text{psych}_{t-1}} \exp(-0.06924 (A_t - A_{t-1})). \quad (5)$$

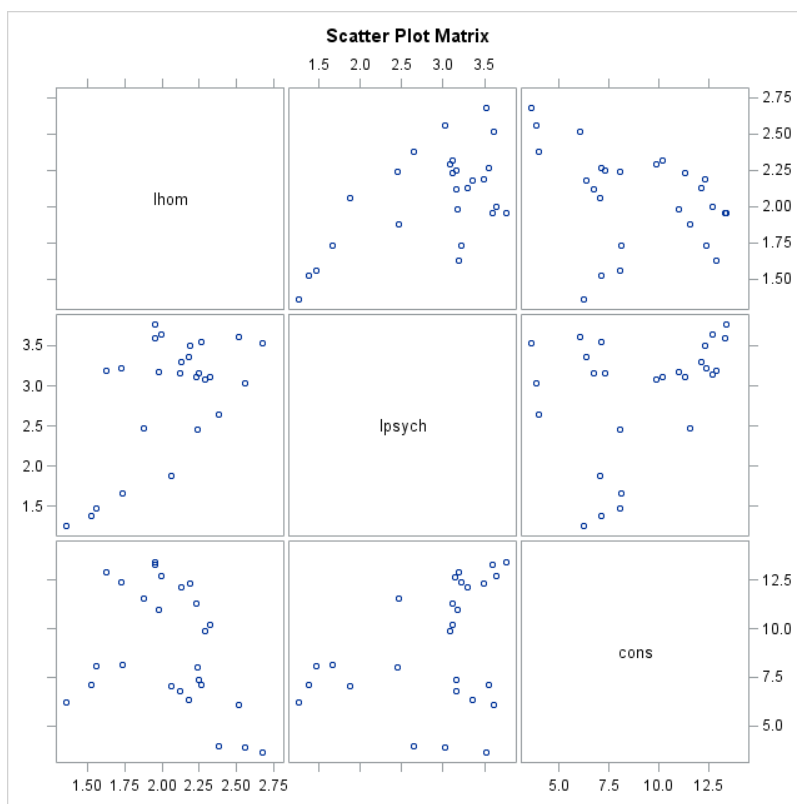
Kadangi N_1 rasti reikia žinoti A_0, psych_0 , tai skaičiuoti galime pradėti tik nuo N_2 . Padaliję visas $N_t, t > 2$ reikšmes iš N_2 gauname bedimensinį nelagalaus alkoholio suvartojimo įvertį, pavaizduotą 3 (a) pav. ir 1 lent.

$$\ln \frac{\text{hom}_t}{\text{hom}_{t-1}} = -0.04458 (A_{t-2} - A_{t-3}) + \ln N_t, \quad (6)$$

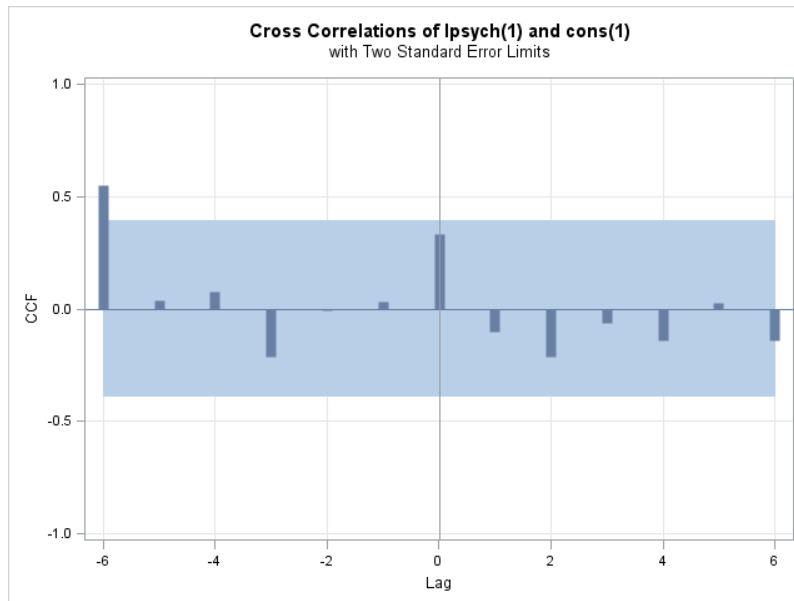
Analogiškai kaip aukščiau aptartu atveju, išsireiškiame N_t . Dabar pirmoji reikšmę kurią galime rasti yra N_4 , o toliau vertinimas vyksta taip pat. Rezultatai pavaizduoti 3 (b) pav. ir 1 lent.

year	alk	alkhom
1985	–	–
1986	100%	–
1987	74%	–
1988	118%	100%
1989	139%	112%
1990	119%	143%
1991	186%	124%
1992	100%	108%
1993	147%	124%
1994	162%	93%
1995	122%	84%
1996	98%	76%
1997	68%	109%
1998	117%	97%
1999	84%	94%
2000	106%	112%
2001	105%	104%
2002	110%	81%
2003	95%	129%
2004	124%	93%
2005	123%	106%
2006	118%	85%
2007	117%	96%
2008	84%	101%
2009	66%	82%
2010	99%	89%
2011	94%	–

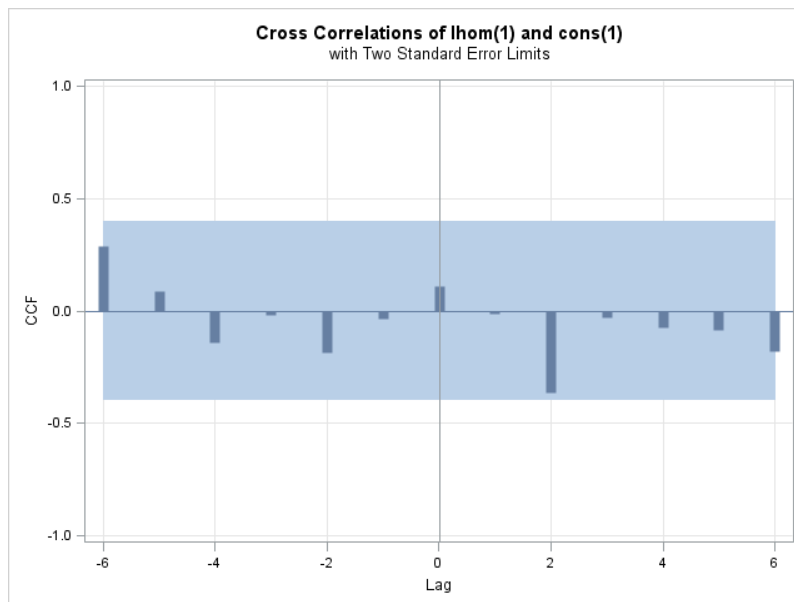
1 lentelė: Santykinis nelegalaus alkoholio įvertinimas pagal Nörstrom metodiką. Statistiškai modeliai nėra labai patikimi, taigi vertinti reikėtų atsargiai.



1 pav.: Sklaidos diagrama tarp analizuojamų kintamųjų. Aiškiaus ryšio tarp $\ln hom_t$, $\ln psych_t$ ir A_t reikšmių nematyti. Pirsono koreliacijos: $\rho_p(\ln hom_t, A_t) = -0.369, p = 0.0636$; $\rho_p(\ln psych_t, A_t) = 0.332, p = 0.09$. Spirmeno koreliacijos: $\rho_s(\ln hom_t, A_t) = -0.455, p = 0.0195$; $\rho_s(\ln psych_t, A_t) = 0.352, p = 0.071$. Tiesinis ryšys silpnas ir nereikšmingas. Monotoniškumo ryšys tarp $\ln hom_t$ ir A_t vidutinio stiprumo ir reikšmingas, tačiau neigiamas (augant alkoholio vartojimui, žmogžudysčių mažėja); likęs monotoniškumo ryšys silpnas ir nereikšmingas.

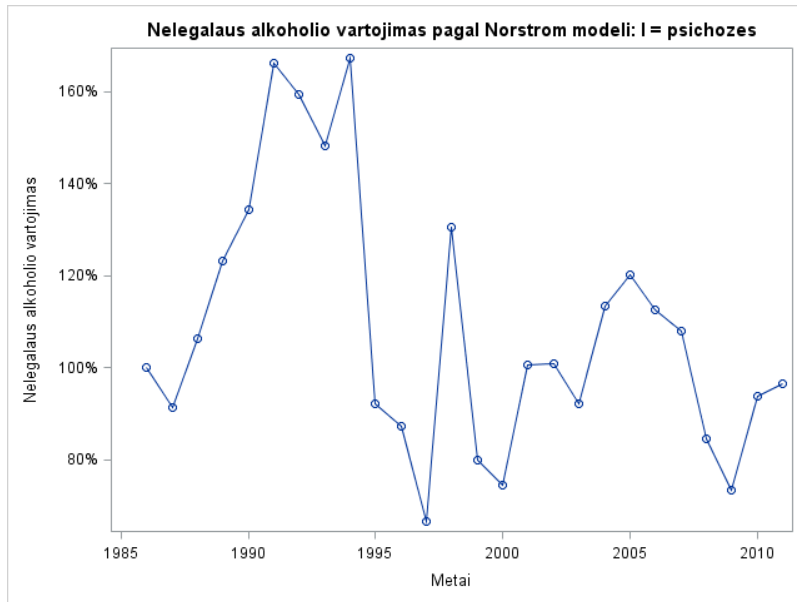


(a)

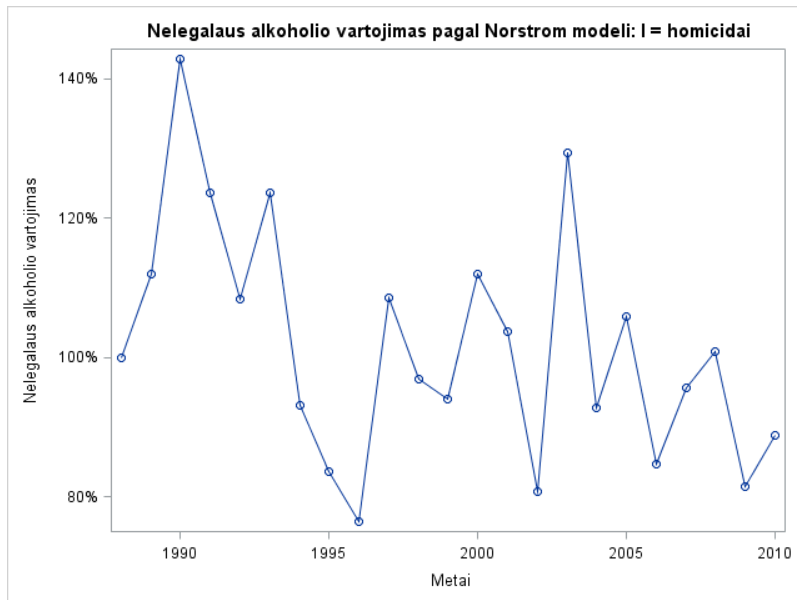


(b)

2 pav.: Koreliacijos įvertinant laiką. Tamsus regionas – pasikliautinis intervalas. Jeigu koreliacijos reikšmė jo neviršija, tuomet koreliacija nereikšminga. Vienintelė reikšminga koreliacija tarp gana toli laike nutolusių reikšmių, taigi tokios eilės modelis bus sunkiai suprantamas.



(a)



(b)

3 pav.: Santykinis nelegalaus alkoholio įvertinimas pagal Nörstrom metodiką. Statistiškai modeliai nėra labai patikimi, taigi vertinti reikėtų atsargiai.