VILNIAUS UNIVERSITETAS

Andrius
BUTEIKIS

Daugiamačiai jungtimis grįstų sveikareikšmių laiko eilučių modeliai: teorija ir taikymai

DAKTARO DISERTACIJOS SANTRAUKA

Gamtos mokslai,
Matematika (N 001)

VILNIUS 2020

Disertacija rengta 2016–2020 metais Vilniaus universitete.

Mokslinis vadovas:
prof. habil. dr. Remigijus Leipus (Vilniaus universitetas, gamtos mokslai, matematika - N 001).

Gynimo taryba:
Pirmininkas – **prof. habil. dr. Alfredas Račkauskas** (Vilniaus universitetas, gamtos mokslai, matematika - N 001).
Nariai:
prof. habil. dr. Vydas Čekanavičius (Vilniaus universitetas, gamtos mokslai, matematika - N 001).
prof. Márton Ispány (Debreceno universitetas (Vengrija), gamtos mokslai, matematika - N 001).
prof. habil. dr. Kęstutis Kubilius (Vilniaus universitetas, gamtos mokslai, matematika - N 001).
doc. dr. Viktor Skorniakov (Vilniaus universitetas, gamtos mokslai, matematika - N 001).

Disertacija ginama viešame Gynimo tarybos posėdyje 2020 m. spalio mėn. 30 d. 14.00 val. VU Matematikos ir informatikos fakulteto 102 auditorijoje.
Adresas: Naugarduko g. 24, LT-03225, Vilnius, Lietuva, tel. +370 5 219 3050; el. paštas mif@mif.vu.lt.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu:
https://www.vu.lt/lt/naujienos/ivykiu-kalendorius

VILNIUS UNIVERSITY

Andrius
BUTEIKIS

Multivariate copula-based integer-valued time series models: theory and applications

SUMMARY OF DOCTORAL DISSERTATION

Life sciences,
Mathematics (N 001)

VILNIUS 2020

This dissertation was written between 2016 and 2020 at Vilnius University.

Academic supervisor:
Prof. Habil. Dr. Remigijus Leipus (Vilnius University, Life sciences, Mathematics - N 001).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:
Chairman - **Prof. Habil. Dr. Alfredas Račkauskas** (Vilnius University, Life sciences, Mathematics - N 001).
Members:
Prof. Habil. Dr. Vydas Čekanavičius (Vilnius University, Life sciences, Mathematics - N 001).
Prof. Dr. Márton Ispány (University of Debrecen, Life sciences, Mathematics - N 001).
Prof. Habil. Dr. Kęstutis Kubilius (Vilnius University, Life sciences, Mathematics - N 001).
Assoc. Prof. Dr. Viktor Skorniakov (Vilnius University, Life sciences, Mathematics - N 001).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 2:00 PM on the 30th of October, 2020 in Room 102 of the Faculty of Mathematics and Informatics, Vilnius University.
Address: Naugarduko st. 24, LT-03225, Vilnius, Lithuania.
Tel. +370 5 219 3050; e-mail: mif@mif.vu.lt.

The text of this dissertation can be accessed at the Library of Vilnius University, as well as on the website of Vilnius University: https://www.vu.lt/lt/naujienos/ivykiu-kalendorius

Summary

The main objective of this thesis is to provide some contributions to the analysis of integer-valued time series models. The main focus is on such models, where the joint innovation distribution can be described by a copula.

Firstly, a class of bivariate integer-valued autoregressive processes of order 1 (BINAR(1)) with copula-joint innovations are analysed. The existing literature on such models is extended by providing a two-step estimation method, where the BINAR(1) model parameters are estimated separately from the dependence parameter of the copula. This procedure reduces the computational time, while providing a similar accuracy of the copula dependence parameter estimate as the conditional maximum likelihood estimation method for all of the model parameters.

Secondly, a univariate integer-valued autoregressive process for seasonality with period d (SINAR$(1)_d$) is introduced, which allows the intra-seasonal dependence of the innovations to be described by a copula. Such a univariate process can also be written as a multivariate specification, for which the BINAR(1) is a special case by taking the seasonal period $d = 2$. The multivariate specification allows the use of an estimation method based on least-squares, which can be generalized to account for the residual dependence. This method is then compared with a two-step likelihood-based estimation method, where emphasis is now put on the accuracy of the parameters of the SINAR$(1)_d$ model itself, rather than the dependence parameter only.

Finally, the estimation methods of the above processes are illustrated using Monte Carlo simulation. Furthermore, the processes studied in this thesis are illustrated using empirical applications in the context of loan defaults (for the BINAR(1) process) and crime data (for the SINAR$(1)_d$ process).

Santrauka

1 Tiriamoji problematika ir jos aktualumas

Sveikareikšmės laiko eilutės stebimos įvairiose ekonominėse ir finansinėse sferose – vertybinių popierių transakcijų skaičius per dieną, nusikaltimų skaičius mieste per valandą, draudimo išmokų skaičius per metus įmonėje, banko išduotų nemokių paskolų skaičius per savaitę, virusu užsikrėtusių žmonių skaičius per dieną ir kt. Skirtingos sveikareikšmės laiko eilutės gali būti tarpusavyje priklausomos. Ši priklausomybė gali būti nusakoma jungties (arba kopulos) funkcija. Pasinaudojant kopulomis galima atskirai modeliuoti marginaliuosius skirstinius, kurie gali būti iš skirtingų skirstinių šeimų, ir jų priklausomybės struktūrą, kurią galima nusakyti jungties funkcija. Genest ir Nešlehová (2007) analizavo jungties funkcijų taikymą diskretiems atsitiktiniams dydžiams (a.d.), lyginant su jų taikymu tolydiems a.d.

Sveikareikšmės laiko eilutės taip pat gali priklausyti nuo savo praeities reikšmių. Šį sąryšį galima nusakyti pirmos eilės sveikareikšmių dydžių autoregresiniu modeliu (INAR(1)), kurį pasiūlė McKenzie (1986) ir Al-Osh ir Alzaid (1987), nusakomas lygtimi:

$$Y_t = \phi \circ Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{1}$$

čia „$\circ$“ vadinamas binominiu retinimo operatoriumi, tokiu, kad $\phi \circ Y_{t-1} = \sum_{i=1}^{Y_{t-1}} B_{i,t}$, kur visiems $t \in \mathbb{Z}$, $\{B_{i,t}, i \in \mathbb{Z}\}$ yra nepriklausomos vienodai pasiskirsčiusios (n.v.p.) Bernulio a.d. sekos su vidurkiu $\phi \in [0,1]$, kurios tarpusavyje nepriklauomos ir nepriklauso nuo n.v.p. neneigiamų sveikareikšmių a.d. sekos $\{\varepsilon_t, t \in \mathbb{Z}\}$. ε_t nepriklauso nuo Y_{t-k}, kai $k > 0$. Dažniausiai laikoma, kad ε_t skirstinys yra Puasono arba neigiamo binominio.

Nagrinėjant diskrečių laiko eilučių porą galima taikyti dvimačius sveikareikšmių dydžių autoregresinius modelius (BINAR), su kuriais galima atsižvelgti į laiko eilučių autokoreliaciją ir į tai, kad laiko eilutes sudaro diskretūs dydžiai. Be to, jungties (*angl.* copula) funkcijos gali būti panaudojamos nusakyti BINAR(1) modelio inovacijas - Karlis ir Pedeli (2013) panaudojo Franko kopulą ir normaliąją kopulą modeliuojant BINAR(1) modelio inovacijų priklausomybę.

Be to, dauguma laiko eilučių pasižymi sezoniškumu (pvz. metų laikai, savaitės diena, rytas/vakaras). Tokie sezoniniai dėsningumai gali būti nusakomi įvairiais būdais, pvz. Brännäs (1995) pristatė INAR(1) modelį su aiškinančiaisiais kintamaisiais, į kuriuos gali būti įtraukiami trendo ir sezoniniai aiškinantieji kintamieji. Šio modelio parametrų vertinimą empiriniu didžiausio tikėtinumo metodu nagrinėjo Ding ir Wang (2016), o Enciso-Mora ir kt. (2009) pristatė aukštesnės eilės INAR modelį su aiškinančiaisiais kintamaisiais. Monteiro ir kt. (2010) pristatė INAR procesą su periodu d (PINAR$(1)_d$), kuris apibrėžiamas lygtimi:

$$Y_t = \phi_t \circ Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{N}, \quad \varepsilon_t \sim \mathrm{Pois}(v_t), \tag{2}$$

čia $\phi_t = \alpha_j \in (0,1)$, $v_t = \lambda_j$, $t = j + kd$, $j = 1,\dots,d$, $k \in \mathbb{N}_0 := \{0,1,2,\dots\}$. Be to, $\phi_t \circ Y_{t-1} = \sum_{i=1}^{Y_{t-1}} B_{i,t}$, kur $\{B_{i,t} = B_{i,t}(\phi_t), t \in \mathbb{Z}\}$ yra periodinė Bernulio a.d. seka, kuri nepriklauso nuo Y_t ir ε_t, o $\mathbb{P}(B_{i,t}(\phi_t) = 1) = \phi_t$. Monteiro ir kt. (2010) taip pat daroma prielaida, kad ε_t yra periodinis Puasono procesas, nepriklausantis nuo Y_{t-1} ir $\phi_t \circ Y_{t-1}$.

Bourguignon ir kt. (2016) pristatė pirmos eilės sezoninį INAR procesą su sezoniškumu d (INAR$(1)_d$), kuris apibrėžiamas lygtimi:

$$Y_t = \phi \circ Y_{t-d} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{3}$$

čia $\phi \in [0,1]$, o ε_t ir $\phi \circ Y_{t-d}$ apibrėžti taip pat, kaip (1) lygtyje. Be to, INAR$(1)_d$ procesas yra sudarytas iš tarpusavyje nepriklausomų INAR(1) procesų.

2 Disertacijos tyrimo objektas ir tikslai

Disertacijoje analizuojami sveikareikšmių dydžių autoregresiniai modeliai, kurių liekanos nusakomos jungties funkcija. Šioje disertacijoje:

- analizuojami literatūroje pristatyti dvimačio sveikareikšmių dydžių autoregresinio proceso parametrų vertinimo metodai ir pristatomas dviejų žingsnių parametrų vertinimo metodas, su kuriuo jungties funkcijos parametras įvertinamas atskirame žingsnyje,

- pristatomas apibendrintas vienmatis neneigiamų sveikareikšmių dydžių autoregresinis procesas su sezoniniais autoregresiniais parametrais bei sezoniškai priklausomomis inovacijomis, kuris gali būti perrašomas daugiamačiu pavidalu. Daugiamatis pavidalas panaudojamas parametrų vertinimui.

3 Disertacijos struktūra

Disertaciją sudaro penki skyriai, išvados, darbe cituojamos literatūros sąrašas ir priedų skyreliai.

Pirmame, įvadiniame, skyriuje pristatomas sveikareikšmių dydžių laiko eilučių aktualumas empiriniuose tyrimuose, aptariami literatūroje nagrinėjami sveikareikšmiai autoregresiniai laiko eilučių modeliai.

Antrame skyriuje pateikiamos pagrindinės binominio retinimo operatoriaus savybės, kurių įrodymai taip pat pateikiami disertacijos priedų skyreliuose.

Trečiame skyriuje pateikiamas jungties funkcijų apibrėžimas ir jų panaudojimas nagrinėjant sveikareišmius atsitiktinius dydžius. Taip pat pateikiami konkrečių kopulų pavyzdžiai.

Ketvirtame skyriuje nagrinėjamas pirmos eilės dvimatis sveikareikšmių dydžių autoregresinis procesas, kurio inovacijos yra aprašomos jungties funkcija. Nagrinėjamos modelio savybės, kurių įrodymai pateikiami prieduose, bei aptariami literatūroje naudojami parametrų vertinimo metodai. Šiame skyrelyje taip pat pristatomas vienas iš disertacijoje gautų mokslinių rezultatų - dviejų žingsnių parametrų vertinimo metodas, kuris jungties parametrą įvertina atskirai nuo kitų modelio parametrų. Šis metodas palyginamas su kitais parametrų vertinimo metodais atliekant Monte Karlo simuliaciją. Skyrelio pabaigoje empiriniams mokių ir nemokių paskolų duomenims sudaromi BINAR(1) modeliai. Naudojamos skirtingos jungčių funkcijų ir marginaliųjų pasiskirstymo funkcijų kombinacijos.

Penktame skyriuje nagrinėjamas antrasis disertacijoje gautas mokslinis rezultatas - vienmatis neneigiamų sveikareikšmių dydžių autoregresinis procesas su sezoniniais autoregresiniais parametrais bei sezono viduje priklausomomis inovacijomis. Parodoma, kaip vienmatė laiko eilutė gali būti užrašyta daugiamačiu pavidalu. Pateikiamos modelio savybės, o jų įrodymai pateikiami atitinkamuose priedų skyreliuose. Nagrinėjami parametro vertinimo metodai, atliekamas jų palyginimas Monte Karlo metodu. Taip pat sudaromas modelis empiriniams Čikagos nusikaltimų duomenims.

4 Disertacijos gautų rezultatų apžvalga

Disertacijos rezultatai gali būti suskirstyti į dvi dalis. Pirmoje dalyje aprašomas literatūroje pristatytas BINAR(1) procesas su jungties funkcija nusakytomis inovacijomis ir šio proceso parametrų vertinimo metodai. Tuomet disertacijoje pasiūlomas minėto proceso parametrų vertinimas dviejų žingsnių metodu. Antroje disertacijos dalyje pristatomas apibendrintas vienmatis neneigiamų sveikareikšmių dydžių autoregresinis procesas su sezoniniais autoregresiniais parametrais bei sezoniškai priklausomomis inovacijomis. Nagrinėjamos modelio savybės, daugiamatis pavidalas ir parametrų vertinimo metodai.

Dvimatis INAR(1) procesas

Disertacijoje nagrinėjamas BINAR(1) procesas, kurį pateikė Pedeli ir Karlis (2011):

1 apibrėžimas. Tegul $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \varepsilon_{2,t}]^\top$, $t \in \mathbb{Z}$ yra nepriklausomų vienodai pasiskirsčiusių dvimačių neneigiamų sveikareikšmių atsitiktinių dydžių seka. Dvimatis sveikareikšmių dydžių pirmos eilės autoregresinis procesas (BINAR(1)), $\mathbf{Y}_t = [Y_{1,t}, Y_{2,t}]^\top$, $t \in \mathbb{Z}$, nusakomas lygtimi:

$$\mathbf{Y}_t = \mathbf{A} \circ \mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \circ \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad t \in \mathbb{Z}, \tag{4}$$

čia $\alpha_j \in [0,1)$, $j = 1,2$, o '$\circ$' yra retinimo operatorius, kuris kartu atlieka ir matricų daugybos operaciją. Taigi, j-asis ($j = 1,2$) elementas atitinka pirmos eilės INAR procesą:

$$Y_{j,t} = \alpha_j \circ Y_{j,t-1} + \varepsilon_{j,t}, \quad t \in \mathbb{Z}, \tag{5}$$

čia $\alpha_j \circ Y_{j,t-1} := \sum_{i=1}^{Y_{j,t-1}} B_{j,t,i}$ ir $B_{j,t,1}, B_{j,t,2}, \ldots$ yra n.v.p. Bernulio a.d. seka su $\mathbb{P}(B_{j,t,i} = 1) = \alpha_j = 1 - \mathbb{P}(B_{j,t,i} = 0)$, $\alpha_j \in [0,1)$. Šios sekos yra tarpusavyje nepriklausomos ir nepriklauso nuo $\boldsymbol{\varepsilon}_t$, $t \in \mathbb{Z}$. $\boldsymbol{\varepsilon}_t$ nepriklauso nuo $\mathbf{Y}_s$, $s < t$, visiems $t \in \mathbb{Z}$.

Taip pat laikoma, kad $\mathbf{Y}_t = [Y_{1,t}, Y_{2,t}]^\top$ proceso inovacijų $[\varepsilon_{1,t}, \varepsilon_{2,t}]^\top$ jungtinis skirstinys su marginaliomis funkcijomis F_1, F_2 yra nusakomas jungties funkcija $C(\cdot,\cdot)$:

$$\mathbb{P}(\varepsilon_{1,t} \le y_1, \varepsilon_{2,t} \le y_2) = C(F_1(y_1), F_2(y_2)).$$

Toliau disertacijoje laikoma, kad $C(u_1,u_2) = C(u_1,u_2;\theta)$, kur θ yra jungties funkcijos priklausomybės parametras.

Disertacijoje pateikiamos (5) lygtimi nusakyto proceso savybės ir jų įrodymai.

BINAR(1) modelio parametrų vertinimas CLS metodu

Sąlyginis mažiausių kvadratų (*angl.* conditional least squares, CLS) įvertinys minimizuoja $\mathbf{Y}_t$ ir jo sąlyginės matematinės vilties skirtumų kvadratų sumą. Disertacijoje remiamasi Silva (2005) gautais INAR(1) modelio parametrų vertinimo rezultatais, ir sukonstruojamas CLS įvertinys BINAR(1) modeliui:

$$\widehat{\alpha}_j^{\mathrm{CLS}} = \frac{\sum_{t=2}^{N}(Y_{j,t} - \bar{Y}_j)(Y_{j,t-1} - \bar{Y}_j^*)}{\sum_{t=2}^{N}(Y_{j,t-1} - \bar{Y}_j^*)^2} \tag{6}$$

ir

$$\widehat{\mu}_{\varepsilon,j}^{\mathrm{CLS}} = \frac{1}{N-1}\left(\sum_{t=2}^{N} Y_{j,t} - \widehat{\alpha}_j^{\mathrm{CLS}} \sum_{t=2}^{N} Y_{j,t-1}\right), \tag{7}$$

kur $\bar{Y}_j := (N-1)^{-1}\sum_{t=2}^{N} Y_{j,t}$ and $\bar{Y}_j^* := (N-1)^{-1}\sum_{t=2}^{N} Y_{j,t-1}$ (žr. Ispány ir kt. (2003))

Asimptotinės INAR(1) modelio parametrų CLS įvertinio savybės pateiktos Latour (1998), Silva (2005), Barczy ir kt. (2010).

Tarkime, kad inovacijos $\varepsilon_{1,t}$ ir $\varepsilon_{2,t}$ turi Puasono marginaliuosius skirstinius su atitinkamais parametrais $\mu_{\varepsilon,1}$ ir $\mu_{\varepsilon,2}$ ir jų jungtinis skirstinys nusakytas jungties funkcija su priklausomybės parametru θ. Pastebint, kad

$$\mathbb{E}(Y_{1,t} - \alpha_1 Y_{1,t-1} - \mu_{\varepsilon,1})(Y_{2,t} - \alpha_2 Y_{2,t-1} - \mu_{\varepsilon,2}) = \mathbb{C}\mathrm{ov}(\varepsilon_{1,t}, \varepsilon_{2,t}), \tag{8}$$

galima įvertinti θ parametrą minimizuojant skirtumų kvadratų sumą:

$$S = \sum_{t=2}^{N} \left(\varepsilon_{1,t}^{\mathrm{CLS}} \varepsilon_{2,t}^{\mathrm{CLS}} - \gamma(\hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}}, \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}}; \theta) \right)^2, \tag{9}$$

čia

$$\varepsilon_{j,t}^{\mathrm{CLS}} := Y_{j,t} - \hat{\alpha}_j^{\mathrm{CLS}} Y_{j,t-1} - \hat{\mu}_{\varepsilon,j}^{\mathrm{CLS}}, \quad j = 1,2$$

ir

$$\begin{aligned}
\gamma(\mu_{\varepsilon,1}, \mu_{\varepsilon,2}; \theta) &:= \mathbb{C}\mathrm{ov}(\varepsilon_{1,t}, \varepsilon_{2,t}) \\
&= \sum_{k,l=1}^{\infty} kl\, c(F_1(k; \mu_{\varepsilon,1}), F_2(l; \mu_{\varepsilon,2}); \theta) - \mu_{\varepsilon,1}\mu_{\varepsilon,2}, \qquad (10)
\end{aligned}$$

o $c(F_1(k;\mu_{\varepsilon,1}), F_2(l;\mu_{\varepsilon,2});\theta)$ yra jungtinė tikimybės masės funkcija:

$$\begin{aligned}
c(F_1(k;\mu_{\varepsilon,1}), F_2(l;\mu_{\varepsilon,2});\theta) &= \mathbb{P}(\varepsilon_{1,t} = k, \varepsilon_{2,t} = l) \\
&= C(F_1(k;\mu_{\varepsilon,1}), F_2(l;\mu_{\varepsilon,2});\theta) \\
&\quad - C(F_1(k-1;\mu_{\varepsilon,1}), F_2(l;\mu_{\varepsilon,2});\theta) \\
&\quad - C(F_1(k;\mu_{\varepsilon,1}), F_2(l-1;\mu_{\varepsilon,2});\theta) \\
&\quad + C(F_1(k-1;\mu_{\varepsilon,1}), F_2(l-1;\mu_{\varepsilon,2});\theta), \\
&\quad k \geq 1,\ l \geq 1. \qquad (11)
\end{aligned}$$

(9) lygtyje pristatytas vertinimo metodas remiasi kovariacijos $\gamma(\hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}}, \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}}; \theta)$ aproksimacija

$$\begin{aligned}
\gamma^{(M_1,M_2)}(\hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}}, \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}}; \theta) = {} & \sum_{k=1}^{M_1} \sum_{l=1}^{M_2} kl\, c(F_1(k; \hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}}), F_2(l; \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}}); \theta) \\
& - \hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}} \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}}. \qquad (12)
\end{aligned}$$

Disertacijoje ši aproksimacija buvo naudojama atliekant simuliacijas Monte Karlo metodu (pvz. paimant $M_1 = M_2 = M = 8$).

Jeigu inovacijų jungties funkcija yra FGM (Farlie–Gumbel–Morgenstern), galima paimti sumos

$$S^{(M_1,M_2)} = \sum_{t=2}^{N} \left(\varepsilon_{1,t}^{\mathrm{CLS}} \varepsilon_{2,t}^{\mathrm{CLS}} - \gamma^{(M_1,M_2)}(\hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}}, \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}}; \theta)\right)^2, \tag{13}$$

išvestinę, prilyginti ją nuliui ir panaudoti (12) lygtį, kad gauti jungties funkcijos priklausomybės parametro įvertinį

$$\hat{\theta}^{\mathrm{FGM}} = \frac{\sum_{t=2}^{N}(Y_{1,t} - \hat{\alpha}_1^{\mathrm{CLS}} Y_{1,t-1} - \hat{\mu}_{\varepsilon,1}^{\mathrm{CLS}})(Y_{2,t} - \hat{\alpha}_2^{\mathrm{CLS}} Y_{2,t-1} - \hat{\mu}_{\varepsilon,2}^{\mathrm{CLS}})}{(N-1)\sum_{k=1}^{M_1} k\widetilde{F}_{1,k} \sum_{l=1}^{M_2} l\widetilde{F}_{2,l}}, \tag{14}$$

čia $\widetilde{F}_{j,k} = F_{j,k}\overline{F}_{j,k} - F_{j,k-1}\overline{F}_{j,k-1}$, o $F_{j,k} := F_j(k; \hat{\mu}_{\varepsilon,j}^{\mathrm{CLS}})$ ir $\overline{F}_{j,k} := 1 - F_{j,k}$, $j = 1,2$. Disertacijos priedų skyrelyje pateiktas (14) lygties išvedimas.

Priklausomai nuo pasirinktos jungties funkcijos šeimos, gali būti sudėtinga suskaičiuoti (11) ir gauti analitinę įvertinio $\hat{\theta}$ išraišką. Disertacijoje, atliekant empirinius skaičiavimus ir Monte Karlo simuliaciją, minimizuojant (9) lygtį naudojama R programos `optim` funkcija. Be to, kitiems marginaliesiems skirstiniams, kurie turi ne tik $\mu_{\varepsilon,j}$, bet ir kitus parametrus, (9) lygtis turėtų būti minimizuojama ne tik pagal θ, bet ir pagal tuos papildomus parametrus.

BINAR(1) modelio parametrų vertinimas CML metodu

BINAR(1) modelio parametrai taip pat gali būti įvertinti sąlyginiu didžiausio tikėtinumo metodu (*angl.* conditional maximum likelihood, CML) (žr. Pedeli ir Karlis (2011) ir Karlis ir Pedeli (2013)). Sąlyginis BINAR(1) proceso skirstinys yra nusakomas sąlygine tikimybe:

$$\begin{aligned}
&\mathbb{P}(Y_{1,t}=y_{1,t},Y_{2,t}=y_{2,t}|Y_{1,t-1}=y_{1,t-1},Y_{2,t-1}=y_{2,t-1})\\
&=\mathbb{P}(\alpha_1\circ y_{1,t-1}+\varepsilon_{1,t}=y_{1,t},\alpha_2\circ y_{2,t-1}+\varepsilon_{2,t}=y_{2,t})\\
&=\sum_{k=0}^{y_{1,t}}\sum_{l=0}^{y_{2,t}}\mathbb{P}(\alpha_1\circ y_{1,t-1}=k)\mathbb{P}(\alpha_2\circ y_{2,t-1}=l)\\
&\times\mathbb{P}(\varepsilon_{1,t}=y_{1,t}-k,\varepsilon_{2,t}=y_{2,t}-l).
\end{aligned}$$

Čia $\alpha_j\circ y$ yra y nepriklausomų Bernulio eksperimentų, todėl tikimybė lygi:

$$\mathbb{P}(\alpha_j\circ y_{j,t-1}=k)=\binom{y_{j,t-1}}{k}\alpha_j^k(1-\alpha_j)^{y_{j,t-1}-k},$$

$k=0,\dots,y_{j,t-1},\ j=1,2$. Tuo atveju, kai BINAR(1) modelio inovacijų liekanos gali būti nusakomo jungties funkcija su Puasono marginaliomis funkcijomis, turime, kad

$$\mathbb{P}(\varepsilon_{1,t}=y_{1,t}-k,\varepsilon_{2,t}=y_{2,t}-l)=c(F_1(y_{1,t}-k,\mu_{\varepsilon,1}),F_2(y_{2,t}-l,\mu_{\varepsilon,2});\theta).$$

Tuomet gauname:

$$\begin{aligned}
&\mathbb{P}(Y_{1,t}=y_{1,t},Y_{2,t}=y_{2,t}|Y_{1,t-1}=y_{1,t-1},Y_{2,t-1}=y_{2,t-1})\\
&=\sum_{k=0}^{y_{1,t}}\sum_{l=0}^{y_{2,t}}\binom{y_{1,t-1}}{k}\alpha_1^k(1-\alpha_1)^{y_{1,t-1}-k}\binom{y_{2,t-1}}{l}\alpha_2^l(1-\alpha_2)^{y_{2,t-1}-l}\\
&\times c(F_1(y_{1,t}-k,\mu_{\varepsilon,1}),F_2(y_{2,t}-l,\mu_{\varepsilon,2});\theta).
\end{aligned}$$

Šiuo atveju, norint įvertinti parametrus $\mu_{\varepsilon,1},\mu_{\varepsilon,2}$, α_1,α_2 ir θ panaudojamas sąlyginės didžiausio tikėtinumo funkcijos logaritmas:

$$\ell(\alpha_1,\alpha_2,\mu_{\varepsilon,1},\mu_{\varepsilon,2},\theta)= \sum_{t=2}^{N}\log\mathbb{P}(Y_{1,t}=y_{1,t},Y_{2,t}=y_{2,t}|Y_{1,t-1}=y_{1,t-1},Y_{2,t-1}=y_{2,t-1})$$

kur $y_{1,1}$ ir $y_{2,1}$ - pradinės reikšmės. Parametrų vertinimui maksimizuojamas sąlyginės didžiausio tikėtinumo funkcijos logaritmas:

$$\ell(\alpha_1,\alpha_2,\mu_{\varepsilon,1},\mu_{\varepsilon,2},\theta)\longrightarrow \max_{\alpha_1,\alpha_2,\mu_{\varepsilon,1},\mu_{\varepsilon,2},\theta}. \tag{15}$$

Disertacijoje tai atliekama skaitiniais metodais panaudojant R programos `optim` funkciją.

Taip pat kaip ir CLS metodui, kitiems marginaliesiems skirstiniams, kurie turi ne tik $\mu_{\varepsilon,j}$, bet ir kitus parametrus (15) lygtis turėtų būti maksimizuojama ir pagal tuos papildomus parametrus. CML įvertinys yra asimptotiškai normaliai pasiskirstęs pagal standartines reguliarumo sąlygas, o jo kovariacijų matrica yra Fišerio matricos atvirkštinė (žr. Pedeli ir Karlis (2011)).

BINAR(1) modelio parametrų vertinimas dviejų žingsnių metodu

Priklausomai nuo parametrų reikšmių ir imties dydžio, CML metodo parametrų vertinimas gali užsitęsti. Kita vertus, α_j ir $\mu_{\varepsilon,j}$ CLS įvertiniai yra nesudėtingai apskaičiuojami, todėl juos galima įsistatyti vietoje atitinkamų nežinomų parametrų (15) lygtyje, naudojant CLS įverčius iš (6) ir (7) lygčių. Tuomet Puasono marginaliųjų inovacijų skirstinių atveju, ℓ reikėtų maksimizuoti tik pagal vieną parametrą, θ. Disertacijoje pateiktas dviejų žingsnių parametrų vertinimo metodas, kurio pirmajame žingsnyje yra įvertinama:

$$(\hat{\alpha}_j^{\text{CLS}}, \hat{\mu}_{\varepsilon,j}^{\text{CLS}}) = \arg\min Q_j(\alpha_j, \mu_{\varepsilon,j}), \quad j = 1,2.$$

Tuomet įvertintos reikšmės užfiksuojamos ir naudojamos vietoje atitinkamų nežinomų parametrų antrajame žingsnyje:

$$\hat{\theta}^{\text{CML}} = \arg\max \ell(\hat{\alpha}_1^{\text{CLS}}, \hat{\alpha}_2^{\text{CLS}}, \hat{\mu}_{\varepsilon,1}^{\text{CLS}}, \hat{\mu}_{\varepsilon,2}^{\text{CLS}}, \theta).$$

Kitiems marginaliesiems skirstiniams papildomi parametrai turėtų būti įvertinami antrajame žingsnyje.

BINAR(1) modelio parametrų vertinimo metodų palyginimas Monte Karlo metodu

Disertacijoje minėti CLS, CML ir dviejų žingsnių parametrų vertinimo metodai buvo palyginti Monte Karlo metodu, atliekant 1000 replikacijų, kai imties dydis 50, ir kai imties dydis 500. Buvo generuojami BINAR(1) procesai, kurių inovacijų jungtinis skirstinys buvo nusakytas FGM, Franko arba Kleitono jungties funkcija, o marginalieji skirstiniai buvo Puasono, arba vienas skirstinys Puasono, o kitas - neigiamas binominis.

Disertacijoje parodyta, kad $\hat{\alpha}_j$, $\hat{\mu}_{\varepsilon,j}$ ir $\hat{\sigma}_2^2$ yra tiksliau įvertinami CML metodu, nei CLS metodu, tačiau naudojant CLS metodu apskaičiuotus $\hat{\alpha}_j$ ir $\hat{\mu}_{\varepsilon,j}$ galima gauti $\hat{\theta}$ įvertį, kuris mažiau skirsis nuo tikrosios θ reikšmės, lyginant su $\hat{\theta}^{\text{CLS}}$, ir nedaug skirsis nuo $\hat{\theta}^{\text{CML}}$.

Paskolų kiekio aprašymas BINAR(1) modeliu

Disertacijoje BINAR(1) modelis taikytas išduotų paskolų kiekiui. Nagrinėtos dvi laiko eilutės - mokių paskolų ir nemokių paskolų skaičiai per savaitę, nuo 2013 m. spalio 21 d. iki 2016 m. sausio 1 d.. Laiko eilutėms įvertintos skirtingos BINAR(1) modelio kombinacijos, kuriose:

- inovacijos buvo nusakytos FGM, Franko arba Kleitono jungties funkcija,
- inovacijų marginalieji skirstiniai buvo Puasono arba neigiamo binominio skirstinių kombinacijos: abiejų inovacijų skirstiniai Puasono, abiejų inovacijų skirstiniai neigiamo binominio, arba vienos iš inovacijų skirstinys Puasono, o kitos - neigimas binominis.

Visais atvejais modelių parametrai buvo įvertinti dviejų žingsnių metodu.

Disertacijoje parodyta, kad nagrinėtiems duomenims Franko ir FGM jungtimis grįsti modeliai yra labai panašūs log-tikėtinumo prasme, nepriklausomai nuo inovacijų marginaliųjų skirstinių. Kita vertus, FGM jungties parametras kai kuriais atvejais įgyjo didžiausią galimą reikšmę (t.y. 1). Modeliai taip pat buvo tikslesni (log-tikėtinumo prasme), jeigu inovacijos buvo nusakomos neigiamais binominiais marginaliaisiais skirstiniais.

Sezoninis INAR(1) procesas (SINAR$(1)_d$)

Disertacijoje pasiūlytas (3) lygties apibendrinimas, leidžiant ϕ parametrui kisti priklausomai nuo sezono, bei darant prielaidą, kad inovacijos gali būti priklausomos sezono viduje:

2 apibrėžimas. Neneigiamų sveikareikšmių a.d. seka $\{Y_t, t \in \mathbb{Z}\}$ vadinama INAR(1) procesu sezoniškumui su periodu d (SINAR$(1)_d$), jeigu ji tenkina lygtį:

$$Y_t = \phi_t \circ Y_{t-d} + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{16}$$

čia $\phi_t = \alpha_j \in [0,1)$, $t = j + kd$, $k \in \mathbb{Z}$ ir $j \in \{1,\ldots,d\}$. $\{\varepsilon_t, t \in \mathbb{Z}\}$ yra sezono viduje priklausomų neneigiamų sveikareikšmių a.d.

seka. Laikoma, kad $\phi_t \circ Y_{t-d} = \sum_{i=1}^{Y_{j+(k-1)d}} B_{i,j,k-1}$, kur $B_{i,j,k-1}$ yra nepriklausomi Bernulio a.d., kurie nepriklauso nuo $Y_{j+(k-1)d}$ ir $\varepsilon_{j+(k-1)d}$ visiems $k \in \mathbb{Z}$, $j \in \{1,\ldots,d\}$ su $\mathbb{P}(B_{i,j,k-1} = 1) = \alpha_j$. $[\varepsilon_{1+kd},\ldots,\varepsilon_{d+kd}]^\top$ yra nepriklausomi nuo $[Y_{1+(k-s)d},\ldots,Y_{d+(k-s)d}]^\top$, $s \in \{1,2,...\}$, o $[\varepsilon_{1+kd},\ldots,\varepsilon_{d+kd}]^\top$ skirstinys gali būti nusakytas daugiamačiu skirstiniu, arba jungties funkcija.

Disertacijoje laikoma, kad 2 apibrėžime esančių ε_{j+kd}, $j \in \{1,\ldots,d\}$ marginalieji skirstiniai tokie, kad $\mathbb{E}(\varepsilon_{j+kd}) = \mu_{\varepsilon,j}$ ir $\mathbb{V}\text{ar}(\varepsilon_{j+kd}) = \sigma_{\varepsilon,jj} < \infty$ visiems k, o ε_t yra priklausomi sezono viduje su kovariacija

$$\mathbb{C}\text{ov}(\varepsilon_{i+kd},\varepsilon_{j+ld}) = \begin{cases} \sigma_{\varepsilon,ij}, & k = l, \\ 0, & k \neq l, \end{cases}$$

t.y. vektoriai $[\varepsilon_{1+kd} - \mu_{\varepsilon,1},\ldots,\varepsilon_{d+kd} - \mu_{\varepsilon,d}]^\top$ sudaro baltojo triukšmo seką.

Disertacijoje pateikiamos 16 lygtimi nusakyto proceso savybės ir jų įrodymai.

1 pastaba. Laikoma, kad 2 apibrėžime nusakytas procesas yra toks, kad $\mathbb{E}(Y_{j+kd}) \neq \mathbb{E}(Y_{i+ld})$ arba $\gamma(j+kd,s) \neq \gamma(i+ld,s)$, kuriam nors $i \neq j$ ir $\forall k,l \in \mathbb{Z}$.

Periodiškai koreliuotos sekos, kurios nusakomos 2 apibrėžimu, yra nestacionarios. Jeigu periodas yra žinomas, tokios sekos gali būti išreiškiamos kaip daugiamatis stacionarus procesas (žr. Hurd ir Miamee (2007)). Tarkime, kad $j \in \{1,...,d\}$, $k \in \{1,...,n\}$, $Y_k^{(j)} := Y_{j+kd}$ ir $\varepsilon_k^{(j)} := \varepsilon_{j+kd}$. Tuomet (16) procesą galima užrašyti kaip atskirus INAR(1) procesus

$$Y_k^{(j)} = \alpha_j \circ Y_{k-1}^{(j)} + \varepsilon_k^{(j)},$$

čia $\varepsilon_k^{(j)}$ and $\varepsilon_k^{(i)}$ yra priklausomi a.d. $\forall i,j \in \{1,...,d\}$. Tuomet galima gauti daugiamatę INAR(1) proceso reprezentaciją perrašius (16) lygties procesą į atskirus INAR(1) procesų blokus. Ši reprezentacija pavaizduota 1 lentelėje.

1 lentelė: SINAR$(1)_d$ procesas, išreikštas atskirais INAR(1) procesais. $Y_\bullet^{(j)}$ atitinka $Y_k^{(j)}, \forall k \in \{1, ..., n\}$.

Periodo elementai (pvz. mėnesiai)	Periodų sk. (pvz. metai)						
	1	2	...	k	...	n	
1							$Y_\bullet^{(1)}$
2							$Y_\bullet^{(2)}$
$\vdots$							$\vdots$
j				(j,k)			$Y_\bullet^{(j)}$
$\vdots$							$\vdots$
d							$Y_\bullet^{(d)}$

SINAR$(1)_d$ proceso daugiamatis pavidalas

Tegul $\{Y_t\}$ procesas yra nusakytas 2 apibrėžimu. Tuomet Y_t galima užrašyti daugiamačiu INAR(1) pavidalu

$$\mathbf{Y}_k = \mathbf{A} \circ \mathbf{Y}_{k-1} + \mathbf{Z}_k, \; k \in \mathbb{Z}, \tag{17}$$

kur:

$$\begin{aligned} \mathbf{Y}_k &= [Y_{1+kd}, \ldots, Y_{d+kd}]^\top, \; \mathbf{Z}_k = [\varepsilon_{1+kd}, \ldots, \varepsilon_{d+kd}]^\top, \\ \mathbf{A} &= \operatorname{diag}(\alpha_1, \ldots, \alpha_d). \end{aligned}$$

Tuomet (17) lygtimi nusakytas procesas $\{\mathbf{Y}_k\}$ yra stacionarus (žr. Gladyshev (1961) ir Hurd ir Miamee (2007)).

Disertacijoje pateiktos daugiamatės proceso reprezentacijos savybės ir jų įrodymai. Be to, jei $d = 2$, tai (17) lygties specifikacija atitinka dvimatį INAR(1) procesą su jungties funkcija nusakytomis inovacijomis, kuris atskirai nagrinėtas disertacijoje (žr. 1 apibrėžimą).

SINAR$(1)_d$ modelio parametrų vertinimas apribotu apibendrintu CLS metodu

$\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ galima perrašyti VAR(1) pavidalu:

$$\mathbf{Y}_k = \boldsymbol{\mu}_\varepsilon + \mathbf{A}\mathbf{Y}_{k-1} + \boldsymbol{e}_k,\ k = 1,...,n, \tag{18}$$

čia $\boldsymbol{e}_k$ yra inovacijų vektorius

$$\boldsymbol{e}_k = \mathbf{Y}_k - \mathbb{E}(\mathbf{Y}_k|\mathbf{Y}_{k-1}) = \mathbf{A}\circ\mathbf{Y}_{k-1} - \mathbf{A}\mathbf{Y}_{k-1} + \mathbf{Z}_k - \boldsymbol{\mu}_\varepsilon$$

su $\mathbb{E}(\boldsymbol{e}_k) = \mathbf{0}$ ir kovariacijų matrica

$$\boldsymbol{\Sigma}_{\boldsymbol{e}} = \operatorname{diag}(\mathbf{B}\boldsymbol{\mu}_{\mathbf{Y}}) + \boldsymbol{\Sigma}_\varepsilon = \left(\sigma_{\varepsilon,ij} + \alpha_i\mu_{\varepsilon,i}\mathbb{1}_{\{i=j\}}\right)_{i,j=1,\ldots,d}. \tag{19}$$

Disertacijoje pateikiamas (19) lygties įrodymas.

Pažymėkime

$$\mathbf{Y} = [\mathbf{Y}_1,\ldots,\mathbf{Y}_n]^\top,\quad \mathbf{X}_k = [\mathbf{Y}_k^\top \vdots 1]^\top,\quad 0\le k\le n-1,$$
$$\mathbf{X} = [\mathbf{X}_0,\ldots,\mathbf{X}_{n-1}]^\top,\ \mathbf{E} = [\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n]^\top,\ \mathscr{B} = [\mathbf{A} \vdots \boldsymbol{\mu}_{\boldsymbol{\varepsilon}}]^\top.$$

Čia $\mathbf{Y}$ ir $\mathbf{E}$ yra $n\times d$ matricos, $\mathbf{X}_k$ yra $(d+1)\times 1$ vektorius, $\mathbf{X}$ yra $n\times(d+1)$ matrica, $\mathscr{B}$ yra $(d+1)\times d$ matrica.

Tuomet (18) galime užrašyti tokiu pavidalu

$$\mathbf{Y} = \mathbf{X}\mathscr{B} + \mathbf{E}. \tag{20}$$

Tuomet $\mathscr{B}$ sąlyginis neapribotas mažiausių kvadratų įvertinys užrašomas pavidalu:

$$\widetilde{\mathscr{B}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y}. \tag{21}$$

CLS įvertinio asimptotinis normalumas įrodomas Latour (1997).

Daugiamatė INAR specifikacija (17) lygtyje apriboja koeficientų matricą **A** taip, kad nediagonaliniai elementai yra nuliniai. Tuo tarpu (21) lygtyje esanti koeficientų įvertinių matrica yra neapribota. Norint pagerinti įvertinių efektyvumą, disertacijoje nediagonaliniams koeficientams uždedami apribojimai. Remiantis Lütkepohl (2007, 5.2 Skyrius), disertacijoje matricai $\mathscr{B}$ uždedami nuliniai apribojimai. Laikoma, kad „$\otimes$“ yra matricų Kronekerio sandauga, o „vec“ yra matricos vektorizavimas.

1 teiginys. *Tegul* $\boldsymbol{\beta} := \mathrm{vec}\left(\mathscr{B}\right) = \mathbf{R}\boldsymbol{\gamma}$, *kur* $\boldsymbol{\gamma} = [\alpha_1, \mu_{\varepsilon,1}, \ldots, \alpha_d, \mu_{\varepsilon,d}]^\top$ *yra neapribotas nežinomų parametrų vektorius, o* $\mathbf{R}^\top = [\mathbf{R}_0^\top, \ldots, \mathbf{R}_{d-1}^\top]$ *yra žinoma apribojimų matrica, kur* $\mathbf{R}_k = \left(r_{i,j}^{(k)}\right)$, $k = 0, \ldots, d-1$ *yra* $(d+1) \times 2d$ *matricos, su elementais*

$$r_{i,j}^{(k)} = \begin{cases} 1, & \textit{jei arba } i = k+1 \textit{ ir } j = 2k+1, \\ & \textit{arba } i = d+1 \textit{ ir } j = 2k+2, \\ 0, & \textit{kitu atveju.} \end{cases} \tag{22}$$

Tuomet apribotas apibendrintas sąlyginis mažiausių kvadratų (REG-CLS) įvertinys gali būti užrašomas pavidalu

$$\widehat{\boldsymbol{\beta}} = \mathbf{R}\widehat{\boldsymbol{\gamma}}, \quad \widehat{\boldsymbol{\gamma}} = \left[\mathbf{R}^\top \left(\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-1} \otimes \left(\mathbf{X}^\top \mathbf{X}\right)\right) \mathbf{R}\right]^{-1} \mathbf{R}^\top \left(\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}^{-1} \otimes \mathbf{X}^\top\right) \mathrm{vec}\left(\mathbf{Y}\right), \tag{23}$$

čia matrica $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}}$ *apskaičiuojamas pagal formulę*

$$\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{e}} = \frac{1}{n-d-1}(\mathbf{Y}-\mathbf{X}\widetilde{\mathscr{B}})^\top(\mathbf{Y}-\mathbf{X}\widetilde{\mathscr{B}}), \tag{24}$$

o $\widetilde{\mathscr{B}}$ *yra neapribotas įvertinys, apskaičiuotas pagal* (21) *lygtį.*

Disertacijoje pateikiamas 1 tTeiginio įrodymas.

2 teiginys. *Tegul* $\{\mathbf{Y}_k\}$ *yra* (17) *lygtimi nusakytas procesas su nepriklausoma baltojo triukšmo seka* $\{\mathbf{Z}_k\}$ *su baigtiniais ketvirtais momentais. Tuomet* (23) *lygtimi nusakytas REG-CLS įvertinys yra suderintas ir asimptotiškai normaliai pasiskirstęs:*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \xrightarrow{d} \mathscr{N}\left(\mathbf{0},\ \mathbf{R}\left[\mathbf{R}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{e}}^{-1}\otimes\boldsymbol{\Gamma}_{\boldsymbol{X}}(0)\right)\mathbf{R}\right]^{-1}\mathbf{R}^\top\right), \tag{25}$$

čia $\dfrac{\mathbf{X}^\top\mathbf{X}}{n} \longrightarrow \boldsymbol{\Gamma}_{\boldsymbol{X}}(0) := \mathbb{E}\left(\mathbf{X}_k^\top\mathbf{X}_k\right)$ *pagal tikimybę.*

Disertacijoje pateikiamas 2 teiginio įrodymas.

SINAR$(1)_d$ modelio parametrų vertinimas CML metodu

Laikome, kad jungtinis vektoriaus $[\varepsilon_{1+kd},\ldots,\varepsilon_{d+kd}]^\top$ skirstinys $F(\cdot)$, gali būti nusakomas d-mate jungties funkcija $C:[0,1]^d\to[0,1]$, tokia, kad

$$F(a_1,...,a_d) = C(F_1(a_1),...,F_d(a_d)),$$

čia $F_j(\cdot)$ yra ε_{j+kd} marginalusis skirstinys, $k\in\mathbb{Z}$.

Remiantis Nikoloulopoulos ir Karlis (2009), vektoriaus $[\varepsilon_{1+kd},...,\varepsilon_{d+kd}]$ tikimybės masės funkcija apskaičiuojama iš lygybės

$$
\begin{aligned}
c(a_1,\ldots,a_d;\boldsymbol{\theta}) :&= \mathbb{P}\left(\varepsilon_{1+kd}=a_1,\ldots,\varepsilon_{d+kd}=a_d\right) \\
&= \sum_{i_1\in\{0,1\}}\cdots\sum_{i_d\in\{0,1\}}(-1)^{i_1+\ldots+i_d} \\
&\times \mathbb{P}\left(\varepsilon_{1+kd}\le a_1-i_1,\ldots,\varepsilon_{d+kd}\le a_d-i_d\right) \\
&= \sum_{i_1\in\{0,1\}}\cdots\sum_{i_d\in\{0,1\}}(-1)^{i_1+\ldots+i_d} \\
&\times C(F_1(a_1-i_1),\ldots,F_d(a_d-i_d);\boldsymbol{\theta}),
\end{aligned}
\tag{26}
$$

kur $\boldsymbol{\theta}=[\theta_1,\ldots,\theta_m]^\top$ yra m nežinomų jungties funkcijos parametrų vektorius, o ε_{j+kd} marginalusis skirstinys turi nežinomą vidurkį $\mu_{\varepsilon,j}$ ir nežinomą dispersiją $\sigma_{\varepsilon,jj}$.

Tegul nežinomų parametrų vektorius yra

$$
\boldsymbol{\gamma} \;=\; [\alpha_1,\mu_{\varepsilon,1},\sigma_{\varepsilon,11},\ldots,\alpha_d,\mu_{\varepsilon,d},\sigma_{\varepsilon,dd},\theta_1,\ldots,\theta_m]^\top .
$$

Tuomet sąlyginė log tikėtinumo funkcija gali būti užrašoma

$$
\ell(\boldsymbol{\gamma}|\mathbf{y}_0)=\sum_{k=1}^{n}\log(\mathbb{P}_{\mathbf{Y}_k|\mathbf{Y}_{k-1}}(\mathbf{y}_k|\mathbf{y}_{k-1})),
\tag{27}
$$

kur $\mathbf{y}_0=[y_1,\ldots,y_d]^\top$ yra pradinės reikšmės. Čia $\mathbb{P}_{\mathbf{Y}_k|\mathbf{Y}_{k-1}}(\mathbf{y}_k|\mathbf{y}_{k-1})$ yra (17) lygtimi nusakomo proceso sąlyginė tikimybė su sezono viduje priklausomomis inovacijomis, kuri užrašoma lygtimi

$$
\begin{aligned}
\mathbb{P}_{\mathbf{Y}_k|\mathbf{Y}_{k-1}}(\mathbf{y}_k|\mathbf{y}_{k-1}) \;:=&\; \mathbb{P}\left(\mathbf{Y}_k=\mathbf{y}_k|\mathbf{Y}_{k-1}=\mathbf{y}_{k-1}\right) \\
&= \mathbb{P}\left(\mathbf{A}\circ\mathbf{y}_{k-1}+\mathbf{Z}_k=\mathbf{y}_k\right) \\
&= \sum_{j_1=0}^{y_{1+kd}}\cdots\sum_{j_d=0}^{y_{d+kd}}\mathbb{P}\left(\varepsilon_{1+kd}=j_1,\ldots,\varepsilon_{d+kd}=j_d\right) \\
&\times\prod_{i=1}^{d}p_i(j_i,y_{i+(k-1)d}),
\end{aligned}
$$

kur $\mathbf{y}_k = [y_{1+kd}, \ldots, y_{d+kd}]^\top$, o $p_i(j_i, y_{i+(k-1)d})$ yra $y_{i+(k-1)d}$ nepriklausomų Bernulio eksperimentų su j_i sėkmių skaičiumi suma:

$$\begin{aligned} p_i(j_i, y_{i+(k-1)d}) &:= \mathbb{P}\left(\alpha_i \circ y_{i+(k-1)d} = y_{i+(k-1)d} - j_i\right) \\ &= \binom{y_{i+(k-1)d}}{y_{i+(k-1)d} - j_i} \alpha_i^{y_{i+(k-1)d} - j_i} (1-\alpha_i)^{j_i} \end{aligned} \tag{28}$$

kur $j_i = 0, \ldots, y_{i+kd}$, $i = 1, ..., d$, $k = 1, ..., n$.

Norėdami įvertinti nežinomus parametrus, maksimizuojame (27) lygtį:

$$\ell(\boldsymbol{\gamma}|\mathbf{y}_0) \longrightarrow \max_{\boldsymbol{\gamma}}. \tag{29}$$

Esant išpildytoms atitinkamoms reguliarumo sąlygoms (žr. Franke ir Seligmann (1993)), galima parodyti, kad iš (29) gautas $\boldsymbol{\gamma}$ CML įvertinys yra asimptotiškai normaliai pasiskirstęs:

$$\sqrt{n}\left(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\gamma})\right), \tag{30}$$

kur $\mathbf{I}(\boldsymbol{\gamma})$ yra $(3d+m) \times (3d+m)$ Fišerio informacijos matrica.

Pagrindinis (27) lygtimi nusakyto vertinimo metodo trūkumas - priklausomai nuo sezono d dydžio, marginaliųjų skirstinių ir jungties funkcijos, įvertinamų parametrų skaičius gali būti per didelis norint įvertinti modelį prakitniuose taikymuose. Į tai galima atsižvelgti pritaikant IFM (*angl.* inference function for margins) vertinimo metodą, t.y. dviejų žingsnių vertinimo metodą, kuris yra remiasi tikėtinumo funkcija.

SINAR$(1)_d$ modelio parametrų vertinimas IFM metodu

Modelio parametrus galima įvertinti IFM metodu dviem žingsniais: (1) įvertinti marginaliųjų skirstinių parametrus; (2) įvertinti jungties funkcijos parametrų vektorių, laikant (1) žingsnyje įvertintus parametrus fiksuotais.

Tegul Y_{j+kd} marginaliojo skirstinio nežinomų parametrų vektorius yra $\boldsymbol{\gamma}_j = [\alpha_j, \mu_{\varepsilon,j}, \sigma_{\varepsilon,jj}]^\top$, $j = 1,...,d$. Pirmame žingsnyje kiekvienam marginaliajam skirstiniui nepriklausomai maksimizuojame sąlyginę log tikėtinumo funkciją

$$\ell_j(\boldsymbol{\gamma}_j|y_j) = \sum_{k=1}^{n} \log\left(\mathbb{P}_{Y_{j+kd}|Y_{j+(k-1)d}}(y_{j+kd}|y_{j+(k-1)d})\right) \to \max_{\boldsymbol{\gamma}_j} \tag{31}$$

su pradinėmis reikšmėmis y_j, $j = 1,...,d$. Čia sąlyginė tikimybė yra

$$\begin{aligned}
&\mathbb{P}_{Y_{j+kd}|Y_{j+(k-1)d}}(y_{j+kd}|y_{j+(k-1)d}) \\
&\qquad = \mathbb{P}(Y_{j+kd} = y_{j+kd}|Y_{j+(k-1)d} = y_{j+(k-1)d}) \\
&\qquad = \mathbb{P}(\alpha_j \circ y_{j+(k-1)d} + \varepsilon_{j+kd} = y_{j+kd}) \\
&\qquad = \sum_{i=0}^{y_{j+kd}} p_j(i, y_{j+(k-1)d})\mathbb{P}(\varepsilon_{j+kd} = i), \quad j = 1,...,d,
\end{aligned} \tag{32}$$

kur $p_j(i, y_{j+(k-1)d})$ yra apibrėžiamas (28) lygtimi, o tikimybė $\mathbb{P}(\varepsilon_{j+kd} = i)$ yra nusakoma arba Puasono arba neigiamu binominiu marginaluoju skirstiniu. Antrame žingsnyje įvertiname jungties funkcijos parametrų vektorių $\boldsymbol{\theta}$ maksimizuodami (27) lygtį, laikant, kad marginaliųjų skirstinių parametrai yra fiksuoti ir gauti iš (31) lygties. IFM asimptotinis efektyvumas ir normalumas aprašytas Joe (2005).

SINAR(1)$_d$ modelio parametrų vertinimo metodų palyginimas Monte Karlo metodu

Disertacijoje REG-CLS ir IFM parametrų vertinimo metodai buvo palyginti Monte Karlo metodu, atliekant 5000 replikacijų, kai $d = 2$ ir 1000 replikacijų, kai $d = 4$, o imties dydis 50 arba 540. Buvo generuojami $\mathrm{SINAR}(1)_d$ procesai, kurių inovacijų marginalieji skirstiniai buvo Puasono arba neigiami binominiai, o jų jungtiniai skirstiniai aprašomi Kleitono arba Franko jungties funkcija. Kadangi REG-CLS metodu neatsižvelgiama į inovacijų skirstinį, Puasono marginaliųjų skirstinių atveju - REG-CLS metodu jungties funkcijos parametras nebuvo vertinamas, bet buvo įvertinami inovacijų dispersijų parametrai. Tais atvejais, kai inovacijų marginalieji skirstiniai yra Puasono, REG-CLS metodu įvertintų dispersijų parametrų reikšmės yra artimos įvertintų vidurkio parametrų reikšmėms, bet vidurkio įverčių vidutinė kvadratinė paklaida buvo mažesnė nei dispersijos įverčių. α_j parametro REG-CLS įverčių tikslumas (vidutinės kvadratinės paklaidos prasme) nepriklausė nuo inovacijų marginaliųjų skirstinių, ar jungties funkcijos. Didesnės imties parametrų įverčiai buvo tikslesni vertinant tiek REG-CLS, tiek IFM parametrų vertinimo metodais.

Atlikus simuliacijas pastebėta, kad REG-CLS metodas parametrus įvertina greičiau nei IFM metodas, ypač kai $d > 2$, nes jungties funkcijos tikimybės masės funkcijos vertinimui reikia 2^d vertinimų apskaičiuoti (26) lygtį bet kuriam reikšmių vektoriui $[a_1, ..., a_d]$ (žr. Panagiotelis ir kt. (2012)).

Nusikaltimų skaičiaus aprašymas $\mathrm{SINAR}(1)_d$ modeliu

Disertacijoje $\mathrm{SINAR}(1)_d$ modelis buvo pritaikytas Čikagos nusikaltimų duomenims. Nagrinėtos nusikaltimų skaičių per keturias valandas laiko eilutės nuo 2016 m. sausio iki 2018 m. kovo mėn. Nagrinėti keturių tipų nusikaltimai - užpuolimas (*angl.* assault), sumušimas (*angl.* battery), įsilaužimas (*angl.* burglary) ir apiplėšimas (*angl.* robbery). Laiko eilutės pasižymi sezoniniu autokoreliuotumu, todėl kiekvienai laiko eilutei atskirai sudaryti $\mathrm{SINAR}(1)_6$ modeliai, o jų parametrai įvertinti REG-CLS metodu.

Sudarytų modelių liekanų grafikai parodė, kad įvertinti modeliai atsižvelgia į laiko eilučių sezoninę autokoreliaciją, tačiau norint atsižvelgti į nesezoninę laiko eilučių autokoreliaciją, šiuos modelius reikėtų papildyti.

5 Rezultatų naujumas

Disertacijoje gauti rezultatai prisideda prie mokslinės literatūros apie sveikareikšmių dydžių laiko eilučių modelius. Iki šiol literatūroje nebuvo nagrinėtas disertacijoje pristatytas dviejų žingsnių parametrų vertinimo metodas, skirtas dvimačio sveikareikšmių dydžių autoregresinio proceso, kurio liekanos nusakomos jungties funkcija, parametrų vertinimui. Taip pat mokslinę literatūrą papildo disertacijoje pristatytas vienmatis neneigiamų sveikareikšmių dydžių autoregresinis procesas su sezoniniais autoregresiniais parametrais bei sezoniškai priklausomomis inovacijomis, ir jo daugiamatis pavidalas.

Disertacijoje atlikti originalūs ir nauji modelių taikymai paskolų skaičiaus ir nusikaltimų skaičiaus duomenims.

6 Aprobacija

Disertacijos rezultatai buvo pristatyti šiose mokslinėse konferencijose:

- *58th Conference of Lithuanian Mathematical Society*, Vilnius, Lietuva, 2017 m. birželio 21–22 d.
- *20th European Young Statisticians Meeting*, Upsala, Švedija, 2017 m. rugpjūčio 14–18 d.
- *11th International Conference on Computational and Financial Econometrics (CFE 2017) and 10th International Conference of the ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on Computational and Methodological Statistics (CMStatistics 2017)*, Londonas, Didžioji Britanija, 2017 m. gruodžio 16–18 d.

- *Inaugural Baltic Economic Conference, organized jointly with the 7th Annual Lithuanian Conference on Economic Research*, Vilnius, Lietuva, 2018 m. birželio 11–12 d.

- *12th International Vilnius Conference on Probability Theory and Mathematical Statistics and 2018 IMS Annual Meeting on Probability and Statistics*, Vilnius, Lietuva, 2018 m. liepos 2–6 d.

- *60th Conference of Lithuanian Mathematical Society*, Vilnius, Lietuva, 2019 m. birželio 19–20 d.

- *13th International Conference on Computational and Financial Econometrics (CFE 2019) and 12th International Conference of the ERCIM Working Group on Computational and Methodological Statistics (CMStatistics 2019)*, Londonas, Didžioji Britanija, 2019 m. gruodžio 14–16 d.

7 Publikacijos

Disertacijos rezultatai yra publikuoti šiuose straipsniuose:

- A. Buteikis ir R. Leipus. A copula-based bivariate integer-valued autoregressive process with application. *Modern Stochastics: Theory and Applications*, 6(2):227–249, 2019.

- A. Buteikis ir R. Leipus. An integer-valued autoregressive process for seasonality. *Journal of Statistical Computation and Simulation*, 90(3):391–411, 2020.

8 Konferencijų pranešimų leidiniai

Disertacijos rezultatai taip pat pateikti kaip santraukos ir trumpas straipsnis šiuose konferencijų leidiniuose:

Trumpas straipsnis:

- A. Buteikis. Copula based BINAR models with applications. *Proceedings of the 20th European young statisticians meeting, Uppsala University, Uppsala, Sweden, 14-–18 August 2017*, ISBN: 978-91-506-2680-3.

Santraukos:

- A. Buteikis ir R. Leipus. Application of copula-based BINAR models in loan modelling. *11th international conference on computational and financial econometrics (CFE 2017) and 10th international conference of the ERCIM (European Research Consortium for informatics and mathematics) working group on computational and methodological statistics (CMStatistics 2017), London, 16–18 December 2017 : programme and abstracts.*

- A. Buteikis ir R. Leipus. Application of copula-based BINAR models in loan modelling. *12th International Vilnius Conference on Probability Theory and Mathematical Statistics and 2018 IMS Annual Meeting on Probability and Statistics, July 2–6, 2018, Vilnius, Lithuania.*

- A. Buteikis ir R. Leipus. An integer-valued autoregressive process for seasonality. *13th International Conference on Computational and Financial Econometrics (CFE 2019) and 12th International Conference of the ERCIM Working Group on Computational and Methodological Statistics (CMStatistics 2019), London, 14–16 December 2019: programme and abstracts.*

9 Išvados

Šioje disertacijoje nagrinėjami sveikareikšmių dydžių autoregresiniai modeliai, kurių inovacijos nusakomos jungties funkcija. Nagrinėjamas pirmos eilės dvimatis sveikareikšmių dydžių autoregresinis procesas (BINAR(1)), kurio inovacijos yra aprašomos jungties funkcija. Pristatomas dviejų žingsnių modelio parametrų vertinimo metodas, kuriame BINAR(1) modelio parametrai ir inovacijų jungties funkcijos parametras vertinami atskirai. Šia procedūra parametrai įvertinami greičiau ir panašiu tikslumu kaip taikant sąlyginį didžiausio tikėtinumo metodą, kuris gali būti taikomas visų modelio parametrų vertinimui.

Disertacijoje pristatytas iki šiol literatūroje nenagrinėtas procesas – vienmatis neneigiamų sveikareikšmių dydžių autoregresinis procesas su sezonu d ir sezono viduje priklausomomis inovacijomis, kurių priklausomybė gali būti nusakoma jungties funkcija (SINAR$(1)_d$). Šis vienmatis procesas taip pat gali būti perrašytas daugiamačiu pavidalu. Daugiamatis proceso pavidalas leidžia modelio parametrus vertinti apibendrintu sąlyginiu mažiausių kvadratų metodu, uždedant tam tikrus apribojimus parametrams ir atsižvelgiant į inovacijų priklausomybę. Šis metodas palyginamas su dviejų žingsnių tikėtinumo funkcija paremtu metodu.

Disertacijoje BINAR(1) ir SINAR$(1)_d$ modelių parametrų vertinimo metodai palyginti Monte Karlo metodu, taip pat modeliai pritaikyti paskolų skaičiaus (BINAR(1)) ir nusikaltimų skaičiaus (SINAR$(1)_6$) duomenims.

Literatūra

Al-Osh, M. ir Alzaid, A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8, 261–275.

Barczy, M., Ispány, M., Pap, G., Scotto, M., ir Silva, M. E. (2010). Innovational outliers in INAR(1) models. *Communications in Statistics - Theory and Methods*, 39(18), 3343–3362.

Bourguignon, M., Vasconcellos, K. L., Reisen, V. A., ir Ispány, M. (2016). A Poisson INAR(1) process with a seasonal structure. *Journal of Statistical Computation and Simulation*, 86(2), 373–387.

Brännäs, K. (1995). Explanatory variables in the AR(1) count data model. *Umeå Economic Studies 381.*

Buteikis, A. ir Leipus, R. (2019). A copula-based bivariate integer-valued autoregressive process with application. *Modern Stochastics: Theory and Applications*, 6(2), 227–249.

Buteikis, A. ir Leipus, R. (2020). An integer-valued autoregressive process for seasonality. *Journal of Statistical Computation and Simulation*, 90(3), 391–411.

Ding, X. ir Wang, D. (2016). Empirical likelihood inference for

INAR(1) model with explanatory variables. *Journal of the Korean Statistical Society*, 45(4), 623–632.

Enciso-Mora, V., Neal, P., ir Rao, T. S. (2009). Integer valued AR processes with explanatory variables. *Sankhyā: The Indian Journal of Statistics, Series B*, 71(2), 248–263.

Franke, J. ir Seligmann, T. (1993). Conditional maximum likelihood estimates for INAR(1) processes and their application to modelling epileptic seizure counts. In T. Subba Rao (Ed.), *Developments in Time Series Analysis* (pp. 310–330). Chapman & Hall: London.

Genest, C. ir Nešlehová, J. (2007). A primer on copulas for count data. *Astin Bulletin*, 37(2), 475–515.

Gladyshev, E. G. (1961). Periodically correlated random sequences. *Sov. Math., Dokl.*, 2, 385–388.

Hurd, H. ir Miamee, A. (2007). *Periodically Correlated Random Sequences: Spectral Theory and Practice*. Wiley Series in Probability and Statistics. Wiley.

Ispány, M., Pap, G., ir van Zuijlen, M. C. A. (2003). Asymptotic behaviour of estimators of the parameters of nearly unstable INAR(1) models. In Y. Haitovsky, Y. Ritov, ir H. R. Lerche (Eds.), *Foundations of Statistical Inference* (pp. 194–206). Heidelberg: Physica-Verlag HD.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2), 401–419.

Karlis, D. ir Pedeli, X. (2013). Flexible bivariate INAR(1) processes using copulas. *Communications in Statistics - Theory and Methods*, 42, 723–740.

Latour, A. (1997). The multivariate GINAR(p) process. *Advances in Applied Probability*, 29(1), 228–248.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis*, 19(4), 439–455.

Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin Heidelberg.

McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, 18(3), 679–705.

Monteiro, M., Scotto, M. G., ir Pereira, I. (2010). Integer-valued autoregressive processes with periodic structure. *Journal of Statistical Planning and Inference*, 140(6), 1529–1541.

Nikoloulopoulos, A. K. ir Karlis, D. (2009). Modeling multivariate count data using copulas. *Communications in Statistics - Simulation and Computation*, 39(1), 172–187.

Panagiotelis, A., Czado, C., ir Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, 107(499), 1063–1072.

Pedeli, X. ir Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical Modelling: An International Journal*, 11(4), 325–349.

Silva, I. M. M. (2005). Contributions to the analysis of discrete-valued time series. *PhD thesis, University of Porto.*

UŽRAŠAMS

Vilniaus universiteto leidykla
Saulėtekio al. 9, LT-10222 Vilnius
El. p. info@leidykla.vu.lt,
www.leidykla.vu.lt
Tiražas 35 egz.