

Ekonominių laiko eilučių komponentių įverčių tikslumo analizė

Vitalija KLOKOVA (VGTU)

el. paštas: vitalija.kloкова@mail.std.lt

1. Įvadas

Dabartiniu metu pasaulyje vis plačiau naudojamos sezoniškai išlygintos laiko eilutės. Tai reiškia, kad laiko eilutėje pašalinama sezoninė komponentė. Kelios priežastys lemia šiuos veiksmus:

- sezoniniai svyravimai apsunkina analitikų darbą – sudėtinga pastebėti trumpalaikes tendencijas, kurios yra labai svarbios;
- sezoniniai svyravimai neleidžia palyginti skirtingus metų periodus, o taip pat skirtingų šalių atitinkamus rodiklius;
- dėl sezoninių svyravimų laiko eilutė yra nestacionari. Daugeliui laiko eilučių analizės metodų reikia stacionarių laiko eilučių.

Todėl labai svarbu gerai įvertinti ir pašalinti sezoninę komponentę. Yra daug programinės įrangos, kuri leidžia tai atlikti. Šiame straipsnyje nagrinėjama programa Demetra. Ši programa buvo parinkta dėl kelių priežasčių. Pirmiausia, ji yra Eurostat'o (centrinis Europos statistikos biuras) produktas. Ji buvo specialiai sukurta sezoninės komponentės išskyrimui. Tai yra patogus vartotojui dviejų programų apvalkalas – TRAMO-SEATS (parametriniai metodai) ir X-12-RegARIMA (nparametriniai metodai). Nuo 2001 metų Statistikos Departamentas privalo siųsti sezoniškai išlygintus duomenis Eurostat'ui. Sezoniniam išlyginimui atlikti Eurostat'as siūlo naudoti Demetrą. Tai yra laisvai platinama nemokama įranga. Beje ši programa vertina ne tik modelio komponentes, bet ir jas prognozuoja. Kuo geriau matematinis modelis aprašo duotą laiko eilutę, tuo geresnių rezultatų galima tikėtis sezoninės komponentės išskyrimo, taip pat ir laiko eilutės prognozavime.

Šiame darbe modeliuotų sekų pagalba buvo atlikta komponentių įvertinimo tikslumo analizė.

2. Rezultatai

Ekonominėms laiko eilutėms aprašyti dažniausiai naudojamas modelis:

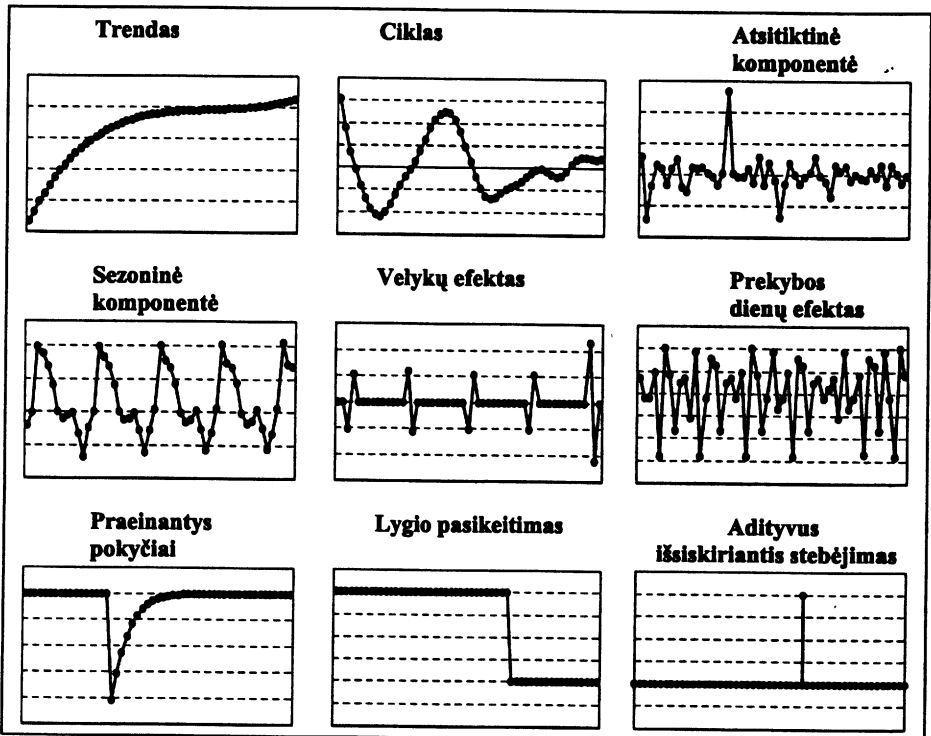
$$X_t = T_t \cdot C_t \cdot S_t \cdot I_t \cdot A_t \quad \text{arba}$$

$$\log(X_t) = \log(T_t) + \log(C_t) + \log(S_t) + \log(I_t) + \log(A_t),$$

čia

- T_t – trendo komponentė (lėtai kintanti, ne ciklinė komponentė aprašanti ilgalaikių faktorių įtaką kurių poveikis atsiliepia palaiapsniui);
- C_t – ciklo komponentė (aprašo santykinio nuosmukio ir pakilimo ilgalaikius periodus. Komponentė sudaryta iš ciklų, kurie yra nepastovios amplitudės ir ilgio);
- S_t – sezoninė komponentė (atspindi pasauliui ir žmogaus veiklai būdingus pasikartojimus laike. Apibendrintai tai yra sezonas, darbo ir prekybos dienų, švenčių ir Velykų efektas);
- I_t – išsiskiriantys stebėjimai (dirbtinis kintamasis, kuris parodo iki šios nestebėtų veiksnių įtaką (skaičiavimo metodikos pasikeitimas, krizės ir t.t). Tai gali būti lygio pokytis, praeinantys pokyčiai, adityvūs išsiskiriantys stebėjimai);
- A_t – atsitiktinė komponentė (visų kitų neištrauktų faktorių įtaka).

Nagrinėta 240 modeliuotų laiko eilučių su skirtingomis komponentėmis ir komponentių parametrais. Nagrinėjami 5 pavidalų trendai, 6 skirtingų amplitudžių sezoninės komponentės (3 determinuotos, 3 kintamos), 3 lygio pokyčio kintamieji, 3 adityvūs išsiskiriančių stebėjimų kintamieji ir praeinančio pokyčio kintamasis. Lygio pokyčio bei adityvūs išsiskiriančių stebėjimų kintamieji (10%, 20% ir 30%) buvo modeliuoti stengiantys imituoti realias laiko, pvz., dėl Rusijos krizės atskirų ekonominės veiklos rūšių produkcijos nuosmukis sudarė nuo 10% iki 30%.



1 pav. Atskirų komponentių pavyzdžiai.

1. $y_t = T_t^{tr} \cdot S_t^{amp} \cdot A_t, tr = 1, \dots, 5, amp = 1, \dots, 6,$
2. $y_t = T_t^{tr} \cdot S_t^{amp} \cdot LP_t^{lyg} \cdot A_t, tr = 1, \dots, 5, amp = 1, \dots, 6, lyg = 10\%, 20\%, 30\%,$
3. $y_t = T_t^{tr} \cdot S_t^{amp} \cdot IS_t^{lyg} \cdot A_t, tr = 1, \dots, 5, amp = 1, \dots, 6, lyg = 10\%, 20\%, 30\%,$
4. $y_t = T_t^{tr} \cdot S_t^{amp} \cdot PP_t \cdot A_t, tr = 1, \dots, 5, amp = 1, \dots, 6,$

čia

T_t – trendo komponentė,

S_t – sezoninė komponentė,

A_t – atsitiktinė komponentė,

LP_t – lygio pokytis,

IS_t – adityvus išsiskiriantis stebėjimas,

PP_t – praeinantis pokytis.

Norint imituoti realias laiko eilutes modeliuotų laiko eilučių ilgis – 44 stebėjimai. Kadangi laiko eilutės nėra ilgos, norint sumažinti vertinamų parametru skaičių į modelių neįtraukta ciklinė komponentė. Modelių vertinimui buvo naudoti pirmieji 40 stebėjimų, o likusieji 4 stebėjimai naudoti įvertinti modelio prognozės paklaidą. Deja programinėje įrangoje nėra numatytos galimybės įvertinti modelio tikslumą naudojant standartinius metodus (pvz., Jack-knife metodą). Todėl vertinant modelio tikslumą naudojamos dvejų tipų paklaidos:

– absoliuti santykinė modelio paklaida:
$$\Delta = \frac{1}{40} \sum_{t=1}^{40} \frac{|y_t - \hat{y}_t|}{y_t} \cdot 100\% \text{ ir}$$

– absoliuti santykinė prognozės paklaida:
$$\Delta^P = \frac{1}{4} \sum_{t=41}^{44} \frac{|y_t - \hat{y}_t|}{y_t} \cdot 100\%.$$

Programoje pateikiami du vertinimo būdai: automatinis režimas ir vartotojo nustatytas režimas. Pirmu atveju pakanka nurodyti pagrindines taisykles ir laiko eilutės skaidymas į komponentes atliekamas Demetra. Antru atveju vartotojas vadovaujasi savo nuožiūra ir skaidymas į komponentes vyksta pagal jo nustatytą tvarką. Šiame darbe pateikiami rezultatai gauti naudojantis automatinio režimu. Kiekvienai laiko eilutei parinkti keturi TRAMO-SEATS ir keturi X-12-RegARIMA modeliai. Bendrai buvo įvertinta 1920 variantų (240 modeliuotų laiko eilučių aštuoniais metodais). Taikomi modeliai skiriasi ARIMA modelio parinkimu pradinei laiko eilutei. Naudojant automatinį režimą galimi tris parinkimo būdai:

1. Automatinis identifikavimas ir įvertinimas (keliems ARIMA modeliams skaičiuojamos paklaidos ir parenkamas tas, kur jos mažiausios);
2. Automatinis parinkimas ir įvertinimas (įvertinimas tik vienas modelis (011)(011));
3. Galima pasirinkti bet koki ARIMA modelį (darbe buvo atliktas skaičiavimas su dviem modeliais (111)(011) bei (111)(111), nes šie modeliai nėra įtraukti į pirmo parinkimo sąrašą).

Priklausomai nuo laiko eilutės struktūros ir automatinio modelio parinkimo metodo gauti skirtingas paklaidas. Darbo tikslas buvo panagrinti paklaidų dydžių priklausomybę nuo šių veiksnių:

- sezoninės komponentės pavidalo;
- laiko eilutės struktūros;
- komponenčių vertinimo metodų.

Deja, ne visos laiko eilutės buvo suskaidytos į komponentes. Demetroje numatytas kriterijų sąrašas, kuriuos turi tenkinti įverčiai. Šiuos kriterijus galime sugrupuoti taip:

- liekanų statistikos;
- prognozės paklaidos;
- išsiskiriančių stebėjimų diagnostika;
- ARIMA modelio parinkimo kriterijai;
- komponentių reikšmingumo tikrinimas.

Modeliai netenkinantys bent vieno kriterijaus buvo atmesti. Tokias laiko eilutės patariama nagrinėti individualiai. Šiame darbe individualių modelių parinkimų atskiroms laiko eilutėms nenagrinėsime.

Žemiau 1, 2, 3 lentelėse pateikti paklaidų skaičiavimo rezultatai.

1 lentelė. Paklaidos pagal sezoninės komponentės pavidalą (determinuota ir kintama) ir išsibarstymo lygį

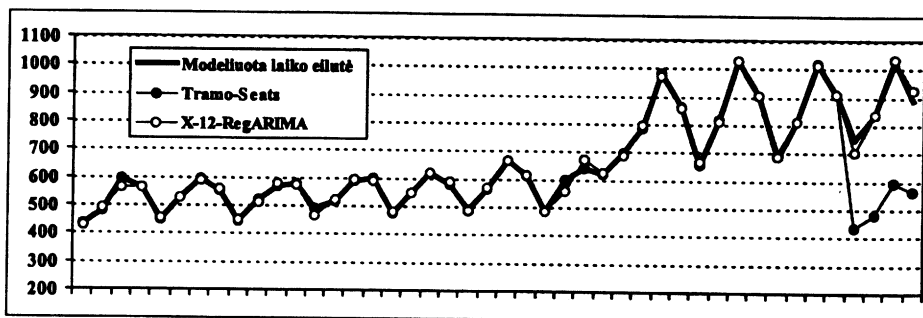
	Sez1	Sez2	Sez3	Sez4	Sez5	Sez6	Vidutinė
Modelio paklaida	2,1%	2,0%	2,0%	1,9%	2,0%	2,0%	2,0%
Prognozės paklaida	14,5%	14,2%	14,8%	13,3%	14,4%	14,8%	14,3%
Įvertintų modelių dalis	79%	85%	81%	79%	88%	84%	83%

2 lentelė. Paklaidos pagal išsiskiriančių stebėjimų formą ir lygį

	Be Iš.s.	LP10%	LP20%	LP30%	IS10%	IS20%	IS30%	PP
Modelio paklaida	1,7%	1,8%	1,8%	1,8%	1,9%	2,8%	3,6%	1,4%
Prognozės paklaida	11,1%	14,7%	17,1%	25,3%	10,2%	12,1%	15,0%	11,0%
Įvertintų modelių dalis	80%	92%	96%	74%	93%	74%	53%	100%

3 lentelė. Paklaidos pagal vertinimo metodus

	T-S 1	T-S 2	T-S 3	T-S 4	X-12 1	X-12 2	X-12 3	X-12 4
Modelio paklaida	2,1%	2,4%	2,5%	2,3%	1,6%	1,5%	1,5%	1,6%
Prognozės paklaida	21,3%	21,0%	21,0%	21,0%	4,3%	4,4%	4,5%	4,3%
Įvertintų modelių dalis	100%	98%	98%	98%	57%	73%	76%	60%



2 pav. Skirtumas tarp įverčio, gauto TRAMO-SEATS ir X-12-RegARIMA.

Labai didelės prognozės paklaidos gautos naudojant TRAMO-SEATS procedūras. Tai atsitiko dėl blogos išsiskiriančių stebėjimų specifikacijos. Realiai buvo sumodeliuotas kintamasis „Lygio pokytis“, bet programa įvertino tai kaip „Išsiskiriantis stebėjimas“. Rezultate gavome labai didelę prognozės paklaidą. Tai pavaizduota 2 pav.

Šio darbo išvados būtų tokios:

- papildomai gavus naują stebėjimą ir perskaičiavus rezultatus, parametriniai metodai pakeičia įverčius daugiau, nei neparimetriniai;
- parametriniais metodais gautų įverčių paklaidos yra didesnės;
- sezoninės komponentės pavidalas (determinuota ar kintama) ir amplitudė nedaro didelės įtakos modelio tikslumui;
- didėjant išsiskiriančių stebėjimų lygiui paklaidos didėja;
- naudojant automatines procedūras Tramo-Seats įvertinta daugiau variantų;
- negalima pasitikėti vien automatinėmis procedūromis, nes kartais gauti rezultatai neatitinka realių situacijų, verta panagrinėti ir vartotojo nustatytą režimą;
- programinė įranga Demetra kol kas negalutinis produktas, nes buvo aptiktos klaidos. Kai kurios iš jų yra pataisytos.
- kadangi ši programinė įranga yra nemokama ir siūloma visom šalim naudoti sezoniniam išlyginimui, išskyla būtinumas toliau nagrinėti ir tobulinti jos galimybes.

Literatūra

- [1] D.F. Findley, B.C. Monsell, W.R. Bell, M.C. Otto, B.C. Chen, New capabilities and methods of the X12 ARIMA seasonal adjustment program, *Journal of Business and Economic Statistics*, 16, 127–177.
- [2] V. Gomez, A. Maravall, Automatic modelling methods for univariate series, *Working Papers 9808*, Research Department, Bank of Spain.
- [3] A. Maravall, Short term analysis of macroeconomic time series, in: A.Gerard-Varet and A. Kirman (eds.), *Economics: beyond the Millennium*, Oxford, Oxford University Press (1999).
- [4] A. Maravall, Unobserved components in economic time series, In H. Pesaran and M. Wickens (eds.), *The Handbook of Applied Econometrics*, vol. 1, Oxford, Basil Blackwell (1995).

Accuracy analysis of different components evaluation of time series

V. Klokova

Mathematical models for decomposing and forecasting of time series are analysed in this paper. Sometimes we have the situations when we don't get good result describing time series by standard components (trend, cyclical, seasonal). It is often happened with economic indicators. For example, the method of evaluation is changed or economic crisis has happened. Inclusion of new variables – intervention variables – can help with these problems. There is investigation of model accuracy depending on accuracy of evaluation of different components. Time series of different structure were simulated for imitation of real situations. Estimation of time series models with intervention variables and traditional components was done by program Demetra.