

SUDĖTINIAI RODIKLIAI – PAPILDOMAS MATAS LIETUVOS EKONOMIKAI VERTINTI

Dokt. Jurga Rukšėnaitė
Vilniaus Gedimino technikos universiteto
Fundamentinių mokslų fakulteto
Matematinės statistikos katedra
Saulėtekio al. 11, 10223 Vilnius
El. p. jurga.ruksenaite@vgtu.lt

Norint apibūdinti šalies ekonomikos vystymąsi, kaip rodiklis labai dažnai pasirenkamas bendrasis vidaus produktas (BVP). Analizuojant šio rodiklio, pvz., vienam gyventojui tenkančio BVP, tendencijas, nuskamba ne tik šalies ekonomikos raida, bet ir atskirų gyventojų turtingumas ar skurdas. Tačiau pastaruoju metu galima rasti vis daugiau straipsnių, kuriuose pagrindžiamas teiginys, kad BVP ne visada atspindi tikrąjį šalies vystymąsi, žmonių gyvenimo kokybę ir technologijų plėtrą [5].

Kaip pavyzdį galima nagrinėti pastarąją ekonominę finansinę krizę, kuri palietė ne tik Europą, bet išplito ir pasauliniu mastu. Lietuvoje, vertinant BVP pokyčius, krizė prasidėjo tik 2008 m. ketvirtąjį ketvirtį – taip rodė statistiniai BVP duomenys. Tačiau tam tikri rodikliai prasidėjusį ekonomikos nuosmukį rodė gerokai anksčiau: pradėjo mažėti statybų, pramonės produkcijos mastas ir t. t.

Taigi, iškyla matematinis uždavinys sukurti sudėtinį rodiklį, kuris leistų tiksliau ir objektyviau įvertinti šalies ekonomikos tendencijas.

Šiame darbe pateikiamas naujas sudėtinis rodiklis, kuris gali būti priemonė analizuojant ir vertinant Lietuvos ekonomikos tendencijas. Darbe apibūdinamas šis rodiklis ir pateikiama jo kūrimo metodika, paremta faktorinės analizės teorija. Šiam rodikliui sudaryti buvo naudojami skirtingi socialiniai ir ekonominiai rodikliai, atspindintys įvairius Lietuvos ekonomikos aspektus.

Sudėtiniai rodikliai (angl. *composite indicator* – CI) padeda įvertinti šalies vystymąsi, yra reikalingi nuostatant finansinius ir politinius prioritetus, suteikia galimybę palyginti atskirų šalių raidą. CI leidžia palyginti skirtingus objektus kompleksiskai, apimant įvairias sritis: ekonomiką, technologinę pažangą, socialinę aplinką ir kitas. Naudojant CI dažniausiai yra vertinamos atskiros valstybės regionų arba pasaulio mastu. Pavyzdžiais galėtų būti konkurencingumo indeksas (angl. *competitiveness index*), žmogaus socialinės raidos indeksas (angl. *human development index*) ir kiti.

Apibrėžkime CI kaip duotą matematinę funkciją-modelį:

$$CI = f(X, W), \quad (1)$$

kur:

X – rodiklių, priklausančių įvairioms sritims, pvz., verslo statistikos, technologijų ir pan., rinkinys, o pačios sritys dažnai skirstomos į mažesnes dalis, pvz., verslo statistika gali būti skirstoma į pramonės, statybos statistiką ir t. t.;

W – svorių matrica, sudaryta iš atskirų rodiklių duomenų svorių.

Sudėtinio rodiklio matematinis modeliavimas susideda iš keleto etapų: atrenkami paplitę rodikliai, išsamiai apibūdinantys Lietuvos ūkį; nustatomas kiekvieno rodiklio svoris; rodikliai agreguojami – taip gaunamas sudėtinis rodiklis. Naujojo rodiklio pokyčiai lyginami su BVP raida.

Šiame darbe daugiau akcentuojamas svorių nustatymas, kartu pateikiamas siūlomo metodo pritaikymas šalies ūkio tendencijoms vertinti.

METODIKOS APŽVALGA

Šiame skyriuje aprašomi CI modeliavimo etapai.

Pirmame ir antrame etape (etapai susieti) pasirenkamas pirminis rodiklių rinkinys X ir sumažinama rodiklių aibė. Modelyje paliekami informatyviausi rodikliai, jų tinkamumui įvertinti naudojama faktorinė analizė. Trečiame etape nustatomas kiekvieno rodiklio svoris formulėje (1).

Rodiklių matricos X nustatymas

Pradedant konstruoti CI, pirmiausia pasirenkamas pirminis rodiklių rinkinys X_i , $i = 1..n^*$. Toliau yra vertinama kiekvieno rodiklio koreliacija su pagrindiniu ekonomikos indikatoriumi – BVP. Rodikliai, kurie silpnai ar vidutiniškai ($\rho_{X_i, BVP} < 0,75$) koreliuoja su BVP, šalinami iš tolimesnės analizės. Atrinkti $n < n^*$ rodikliai yra standartizuojami – įgauna nulinį vidurkį ir vienietinę dispersiją.

Apibrėžkime tiesinį faktorinės analizės (FA) modelį:

$$X = LF + E, \quad (2)$$

čia L – svorių (angl. loadings) matrica, F – latentinių faktorių matrica, E – charakteringųjų faktorių vektorius.

FA modelio prielaidos:

1. bendrieji faktoriai F_j nekoreliuoti, o dispersija $DF_j = 1$;
2. charakteringieji faktoriai E_i nekoreliuoti, o $DE_i = \tau_i$;
3. faktoriai F_j ir E_i nekoreliuoti, čia $j = \overline{1..p}$.

FA modelyje kiekvieno pradinio kintamojo X_i dispersija skaidoma į dvi dalis – dispersiją, kurią sąlygoja latentiniai bendrieji faktoriai (bendrumą h_i^2), ir dispersiją, kurią lemia bendraisiais faktoriais nepaaiškinama paties kintamojo variacija (specifiškumą τ_i) [2]. (Norint atskirti FA svorius nuo galutinių CI svorių, kai kuriose straipsnio vietose skliausteliuose bus nurodytas ir angliškas terminas *loadings*.)

Naudojant FA modelį reikia atsižvelgti į faktorių pasirinkimo problemas:

1. faktorių išskyrimas priklauso nuo pasirinkto metodo;
2. faktorių rinkinių gali būti daugiau nei vienas;
3. nevienareikšmiškas interpretavimas.

Darbe naudoti Bartleto sferiškumo (angl. *Bartlett's test of sphericity*) ir Kaizerio, Mejerio ir Olkino (KMO) duomenų tikimo faktorinei analizei testai.

Bartleto sferiškumo testas tikrina hipotezę H_0 , ar korelacių matrica yra vienetinė: $R = I$ (visi stebimi kintamieji nekoreliuoti). Jeigu hipotezė H_0 priimama, reiškia, kad FA modelio taikyti negalima.

KMO matas vertina, ar duomenys tinkami FA modeliui:

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}}{\sum \sum_{i \neq j} r_{ij} + \sum \sum_{i \neq j} r_{ij}^2}, \quad (3)$$

čia r_{ij} – korelacijos koeficientas, r_{ij}^2 – dalinės korelacijos koeficientas (tarpusavio korelacija nustatoma pašalinus likusių rodiklių įtaką). Kuo didesnė KMO reikšmė, tuo duomenys labiau tinka FA modeliui. Jeigu $KMO < 0,5$ – duomenys netinkami modeliui.

Panašiai įvertinamas kiekvieno rodiklio tinkamumas FA:

$$MSA_i = \frac{\sum_{i \neq j} r_{ij}}{\sum_{i \neq j} r_{ij} + \sum_{i \neq j} r_{ij}^2}, \quad (4)$$

Rodikliai, kurių MSA_i reikšmė žema, po viena pašalinami iš tolimesnės analizės.

Faktorių išskyrimas ir sukimas

Faktoriams išskirti pasirinktas pagrindinių komponentių analizės (PKA) metodas. Šiuo atveju faktoriais F_j laikomos sunormuotos pagrindinės komponentės. Suformuluojamas PKA uždavinys: rasti pagrindinių komponentių tiesines daugdaras:

$$Y_1 = \sum_{j=1}^m \alpha_{1j} X_j, \quad \dots, \quad Y_m = \sum_{j=1}^m \alpha_{mj} X_j, \quad (5)$$

su sąlygomis:

$$1) \text{cov}(Y_i, Y_j) = 0, \quad i, j = \overline{1..m}, \quad i \neq j,$$

$$2) DY_1 \geq DY_2 \geq \dots \geq DY_m,$$

$$3) \sum_{i=1}^m DY_i = \sum_{i=1}^m DX_i^{def} = S. \quad (5')$$

Išsprendus PKA uždavinį, gaunami koeficientų α_{ij} , $i, j = \overline{1..m}$ įverčiai a_{ij} ir m pagrindinių komponentių įverčiai:

$$\hat{Y}_i = \sum_{j=1}^m a_{ij} X_j, \quad i = \overline{1..m} \quad (6)$$

Pasirenkama prielaida palikti pirmąsias p komponentes Y_1, \dots, Y_p , kurios paaiškina ne mažiau kaip 95 proc. bendrosios dispersijos. Taigi ir faktorių skaičius atitinkamai bus p .

Pradiniai rodikliai pagrindinėmis komponentėmis išreiškiami taip:

$$X_i = \sum_{j=1}^m \alpha_{ji} Y_j, \quad i = \overline{1..m}. \quad (7)$$

Turint korelacių matricą R_{XX} , galima apskaičiuoti bendruosius faktorius F_j , faktorių svorių l_{ij} , bendrumų h_i^2 ir specifiškumą τ_i įverčius. Pačių faktorių ir jų specifiškumų įverčiai nenaudojami, todėl jų įvertinimas nebus akcentuojamas.

Faktorių svorių įverčiai apskaičiuojami taip:

$$\hat{l}_{ij} = a_{ji} \sqrt{s^2(\hat{Y}_j)}, \quad i = \overline{1..m}, \quad j = \overline{1..p_\lambda} \quad (8)$$

Naudojami bendrumų h_i^2 įverčiai h_i^2 , norint įvertinti rodiklio X_i tinkamumą modelyje:

$$\hat{h}_i^2 = \sum_{j=1}^k \hat{l}_{ji}^2 = \sum_{j=1}^k a_{ji}^2 s^2(\hat{Y}_j), \quad (9)$$

čia ir (8) $s^2(\hat{Y}_j)$ – i -osios komponentės dispersijos DY_i įvertis.

FA modelis (2) taikomas keletą kartų, kiekvieną kartą pašalinant tą rodiklį X_i , kurio bendrumų įvertis h_i^2 (9) buvo mažiausias rinkinyje: $h_i^2 < 0,8$. Procedūra kartojama tol, kol gaunamas rodiklių rinkinys X , kur išskirti faktoriai gerai paaiškina kiekvieną X_i . Tokiu būdu sumažinama rodiklių aibė: $R^h \rightarrow R^m$.

Gautoji faktorių svorių matrica L yra nevienareikšmė, t. y. egzistuoja aibė matricų, tenkinančių FA modelio (2) prielaidas.

Iš to kyla teorinis uždavinys – faktorių svorių matricai suteikti paprastosios struktūros (angl. *simple structure*) pavidalą (šis terminas gana senas, jį 1947 m. aprašė Thurstonas). Paprastoji struktūra reiškia, kad kiekvienas rodiklis yra susijęs tik su vienu faktoriumi.

Praktiškai tai sunkiai įgyvendinama, todėl stengiamasi rasti struktūrą, panašią į paprastąją.

Šiam tikslui įgyvendinti darbe naudotas ortogonalusis faktorių sukimas (*Varimax* metodas), nes šis metodas po sukimo išlaiko faktorių ortogonalumą, t. y. faktoriai išlieka nekoreliuoti (po faktorių sukimo gaunama nauja svorių (angl. *loadings*) matrica L^*). *Varimax* metodas dažnai naudojamas ekonominiuose tyrimuose [1, 3]. Šiuo metodu ieškoma kuo paprastesnės struktūros L , kad faktorių bendroji dispersija būtų kuo didesnė. Maksimizuojamas paprastumo matas yra apibrėžiamas lygybe:

$$\max(V) = \frac{1}{m^2} \left(\sum_{j=1}^p m \sum_{i=1}^m l_{ij}^4 - \sum_{j=1}^p \left(\sum_{i=1}^m l_{ij}^2 \right)^2 \right), \quad (10)$$

čia l_{ij}^2 – svoriai i -ojo kintamojo j -ajam faktoriui.

Svorių įvertinimas ir agregavimas

Indikatorių svorių parinkimas yra vienas iš esminių žingsnių, konstruojant CI, tiesiogiai įtakojantis CI reikšmes. Paprastai pasukus faktorius negaunama paprastosios svorių matricos L struktūros, kurią būtų galima panaudoti CI svoriams vertinti, todėl reikalingas teorinis metodas, leidžiantis įvertinti svorius iš turimos pasuktosios matricos L^* . Toliau šiame darbe bus naudojamas Nicoletti pasiūlytas metodas (Nicoletti et al., 2000), kurio esmę sudaro didžiausius svorius (angl. *loadings*) turinčių faktorių grupavimas į tarpinius sudėtinius rodiklius (angl. *intermediate composites*) (IC_k^i), $k = 1..p$ [4]. Paprastai sakant, šis metodas didesnius svorius suteikia tiems rodikliams, kurie labiau koreliuoja su atitinkamu faktoriumi.

CI svoriams įvertinti naudojamas algoritmas (formulėse yra anksčiau apibrėžti rodikliai):

$$\begin{aligned} 1) \tilde{l}_{ij} &= \frac{l_{ij}^2}{S_j}, \\ 2) w_{IC_k}^i &= \begin{cases} \max_i \tilde{l}_{ij}, & \text{jeigu } j=k - \text{svoriai } IC_k^i, \\ 0, & j \neq k \end{cases} \\ 3) w_{IC_k}^{*i} &= w_{IC_k}^i \cdot S_j / S - \text{korekcija,} \\ 4) w_i &= w_{IC_k}^{*i} / \sum_i w_{IC_k}^{*i} - \text{normalizacija.} \quad (11) \end{aligned}$$

Taigi, gautas svorių vektorius W naudojamas formulėje (1) $CI = f(X, W)$.

CI gauti buvo pasirinkta tiesinė agregavimo technika:

$$CI(t, w) = \sum X_i(t) \cdot w_i, \quad t = \overline{1..T}. \quad (12)$$

PRAKTINIS METODO TAIKYMAS

Darbe naudoti Statistikos departamento Rodiklių duomenų bazės duomenys. Pasirinktas 2000–2009 m. laikotarpis, norint išvengti 1998 m. Rusijos krizės įtakos kai kuriems statistiniams rodikliams iki 2000 m. Ketvirtinis rodiklių periodiškumas buvo pasirinktas tam, kad gautus rezultatus būtų lengviau interpretuoti ir lyginti su BVP.

Iš socialinės-ekonominės srities buvo pasirinkti įvairūs rodikliai ($n^* = 41$), kurie pastebimai koreliuoja su BVP. 1 lentelėje yra pateikta keletas rodiklių, kurių koreliacija su BVP yra didžiausia.

1 lentelė. Stipriausiai su BVP koreliuojantys rodikliai X_i

Sritis	Rodikliai X_i	Koreliacijos koeficientas $\rho_{X_i, BVP}$
Verslo statistika	Mažmeninės prekybos apyvarta, išskyrus variklinių transporto priemonių ir motociklų prekybą, automobilių degalų prekybą	0,996
Užsienio prekyba	Eksportas	0,974
Verslo statistika	Mažmeninės prekybos maitinimo ir gėrimų teikimo veiklos apyvarta	0,971
Ekonomika ir finansai	Tiesioginės užsienio investicijos	0,965
Verslo statistika	Pramonės produkcija	0,957
	Apdirbamosios gamybos pramonės produkcija	0,957
Užsienio prekyba	Importas	0,955

* Pearsono koreliacijos koeficientas, reikšmingumo lygmuo – 0,01.

Pagrindiniai atskirų analizės etapų rezultatai

1. Pirminė statistinių duomenų analizė ir faktorių išskyrimas.

Pirmiausia buvo atlikta pirminė statistinių duomenų analizė: įrašytos trūkstamos reikšmės, duomenys sezoniškai išlyginti bei standartizuoti. Palikti tik tie X_i , kurie stipriai koreliuoja su BVP ($\rho_{X_i, BVP} \geq 0,75$), likusieji buvo pašalinti.

Duomenų kokybė ir tinkamumas FA buvo patikrinti naudojant teorinėje dalyje aprašytus testus. Bartlettto sferiškumo testas atmetė hipotezę H_0 , kad pasirinkti rodikliai tarpusavyje koreliuoja. KMO testu buvo gauta reikšmė 0,809, reiškianti, kad duomenys tinkami FA. Taip pat buvo analizuojami atskirų rodiklių bendrumai h_i^2 ir buvo palikti tik tie X_i , kurių $h_i^2 > 0,8$.

Po analizės likusių rodiklių X_i ($m = 13$), kurie naudojami darbe, reikšmės:

EM – užimtieji;

MAN – apdirbamosios gamybos pramonės produkcija;

RTEM – mažmeninės prekybos apyvarta, išskyrus variklinių transporto priemonių ir motociklų prekybą; automobilių degalų prekybą;

SMV – variklinių transporto priemonių ir motociklų didmeninės ir mažmeninės prekybos bei remonto apyvarta;

IMP – importas;

EXP – eksportas;

GRT – pervežta krovinių geležinkeliu, tūkst. tonų;

FDI – tiesioginės užsienio investicijos;

CPI – vartotojų kainų indeksas;

PPI – gamintojų kainų indeksas;

NRB – negyvenamieji pastatai, apimtis to meto kainomis;

CES – inžineriniai statiniai, apimtis to meto kainomis;

FBS – mažmeninės prekybos maitinimo ir gėrimų teikimo veiklos apyvarta.

Proceso pabaigoje buvo išskirtos dvi pagrindinės komponentės, kurios paaiškina 96,02 proc. bendros dispersijos. Taip pat buvo įvertinta svorių (angl. *loadings*) matrica L .

2. Svorių vektoriaus įvertinimas.

Norint supaprastinti matricos L struktūrą, faktoriams sukurti buvo pasirinktas *Varimax* metodas. Kaip įprasta, pasukus faktorius nebuvo gauta matricos L^* paprastosios struktūros (2 lent.). Dėl šios priežasties toliau buvo taikomas svorių įvertinimo metodas.

Pritaikant CI svorių vertinimo metodą, algoritme naudojamos rodiklių X_i dispersijų dedamosios sumos S ($S_1 = 6,702$ ir $S_2 = 5,781$), kurias paaiškina faktorai $j = 1, 2$ (žr. (5)), S įvertis $S = 13$).

Trumpai aprašysime algoritmo etapus:

1. Naudojant L^* svorius ir dydžius S_j įvertinami koreguotieji faktorių svoriai (angl. *loadings*);

2. Atrenkami tarpiniai rodikliai, kurių koreguotieji svoriai yra didžiausi;

3. Panaudojant santykius S_j / S gaunami agreguoti tarpiniai rodikliai;

4. Svoriai normalizuojami, gaunamas rezultatas $\sum w_i = 1$ (3 lentelė).

2 lentelė. Svorių (loadings) matrica L^* pasukus faktorius

Rodikliai	Faktoriai	
	1	2
EM	0,940	0,246
MAN	0,715	0,676
RTEM	0,647	0,759
SMV	0,875	0,434
IMP	0,766	0,626
EXP	0,664	0,738
GRT	0,880	0,348
FDI	0,554	0,810
CPI	0,148	0,980
PPI	0,488	0,847
NRB	0,883	0,416
CIS	0,745	0,627
FBS	0,648	0,737

3 lentelė. Svorių vertinimo rezultatai

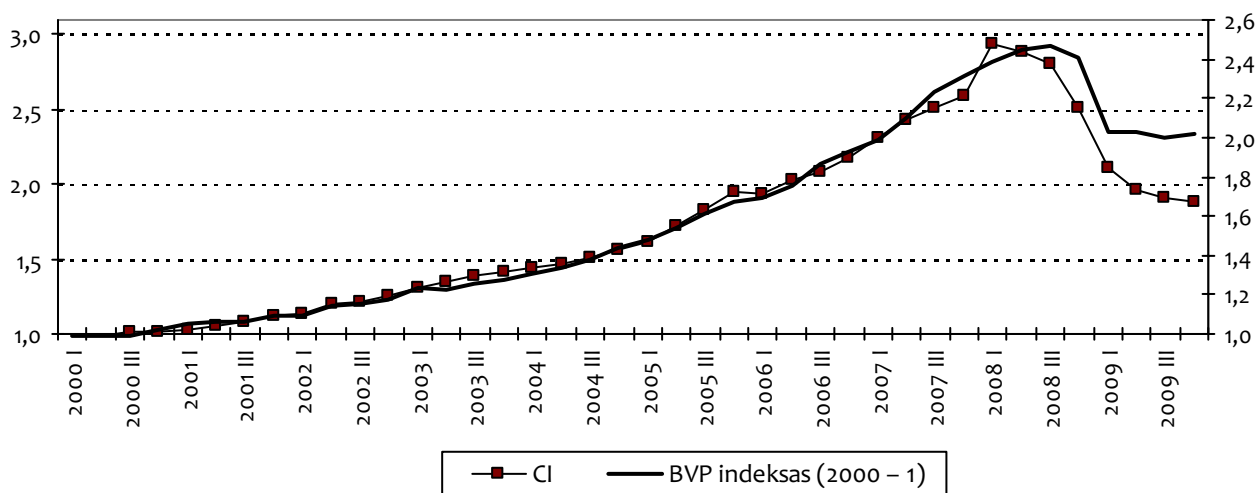
Rodikliai	CI svoriai W_i
EM	0,093
MAN	0,056
RTEM	0,070
SMV	0,080
IMP	0,062
EXP	0,066
GRT	0,081
FDI	0,080
CPI	0,117
PPI	0,087
NRB	0,082
CIS	0,058
FBS	0,066

CI IR BVP Palyginimo rezultatai

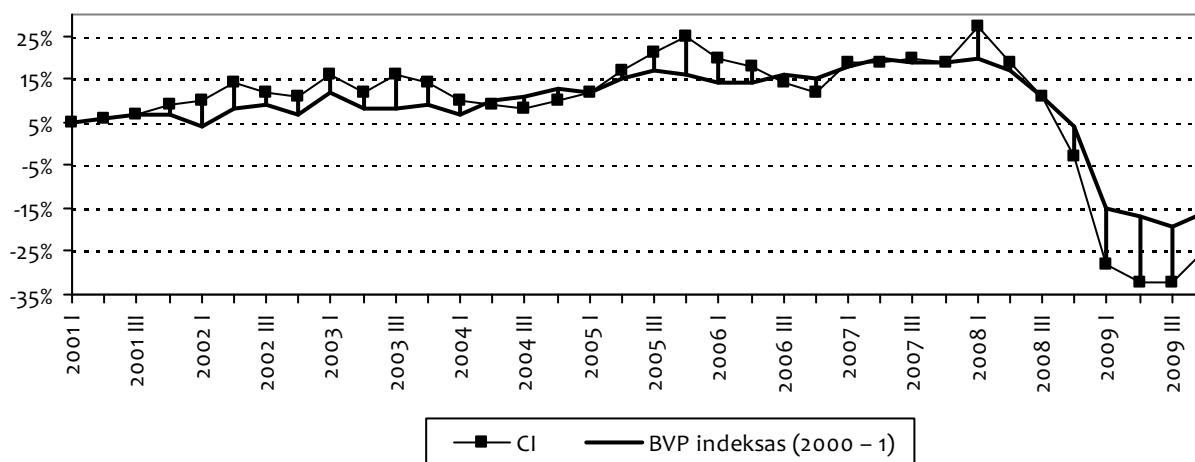
Sukonstruoto CI ir BVP indekso lyginamosios analizės grafinis vaizdas pateikiamas 1 pav. Analizėje naudojamas BVP yra standartizuotas (2000 m. vidurkis atitinka 1), sezono įtaka pašalinta.

Iš 1 pav. matyti, kad sukonstruoto CI ir BVP 2000–2008 m. kitimo tendencijos buvo labai panašios. Išsiskiria 2009 m., kai Lietuvoje ekonomikos krizė buvo pasiekusi viršūnę. Taip pat galima pastebėti, kad CI prasidedantį ekonomikos smukimą rodo ketvirčiu anksčiau nei BVP. Dėmesys atkreiptinas į tai, kad naudojamos kiek skirtingos skalės, BVP skalė – dešinėje.

1 pav. Sukonstruotas CI ir BVP indeksas



2 pav. Pokyčiai, palyginti su praėjusių metų atitinkamu ketvirčiu, procentais



Sukonstruotas CI traktuojamas kaip bendrasis ūkio plėtros rodiklis (BŪPR). Lyginant jo ir BVP dinamiką matyti, kad tolygios ekonominės raidos laikotarpiu jų trajektorijos beveik sutampa, tačiau 2009 m. skirtumas buvo ryškus (2 pav.). CI pokyčiai labiau koreliuoja su Statistikos departamento skelbiamais ekonominių vertinimų rodikliais, todėl galima sakyti, kad pastaraisiais metais Lietuvos BVP ne visai tiksliai atspindėjo bendrą šalies ekonomikos būklę.

Atlikta analizė leidžia teigti, kad vertinant šalies ekonomiką būtini papildomi sudėtiniai rodikliai, kurie, analizuojami lygiagrečiai su BVP, duotų bendresnį ir aiškesnį šalies ekonomikos vaizdą.

LITERATŪRA

1. Costello, A. B., Osborne, J. W. 2005: Best Practices in Exploratory Factor Analysis: Four

Recommendations for Getting the Most From Your Analysis, *Practical assessment, research & evaluation*, vol. 10, e. journal, p. 1–9.

2. Čekanavičius, V., Murauskas, G. 2002: *Statistika ir jos taikymai II*, Vilnius: TEV.

3. Kim, J. O., Muller, C. W. 1989: *Factor, discriminant and cluster analysis* (in Russian). Moscow: Finance and statistics, p. 25–35.

4. Nicoletti, G., Scarpetta, S. and Boylaud, O. 2000: *Summary Indicators of Product Market Regulation with an Extension to Employment Protection Legislation*, OECD, Economics Department Working Papers, No. 226.

5. Stiglitz J. E., Sen A., Fitoussi J. P. 2009: *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Interneto prieiga: http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.