



Munich Personal RePEc Archive

Measuring GDP forecasting accuracy using factor models: aggregated vs. disaggregated approach

Bessonovs, Andrejs

University of Latvia

2 April 2010

Online at <https://mpra.ub.uni-muenchen.de/30386/>
MPRA Paper No. 30386, posted 20 Apr 2011 10:58 UTC

Faktoru modeļu agregēta un dezagregēta pieeja IKP prognožu precizitātes mērīšanā

Measuring GDP forecasting accuracy using factor models: aggregated vs. disaggregated approach

Andrejs Bessonovs

Latvijas Banka

K. Valdemāra 2a, Rīga, LV-1050

E-pasts: Bessonovs@gmail.com

Anotācija

Šī raksta mērķis ir noteikt, vai dezagregēta pieeja ir precīzāka iekšzemes kopprodukta (turpmāk – IKP) prognozēšanā nekā agregēta pieeja. Lai pārbaudītu šo apgalvojumu, izmantots Stoka-Vatsona faktoru modelis. IKP tiek dezagregēts gan no izlietojuma puses, gan no ražošanas puses. Līdz ar to divu veidu pieejas sniedz plašāku skatu uz dezagregēto IKP un salīdzināšanas iespējām. Modelēšanas procesā izmantoti operatīvie mēneša dati, t. i., apsteidzošie indikatori IKP prognozēšanā.

Ārpus izlases prognozēšanai tiek izmantotas reālā laika datubāzes, un precizitātes mērīšanai autors izmanto plaši lietojamo rādītāju – vidējo kvadrātisko prognozēšanas kļūdu.

Atslēgvārdi: faktoru modelis, ārpus izlases prognozēšana, dezagregēta pieeja, reālā laika datubāze.

levads

Dezagregēta IKP komponentes neapšaubāmi satur vairāk informācijas par valsts ekonomisko aktivitāti nekā agregētais IKP pats par sevi. Šo papildu informāciju varētu izmantot IKP analīzē un prognozēšanā. Jautājums ir tikai, vai šī papildu informācija sniedz būtisku ieguldījumu un/vai dezagregētais IKP uzlabo prognozēšanas spēju, modelējot un prognozējot katru komponenti atsevišķi.

Lai pārbaudītu šo apgalvojumu, pētījumā analizēts Latvijas IKP. Prognozēšanā izmantots plaši lietojamais Stoka-Vatsona faktoru modelis. IKP tiek dezagregēts no izlietojuma un ražošanas pusēm. Līdz ar to divu veidu pieejas sniedz plašāku skatu uz dezagregēto IKP un salīdzināšanas iespējām.

Modelēšanas procesā izmantoti operatīvie mēneša dati par galvenajiem makroekonomiskajiem rādītājiem valstī: plašo naudu, rūpniecības izlaidi, tirdzniecības apgrozījumu, inflāciju un bezdarba līmeni, valsts konsolidētā kopbudžeta datiem, eksportu un importu, biznesa un patērētāju aptaujām. Lai veiktu ārpus izlases prognozēšanu, sastādīta reālā laika datubāze, kurā iekļautas reālā IKP dažādu periodu laika rindas.

Pētījuma struktūra ir šāda: otrajā nodaļā apskatīts Stoka-Vatsona faktoru modeļa matemātiskais aparāts. Trešajā nodaļā apspriesti dati, kuri izmantoti modelēšanā un prognozēšanā. Ceturtajā nodaļā autors sniedz ieskatu par lietotajām metodēm un pieņēmumiem. Piektā nodaļa veltīta empīriskiem rezultātiem.

Dati

Faktoru modeļu analīzē un prognozēšanā kā izskaidrojošie mainīgie izmantoti mēneša dati par galvenajiem makroekonomiskajiem rādītājiem valstī: plašo naudu, rūpniecības izlaidi, tirdzniecības apgrozījumu, inflāciju un bezdarba līmeni, valsts konsolidētā kopbudžeta datiem, eksportu un importu, biznesa un patērētāju aptaujām. Kopumā izmantoti 30 mainīgie laika periodā no 1996. gada janvāra līdz 2009. gada novembrim. Pārsvārā tā ir operatīvā informācija, kas ir pieejama pietiekami ātri no pāris dienām līdz pusotram mēnesim kopš attiecīgā mēneša beigām. K. Beņkovskis (2008) sniedz visaptverošu informāciju par mēneša datu izdevumiem un operatīvo datu pieejamību valstī, kā arī diskutē par savlaicīgas informācijas nepieciešamību un nozīmi tautsaimniecības analīzē.

Faktoru modeļu pētījumos parasti tiek izmantots liels skaits makroekonomisko rādītāju (vairāk par 100), piemēram, Dž. Stoka un J. Vatsona (2002) pētījumā izmantoti 215 mainīgie, V. Ajevskā un G. Dāvidsona (2008) – 126 mainīgie. Liels skaits makroekonomisko rādītāju

tiek izmantots, lai pēc iespējas maksimāli aptvertu visus ekonomiskās aktivitātes avotus. Tomēr virkne pētījumu rāda, ka lielāks skaits mainīgo neviennozīmīgi ietekmē prognozēšanas precizitāti – ne tikai paaugstina, bet arī samazina to. Nesen veiktie pētījumi pierāda šo apgalvojumu [4; 5]. Tā kā optimāla mainīgo skaita noteikšana un izvēle nav šī pētījuma mērķis, autors izvēlējies gan reālā sektora, gan finanšu sektora galvenos makroekonomiskos rādītājus par ekonomiskās aktivitātes dzinējspēku.

Agregēta un dezagregēta IKP dati ir pieejami ceturkšņu griezumā laika periodam no 1996. gada līdz 2009. gada 3. ceturksnim. Tā kā pētījuma uzdevums ir salīdzināt dažādu pieeju (t. i., gan agregēto, gan dezagregēto) spēju prognozēt IKP, ir izveidota IKP komponentu datubāze gan no izlietojuma, gan no ražošanas puses:

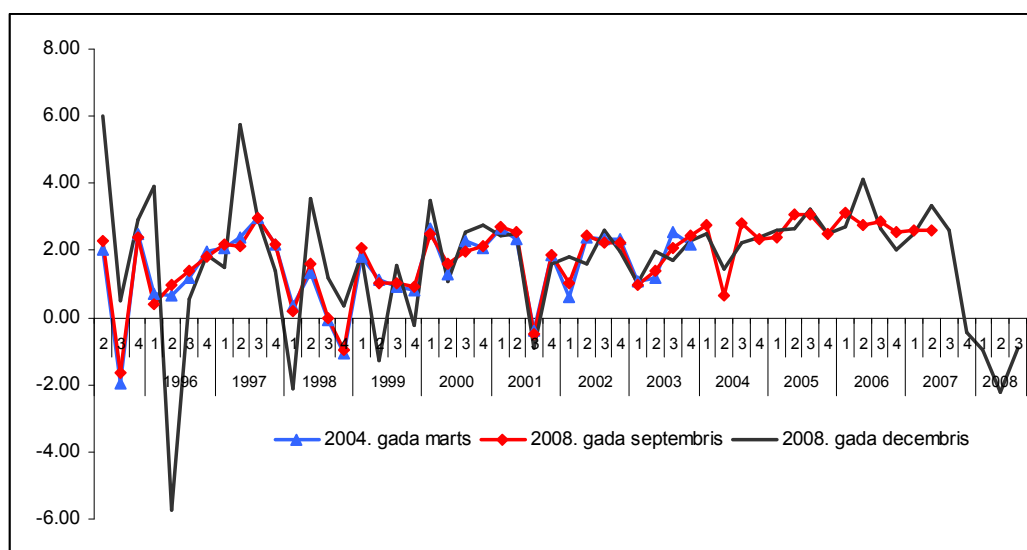
- no izlietojuma puses: privātais patēriņš (C), valsts patēriņš (G), bruto investīcijas (ieskaitot krājumu izmaiņas) (I), eksports (X) un imports (M);
- no ražošanas puses pēc NACE 1.1 klasifikācijas: lauksaimniecība (A), zvejniecība (B), ieguves rūpniecība (C), apstrādes rūpniecība (D), elektroenerģija, gāzes un ūdens apgāde (E), būvniecība (F), vairumtirdzniecība un mazumtirdzniecība (G), viesnīcas un restorāni (H), transports un sakari (I), finanšu starpniecība (J), operācijas ar nekustamo īpašumu (K), valsts pārvalde (L), izglītība (M), veselība (N), pārējie komunālie, sociālie un individuālie pakalpojumi (O).

Dažos pētījumos jau vērtēta dažādu modeļu prognozētspēja Latvijas reālā IKP noteikšanā, izmantojot pseidoreālā laika analīzi [1; 7]. Lai gan tajos daļēji simulētas reālā laika situācijas, ir ignorēta IKP un citu laika rindu pārskatīšanas iespēja. Taču ekonomiskās politikas veidotāji un uzņēmēji izmanto prognozes, kas sagatavotas, pirms kļūst zināmi datu pārskatīšanas rezultāti. Tāpēc datu pārskatīšanas ietekmes novērtēšana ir būtiska īstermiņa prognožu ticamības pārbaudei.

Latvijas reālā IKP ceturkšņa dati tiek publiskoti 70. dienā pēc pārskata perioda beigām. Tas nozīmē, ka IKP rādītāji kļūst pieejami ar vairāk nekā divu mēnešu novēlošanos. Turklāt IKP dati pakļauti pārskatīšanai pēc Nacionālās kontu sistēmas atlikumu gada izlīdzināšanas. Sākot ar 2007. gadu, Centrālā statistikas pārvalde (CSP) publicē arī reālā IKP gada pieauguma ātro novērtējumu, kas balstīts uz pieejamiem statistiskajiem datiem un ekonometriskajiem modeļiem. Šādi novērtējumi tiek publiskoti 40. dienā pēc pārskata perioda beigām un tādējādi pieejami vienu mēnesi pirms pirmās datu publiskošanas. Tomēr ātro

novērtējumu vēsture ir pārāk īsa, lai tos izmantotu sistemātiskā analīzē, un reālā IKP provizoriskie novērtējumi šajā pētījumā nav ņemti vērā.

Lai veiktu reālā laika prognozētspējas analīzi, izveidota reālā laika datubāze, kurā iekļautas reālā IKP dažādu periodu laika rindas. Tātad netiek izmantota tikai viena pēdējā IKP datu laika rinda, bet katram ceturksnim tiek veidota sava IKP datu laika rinda. Šī datubāze palīdz atrast attiecīgā laika posma analīzei nepieciešamos vēsturiskos IKP rādītājus. Turklāt šī reālā laika datubāze atspoguļo IKP datu korekcijas un to veikšanas laiku. Dažādu IKP datu publikāciju salīdzinājums dots 1. attēlā.



1. attēls. Dažādu IKP datu publikāciju salīdzinājums. IKP Ceturkšņu pieauguma tempi salīdzināmās cenās, sezonāli izlīdzināti dati
Comparison of different GDP data vintages. Quarter GDP growth rates at constant prices, seasonally adjusted data.

Avots: zīmējis autors.

Attēls atspoguļo dažādu IKP publikāciju revīzijas un tās lielumu. Vislielākā IKP datu revīzija notika 2008. gada decembrī, kad CSP pārrēķināja datus ķēdēto rindu veidā.

Datubāze tiek sagatavota, izmantojot CSP ceturkšņa biļetenu "Latvijas makroekonomiskie rādītāji". Tā ietver 24 dažādu periodu reālā IKP ceturkšņa datus, sākot ar datiem, kas bija pieejami 2004. gada janvāra sākumā (1995. gada 1. cet. – 2003. gada 3. cet.), līdz 2009. gada decembra sākumā pieejamiem datiem (1995. gada 1. cet. – 2009. gada 3. cet.). Līdzīga datubāze tiek izveidota arī reālā laika datiem pa IKP komponentēm no izlietojuma un ražošanas puses.

Modelis

Pētījumā izmantots Stoka-Vatsona faktoru modelis [8; 9]. Faktoru modeļa galvenā ideja ir reducēt lielo skaitu mainīgo N faktoru skaitā q tā, lai q būtu daudz mazāks nekā mainīgo skaits N ($q \ll N$). Līdz ar to katru mainīgo $i = 1, \dots, N$ faktoru analīzē mēs varam izteikt kā divu nenovērojamo ortogonālo elementu summu: kopējā komponente un idiosinkrātiskā komponente (2.1). Kopējā komponente tiek izskaidrota ar izvēlētiem faktoriem q , savukārt idiosinkrātiskā komponente ir katra mainīgā specifiskais šoks:

$$X_{it} = \lambda_i(L)f_t + e_{it} \quad (2.1)$$

kur

$X_{it} = (X_{1t}, X_{2t}, \dots, X_{Nt})'$ vektors ar dimensiju $(N \times 1)$;

f_t ir $(q \times 1)$ faktoru vektors;

$\lambda_i(L)$ ir $(1 \times q)$ faktoru svaru laga operators;

e_{it} ir idiosinkrātisko komponentu kļūda;

$e_{it} = (e_{1t}, e_{2t}, \dots, e_{Nt})'$ ir $(N \times 1)$ vektors.

Vienādojums (2.1) ir dinamiskā faktoru modeļa pieraksts, jo tiek izmantotas faktoru novēlotās vērtības un parametri ir atkarīgi no laika.

Sekojošā Dž. Stoka un M. Vatsona metodoloģijai, pieņemam modeļa formu vienādojumos (2.1) un (2.2), lai prognozētu y laika rindu:

$$y_{t+j} = \beta(L)f_t + \gamma(L)y_t + \varepsilon_{t+j} \quad (2.2)$$

kur

y_{t+j} ir j -tā perioda prognozējamā skalārā vērtība laika periodam t , kur $t = 1, \dots, T$;

$\beta(L), \gamma(L)$ ir laga operatori;

$E(\varepsilon_{t+j} | X_t, f_{t-1}, y_{t-1}, X_{t-1}, f_{t-2}, y_{t-2}, \dots) = 0$.

Galvenā problēma faktoru modeļu novērtēšanā ir faktoru skaita q noteikšana no datu kopas X_t . Tāpēc, lai novērtētu modeli, parasti pieņem galīgas kārtas polinomus un pārformulē modeli statistiskā formā vienādojumos (2.3) un (2.4) (pilno statistiskā modeļa izvedumu sk. [6]).

$$\beta(L) = \sum_{j=0}^p \beta_j L^j$$

$$\gamma(L) = \sum_{j=0}^p \gamma_j L^j$$

$$\lambda_i(L) = \sum_{j=0}^p \lambda_{ij} L^j$$

Galīgā polinoma noteikšana ļauj modeli pārrakstīt šādā formā:

$$y_{t+j} = \beta' F_t + \gamma(L)y_t + u_{t+j} \quad (2.3)$$

$$X_t = \Lambda F_t + e_t \quad (2.4)$$

kur

$F_t = (f'_t, f'_{t-1}, \dots, f'_{t-p})'$ ir $(r \times 1)$ dimensiju vektors ar $r = q(p+1)$, kurš satur visus faktorus un to novēlotās vērtības;

Λ ir $(N \times r)$ faktoru svaru matrica, kur i -tā rinda ir $(\lambda_{i0}, \lambda_{i1}, \dots, \lambda_{ip})$;

$\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ ir $(1 \times r)$ dimensiju vektori;

e_t ir $(N \times 1)$ idiosinkrātiskās kļūdas.

Statiskā modeļa priekšrocība ir tā, ka to var novērtēt, izmantojot galveno komponentu metodi. Ar galveno komponentu metodes palīdzību mainīgo kopas dimensija N tiek reducēta uz r dimensiju ($r \leq N$). Galveno komponentu analīze risināma ar īpašvektoru palīdzību, kuri atbilst r lielākajām, dilstošajā secībā sakārtotajām īpašvērtībām.

Tā kā šajā pētījumā galvenā ir laika rindas prognozēšana, jādefinē prognozējamās laika rindas vienādojums:

$$y_{t+1}^1 = \alpha_1 + \beta_1(L)F_t + \gamma_1(L)y_t \quad (2.5)$$

$$y_{t+h}^h = \alpha_h + \beta_h(L)F_t + \gamma_h(L)y_t \quad (2.6)$$

Vienādojums (2.5) parāda sakarību y vērtības prognozēšanai vienam periodam uz priekšu, bet vienādojumā (2.6) tiek parādīta sakarība h -periodu prognozēšanai uz priekšu. Prognozējamā y_{t+h} vērtība ir atkarīga no faktoru kopas F_t un no pašas y_t laga vērtībām.

Metodes un pieņēmumi

IKP no ražošanas puses modelēts, izmantojot mazāku skaitu kategoriju, tādā veidā apvienojot pēc funkciju vai darbības veida līdzīgas nozares vienā. Pamatojums šādai pieejai ir tāds, ka ne vienmēr var atrast izskaidrojošos mainīgos vai pietiekami operatīvu informāciju konkrētai nozarei modelēšanas vajadzībām. Apvienojot nozares mazākās kategoriju grupās un tālāk lietojot prognozēšanas metodes, teorētiski ir iespēja samazināt prognozēšanas kļūdas nozaru trūkstošo izskaidrojošo mainīgo dēļ. Pētījumā tika izmantotas astoņas kategoriju grupas:

- zemkopība: lauksaimniecība un zvejniecība (AB);
- kopējā ražošana: ieguves rūpniecība; apstrādes rūpniecība; elektroenerģija, gāzes un ūdens apgāde (CDE);
- būvniecība (F);
- vairumtirdzniecība un mazumtirdzniecība (G);
- transports un sakari (I);
- sabiedriskie pakalpojumi: valsts pārvalde, izglītība, veselība (LMN);
- pārējie komercpakalpojumi: viesnīcas un restorāni, finanšu starpniecība, operācijas ar nekustamo īpašumu, pārējie komunālie, sociālie un individuālie pakalpojumi (HJKO);
- tīrie nodokļi: produktu nodokļi mīnus produktu subsīdijas (TS).

Visi dati ir sezonāli izlīdzināti ar X-12-ARIMA¹ metodi. Turklāt IKP un to komponentu datiem tiek izmantota X-12-ARIMA metode ar specifiskiem parametriem, kurus izmanto CSP.² Parametri attiecas uz sezonālo un trenda filtriem, datu transformāciju, ARIMA modeļa specifikāciju un Lieldienu efekta korekciju. Kā rāda prakse, specifiskās sezonālās izlīdzināšanas metodes izmantošana prognozēšanā ļauj samazināt prognozēšanas kļūdas un paaugstināt precizitāti.

Sezonāli izlīdzinātās IKP komponentes no ražošanas puses astoņām jaunizveidotajām kategorijām tiek iegūtas netiešā veidā, t. i., sezonāli izlīdzinot katru komponenti atsevišķi un tālāk sasummējot tās vienā.

¹ Pieejams: <http://www.census.gov/srd/www/x12a/>

² Logaritmiskā datu transformācija; vidējā korekcija; Lieldienu efekta korekcija (6 dienas); ARIMA modeļa specifikācija: (010) (011); X-11 sezonālais filtrs: 3x3 slidojošais vidējais; X-11 trenda filtrs: 5-elementu Hendersona slidojošais vidējais.

Visi eksogēnie mainīgie ir diferencēti vai logaritmēti un diferencēti atkarībā no datiem, padarot tos par stacionāriem. Mainīgajiem tiek izņemta sezonālā komponente, un tie tiek standartizēti ar vidējo nulli un standartnovirzi viens.

Ņemot vērā datu publicēšanas īpašības, t. i., datu publicēšanas laiku no attiecīgā mēneša beigām un laiku, no kura tiek publicēta dota laika rinda, katrā laika momentā ir pieejama nesabalansēta datu kopa. Piemēram, novērojumi par plašo naudu ir pieejami mēnesi pirms tam, kad kļūst pieejami dati par eksportu un importu, vai dažas laika rindas ir novērotas daudz vēlāk par 1996. gadu. Lai atrisinātu nesabalansētas datu kopas problēmu, tiek izmantota gaidu maksimizēšanas metode (angl. *Expectations-Maximization*). Sīkāk par gaidu maksimizēšanas metodi skatīt Dž. Stoka un J. Vatsona pētījumu [9].

Gaidu maksimizēšanas metodes galvenā ideja ir iteratīvs process un trūkstošo novērojumu novērtēšana ar galveno komponentu metodi. Trūkstošo novērojumu vērtības ir pielīdzinātas laika rindas vidējai vērtībai (šajā gadījumā tās ir nulles, jo laika rindas ir standartizētas ar vidējo nulli), lai iegūtu sabalansētu datu kopu. Tālāk datu kopa tiek novērtēta ar galveno komponentu metodi un tiek iegūti faktori un faktoru svāri. Pēc tam datu kopa tiek pārvērtēta, izmantojot faktoros un faktoru svarus un trūkstošie novērojumi tiek aizstāti ar vērtībām, kas iegūtas iepriekšējā solī. Šādā veidā sākotnējiem datiem paliekot nemainīgiem, bet mainoties trūkstošām vērtībām, kas nosacīti kļūst atkarīgas no novērojamās datu kopas, tiek veikts iteratīvs process, tikmēr trūkstošo vērtību izmaiņas kļūst mazākas kā brīvi izvēlēts konverģences kritērijs, t. i., visai minimālas.

ARIMA (*Auto Regressive Integrated Moving Average*) modeļu izvēle ir balstīta uz Švarca informatīvo kritēriju (SIC). Autors ir norobežojies ar AR un MA komponentēm maksimāli līdz četriem katras komponentes lagiem. Tātad par labāko ARIMA modeli izvēlēts modelis ar mazāko SIC vērtību ārpus izlases sākuma, pieņemot, ka ir zināma visa informācija tikai līdz ārpus izlases prognozēšanas eksperimentam. Pētījumā izmantoti divu veidu ARIMA modeļi: ar līdzīgu lagu struktūru un dažādu lagu struktūru. Kā arī reprezentatīvos nolūkos tiek parādīts arī RW modelis (angl. *Random Walk*). ARIMA modelis ar fiksētu lagu struktūru ir modelis, kurā katrai komponentei tiek izmantota tā pati lagu struktūra kā agregētam ARIMA modelim. Ar ARIMA modeli ar mainīgu lagu struktūru tiek domāts modelis, kurā katrai komponentei ir sava specifiska lagu struktūra. Specifiskā lagu struktūra tiek iegūta arī ar SIC kritēriju palīdzību ārpus izlases perioda sākumā. Katrai komponentei tiek piemēklēts modelis ar mazāko SIC kritēriju, un agregētais IKP tālāk izmantots ārpus izlases prognozēšanā. Lagu struktūra parādīta 1.a un 1.b tabulā. IKP agregētai laika rindai tiek piemērots ARIMA (2,1,0) modelis.

1.a tabula
Dezagregētā pieeja no ražošanas puses
Output approach

Komponents	Specifikācija ar fiksētu lagu struktūru	Specifikācija ar mainīgu lagu struktūru
Zemkopība (A, B)	ARIMA (2,1,0)	ARIMA (3,1,3)
Kopējā ražošana (C, D, E)	ARIMA (2,1,0)	ARIMA (1,1,3)
Būvniecība (F)	ARIMA (2,1,0)	ARIMA (4,1,1)
Vairumtirdzniecība un mazumtirdzniecība (G)	ARIMA (2,1,0)	ARIMA (3,1,1)
Transports un sakari (I)	ARIMA (2,1,0)	ARIMA (2,1,4)
Sabiedriskie pakalpojumi (L, M, N)	ARIMA (2,1,0)	ARIMA (4,1,0)
Pārējie komercpakalpojumi (H, J, K, O)	ARIMA (2,1,0)	ARIMA (2,1,2)
Tīrie nodokļi (T, S)	ARIMA (2,1,0)	ARIMA (2,1,4)

1.b tabula
Dezagregētā pieeja no izlietojuma puses
Expenditure approach

Komponents	Specifikācija ar fiksētu lagu struktūru	Specifikācija ar mainīgu lagu struktūru
Privātais patēriņš	ARIMA (2,1,0)	ARIMA (1,1,3)
Valsts patēriņš	ARIMA (2,1,0)	ARIMA (4,1,4)
Investīcijas	ARIMA (2,1,0)	ARIMA (4,1,1)
Eksports	ARIMA (2,1,0)	ARIMA (3,1,1)
Imports	ARIMA (2,1,0)	ARIMA (0,1,4)

Ārpus izlases prognozēšanas eksperiments

Šajā pētījuma daļā tiek apskatīti ārpus izlases prognozēšanas rezultāti. Tas nozīmē, ka tiek pārbaudīta faktoru modeļa prognozētspēja uz aktuālajiem datiem. Ņemot vienu izlases daļu no pieejamiem datiem modeļa novērtēšanai, tiek prognozētas nākotnes vērtības, bet pārējā datu kopa salīdzināta ar prognozētām vērtībām, un tiek iegūts prognožu precizitātes rādītājs. Prognozētspējas pārbaudei izmantots standartu rādītājs – kvadrātsaknes vidējā kvadrātiskā prognozēšanas kļūda:

$$RMSFE = \sqrt{\frac{1}{T} \sum_T (\hat{y}_{t+h|t} - y_{t+h})^2} \quad (5.1)$$

kur

T – ārpus izlases periodu skaits;

$\hat{y}_{t+h|t}$ – prognozētā y vērtība laikā $t+h$ -periodam uz priekšu;

y – īstā vērtība.

Kā bieži minēts literatūrā, ārpus izlases prognozēšanai izmanto 1/3 no pieejamās datu izlases. Šajā gadījumā ārpus izlases sākuma periods ir 2004. gada 4. ceturksnis līdz 2009. gada 3. ceturksnim. Pētījumā tiks nodalīti 3 prognozējamie periodi. Vidējā kvadrātiskā prognozēšanas kļūda parādīs, kāda ir modeļu kļūda, prognozējot vienu, divus un trīs ceturkšņus uz priekšu, tātad $h = 1, 2, 3$.

2008. gada decembrī CSB nomainīja IKP aprēķināšanas metodoloģiju un aprēķina IKP t. s. ķēdes indeksu veidā. Par aprēķinu bāzi ķēdētiem datiem tiek lietots iepriekšējais kalendārais gads, un aprēķiniem izmantota “gada vidējā metode”, tas ir, esošo ceturksni rēķina iepriekšējā gada vidējās cenās. Datu dinamika dažādu publikāciju laikā ir parādīta 1. attēlā. Šāda datu aprēķināšanas metodes maiņa atstāja būtisku iespaidu uz prognozēšanas procesu. Līdz ar to turpmāk apskatīti divi ārpus izlases prognozēšanas periodi – līdz brīdim kad tika ieviesta jauna metode, un pēc tā brīža. Kā rāda rezultāti, starpība ir pietiekami būtiska, lai apskatītu abus gadījumus.

Faktoru modeļa un vienkāršo modeļu agregēta un dezagregēta IKP prognozētspējas rezultāti laika periodam no 2004. gada decembra līdz 2009. gada decembrim (t. i., periods, ņemot vērā datus ar jauno IKP aprēķināšanas metodi) parādīti 2. tabulā. Prognozētspējas rezultāti laika periodam no 2004. gada marta līdz 2008. gada novembrim (t. i., periods līdz brīdim, kad tiek ieviesta jauna IKP aprēķināšanas metode) doti 3. tabulā. Rezultāti atspoguļo standartnovirzi no gada pieaugumu tempiem procentpunktos trīs prognozēšanas periodiem uz priekšu.

2. tabula

Ārpus izlases vidējās kvadrātiskas prognozēšanas kļūdas rezultāti laika periodam no 2004. gada decembra līdz 2009. gada decembrim

Root mean squared error's results for out-of-sample period December 2004 – December 2009.

	IKP	IKP ^{izlietojums}	IKP ^{izlaide}
+1 periods			
RW*	4,19	3,91	4,29
ARIMA ^{fiks**}	3,68	4,07	4,52
ARIMA ^{main***}	-	4,19	4,47
FACTOR			
q = 1	2,65	2,41	2,94
q = 2	2,7	2,09	2,93

q = 3	2,62	2,34	2,88
q = 4	2,61	2,39	2,83
+2 periodi			
RW	7,30	7,77	7,97
ARIMA ^{fiks}	6,48	6,69	8,12
ARIMA ^{main}	-	7,56	8,75
FACTOR			
q = 1	4,1	4,02	4,83
q = 2	4,03	3,81	4,65
q = 3	4,19	4,20	4,88
q = 4	4,26	4,12	4,97
+3 periodi			
RW	10,39	10,97	10,98
ARIMA ^{fiks}	9,36	9,57	10,93
ARIMA ^{main}	-	11,26	11,94
FACTOR			
q = 1	6,64	6,61	7,03
q = 2	6,42	6,42	6,70
q = 3	6,63	6,65	6,85
q = 4	6,74	6,95	6,98

*RW – *Random Walk* modelis; **ARIMA specifikācija ar fiksētu lagu struktūru; ***ARIMA specifikācija ar mainīgu lagu struktūru.

Piezīme: iekrāsotais skaitlis nozīmē labāko rezultātu konkrētā prognozētā periodā.

2. tabulā iekļauti rezultāti par 3 vienkāršajiem modeļiem un faktoru modeli ar 1 līdz 4 faktoriem. Otrajā kolonnā doti agregētā IKP rezultāti, bet trešajā un ceturtajā kolonnā – dezagregētā IKP rezultāti attiecīgi no izlietojuma un ražošanas pusēm.

No 2. tabulas var izdarīt trīs secinājumus. Pirmkārt, sekojot līdzi Dž. Stoka un J. Vatsona pētījumam faktoru skaita noteikšanā pēc vidējās kvadrātiskas kļūdas rezultātiem, varam secināt, ka optimālais faktoru skaits faktoru modelī nepārsniedz vienu vai divus faktorus, jo lielāks faktoru skaits neuzlabo rezultātus.

Otrkārt, var secināt, ka faktora modeļa prognozēšanas spēja ir labāka nekā vienkāršajiem viendimensijas modeļiem. Faktoru modeļu prognozēšanas kļūdas ir mazākas nekā vienkāršajiem ARIMA vai *Random Walk* modeļiem. It īpaši būtisks uzlabojums redzams pirmajos prognozēšanas horizonta periodos, kad uzlabojums sasniedz 20–30% (respektīvi, vidējā kvadrātiskā kļūda ir 70–80% no vienkāršo modeļu vidējās kvadrātiskās kļūdas). Tam ir labs pamatojums, ka faktoru modelī tiek izmantoti apsteidzošie mainīgie mēneša griezumā, kuri ļauj operatīvāk un precīzāk vērtēt IKP dinamiku tuvā nākotnē.

Treškārt, ārpus izlases vidējās kvadrātiskās prognozēšanas kļūdas rezultāti laika periodam no 2004. gada marta līdz 2008. gada novembrim (3. tabula), t. i., datu izlase pirms IKP aprēķināšanas pārejas uz ķēdētiem datiem, liecina, ka metodoloģijas izmaiņas atstāja būtisku

iespaidu uz prognozētspējas precizitāti. Rezultāti rāda, ka prognozēšanas kļūdas pēc metodoloģijas izmaiņām ir 1,5–2 reizes lielākas nekā pirms tam.

3. tabula

Ārpus izlases vidējās kvadrātiskās prognozēšanas kļūdas rezultāti laika periodam no 2004. gada marta līdz 2008. gada novembrim

Root mean squared error's results for out-of-sample period March 2004 – November 2009.

	IKP	IKP ^{izhetojums}	IKP ^{izlaide}
+1 periods			
RW*	2,00	2,76	1,91
ARIMA ^{fiks**}	1,87	2,84	1,97
ARIMA ^{main***}	-	3,23	1,98
FACTOR			
q = 1	1,67	2,31	1,51
q = 2	1,66	2,15	1,48
q = 3	1,84	2,41	1,67
q = 4	1,88	2,78	1,71
+2 periodi			
RW	3,04	3,71	2,81
ARIMA ^{fiks}	2,85	3,57	2,89
ARIMA ^{main}	-	4,98	3,12
FACTOR			
q = 1	2,54	3,15	2,43
q = 2	2,47	2,93	2,25
q = 3	2,70	3,09	2,49
q = 4	2,74	3,03	2,51
+3 periodi			
RW	3,92	4,27	3,64
ARIMA ^{fiks}	3,75	4,25	3,76
ARIMA ^{main}	-	6,02	3,92
FACTOR			
q = 1	3,42	3,40	3,14
q = 2	3,60	3,30	3,25
q = 3	4,00	3,55	3,63
q = 4	4,23	4,07	3,86

*RW – *Random Walk* modelis; **ARIMA specifikācija ar fiksētu lagu struktūru; ***ARIMA specifikācija ar mainīgu lagu struktūru.

Piezīme: iekrāsotais skaitlis nozīmē labāko rezultātu konkrētā prognozētā periodā.

Datu izlasei pirms datu maiņas tika saglabāts ārpus izlases garums un modeļu specifikācijas, lai rezultāti būtu pēc iespējas maksimāli salīdzināmi.

Ceturtkārt, lai gan 1. tabulas rezultāti rāda, ka dezagregēta IKP prognozēšanas pieeja ir labāka nekā agregēta, tomēr rezultātu starpība nav pārliecinoši liela un prognozētspējas uzlabojums ir ļoti minimāls, lai pilnībā spriestu par vienas vai otras pieejas priekšrocību. Lai arī 3. tabulas vidējās kvadrātiskās prognozēšanas kļūdas rezultāti laika periodam no 2004.

gada marta līdz 2008. novembrim (periods ar mazāko prognozēšanas kļūdu volatilitāti) apstiprina apgalvojumu par iespējamo dezagregēto pieeju priekšrocību prognožu novērtēšanā, tomēr joprojām uzlabojums ir relatīvi mazs. Dezagregēto pieeju uzlabojumu prognozēšanai varētu arī sniegt katras IKP komponentes specifiskācija – faktoru skaits, faktoru lagu skaits vai prognozējamo mainīgo lagu skaits. Šajā analīzē pieņemts, ka specifiskācija ir identiska katrai komponentei.

Secinājumi

Pirmkārt, lai gan rezultāti rāda, ka dezagregēta IKP prognozēšanas pieeja ir labāka nekā agregēta, tomēr rezultātu starpība nav pārlicinoši liela un prognozētspējas uzlabojums ir ļoti minimāls, lai pilnībā spriestu par vienas vai otras pieejas priekšrocību.

Otrkārt, balstoties uz iegūtiem rezultātiem, nešaubīgi var teikt, ka faktoru modelim ir priekšrocības salīdzinājumā ar vienkāršajiem modeļiem. Turklāt faktoru skaitu modelī varētu ierobežot viens vai divi faktori, jo lielāks faktoru skaits nesniedz uzlabojumu prognozēšanas precizitātes novērtēšanā.

No rezultātiem var secināt: pat ja dezagregēto pieeju aprēķini nesniedz būtisku priekšrocību salīdzinājumā ar agregētiem rezultātiem, tad dezagregētas modeļu specifiskācijas var izmantot prognožu kombinēšanas analīzē. Modeļu prognožu kombinācijas savukārt ir alternatīva pieeja lielā modeļu klāsta izmantošanai precīzāku prognožu izstrādāšanā.

Literatūra

1. Ajevskis Viktors, Dāvidsons Gundars. Dinamisko faktoru modeļu lietojums Latvijas iekšzemes kopprodukta prognozēšanā. Latvijas Banka, 2008. Pētījums 2/2008.
2. Bai Jushan, Ng Serena. Determining Number of Factors in Approximate Factor Models. *Econometrica*, vol. 70, No. 1, 2002, p. 191–221.
3. Beņkovskis Konstantīns. Mēneša rādītāju izmantošana Latvijas reālā iekšzemes kopprodukta pieauguma īstermiņa prognozēšanā. Latvijas Banka, 2008. Pētījums 5/2008.
4. Boivin Jean, Ng Serena. Are More Data Always Better For Factor Analysis? NBER Working Paper 9829, July 2003.
5. Caggiano Giovanni, Kapetanios George, Labhard Vincent. Are More Data Always Better For Factor Analysis? Results For Euro Area, The Six Largest Euro Area And The UK. ECB Working Paper No. 1051, May 2009.

6. Dreger Christian, Schumacher Christian. Estimating large-scale factor models for economic activity in Germany: Do they outperform simpler models? HWWA discussion paper No. 199, Hamburg Institute of International Economics, 2002.
7. Meļihovs Aleksejs, Rusakova Svetlana. Short-Term Forecasting of Economic Development in Latvia Using Business and Consumer Survey Data. Bank of Latvia Working Paper, No. 4, 2005.
8. Stock James H., Watson Mark W. Diffusion Indexes. NBER Working Paper No. 6702, August 1998.
9. Stock James H., Watson Mark W. Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business & Economic Statistics, vol. 20, No. 2, April 2002.

Summary

The purpose of this paper is to conduct whether the disaggregated data of GDP gives us any additional information in the sense of forecasting accuracy. To test latter hypothesis author employs Stock-Watson factor model. GDP is disaggregated both on expenditure basis and on output basis. Thus both approaches should widen overlook to comparison's capability. In order to measure forecasting accuracy root mean squared error measure was employed. Author concludes that disaggregated approach outperforms aggregated data but at very little extent. In addition, factor model showed better results in the sense of forecasting accuracy and outperformed univariate models on average by 20-30%.

Keywords: Factor model, out-of-sample forecasting, disaggregated approach, real-time database.