FAKTORSKE PROGNOZE INFLACIJE U HRVATSKOJ

Davor KUNOVAC[1]
Hrvatska narodna banka, Zagreb

Izvorni znanstveni članak*
UDK: 336.748.12(497.5)
JEL: E44

Sažetak

U radu je provjereno može li se informacija 144 ekonomske varijable (prezentirana uz pomoć nekoliko konstruiranih faktora) iskoristiti za prognoziranje kretanja potrošačkih cijena u Hrvatskoj. Dobiveni rezultati govore da uporaba jednog faktora povećava preciznost prognoza inflacije benchmark modela. Korištena je metodologija dovoljno općenita da može biti izravno primijenjena i za prognoziranje ostalih ekonomskih varijabli.

Ključne riječi: faktorski modeli, analiza vremenskih nizova, inflacija, prognoziranje

1. Uvod

1.1. Motivacija

U radu želimo ispitati mogu li neki elementi faktorske analize i srodne matematičke metode – analize glavnih komponenti (engl. *Principal Component Analysis* – PCA) pomoći u prognoziranju i boljem razumijevanju ekonomskih pojava u Hrvatskoj.

Ideja da se korelacijska struktura velikog broja varijabli (više stotina ili tisuća) što točnije aproksimira sa samo nekoliko (često samo pet ili šest) komponenti – faktora, zastupljena je u radu istraživača, kako prirodnih, tako i društvenih disciplina. Korijene faktorske analize nalazimo u psihologiji, u kojoj početkom prošloga stoljeća Charles Spearman, 1904) na temelju rezultata testova iz matematike, stranih jezika i sl. ekstrahira faktore u kontek-

* Primljeno *(Received)*: 1.6.2007.
Prihvaćeno *(Accepted)*: 10.12.2007.

[1] Za stajališta iznesena u radu odgovoran je autor i ta stajališta nisu nužno istovjetna onima Hrvatske narodne banke. Autor zahvaljuje anonimnim recenzentima na korisnim komentarima.

stu mjerenja inteligencije. Osim psiholozima, faktorska analiza u istraživanju najčešće još pomaže kemičarima – kemometričarima, te eksperimentalnim fizičarima.

Standardnim ekonometrijskim modelima, poput VAR modela ili sustava simultanih jednadžbi, zbog kratkoće dostupnih serija podataka možemo simultano modelirati međudjelovanje samo nekoliko varijabli, obično manje od deset. Zato se u posljednjih nekoliko godina intenzivno istražuje mogućnost *kompresije* informacija u ekonomskim i financijskim podacima. U kontekstu prognoziranja ključnih makroekonomskih varijabli američki ekonomisti Stock i Watson (1998, 1999, 2002) serijom radova istražuju iskoristivost modela faktorske analize. Oni, naime, iz velikog broja makroekonomskih varijabli procjenjuju nekoliko faktora koji u izmjerivoj količini sažimaju informaciju ukupnoga američkoga gospodarstva. U drugom koraku procijenjenim se faktorima koriste za prognoziranje rasta proizvodnje, inflacije i sl. Dobivene faktorske prognoze uvelike nadmašuju prognoze standardnih ekonometrijskih modela – univarijatnih regresija, autoregresija, VAR modela te modela utemeljenih na prethodećim indikatorima, što je potaknulo daljnje ispitivanje iskoristivosti faktorskih modela (npr. Matheson, 2006; Camacho i Sancho, 2003; Camba-Méndez i Kapetanios, 2005). Dobiveni su rezultati obećavajući, no ne smiju automatski biti preslikani na druga gospodarstva bez izravne empirijske provjere i eventualne prilagodbe. Posebno bi pogrešno bilo da te rezultate nekritički preuzimaju mlade, otvorene ekonomije, ponajprije zbog razlike u kvaliteti i količini dostupne statistike te naravi funkcioniranja mehanizama pojedinih ekonomija. Stoga je ova analiza provedena kako bi se provjerila korisnost jedne relativno nove statističke tehnike pri modeliranju i prognoziranju ekonomskih procesa u Hrvatskoj. Nadalje, jasan je i doprinos koji takve analize mogu imati u kontekstu šireg sagledavanja prognostičkih sposobnosti, za sada nestandardnih, modela faktorske analize.

1.2. Primjena faktorskih modela u makroekonomskim analizama

U posljednjih desetak godina neprekidno se povećava broj radova koji se bave faktorskim modelima u ekonomiji. Donosimo pregled recentnih radova kojima su određeni ključni smjerovi razvoja ekonomskih faktorskih modela, kako u smislu metodologije, tako i u empirijskoj analizi.

Osim spomenutih radova Stocka i Watsona, treba spomenuti i dva rada o razvoju metodologije. Forni i sur. (2000) daju metodu procjene faktora koja se zasniva na elementima spektralne analize i smatra se komplementarnom standardnoj Stock-Watsonovoj metodologiji. Usto Bai i Ng (2002) definiraju procjenitelj broja nepoznatih faktora koji se rabi u prognostičkome modelu, što se standardno smatra jednim od osnovnih problema u dizajnu faktorskih modela.

Bernanke i Boivin (2003) rezultate faktorske analize testiraju u kontekstu potreba američke središnje banke i pritom daju zanimljivu kritiku empirijskih analiza monetarne politike koje gotovo redovito predviđaju da vodstvo te banke donosi odluke na temelju samo nekoliko ključnih makrovarijabli poput inflacije ili BDP-a, dok je u praksi situacija ipak drugačija i broj varijabli koje se ozbiljno analiziraju mnogo je veći. Na toj liniji Bernanke i Boivin procjenjuju reakcijsku funkciju Feda, determiniranu ukupnom informacijom velikog broja, za SAD, relevantnih ekonomskih varijabli. Ta je ukupna informacija prezentirana s nekoliko procijenjenih faktora koji se tumače kao silnice što generiraju cjelokupno gospodarstvo.

Faktorski su se modeli pokazali vrlo korisnima i pri *brzoj procjeni* (engl. *flash estimate*) BDP-a. Naime, poznato je kako službeni podaci za BDP dolaze s više mjeseci zakašnjenja, no važno je znati trenutačno stanje ekonomske aktivnosti koje zbog spomenutog kašnjenja – procjenjujemo. U tom kontekstu važan su doprinos analize Schumachera i Breitunga (2006) te Giannonea i sur. (2005).

Većina ekonomskih aplikacija faktorskih modela pretežno je *ateoretska* – odnosi se na prognoziranje. Jedna od iznimaka je rad Boivina i Giannonija (2005), u kojemu informacija velikog broja varijabli služi za procjenu parametara teoretskih DSGE[2] modela.

Za nas su posebno važne analize kojima se testira vjerodostojnost faktora konstruiranih od relativno velikog broja serija, no na kratkom vremenskom intervalu. U tom kontekstu ističemo dvije analize. Banerjee i sur. (2006) ispituju postojanost faktorskih prognoza na kratkim uzorcima sa strukturnim lomovima, u čemu se Slovenija uzima kao reprezentant skupine novih zemalja članica EU. Zatim Boivin i Ng (2005) Monte Carlo simulacijama ispituju osjetljivost kvalitete faktorskih prognoza na duljinu uzorka i ta analiza ohrabruje provedbu faktorske analize čak i u uvjetima kratkih vremenskih serija poput hrvatskih.

Od značajnih primjera primjene faktorske analize u ekonomiji izdvajamo još analizu Europske komisije (Grenouilleau, 2006) i ekonomski indikator Chicago Feda (Chicago Fed Letter, 2000).

Konačno, pregled rezultata faktorskih prognoza inflacije i realne aktivnosti iz velikog broja relevantnih članaka, čak 46, dan je u metaanalizi Eickmeier i Zieglera (2006).

1.3. Struktura analize

Struktura rada je ovakva: u sljedećem poglavlju dajemo glavne matematičke rezultate nužne za ispravnu primjenu faktorske analize. Najprije iznosimo osnove analize glavnih komponenti te izračunavamo glavnu komponentu varijabli iz grupe cijena i tečaja.

U trećem poglavlju primjenjujemo *Stock-Watsonovu* metodologiju prognoziranja na hrvatsku inflaciju, pri čemu faktore i parametre procjenjujemo uz pomoć glavnih komponenata. Osim toga, navodimo mogućnosti za daljnje unaprjeđenje modela.

U posljednjem, četvrtom poglavlju dajemo zaključak.

2. Glavne komponente

2.1. Definicija

U ovom ćemo dijelu definirati pojam i primjerom ilustrirati *metodu glavnih komponenti*.

Pretpostavimo da su dana opažanja za N varijabli $X_1 K, X_N$, čiju kovarijacijsku strukturu želimo istražiti.[3] U našem kontekstu ekonomiju ili neki njezin dio u svakom trenutku aproksimiramo slučajno uređenom N-torkom, dakle N-dimenzionalnim slučajnim vektorom.

[2] *Dynamic Stochastic General Equilibrium.*

[3] U faktorskoj analizi proučavamo kovarijance i varijance podataka, dakle samo informaciju drugog momenta. Eventualna korisna informacija trećega ili viših momenata ignorira se.

Da bismo u potpunosti opisali kovarijacijsko-varijančnu strukturu tih serija potrebno je, naravno, svih N serija. No često je moguće *velik* dio varijabilnosti u podacima opisati sa k ($< N$) komponenti – glavnih komponenti. Formalizirajmo to.

Neka je $X^\tau = (X_1, ..., X_N)$ slučajni vektor s kovarijacijskom matricom $\sum$ kojoj su brojevi $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$ svojstvene vrijednosti. Za proizvoljne realne vektore a_i, $i = 1, ..., N$ možemo definirati linearne kombinacije:

$$Y_i = a_i^\tau X = a_{i1}X_1 + a_{i2}X_2 + \ldots + a_{iN}X_N, \quad i = 1, \ldots, N. \tag{1}$$

Te linearne kombinacije varijabli $X_1, ..., X_N$ također su slučajne varijable s varijancom i kovarijancom:

$$VarY_i = a_i^\tau \Sigma a_i, \quad i = 1, \ldots, N \tag{2}$$

$$Cov(Y_i, Y_k) = a_i^\tau \Sigma a_k, \quad i, k = 1, \ldots, N. \tag{3}$$

Glavne komponente definiramo rekurzivno kao nekorelirane linearne kombinacije $Y_1, Y_2, ..., Y_N$ iz (1) kojima je varijanca najveća moguća. U svrhu jedinstvenosti, u definiciji glavnih komponenti dodatno zahtijevamo da vektori pondera a_i budu norme jedan, pa je $a_i^\tau a_i = 1$:

- prva glavna komponenta $PC1$ linearna je kombinacija $a_1^\tau X$ maksimalne varijance uz uvjet $a_1^\tau a_1 = 1$,
- druga glavna komponenta $PC2$ linearna je kombinacija $a_2^\tau X$ maksimalne varijance uz uvjete $a_2^\tau a_2 = 1$ i $Cov(a_1^t X, a_2^\tau X) = 0$,

$\vdots$

- i-ta glavna komponenta (za $i \leq N$) PCi linearna je kombinacija $a_1^\tau X$ maksimalne varijance uz uvjete $a_i^\tau a_i = 1$ i $Cov(a_i^\tau X, a_k^\tau X) = 0$, za sve $k < i$.

Dakle, polaznom definicijom iz postojećih podataka generiramo nekorelirane komponente maksimalne varijance. U praksi se nameću dva pitanja. Prvo, kako efikasno izračunati glavne komponente, i drugo, koji dio ukupne varijance, tj. udio u broju $\sum_{i=1}^{N} VarX_i$, objašnjava prvih k glavnih komponenti. S tim u vezi iznosimo dva osnovna rezultata, no bez dokaza (dokaz se može naći npr. u Johnson i Wichern, 1998).

Rezultat 1. Neka je $\sum$ kovarijacijska matrica vektora $X^\tau = (X_1, ..., X_N)$. Neka su $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$ svojstvene vrijednosti, a $x_1, ..., x_N$ odgovarajući svojstveni vektori (jedinične norme) od $\sum$. Tada je i-ta glavna komponenta dana sa:

$$Y_i = x_i^\tau X, \quad i = 1, \ldots, N. \tag{4}$$

Nadalje vrijedi:

$$VarY_i = \lambda_i, \quad i = 1, \ldots, N \tag{5}$$

$$Cov(Y_i, Y_k) = 0, \quad i \neq k. \tag{6}$$

Rezultat 2. Neka su $Y_1 = x_1^\tau X, \ldots, Y_N = x_N^\tau X$ glavne komponente (iz rezultata 1). Tada je:

$$\sum_{i=1}^{N} VarX_i = \lambda_1 + \ldots + \lambda_N = \sum_{i=1}^{N} VarY_i. \tag{7}$$

Prvi rezultat kaže da glavne komponente izračunavamo kao produkt svojstvenih vektora kovarijacijske matrice i vektora *X*, a drugi rezultat osigurava jednakost zbroja varijanci komponenti originalnog vektora *X* i konstruiranih glavnih komponenti. Nadalje, udio *k*-te glavne komponente u ukupnoj varijanci iznosi:

$$\frac{\lambda_k}{\lambda_1 + \ldots + \lambda_N}, \quad k = 1, \ldots, N \tag{8}$$

pa udio prvih *l* komponenti iznosi:

$$\frac{\lambda_1 + \ldots + \lambda_l}{\lambda_1 + \ldots + \lambda_N}, \quad l \leq N. \tag{9}$$

Operativno je dovoljno izračunati (uzoračku) kovarijacijsku matricu podataka te njezine svojstvene vektore i vrijednosti. Obično je riječ o računanju svojstvenih vrijednosti matrica velikih dimenzija pa je nužna odgovarajuća programska potpora. Sa stajališta numeričke stabilnosti umjesto direktnog računanja svojstvenih vrijednosti i vektora bolje je provesti *dekompoziciju svojstvenih vrijednosti* (*Singular Value Decomposition*). O tome se detaljno može naći u nekom od udžbenika (numeričke) linearne algebre.

Metoda glavnih komponenti to je korisnija što je izraz (9) bliži jedinici za relativno mali *l*, tj. ako sa što manje komponenti uspijemo u značajnoj mjeri *pročitati* varijabilnost polaznih podataka. Koji je *l* dovoljno malen i koliko je blizu jedinice zaista dovoljno blizu, uglavnom ovisi o naravi problema i subjektivnoj odluci pojedinca, iako postoje i određene formalne procedure koje u tome pomažu.

U primjeni redovito imamo *T* realizacija vektora *X* koje promatramo kao matricu *X* dimenzija, $N \times T$. Neka je Y $K \times T$ matrica prvih *K* glavnih komponenti. Ako je za neko *K* značajan dio varijance podataka objašnjen u prvih *K* komponenti ($K<N$), tada *N*-dimenzionalni vektor originalnih podataka $X^\tau = (X_1, \ldots, X_N)$ reprezentiramo *K*-dimenzionalnim vektorom $Y^\tau = (Y_1, \ldots, Y_N)$. Tako komprimiranim podacima tada se možemo koristiti u daljnjoj statističkoj analizi. Dakle, smanjenjem dimenzije svjesno se odričemo određene količine informacija kako bismo informaciju potencijalno vrlo velikog broja vremenskih serija mogli iskoristiti u modeliranju. Isplativost takvog *trade-offa* ispituje se od slučaja do

slučaja. To je posebno korisno u uvjetima kratkih uzoraka, u kojima se, na primjer, regresijskom analizom može mjeriti utjecaj velikog broja varijabli (svedenih na mali broj konstruiranih komponenti) na ključne serije. U tom kontekstu smatramo da je uputno ispitati potencijal metode glavnih komponenti pri ekonomskim istraživanjima u Hrvatskoj.

2.2. Prva glavna komponenta serija cijena i tečaja

Radi ilustracije analizirat ćemo prvu glavnu komponentu grupe od 32 vremenske serije cijena i tečajeva koje smatramo relevantnima za kretanje inflacije hrvatskog indeksa potrošačkih cijena (IPC-a). U dodatku je dan potpun popis svih serija grupe.

Podaci u analizi su mjesečni, i to za razdoblje od siječnja 1998. do rujna 2007. godine.[4] Prije izračunavanja kovarijacijske matrice podaci su transformirani na sljedeći način.

1. Programom X12 – ARIMA filtriramo podatke[5] i u daljnjoj analizi uzimamo u obzir trend-ciklus[6] komponente podataka.
2. Logaritmiramo[7] pa diferenciramo podatke. Razlike logaritama aproksimiraju mjesečne stope promjene.
3. Standardiziramo podatke na način da sve serije imaju nulto očekivanje i standardnu devijaciju jednaku 1. Tako ih, neovisno o *mjernoj jedinici*, stavljamo u istu skalu. Primijetimo da je kovarijacijska matrica standardiziranih podataka – korelacijska matrica, pa je analiza glavnih komponenti standardiziranih podataka analiza korelacijske strukture podataka.

Nakon takve transformacije podaci su pripremljeni za klasičnu analizu glavnih komponenti. Najprije izračunavamo kovarijacijsku (korelacijsku) matricu podataka te njezine svojstvene vrijednosti i vektore. Iz tih matematičkih objekata po već opisanom postupku izračunavamo glavne komponente te komentiramo sastav prve komponente radi detaljnijeg uvida u razinu koreliranosti među varijablama grupe. U tablici 1. dajemo osnovne rezultate za prvih nekoliko glavnih komponenti.

Tablica 1. Prvih šest glavnih komponenti serija cijena i tečajeva (PCi, i = 1,...,6, oznaka je za i-tu komponentu)

Komponenta	PC1	PC2	PC3	PC4	PC5	PC6
svojstvena vrijednost	8,49	5,46	4,07	2,62	1,82	1,58
udio u ukupnoj varijanci	0,28	0,18	0,14	0,09	0,06	0,05
kumulativ udjela u ukupnoj varijanci	0,28	0,47	0,60	0,69	0,75	0,80

Izvor: izračun autora

[4] Državni zavod za statistiku počeo je objavljivati indeks potrošačkih cijena u veljači 2004. godine. Naknadno je izračunana retrogradna serija indeksa s početkom od siječnja 1998. godine.

[5] To je službeni program za desezoniranje Američkoga statističkog ureda. Detalji vezani za metodu mogu se pronaći na web stranici: http://www.census.gov/srd/www/x12a/

[6] Pretpostavlja se aditivna dekompozicija serije $Y = TC + S + I$, gdje je TC trend-ciklus komponenta (T-trend, možda ciklus izrazito niske frekvencije, C je ciklus, tj. periodička komponenta srednjeg roka), S je sezonska komponenta, a I je ostatak, tzv. iregularna komponenta. Korištenje trend-ciklus komponente, dakle ignoriranje i sezonske i iregularne komponente motivirano je raspravom u Camacho i Sancho (2003) u kontekstu primjene faktorskih modela sa španjolskim podacima.

[7] Koristimo se prirodnim logaritmom.

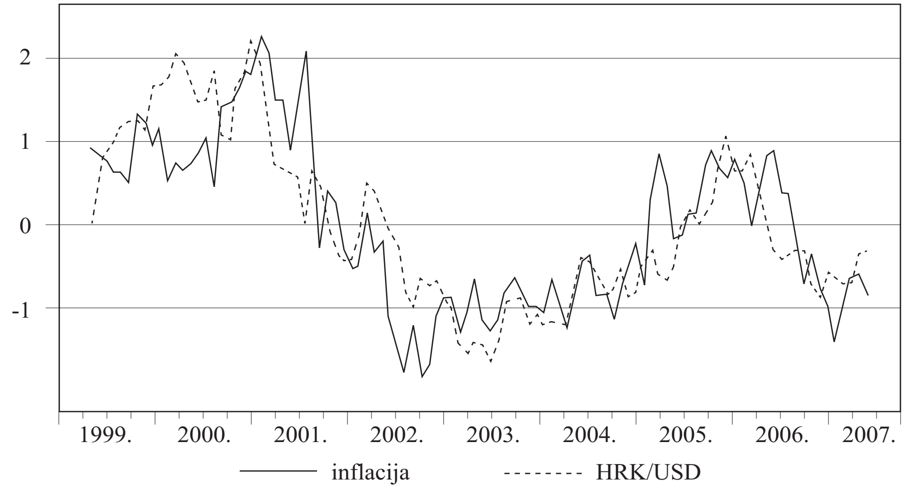
Prvih šest komponenti objašnjava 80% ukupne varijance podataka, a prvih deset komponenti objašnjava čak 92% varijance. Prema konstrukciji, prva među komponentama objašnjava najveći dio varijance – u promatranom primjeru 28%. Radi boljeg uvida, napominjemo da već deseta komponenta objašnjava samo 2% varijance podataka. Na temelju izračunanih vrijednosti zaključujemo da postoji relativno jaka korelacija stopa rasta serija unutar grupe. Prema definiciji, izračunana je prva glavna komponenta linearna kombinacija maksimalne varijance svih serija promatrane grupe, pa je u određenom smislu možemo smatrati indeksom konstruiranim iz 32 vremenske serije cijena i tečajeva. Radi interpretacije, potrebno je analizirati pondere, tj. svojstvene vektore korelacijske matrice, takvog indeksa. Primijetimo da predznak pojedinog pondera u tom kontekstu ne pomaže jer su glavne komponente prema definiciji identificirane samo do na predznak. Zbog toga se koncentriramo na apsolutni iznos pondera.

Tablica 2. Ponderi u prvoj komponenti i korelacije za četiri najznačajnije serije

	IPC	Dobra (bez energije)	Dobra	HRK/USD
ponderi u komponenti	0,31	0,29	0,29	0,25
korelacija s komponentom	0,85	0,80	0,80	0,70

Izvor: izračun autora

Slika 1. Standardizirane godišnje stope rasta prosječnoga mjesečnog tečaja HRK/USD i indeksa potrošačkih cijena

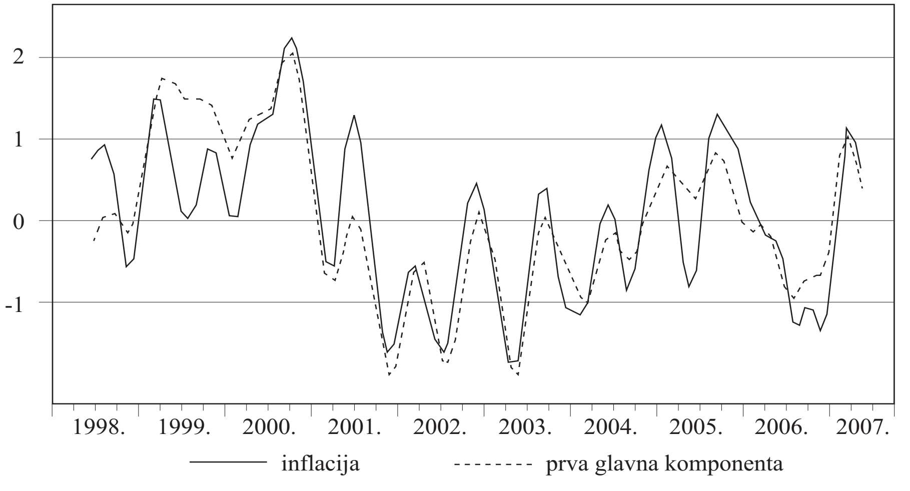


Izvor: izračun autora; HNB

U tablici 2. dani su podaci za četiri serije iz promatrane grupe cijena i tečajeva koje ulaze u konstruirani indeks (prvu glavnu komponentu) s najvećim ponderima. Osim toga, prikazujemo i korelacije tih serija s izračunanim indeksom. Tri serije s najvećim udjelima indeksi su cijena – ukupni *indeks potrošačkih cijena* (0,31), serija *DOBRA bez električne energije* (0,29) te *DOBRA* (0,29). To smo i očekivali jer velik dio varijabli promatrane grupe čine podindeksi IPC-a te su oni u određenoj mjeri i korelirani s njim. Prema definiciji, od svih mogućih linearnih kombinacija danih varijabli cijena i tečajeva, njihova prva glavna komponenta optimalno aproksimira ukupnu varijancu grupe. Ako postoji snažna korelacija među njezinim članovima, može se očekivati da glavna komponenta također prati tu zajedničku dinamiku. Zanimljivo, četvrta varijabla po redu[8] srednji je mjesečni tečaj *HRK/USD*. Radi ilustracije veze kretanja potrošačkih cijena i tečaja kune prema dolaru, na slici 1. prikazujemo standardizirane godišnje stope rasta prosječnoga mjesečnog tečaja *HRK/USD* i našeg indeksa potrošačkih cijena. Detaljnije o vezi tečaja kune prema američkom dolaru i inflacije može se naći u Jankov i sur. (2007).

Na slici 2. prikazujemo izračunanu prvu glavnu komponentu serija cijena i tečajeva zajedno s mjesečnom stopom rasta trend-ciklus komponente IPC-a.

Slika 2. Prva glavna komponenta i standardizirane mjesečne stope rasta indeksa potrošačkih cijena



Izvor: izračun autora; HNB

Glavna je komponenta po konstrukciji linearna kombinacija mjesečnih stopa rasta trend-ciklusa serija promatrane grupe. Stoga takav *indeks*, koji kao sastavnice, osim komponenti *IPC-a*, ima i niz drugih varijabli koje su bitne za kretanje inflacije, eksplicitno i izmjerivo uvažava njihovu dinamiku.

Tim je primjerom ilustrirano kako se metode kompresije informacija mogu primijeniti za izradu indeksa koji je prema jasnom kriteriju optimalna jednodimenzionalna reprezentacija grupe varijabli. Indeksi tog tipa služe za konstrukciju indikatora temeljne inflacije, razine ekonomske aktivnosti i sl.

[8] Uočimo (tabl. 2) da isti poredak kao i u pondera ostaje i ako promatramo korelacije, no nije uvijek tako (v. npr. Johnson i Wichern, 1998).

3. Faktorske prognoze inflacije

3.1. Uvod

U uvodu smo spomenuli da se pri prognoziranju ključnih varijabli u središnjim bankama vodi briga o potencijalno vrlo velikom broju ekonomskih varijabli. Pritom je glavno pitanje kako tako velik broj serija na neki način komprimirati a da se glavnina informacije ipak sačuva. Među ostalima, formalni okvir za takvo sažimanje informacija u faktore te za njihovu primjenu u daljnjoj statističkoj analizi predlažu i Stock i Watson (1998, 1999, 2002).

U ovom poglavlju najprije definiramo jednu varijantu faktorskog modela i objašnjavamo kako općenito faktorska analiza pomaže u analizi velikog broja više ili manje povezanih varijabli. Nakon toga implementiramo model na hrvatske podatke.

3.2. Definicija i osnovna svojstva

U prošlom smo poglavlju dali osnovne činjenice analize glavnih komponenti. Faktorska se analiza zasniva na nešto drugačijoj ideji, no uz određene pretpostavke parametri se u njoj procjenjuju metodom glavnih komponenti.

Pretpostavimo da su $T \times N$ matricom X, dane realizacije N varijabli $X_1, ..., X_N$ tijekom T vremenskih razdoblja. Faktorska se analiza temelji na pretpostavci da postoji r ($<N$) varijabli – *faktora*, $f_1, ..., f_r$, koji su dovoljni za modeliranje dinamike svih N promatranih serija. Pretpostavimo linearnu ovisnost originalnih podataka iz X o faktorima gdje su u matrici Λ spremljeni koeficijenti. Dakle, pretpostavljamo da vrijedi:

$$X = F\Lambda^{\tau} + \varepsilon, \tag{10}$$

gdje je X $T \times N$ matrica standardiziranih originalnih podataka, F je $T \times r$ matrica faktora, Λ *$N \times r$ matrica faktorskih opterećenja* (engl. *factor loadings*) i ε matrica slučajnih pogrešaka. Po elementima je:

$$X_{ti} = F_t \lambda_i^{\tau} + \varepsilon_{ti} \quad t = 1, \ldots, T,\ i = 1, \ldots, N, \tag{11}$$

gdje je F_t r-dimenzionalni vektor faktora u trenutku t, a λ_i vektor koeficijenata kojim je uređena ovisnost i-te varijable o faktorima. To je tzv. statička reprezentacija faktorskog modela, u kojoj pretpostavljamo ovisnost podataka samo o trenutačnim vrijednostima faktora.[9]

Vidimo da se u faktorskome modelu pretpostavlja da se svaka varijabla može dekomponirati na linearnu kombinaciju faktora i slučajnu pogrešku. Faktori u tom kontekstu predstavljaju silnice koje generiraju ekonomiju i zajednički su svim promatranim varijablama u X. Ono što je specifično za pojedinu varijablu nalazi se u izrazu za pogrešku, dakle i pogreška u pravom smislu i, eventualno, faktori specifični baš za tu varijablu. U

[9] Osim statičkih faktorskih modela, postoje i dinamički faktorski modeli u kojima se pretpostavlja da elementi od X osim trenutačnih vrijednosti faktora dodatno ovise i o njihovim prošlim vrijednostima. No dinamički modeli imaju statičku reprezentaciju tipa (10) (Stock i Watson, 2002), stoga se ovdje ograničavamo samo na tu klasu modela.

(11) je dopuštena određena korelacija među komponentama od ε, zbog čega predstavljeni model nazivamo i aproksimativnim faktorskim modelom (Bai i Ng, 2002).

U jednadžbama (10) i (11) nepoznato je sve osim matrice podataka X. Dakle, da bismo procijenili i matricu Λ ($N \times r$ parametara) i faktore ($T \times r$ parametara), nužno je uvesti određene restrikcije na parametre. Pretpostavimo stoga[10] da je $\Lambda^\tau \Lambda = I$.

Parametre procjenjujemo metodom najmanjih kvadrata, pri čemu tražimo one procjene $\hat{F}$ i $\hat{\Lambda}$ za koje je pogreška u (11) najmanja (u smislu 2-matrične norme), tj. za koje je $\sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{ti}^2$ minimalno, uz uvjet $\Lambda^\tau \Lambda = I$. Pokazuje se da je to klasičan problem glavnih komponenti[11] koji rješavamo uzimajući za $\hat{\Lambda}$ svojstvene vektore pridružene prvim r najvećim svojstvenim vrijednostima matrice $X^\tau X$. Tada je $\hat{F} = X\hat{\Lambda}$. Detalji o procjeni dani su u dodatku, dok se o svojstvima tih procjenitelja može naći u Stock i Watson (1998).

Dakle, sam postupak procjene faktora metodom glavnih komponenti krajnje je jednostavan, no pritom se postavlja nekoliko pitanja. Prije svega, kako odabrati broj faktora, tj. procijeniti r (Bai i Ng, 2002)? Ponajprije nije jasna ni konzistentnost podataka s pretpostavljenom strukturom (10). Zatim se pojavljuje pitanje identifikacije u (10), što je u uskoj vezi s interpretacijom faktora. Naime, za $r \times r$ matricu O, za koju je $OO^\tau = I$, (10) se može napisati i kao:

$$X = (FO)(O^\tau \Lambda^\tau) + \varepsilon, \tag{12}$$

i tada su *novi* faktori dani s FO, a *faktorska opterećenja* s $O^\tau \Lambda^\tau$. No to nije problem ako se faktorska analiza primjenjuje samo u fazi pripreme podataka za prognostički model jer svaki par FO i $O^\tau \Lambda^\tau$ generira isti X. Problem nastaje ako faktorima želimo dodijeliti strukturnu interpretaciju jer nije apriori jasno jesu li sa F dani *pravi* faktori, tj. silnice koje generiraju cjelokupnu ekonomiju ili samo neka njihova teško interpretabilna transformacija.

3.3. Prognostički model

U ovom ćemo dijelu implementirati *Stock-Watsonov* prognostički model na hrvatske ekonomske serije. Cilj je, uz dostupne podatke do trenutka T, prognozirati vrijednost varijable y u trenutku $T + h$, dakle prognozirati y_{T+h}, za neko h. U promatranom primjeru pokušavamo odgovoriti na pitanje može li nam poznavanje vrijednosti 144 tromjesečne serije danas (vidjeti dodatak), pomoći u prognozi inflacije h kvartala od danas.

Model je ovakav:

$$X_{ti} = F_t \lambda_i^\tau + \varepsilon_{ti} \tag{13}$$

$$y_{t+h} = \alpha_h + \beta(L) F_t + \gamma(L) y_t + \varepsilon_{t+h}, \tag{14}$$

[10] Alternativno možemo pretpostaviti ortogonalnost faktora – $F^\tau F = I$, što bi vodilo istom rješenju (dodatak).

[11] Osim glavnim komponentama, parametri faktorskog modela mogu se procijeniti i iz tzv. *state space* reprezentacije modela metodom maksimalne vjerodostojnosti (engl. *maximum likelihood*). No maksimizacija postaje vrlo zahtjevna kako N (broj varijabli iz kojih izlučujemo faktore) raste. Stoga se takva procjena primjenjuje samo pri ekstrakciji komponenti iz manjeg broja varijabli.

pri čemu je α_h konstanta, $\beta(L)$ i $\gamma(L)$ polinomi konačnog reda u varijabli L, gdje je L operator pomaka. X_{ti} je vrijednost i-te varijable u trenutku t i F_t je r-dimenzionalni vektor faktora u trenutku t. Dakle, u (14) pretpostavljamo da postoji linearna veza skalarne varijable *y u trenutku t + h* i faktora *do trenutka t*. Dodatno pretpostavljamo da se o budućnosti varijable od interesa može nešto zaključiti i iz njezine prošlosti, pa je uključen i izraz $\gamma(L)\gamma_t$.

Ne možemo odmah započeti procjenu parametara u (14), jer nisu poznati svi prediktori, tj. faktori F_t. Prave vrijednosti faktora ne možemo opažati, ali ih procjenjujemo iz (13), i to po već opisanom postupku. Dakle, strategija je sljedeća.

1. *Procjena faktora.* Iz jednadžbe (13) procjenjujemo vektor faktora kojim aproksimiramo podatke iz X.
2. *Prognoziranje.* Poznati su podaci do trenutka T. Koeficijente regresije (14) procijenimo na podacima od trenutka 1 do T (metodom najmanjih kvadrata) i tako izračunamo procjene $\hat{\alpha}_h$, $\hat{\beta}(L)$ i $\hat{\gamma}(L)$. Nakon toga prognozu izvodimo kao $\hat{y}_{t+h} = \hat{\alpha}_h + \hat{\beta}(L)\hat{F}_t + \hat{\gamma}(L)y_t$. Da bi to bila najbolja prognoza u smislu srednje kvadratne pogreške, moramo pretpostaviti da je $E(\varepsilon_{t+h} \mid I_t) = 0$, gdje je I_t informacija dostupna do trenutka t, dakle i faktori i y_i, za $i \leq t$.

Faza *prognoziranja* zahtijeva obrazloženje. Prognoze se izvode izravnom projekcijom iz podataka dostupnih do trenutka t, dakle kao $\hat{y}_{t+h} = \hat{\alpha}_h + \hat{\beta}(L)\hat{F}_t + \hat{\gamma}(L)y_t$. To je pogodno zato što tada nije potrebno prognozirati faktore (npr. vektorskim autoregresijama) da bi se našlo $\hat{y}_{t+h}$. Koja metoda daje bolje rezultate, nije sasvim jasno, no provedeni izračun svakako je jednostavniji i, što je bitnije, zahtijeva procjenu manjeg broja parametara.

Podaci iz kojih izvodimo analizu sektorski su podijeljeni u pet grupa – *Cijene i tečaj* (32 serije), *Tržište rada* (24 serije), *Eksterni sektor* (28 serija), *Realni sektor* (26 serija) i *Novac* (34 serije), što ukupno iznosi 144 serije. Popis svih serija nalazi se u dodatku.

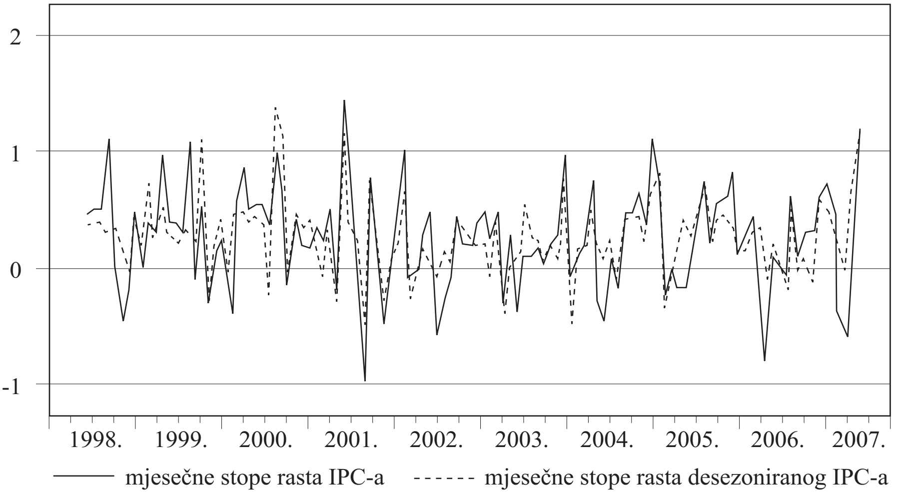
Kao i pri računanju glavnih komponenti, i u ovom postupku najprije transformiramo serije. Teorija zahtijeva da serije koje ulaze u analizu budu stacionarne. Uz pretpostavku da su originalni podaci integrirani reda jedan (dakle, i stacionarni nakon što jednom budu diferencirani), provedenom transformacijom aproksimiramo stacionarnost.[12] Desezoniranjem *čistimo* sezonsku komponentu, logaritmiranjem reduciramo eventualnu heteroskedastičnost, a diferenciranjem eliminiramo trend.

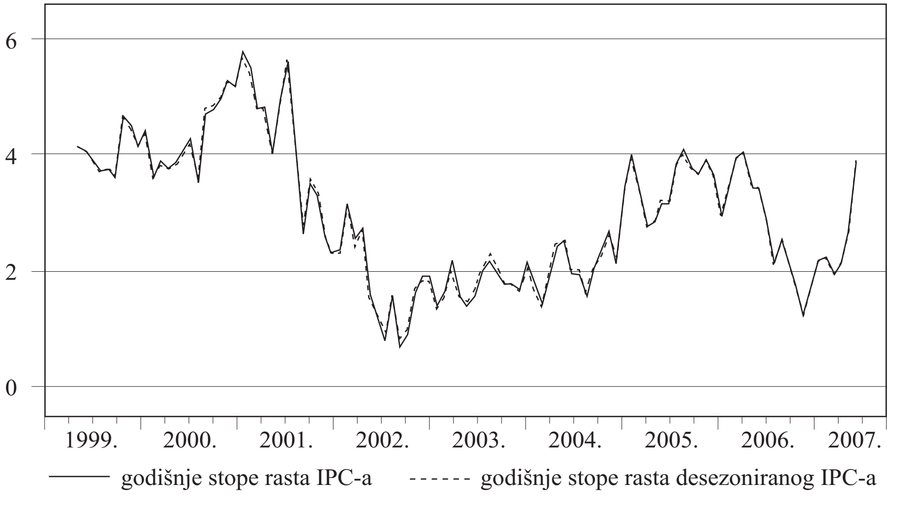
Najbitnije su nam prognoze inflacije *indeksa potrošačkih cijena*. Taj indeks ima izraženu sezonsku komponentu, pa prije analize stopa seriju valja desezonirati. Na slici 3. dane su mjesečne i godišnje stope originalne i desezonirane serije indeksa potrošačkih cijena.

[12] Provedenim transformacijama zaista samo aproksimiramo stacionarnost. Najproblematičnije je diferenciranje i teško je sa sigurnošću utvrditi koliko puta je potrebno diferencirati pojedinu seriju. Sve serije koje ulaze u izračun (osim tečajeva) posjeduju jasne rastuće ili padajuće trendove, što sugerira da ih je barem jedanput potrebno diferencirati. Eventualnu potrebu za daljnjim diferenciranjem teško je utvrditi u uvjetima kratkih vremenskih serija, čemu je glavni uzrok mala snaga *unit-root* testova.

Vidimo kako mjesečne stope uvelike ovise o činjenici je li serija desezonirana, dok je za godišnje stope to sve manje bitno što je sezonska komponenta bliža nekoj periodičkoj funkciji.[13]

Slika 3. Stope rasta originalne serije i desezonirane serije indeksa potrošačkih cijena. Prva slika pokazuje mjesečne, a druga godišnje stope





Izvor: izračun autora; HNB

[13] Pretpostavimo da logaritam neke varijable, snimane na mjesečnoj frekvenciji, dekomponiramo na sezonsku komponentu (S) i preostali dio (N): $lnY_t = S_t + N_t$. Godišnje stope promjene aproksimiramo sa $lnY_t - lnY_{t-12} = S_t + N_t - (S_{t-12} + N_{t-12}) = S_t - S_{t-12} + (N_t - N_{t-12})$, a ako je S periodička funkcija perioda 12, tj. $S_t = S_{t-12}$, godišnje stope promjene ne ovise o sezonskoj komponenti, pa nije bitno je li serija desezonirana (sl. 3).

Uz podatke poznate do trenutka t za, primjerice, varijablu z zanima nas očekivana stopa promjene u trenutku $t + h$ (u odnosu prema t), dakle, očekivana vrijednost od:

$$y_{t+h}^{h} = \frac{400}{h} \ln \frac{z_{t+h}}{z_t}. \tag{15}$$

Tako definirane stope su anualizirane, pa su prognoze za razne horizonte h izravno usporedive. Dodatno definirajmo (anualizirane) tromjesečne stope serije od interesa $y_t = 400 \ln \frac{z_t}{z_{t-1}}$. Primijetimo da su sa (15) dane akumulirane tromjesečne stope za razdoblje od $t + 1$ do $t + h$.

Naš su krajnji cilj faktorske prognoze (15) *indeksa potrošačkih cijena*. Radi što točnije evaluacije njihove kvalitete, potrebno je u najvećoj mogućoj mjeri simulirati realne uvjete prognoziranja.

Podaci su tromjesečni, i to iz intervala od prvog tromjesečja 1998. godine do drugog tromjesečja 2007. godine. Za prognoze uz horizont h, intervalom od 1. tromjesečja 1998. do (4 - h + 1)-og tromjesečja 2003. godine koristimo se isključivo za procjenu, dok ostalo razdoblje služi i za prognoze, i to rekurzivno, na sljedeći način.

1. Metodom glavnih komponenti procjenjujemo faktore (tj računamo $\hat{F}_t$).
2. *Akaike* informacijskim kriterijem odredimo broj *pomaka* za faktore i zavisnu varijablu.
3. Metodom najmanjih kvadrata procjenjujemo parametre regresije $y_{t+h}^{h} = \alpha_h + \beta(L)F_t + \gamma(L)y_t + \varepsilon_{t+h}$.
4. Konstruiramo prognoze $\hat{y}_{t+h}^{h} = \hat{\alpha}_h + \hat{\beta}(L)\hat{F}_t + \hat{\gamma}(L)y_t$.
5. Dodajemo novi podatak u uzorak i idemo na 1 (osim ako nismo iscrpili uzorak; tada je postupak završen).

Na taj način za proizvoljno h izračunamo prognoze za interval od prvog tromjesečja 2004. do drugog tromjesečja 2007. Algoritam ostavlja mogućnost definiranja više varijanti modela. Na našim podacima testiramo kvalitetu sljedećih dvaju.

Autoregresivni model – AR. Pretpostavljamo da stope y_{t+h}^{h} ovise isključivo o tromjesečnim stopama (i njezinim prethodnim vrijednostima), $y_t = 400 \ln \frac{z_t}{z_{t-1}}$. Broj pomaka od 0 do 4 biramo AIC-om. Taj model služi kao *benchmark*. Diskusija o kvaliteti prognoza takvog modela može se naći u Kapetanios i sur. (2007).

Autoregresivni faktorski model – ARF. Pretpostavljamo da stope y_{t+h}^{h} ovise o sadašnjim i prethodnim vrijednostima i faktora i stopa y_t. Broj pomaka biramo AIC-om (od 0 do 4 za y i od 0 do 3 za faktore). Definiramo model za k faktora, k = 1,2, dakle dva ARF modela.

3.4. Kvaliteta dobivenih prognoza

Kvalitetu prognoza mjerimo standardnom mjerom, *Mean Squared Error* (MSE), definiranom na sljedeći način. Neka su $x_1, ..., x_N$ ostvarenja, a $\hat{x}_1, \ldots, \hat{x}_n$ prognozirane vrijednosti procesa x. Tada je:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{16}$$

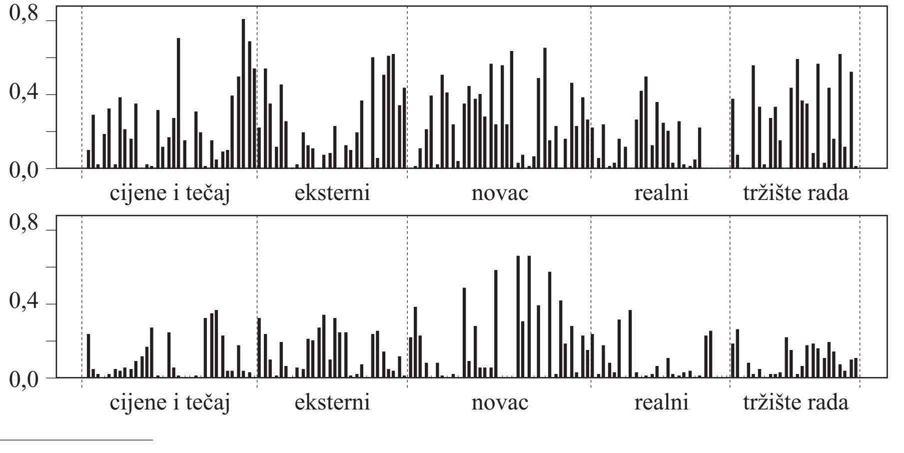
U tablici 3. dajemo MSE statistike prognoza inflacije za horizonte od jednog do četiri tromjesečja. Statistike su dane kao omjeri s obzirom na MSE autoregresivnog (AR) modela. Dakle, ako je vrijednost veća od jedinice, model je (u MSE smislu) lošiji od *benchmark* modela, no ovdje definiran *benchmark* nije naivno određen (npr. kao *random walk* model i sl.) te ga stoga nije nužno lako nadmašiti. Rezultati pokazuju da faktori izvučeni iz promatrane grupe serije imaju određeni potencijal za prognoziranje dinamike indeksa potrošačkih cijena. Dodavanjem jednog faktora (model ARF1) prognoze su nešto bolje od *benchmarka*. No daljnje dodavanje faktora ne poboljšava prognoze.[14] Taj je rezultat na tragu rezultata iz Stock i Watson (2002) ili Matheson (2006).

Tablica 3. Rezultati prognoza inflacije indeksa potrošačkih cijena. MSE statistike prikazane su kao omjeri s obzirom na benchmark

Horizont	1	2	3	4
ARF1	0,91	0,81	0,72	0,92
ARF2	0,99	1,05	3,55	2,87

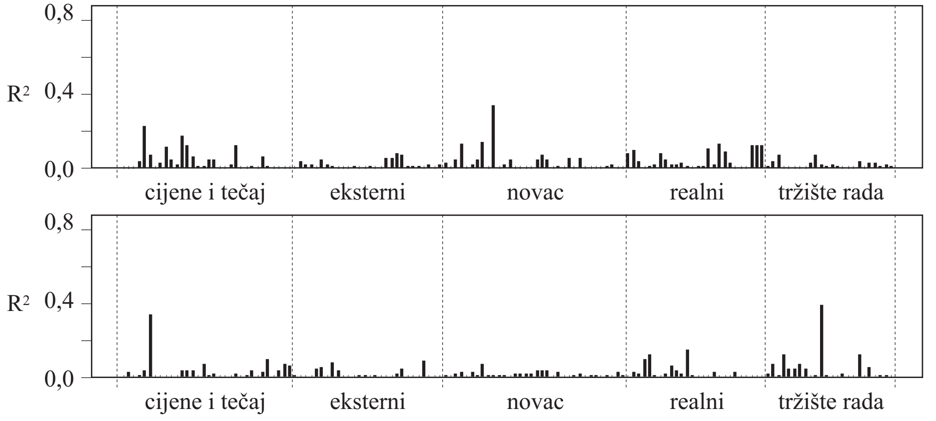
Izvor: izračun autora

Slika 4. Povezanost prvog, drugog, desetog i jedanaestog faktora sa 144 varijable iz kojih se izlučuju faktori



[14] Analiza je provedena i na podacima uz mjesečnu frekvenciju, pri čemu su konstruirane prognoze na horizontima do 12 mjeseci. Rezultati te analize sugeriraju da su faktorske prognoze, uz korištenje do dva faktora, bolje od *benchmarka* na kratkim horizontima, do 7 ili 8 mjeseci. Iako su uz mjesečnu frekvenciju dostupne dulje serije podataka, cilj analize je konstrukcija prognoza inflacije do godinu dana unaprijed. Uz danu duljinu dostupnih podataka vjerodostojnijima se čine kvartalne prognoze za do četiri razdoblja unaprijed nego mjesečne prognoze za do dvanaest razdoblja unaprijed.

Slika 4. nastavak



Napomena: Stupci predočuju koeficijente determinacije iz regresija svake od 144 serija na prvi (prva slika), drugi (druga slika), deseti (treća slika) i jedanaesti (četvrta slika) faktor. Isprekidane vertikalne linije odvajaju serije po sektorima. Statistike (visina stupaca) za svaki faktor određuju koliki udio varijance pojedine serije on objašnjava. Prva dva faktora kumulativno objašnjavaju oko 45% ukupne varijabilnosti svih serija. Za ilustraciju, deseti i jedanaesti faktor zajedno objašnjavaju samo 4% varijance.

Izvor: izračun autora

3.5. Mogućnosti unaprjeđenja modela

Rad sa samo jednim faktorom pri prognoziranju povećava preciznost prognoza inflacije u odnosu prema prognozi *benchmark* modela. No prostora za unaprjeđenje u ovom radu implementiranoga faktorskog modela zaista je mnogo. Nabrojit ćemo neke segmente analize koje smo ovdje sasvim ili djelomice ignorirali. Stoga ovime dajemo i smjernice za daljnji rad.

- *Određivanje broja faktora koji ulaze u prognostički model.* Broj faktora trebao bi se tražiti nekim od informacijskih kriterija razvijenima za tu namjenu. Svakako je najreprezentativniji od njih Bai-Ng procjenitelj (Bai i Ng, 2002). U ovoj smo analizi testirali modele u koje smo uključili do dva faktora. Oni zajedno opisuju 45% ukupne varijance podataka. Slika 4. ilustrira relativno snažnu koreliranost prvih dvaju faktora sa svim serijama te vrlo nisku koreliranost desetoga i jedanaestog faktora sa serijama. To djelomično opravdava *ad hoc* metodu odabira broja faktora.
- *Tretiranje nedostajućih vrijednosti.* U analizi smo se koristili sa 144 serije koje nisu imale nedostajućih vrijednosti, dakle u samom početku izbor serija bio je znatno sužen. Za procjenu nedostajućih vrijednosti standardno se koristi *expectation maximization* (EM) algoritam (Stock i Watson, 2002). To je posebno korisno u tzv. *nowcastingu*, npr. BDP-a (Schumacher i Breitung, 2006. te Giannone i sur., 2005), u kojemu podaci za njegov izračun stižu sa zakašnjenjem.

- *Tretiranje outliera.* Budući da se analiza odnosi na korelacijsku strukturu podataka, trebalo bi izbaciti ili na neki način korigirati vrijednosti neuobičajeno daleko od očekivanih (ovdje stacionarnih) analiziranih serija. To se radi tako da se serije prije analize skeniraju i outliere proglasimo nedostajućim vrijednostima, a onda ih tretiramo kao u prethodnoj točki.
- *Podaci različitih frekvencija.* Pokatkad je korisno u analizu uključiti i mjesečne i kvartalne podatke. To se također radi primjenom EM algoritma.
- *Desezoniranje.* Ovdje smo desezonirali sve serije iz kojih smo ekstrahirali faktore, no to nije bilo nužno. Zatim smo se koristili samo trend-ciklusom komponente serija u analizi, pa bi valjalo provesti postupak i s klasično desezoniranim podacima. Konačno, analizu bi trebalo provesti uz različite metode desezoniranja.
- *Odabir serija.* Serije koje bi ulazile u analizu mogle bi biti rigoroznije odabrane, npr. na temelju neke vrste *expert knowledge* informacija. Zatim se može testirati iskoristivost sektorskih faktora u prognoziranju, dakle faktora izvučenih iz serija neke od pet u radu definiranih grupa (dodatak).

4. Zaključak

Opisali smo osnovne pojmove, ciljeve i mogućnosti metoda sažimanja velikog broja podataka u samo nekoliko varijabli. Najprije smo iznijeli osnove metode glavnih komponenti i predstavili je kao alat za analizu kovarijacijske strukture grupe varijabli. Nakon toga smo predstavili jedan prognostički model faktorske analize koji se u fazi procjene oslanja na metodu glavnih komponenti.

Provjerili smo iskoristivost tih metoda na hrvatskim podacima. U prvom smo koraku informaciju (drugog momenta!) 144 varijable bitne za hrvatsko gospodarstvo saželi u mali broj faktora i provjerili može li se tako komprimirana iskoristiti za prognoziranje akumuliranih stopa rasta indeksa potrošačkih cijena na horizontima do četiri kvartala. Korištenje samo jednog faktora pri prognoziranju povećava preciznost prognoza inflacije u odnosu prema prognozama *benchmark* modela, što potiče na daljnja istraživanja modela faktorske analize.

Provedena je analiza preliminarna i stoga nije sasvim rigorozna. Prostora za poboljšanje prikazanih modela ima dovoljno te su eksplicitno nabrojene ključne neriješene točke, čime su određeni i smjerovi daljnje analize.

Faktorski su modeli u radu testirani na problemu prognoziranja inflacije, pri čemu se pokazalo da informacija izlučena iz velikog broja ekonomskih varijabli može pomoći u prognoziranju budućih stopa rasta potrošačkih cijena. No metodologija je iznesena dovoljno općenito da može biti izravno primijenjena i na prognoziranje ostalih ekonomskih varijabli od interesa za koje se vjeruje da su pod utjecajem cjelokupnog stanja u ekonomiji (npr. pokazatelji realne aktivnosti ili varijable financijskih tržišta).

Dodatak

Dodatak A

Procjena parametara u faktorskome modelu korištenjem glavnih komponenti

Dan je model:

$$X_{ti} = F_t \lambda_i^\tau + \varepsilon_{ti} \quad t = 1, \ldots, T, i = 1, \ldots, N, \tag{17}$$

u kojemu je F_t r-dimenzionalni vektor faktora u trenutku t, a λ_i vektor koeficijenata kojim je uređena ovisnost i-te varijable o faktorima. Matrično to možemo napisati:

$$X = F\Lambda^\tau + \varepsilon, \tag{18}$$

ili, analogno:

$$X^\tau = \Lambda F^\tau + \varepsilon^\tau. \tag{19}$$

Metodom najmanjih kvadrata treba procijeniti vektor faktora F_t i matricu Λ, no problem je to što su oni, dakle i faktori i Λ, neopazivi. U primjeni, uz poznate podatke X ($N\times T$ podataka) treba procijeniti $N\times r$ elemenata matrice Λ i $r\times T$ elemenata faktora, tj. $(N + T)\times r$ parametara. Dakle, da bismo procijenili parametre, nužno je uvesti neka ograničenja na parametre – ili na faktore ili na matricu Λ. Razlikujemo dva slučaja: broj vremenskih opservacija T veći je od broja promatranih varijabli u X, tj. $T > N$, a druga je mogućnost razmatranje obrnute situacije, dakle $T < N$. Strategija procjene je različita, ovisno o tim relacijama.

Slučaj 1. $T > N$

Uz pretpostavku $\Lambda^\tau\Lambda = I$ minimiziramo funkciju:

$$V(F,\Lambda) = \sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \lambda_i F_t^\tau)^2 = tr(X^\tau - \Lambda F^\tau)^\tau (X^\tau - \Lambda F^\tau), \tag{20}$$

gdje je tr operator – trag matrice.[15] Iz (19) (multivarijatnom višestrukom) metodom najmanjih kvadrata procjenjujemo faktore:

$$\hat{F}^\tau = (\Lambda^\tau\Lambda)^{-1}\Lambda^\tau X^\tau = (uvjet\ na\ \Lambda) = \Lambda^\tau X^\tau, \tag{21}$$

iz čega slijedi:

$$\hat{F} = X\Lambda. \tag{22}$$

[15] Trag matrice definiran je kao zbroj elemenata njezine dijagonale. Pokazuje se da je s $tr(A^\tau A) = \sum_{i=1}^{N}\sum_{j=1}^{T} a_{ij}^2 = \| A \|_F^2$ dan kvadrat tzv. Frobeniusove norme matrice A. Dakle, u matričnoj metodi najmanjih kvadrata dovoljno je minimizirati $tr(\varepsilon^\tau \varepsilon)$.

Uvrštenjem (22) u (20) i koristeći se svojstvima operatora *tr*, dobivamo:

$$\begin{aligned} V(\hat{F},\Lambda) &= \| X^\tau - \Lambda \hat{F}^\tau \|_F^2 \\ &= \| X^\tau - \Lambda\Lambda^\tau X^\tau \|_F^2 \\ &= tr((X - X\Lambda\Lambda^\tau)(X^\tau - \Lambda\Lambda^\tau X^\tau)) \\ &= tr(XX^\tau - X\Lambda\Lambda^\tau X^\tau - X\Lambda\Lambda^\tau X^\tau + X\Lambda(\Lambda^\tau\Lambda)\Lambda^\tau X^\tau) \\ &= \| X \|_F^2 - tr(\Lambda^\tau X^\tau X \Lambda). \end{aligned} \tag{23}$$

Kvadrat $\| X \|_F^2$ ne ovisi o Λ, pa će pogreška $V(\hat{F},\Lambda)$ dosegnuti minimum za onaj Λ za koji je $-tr(\Lambda^\tau X^\tau X\Lambda)$ najmanje ili, analogno, za koji je $tr(\Lambda^\tau X^\tau X\Lambda)$ najveće. No to je analogno problemu u kojemu tražimo linearne kombinacije komponenti od X (dakle, $X\Lambda$) maksimalne varijance (maksimalnog zbroja dijagonalnih elemenata kovarijacijske matrice) uz normiranje $\Lambda^\tau\Lambda = I$. Dakle, to je problem glavnih komponenti opisan u drugom poglavlju, u kojemu smo pokazali da se problem rješava tako da za $\hat{\Lambda}$ uzmemo prvih nekoliko svojstvenih vektora matrice $X^\tau X$ (dimenzija $N\times N$), pa iz (22) izračunamo faktore kao:

$$\hat{F} = X\hat{\Lambda}. \tag{24}$$

Slučaj 2. T<N

Slučaj je analogan prvome slučaju, no primjena jednakosti (18) i restrikcija $F^\tau F = I$ vodi izračunu svojstvenih vrijednosti matrice XX^τ (dimenzija $T\times T$).

Ovisno o dimenziji, biramo jedan od dva predočena slučaja.

Dodatak B

Popis serija korištenih u analizi

Cijene i tečaj	Eksterni sektor	Novac
Nafta (prosjek, brent, WTI, dubai fateh)	opIzvoz (fob)	Neto inozemna aktiva
IPC (ukupno)	opUvoz (cif)	Inozemna aktiva HNB-a (neto)
Prehrana i bezalkoholna pića	opSaldo	Inozemna aktiva poslovnih banaka
Alkoholna pića i duhan	stuIzvoz (fob)	Inozemna pasiva poslovnih banaka
Odjeća i obuća	stuUvoz (cif)	Neto domaća aktiva
Stanovanje, voda, energija, plin i druga goriva	stuSaldo	Plasmani središnjoj državi (neto)
Oprema za stan	otoIzvoz (fob)	Plasmani nebankovnom sektoru
Zdravstvo	otoUvoz (cif)	Plasmani poduzećima
Promet	cvIzvoz (fob)	Plasmani stanovništvu
Komunikacije	cvUvoz (fob)	Ostala aktiva (neto)
Rekreacija i kultura	cvSaldo	Ukupna likvidna sredstva
Obrazovanje	nndIvoz (fob)	Novčana masa
Ugostiteljske usluge	nndUvoz (fob)	Gotov novac izvan banaka
Ostali proizvodi i usluge	nndSaldo	Depozitni novac
Dobra	nnIzvoz (fob)	Depozitni novac poduzeća
Usluge	nnUvoz (cif)	Depozitni novac stanovništva
Ukupne, bez energije	izTotal	Kvazinovac
Ukupne, bez energije i prehrane	izEnergija	Kunski depoziti
Dobra bez električne energije	izIntermedijarni	Kunski depoziti poduzeća
Prehrana, piće i duhan	izKapitalni	Kunski depoziti stanovništva
Dobra bez prehrane, pića i duhana	izTrajni pr.	Devizni depoziti
Temeljna inflacija	izNetrajni pr.	Devizni depoziti poduzeća
HWWA indeks (USD) – total	izNeraspoređeno	Devizni depoziti stanovništva
HWWA indeks (USD) – total excl, energy	uvTotal	Obveznice i instrumenti tržišta novca
HWWA ind.Raw. materials	uvEnergija	Efektivna ukupna likvidna sredstva
HWWA crude oil	uvIntermedijarni	Efektivni plasmani nebankovnom sektoru
PPI	uvKapitalni	Efektivni devizni depoziti
HRK/EUR	uvTrajni pr.	Krediti stanovništvu
HRK/USD		Međunarodne pričuve (mil. EUR)
INET		Neto raspoložive međunarodne pričuve (mil. EUR)
IRET CPI		Primarni novac (MO)
IRET PPI		Ukupna obračunata obvezna pričuva
		Obračunata obvezna pričuva u kunama
		Izdvojena obvezna pričuva u kunama

Realni sektor	Tržište rada
Indeks industrije proizvodnje, ukupno	Registrirana nezaposlenost
Rudarstvo i vađenje	Novoprijavljeni u evidenciji
Prerađivačka industrija	Zaposleni iz evidencije
Opskrba električnom energijom, plinom i vodom	Brisani iz evidencije
Energija	Ukupna zaposlenost
Intermedijarni proizvodi	Zaposleni u pravnim osobama
Trajni za široku potrošnju	Zaposleni u pravnim osobama u javnoj upravi (LMN)
Netrajni za široku potrošnju	Zaposleni u pravnim osobama u industriji (CDE)
Indeks fizičkog obujma građevinskih radova – ukupno	Zaposleni u obrtu i slobodnim profesijama
Indeks fizičkog obujma građevinskih radova – zgrade	Aktivni osiguranici – individualni poljoprivrednici
Indeks fizičkog obujma građevinskih radova – ostale građevine	Aktivno stanovništvo
Indeksi novih narudžaba – ukupno	Administrativno stanje nezaposlenosti
Indeksi novih narudžaba – zgrade	Nominalna neto plaća
Indeksi novih narudžaba – ostale građevine	Realna neto plaća
Dolasci turista – ukupno	Nominalna bruto plaća
Dolasci turista – domaći	Realna bruto plaća
Dolasci turista – strani	Nominalna bruto plaća (LMN)
Noćenja turista – ukupno	Nominalna bruto plaća (CDE)
Noćenja turista – domaći	Realna neto plaća (LMN)
Trgovina	Realna neto plaća (CDE)
BDP	Nominalna neto plaća (LMN)
Investicije	Nominalna neto plaća (CDE)
Potrošnja	Realna neto plaća (LMN)
Uvoz	Realna neto plaća (CDE)
Izvoz	

LITERATURA

Angelini, E., Henry, J. and Mestre, R., 2001. "Diffusion Index-Based Infation Forecasts for the Euro Area". *European Central Bank Working Paper,* No. 61.

Artis, M. J., Banerjee, A. and Marcellino, M., 2005. "Factor forecasts for the UK". *Journal of Forecasting*, 24, 279-298.

Bai, J. and Ng, S., 2002. "Determining the number of factors in approximate factor models". *Econometrica,* 70, 191-221.

Banerjee, A., Marcellino M. and Masten, I., 2006. *Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change*. Mimeo.

Bernanke, B. and Boivin, J., 2003. Monetary Policy in a Data-Rich Environment. *Journal of Monetary Economics,* 50, 525-546.

Boivin, J. and Ng S., 2005. "Understanding and comparing factor-based forecasts". *International Journal of Central Banking*, (1), 117-151.

Camacho, M. and Sancho, I., 2003. "Spanish diffusion indexes". *Spanish Economic Review*, (5), 173-203.

Camba-Méndez, G. and Kapetanios, G., 2005. "Forecasting euro area inflation using dynamic factor measures of underlying inflation". *Journal of Forecasting*, 25, 491-503.

Chicago Fed Letter, 2000. *Forecasting inflation with a lot of data The Federal Reserve Bank Of Chicago*, March 2000, Number 151.

Eickmeier, S. and Ziegler, C., 2006. "How good are dynamic factor models at forecasting output and inflation? A meta-analytic approach". *Discussion Paper Series 1: Economic Studies*, 42.

Forni, M. [et al.], 2000. "The Generalized Dynamic Factor Model: Identification and Estimation". *The Review of Economics and Statistics*, 82 (4), 540-552.

Forni, M. and Reichlin, L., 1996. "Dynamic Common Factors in Large Cross-Sections". *Empirical Economics*, 21, 27-42.

Giannone, D., Reichlin, L. and Small, D., 2005. "Nowcasting GDP and inflation: the real-time informational content of macroeconomic data releases". *Finance and Economics Discussion Series,* No. 2005-42.

Giannoni, M. and Boivin, J., 2005. "DSGE Models in a Data-Rich Environment". *Computing in Economics and Finance,* No. 431.

Grenouilleau, D., 2006. "The Stacked Leading Indicators Dynamic Factor Model: A Sensitivity Analysis of Forecast Accuracy Using Bootstrapping, European Commission Directorate-General for Economic and Financial Affairs". *Economic Papers*, No 249.

Jankov, Lj. [et al.], 2007. *The Impact of USD/EUR Exchange Rate on Inflation in CEE Countries*. Dubrovnik: The Thirteenth Dubrovnik Economic Conference.

Johnson, R. A. and Wichern, D. W., 1998. *Applied Multivariate Statistical Analysis*. New York: Prentice Hall.

Kapetanios, G. and Marcellino, M., 2003. "A Parametric Estimation Method for Dynamic Factor Models of Large Dimensions, Queen Mary". *University of London Working Paper,* No 489.

Kapetanios, G., Labhard, V. and Price, S., 2007. "Forecast combination and the Bank of England's suite of statistical forecasting models". *Bank of England working papers,* No. 323.

Marcellino, M., Stock J. H. and Watson, M. W., 2005. "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-Steps Ahead". *Journal of Econometrics*, 26 (7), 527-549.

Matheson, T. D., 2006. "Factor model forecasts for New Zealand". *International Journal of Central Banking*, 2, 169-237.

Schumacher, C. and Breitung, J., 2006. "Real-time forecasting of GDP based on a large factor model with monthly and quarterly data". *Bundesbank Discussion Paper, Series 1*, No. 33.

Spearman, C., 1904. "General Intelligence, Objectively Determined and Measured". *American Journal of Psychology,* 15, 201-293.

Stock, J. H. and Watson, M. W., 1989. "New Indexes of Coincident and Leading Economic Indicators". *NBER Macroeconomics Annual*, 351-393.

Stock, J. H. and Watson, M. W., 1998. "Diffusion indexes". *NBER Working Paper,* No. 6702.

Stock, J. H. and Watson, M. W., 1999. "Forecasting Inflation". *Journal of Monetary Economics,* 44, 293-335.

Stock, J. H. and Watson, M. W., 2002. "Macroeconomic Forecasting Using Diffusion Indexes". *Journal of Business and Economic Statistics,* 20, 147-162.

D a v o r K u n o v a c
Factor Forecasting of Inflation in Croatia

Abstract

This paper tests whether information derived from 144 economic variables (represented by only a few constructed factors) can be used for the forecasting of consumer prices in Croatia. The results obtained show that the use of one factor enhances the precision of the benchmark model's ability to forecast inflation. The methodology used is sufficiently general to be able to be applied directly for the forecasting of other economic variables.

Key words: factor models, time series analysis, inflation, forecasting