

Draženka Čizmić\*

UDK 311.141:338.5

Pregledni članak

## PLANOVI UZORAKA PRI FORMIRANJU INDEKSA POTROŠAČKIH CIJENA I NJIHOV UTJECAJ NA GREŠKU UZORKA

*Indeks potrošačkih cijena u svojim statističkim sustavima imaju sve zemlje članice Europske unije, a služi kao pokazatelj stope inflacije i za usporedbu stabilnosti cijena među članicama. Budući da indeks troškova života po svojoj metodološkoj koncepciji bolje odgovara indeksu potrošačkih cijena, nego indeks cijena na malo, Državni zavod za statistiku Republike Hrvatske ima namjeru u godini 2004. pokazatelj, koji se sada naziva indeksom troškova života, transformirati metodološki i u nazivu u indeks potrošačkih cijena.*

*Kvaliteta indeksa potrošačkih cijena uvjetovana je kvalitetom podataka, i posebno reprezentativnošću uzorka prodavaonica kojim se koristi za snimanje cijena i izborom artikala. U izboru uzoraka praksa u Republici Hrvatskoj ne razlikuje se u većoj mjeri u odnosu na Europsku uniju, jer se u zemljama članicama EU također uglavnom koriste namjernim izborom, koji se približava nekoj vrsti "cut-off" odabira u kojem se veliki dio populacije (obično artikli s najnižim izdacima) ne promatra. Jedine zemlje u EU koje nastoje uvesti potpuni "probability" uzorak jesu Ujedinjeno Kraljevstvo i Švedska. Glavni je problem s "nonprobability" uzorkom to što ne omogućuje precizno procjenjivanje greške uzorka. Poznavanje greške uzorka nije značajno samo za korisnike, već također i za statističare. Optimizacija veličine uzorka jedino je moguća u slučaju "probability" uzorka. Zemlje koje optimiziraju alokaciju uzorka koriste se sljedećim kriterijima: varijancom indeksa*

---

\* D. Čizmić, dr. sc., viši asistent na Ekonomskom fakultetu Sveučilišta u Zagrebu. Članak primljen u uredništvu: 8. 10. 2002.

*cijena, troškovima skupljanja cijena, troškovima putovanja i ponderima izdataka.*

*Potpuni "probability" uzorak zasad se jedino primjenjuje u Sjedinjenim Američkim Državama. Ovaj plan može se smatrati krajnjim ciljem za svaki statistički ured. Najteži problem u odabiru potpunog "probability" uzorka za indekse potrošačkih cijena jest nedostatak prikladnih okvira izbora za artikle.*

## Uvod

Indeks potrošačkih cijena u svojim statističkim sustavima imaju sve zemlje članice Europske unije, a služi kao pokazatelj stope inflacije i za usporedbu stabilnosti cijena među članicama. Budući da indeks troškova života po svojoj metodološkoj koncepciji bolje odgovara indeksu potrošačkih cijena, nego indeks cijena na malo<sup>1</sup>, Državni zavod za statistiku Republike Hrvatske ima namjeru u godini 2004. pokazatelj, koji se sada naziva indeksom troškova života, transformirati metodološki i u nazivu u indeks potrošačkih cijena. Taj će indeks tada preuzeti ulogu pokazatelja porasta maloprodajnih cijena, tj. inflacije, tako da će izračunavanje indeksa cijena na malo u tu svrhu biti napušteno.

Budući da nije svrsishodno pribaviti cijene za sve transakcije, indeksi se procjenjuju iz uzorka. Da bi se uzorak mogao primijeniti, on mora biti reprezentativan. Do reprezentativnog uzorka dolazi se ispravnim izborom elemenata za uzorak. Procedura kojom se izabire uzorak jedinica iz populacije naziva se planom uzorka. Planove uzorka možemo podijeliti u dvije skupine.

Prvu skupinu čine planovi u kojima je izbor uzorka slučajan, uz poznatu vjerojatnost izbora elemenata populacije u uzorak ("probability" uzorci). Vjerojatnost izbora u tim planovima ne mora biti jednaka za svaki element, ali mora biti poznata. U tim se planovima mogu primijeniti zakoni vjerojatnosti, jer se oblik teorijske distribucije može identificirati kao oblik sampling distribucije nekog pokazatelja, uz pomoć koje se mogu kontrolirati i vrednovati sampling varijacije.

U drugoj su skupini planovi uzorka u kojima je izbor elemenata populacije za uzorak namjeran ("nonprobability" uzorci). Namjernim izborom teško je izabrati reprezentativan uzorak. Za namjeran izbor potrebno je dobro poznavati populaciju, zato što se jedinice osnovnog skupa za tu vrstu uzorka izabiru prema nahođenju istraživača. Prema tome, u namjernom izboru nije poznata vjerojatnost kojom će

---

<sup>1</sup> Indeks cijena na malo pokazuje opće kretanje cijena u prometu na malo tj. njihovu tendenciju. Indeks troškova života pokazuje koliko su porasle ili pale cijene proizvoda ili usluga, koje služe za osobnu potrošnju, tj. njihovo opće kretanje cijena.

neki element biti izabran za uzorak. Zbog toga se u tom slučaju ne mogu primijeniti načela teorije vjerojatnosti na osnovi kojih se može odrediti i kontrolirati greška učinjena pri procjeni karakteristika populacije.

### **Izbor uzorka geografskih područja i prodavaonica**

Za najveći broj artikala procjenjivanje indeksa cijena uključuje biranje uzorka prodavaonica. Metode izbora uzorka variraju od zemlje do zemlje. One ponajčešće ovise o raznolikosti cijena, o strukturi trgovine na malo i o trgovačkim pravilima koja su svojstvena svakoj zemlji. U većini zemalja članica Europske unije primjenjuje se dvoetajni uzorak. Prvo se bira uzorak geografskih područja, a nakon toga se u svakom izabranom geografskom području bira uzorak prodavaonica. Razmjer poslovne djelatnosti, bilo grubo procijenjen ili zasnovan na opaženim vrijednostima, faktor je koji se često uzima u obzir. Tako se vjerojatnostima izbora proporcionalnima broju stanovnika često koriste u prvoj etapi izbora. Uz to, prodavaonice se biraju prema njihovoj veličini, bilo na osnovi reprezentativnosti ili vjerojatnosti. U prvom se slučaju izbor prodavaonica zasniva na prosudbi stručnjaka ili alternativno na korištenju "cut-off" postupaka, npr. biraju se prodavaonice s najvećom prodajom.

Samo se četiri zemlje članice Europske unije koriste metodom vjerojatnosti za izbor prodavaonica. U Finskoj trgovine mješovitom robom biraju koristeći se uzorkom s vjerojatnošću proporcionalnom veličini bez ponavljanja ( $\pi$ ps uzorak), sa prometom kao mjerom veličine. Druge prodavaonice biraju skupljači cijena, a poduzeća za narudžbe poštom i diskonti biraju se "cut-off" metodom. U Nizozemskoj se uzorak prodavaonica do određene mjere zasniva na metodama vjerojatnosti. Točan plan uzorka ne može se utvrditi. Uzorak je poststratificiran prema tipu prodavaonice i prema veličini. Kada prodavaonica iz uzorka nestane, bira se nova prodavaonica slučajnim izborom s jednakim vjerojatnostima iz poslovnog registra. U Švedskoj se izbor prodavaonica uglavnom provodi metodama vjerojatnosti. Procedura izbora kojom se koriste na regionalnoj razini jest stratificirani  $\pi$ ps uzorak. Brojem zaposlenih koristi se kao mjerom veličine, a prodavaonice se stratificiraju prema tipu. Na centralnoj razini koriste se s nekoliko planova uzorka. U Ujedinjenom Kraljevstvu prodavaonice su podijeljene prema tipu tako da formiraju stratume. Iz svakog stratuma prodavaonice se biraju jednostavnim slučajnim izborom ili  $\pi$ ps izborom. Prodajnim prostorom koristi se kao mjerom veličine.

U Republici Hrvatskoj provodi se namjerni izbor. Kriteriji za izbor gradova jesu: broj stanovnika i udio i značaj u potrošnji. Izbor mjesta snimanja obavlja statistički ured koji snima cijene na osnovi vlastitog saznanja na terenu.

## Izbor uzorka artikala i vrsta

Uz iznimku Švedske, u zemljama Europske unije artikli se biraju u središnjim statističkim uredima, koristeći se kriterijima koji se ne zasnivaju na metodama vjerojatnosti, već na reprezentativnosti. Općenito uzevši, izbor je artikala dvoetapan. U prvoj se etapi podskupine artikala biraju koristeći "cut-off" metodu. Biraju se samo podskupine s najvećim tržišnim udjelima. U drugoj etapi biraju se specifični artikli iz svake skupine namjernim izborom. U nekim slučajevima središnji uredi određuju širu specifikaciju artikala. Onda, skupljači cijena obično izabiru najprodavaniju vrstu koja odgovara specifikaciji.

U Švedskoj izbor artikala provodi se "probability" metodama za dnevne potrepštine i namještaj<sup>2</sup>, a namjernim se izborom koriste za odjeću i skupine artikala koje pripadaju lokalnom sustavu cijena (isključujući namještaj). Za druge skupine artikala ili se uopće ne provodi izbor uzorka (obuhvaćeni su svi proizvodi) ili su postupci izbora uzorka mješavina "probability" i namjernog odabira (npr. transport i komunikacije). Od skupljača cijena traži se da pronađu vrstu koja se najbolje prodaje u danoj prodavaonici, a koja udovoljava specifikaciji.

U Republici Hrvatskoj koristi se namjernim izborom, tj. za snimanje cijena biraju se artikli s najvećim udjelom u prometu.

## Izbor uzorka kućanstava

Najveći broj istraživanja potrošnje kućanstava u Europskoj uniji zasniva se na odabiru uzorka u više etapa. Korištenje višestapnog uzorka skupina ima čitav niz prednosti. Koncentriranjem jedinica, umanjuju se troškovi putovanja i drugi troškovi skupljanja podataka. Zbog istih razloga, mogu se poboljšati obuhvat, kontrola, praćenje i drugi aspekti koji određuju stope odaziva i kvalitetu skupljenih podataka. Provođenje istraživanja praktičnije je. Odabir uzorka u nekoliko etapa može umanjiti troškove pripremanja i održavanja okvira izbora zato što je popise potrebno održavati samo za odabrana područja. Posao odabira uzorka može se također umanjiti.

Glavni je nedostatak opadanje reprezentativnosti uzorka zbog formiranja skupina. Složenost plana i analize također se povećava. Općenito, što je sakupljanje podataka po jedinici (kućanstvo, pojedinac) složenije i skuplje, što se više povećava potreba izbjegavanja gubitka reprezentativnosti. Zbog toga razloga veličine skupina

---

<sup>2</sup> Za dnevne potrepštine primjenjuje se postupak izbora s vjerojatnostima proporcionalnima veličini, a u slučaju namještaja primjenjuje se jednostavan slučajni uzorak.

kojima se koriste u istraživanju potrošnje kućanstava obično su male u usporedbi s onima kojima se koriste u mnogim drugim, manje složenim istraživanjima.

Najčešće se koristi dvoetafnim planom. Prvo se izabire stratificirani uzorak prikladnih područja, obično s vjerojatnostima proporcionalnima veličini nakon stratifikacije prema geografskim i drugim varijablama. Druga etapa obuhvaća odabir kućanstava u okviru područja.

Samo se Grčka i Portugal koriste troetafnim planom. Ti planovi uključuju dvije etape za izbor područja: prvo se odabiru veća područja (primarne jedinice izbora), nakon toga izabire se jedno ili više manjih područja, i na kraju se izabiru kućanstva.

Sa druge strane, Danska, Luksemburg, Austrija i Švedska koriste se jednoetafnim izborom uzorka kojim se izravno izabiru kućanstva. Isto se može reći i za Njemačku, premda je u njezinom slučaju situacija složenija. Uzorak se zasniva na kvotama koje se određuju prema posljednjim popisima, a stvarni se odabir provodi korištenjem lista iz različitih izvora.

Izravni izbor kućanstava u jednoj etapi za dio uzorka prilično je često obilježje planova. U najvećem broju slučajeva određena se područja automatski uključuju u uzorak, zaobilazeći na taj način prvu etapu odabira. Jednoetafnim odabirom također se koristi za dodatke glavnom višetafnom uzorku.<sup>3</sup>

U Republici Hrvatskoj uzorak je dvoetafni sa segmentima (prema podacima popisa stanovništva godine 1991.) kao primarnim i stanovima kao sekundarnim jedinicama izbora. Segmenti se biraju sustavno s vjerojatnošću proporcionalnom veličini. Stanovi se biraju jednostavnim slučajnim izborom.

### **Procjenjivanje indeksa cijena Laspeyresova tipa**

Premda je Laspeyresova formula izuzetno jednostavna, postupci procjene kojima se koriste za indekse potrošačkih cijena čine je prilično složenom, i to zato što procjena uključuje tri različite vrste uzoraka (uzorke kućanstava kojima se koriste za procjenjivanje pondera, uzorke prodavaonica i uzorke artikala).

Pretpostavimo da se skupina A sastoji od  $N$  artikala;  $g \in A$  znači da artikl  $g$  pripada skupini A. Pretpostavlja se da je skupina A stalna kroz vrijeme. Premda u

---

<sup>3</sup> Npr., u Irskoj slučajni uzorak poljoprivrednih kućanstava iz Nacionalnog istraživanja poljoprivrednih kućanstava dodaje se istraživanju potrošnje kućanstava; slično se u Ujedinjenom Kraljevstvu uzorak izabire u dvije etape za Veliku Britaniju, a izravno u jednoj etapi za Sjevernu Irsku. U Grčkoj se neke urbane regije uvijek uključuju u uzorak, i isto tako u Italiji 150 velikih komuna automatski uključuje u uzorak.

stvarnosti neki proizvodi mogu nestati sa tržišta i novi se pojaviti, pretpostavka stalne skupine omogućuje da se usredotoči na aspekte izbora uzorka. Indeks cijena Laspeyresova tipa s fiksnim ponderima za skupinu A u razdoblju t definira se kao<sup>4</sup>:

$$P^t = \frac{\sum_{g \in A} e_g^0 P_g^t}{\sum_{g \in A} e_g^0} = \sum_{g \in A} w_g^0 P_g^t \quad (1)$$

gdje  $P_g^t$  označuje indeks cijena artikla g (za koji se pretpostavlja da je dan),  $e_g^0$  je izdatak na g kroz bazno razdoblje 0 i  $w_g^t$  je odgovarajući udio izdataka na g unutar skupine A. U baznom razdoblju uzorak  $\hat{A}$  s fiksnom veličinom n uzima se iz A da bi se procijenilo  $P^t$ . Budući da se pretpostavlja da je A stalno u vremenu, isto vrijedi i za  $\hat{A}$ .

### Jednostavan slučajni uzorak

Kod ovog uzorka svi mogući uzorci imaju jednake vjerojatnosti izbora. Zbog toga je Horvitz-Thompsonov procjenitelj  $\hat{P}_A^t = (N/n) \sum_{g \in \hat{A}} w_g^0 P_g^t$ .

nepristran za  $P^t$ , jer je  $E(\hat{P}_A^t) = P^t$ , gdje očekivanje  $E(\cdot)$  označuje aritmetičku sredinu svih mogućih uzoraka danog plana. Usprkos njegovoj nepristranosti,  $\hat{P}_A^t$  ne koristi se u praksi zbog dvaju nepoželjnih svojstava. Ako bi indeksi cijena za sve artikle u uzorku bili jednaki, procijenjeni bi se indeks skupine razlikovao od te vrijednosti, osim ako bi se aritmetička sredina populacije i aritmetička sredina uzorka izdataka podudarale. Značajnije je to što  $\hat{P}_A^t$  pokazuje izuzetno veliku varijancu. Da bi se svladale obje poteškoće,  $P^t$  se procjenjuje, uzimajući nepristrane procjenitelje brojnika i nazivnika:

$$\hat{P}_B^t = \frac{(N/n) \sum_{g \in \hat{A}} e_g^0 P_g^t}{(N/n) \sum_{g \in \hat{A}} e_g^0} = \sum_{g \in \hat{A}} \hat{w}_g^0 P_g^t \quad (2)$$

gdje je  $\hat{w}_g^0$

udio izdatka na artikl g u uzorku. Koristeći se Taylorovom linearizacijom prvoga reda, varijanca od  $\hat{P}_B^t$

<sup>4</sup> Formule u poglavlju 5 preuzete su iz: Haan J., Opperdoes E., Schut C.; Item sampling in the Consumer Price Index: A Case Study Using Scanner Data; Joint ECE/ILO Meeting on Consumer

može se procijeniti iz podataka iz uzorka. Taylorova linearizacija, međutim dovodi do podcjenjenih varijanci za male uzorke. Varijanca od  $\hat{P}_B^t$  jednaka je:

$$V(\hat{P}_B^t) \approx V(\hat{P}_A^t) - (P^t)^2 (2\mu^t \rho^t - 1)(\sigma^0)^2 \quad (3)$$

gdje  $\sigma^0$

označuje koeficijent varijacije ili relativnu standardnu grešku aritmetičke sredine uzorka izdataka baznog razdoblja,  $\rho^t$

je omjer relativne standardne greške prosječnih izdataka baznog razdoblja izraženih u cijenama razdoblja t i 0, i  $\rho^t$

je koeficijent korelacije između prosječnih izdataka baznog razdoblja u cijenama od t i 0. Izbor  $\hat{P}_B^t$

umjesto  $\hat{P}_A^t$

može se objasniti činjenicom da se prvospomenuti koristi panel karakterom uzorka; sa  $\rho^t > 1/(2\mu^t)$ , pa se očekuje značajno smanjenje varijance. Alternativni izraz za varijancu od  $\hat{P}_B^t$  jest:

$$V(\hat{P}_B^t) \approx \frac{1-f}{n} \frac{N^2}{N-1} \sum_{g \in A} (w_g^0)^2 (P_g^t - P^t)^2 \quad (4)$$

koja se može procijeniti, koristeći se podacima iz uzorka uz uvjet da je poznata frakcija izbora  $f=n/N$ . Ta formula pokazuje da varijanca ovisi o disperziji unutar skupine. Iz toga proizlazi da se varijanca može umanjiti bilo formiranjem skupina koje se sastoje od artikala koji imaju slične cjenovne promjene, ili povećavanjem uzorka.

Procjenitelj  $\hat{P}_B^t$

podložan je maloj pristranosti uzorka od približno  $o(1/n)$ . S malim uzorkom artikala i s velikom varijabilnošću izdataka baznog razdoblja pristranost  $\hat{P}_B^t$

može biti nezanemariva u odnosu na njegovu standardnu grešku. Indeks koji obuhvaća sve artikle vjerojatno neće biti pristran u većoj mjeri, jer je pristranost vagani prosjek pozitivnih i negativnih pristranosti indeksa različitih skupina artikala.

## Uzorak s vjerojatnošću proporcionalnoj veličini

Ovaj uzorak ima tu prednost što najznačajniji artikli imaju najveću vjerojatnost da budu izabrani u uzorak. U slučaju biranja uzorka artikala s vjerojatnošću proporcionalno veličini najčešće se bira plan fiksne veličine bez ponavljanja u kombinaciji s Horvitz-Thompsonovim procjeniteljem. Takav se plan ponekada naziva "πps-sampling". Izdatak baznog razdoblja uzima se kao mjera veličine, pa tražena vjerojatnost uključivanja prvog reda za artikl  $g$  jest  $\cdot ne_g^0 / \sum_{g \in A} e_g^0 = nw_g^0$

$$\text{Tako je, } \sum_{g \in A} P_g^t / n$$

nepriistrani procjenitelj za  $P^t$ .

Budući da je najveći broj postojećih planova za biranje uzorka fiksne veličine s vjerojatnošću proporcionalnoj veličini prilično kompliciran, umjesto njih se koristi ssustavnom verzijom. Za skupine artikala s velikom varijabilnošću izdataka u baznom razdoblju, neće uvijek biti moguće izabrati uzorak strogo proporcionalan izdacima. Konflikt se rješava na sljedeći način. Prvo, podskupina  $A_H$  veličine  $N_{\hat{H}}$  ( $< n$ ) s najvišim izdacima u baznom razdoblju izabire se iz  $A$ . Zatim se uzorak  $A_L$  veličine  $n_L$

( $= n - N_H$ ) bira strogo proporcionalno izdacima iz ostatka podskupine  $A_L$  sa niskim izdacima. Rezultirajući nepriistrani procjenitelj jest vagani prosjek izdataka od  $P^t(H) = \sum_{g \in A_H} e_g^0 P_g^t / \sum_{g \in A_H} e_g^0$

$$\text{stvarnog indeksa cijena za } A_H, \text{ i } \sum_{g \in A_L} P_g^t / n_L$$

procijenjenog indeksa cijena za  $A_L$ .

## Stratificirani uzorak

Očita je prednost jednostavnog slučajnog uzorka u odnosu na uzorak s vjerojatnošću proporcionalnom veličini u tome što, osim popisa artikala koji služi kao okvir izbora, nisu potrebni drugi podaci. S veoma nejednoliko distribuiranim izdacima postoji velika vjerojatnost da tržišni lideri neće ući u uzorak. Umanjenje varijance može se postići ako postoji mogućnost da se skupina artikala stratificira



u homogene podskupine prema njihovim cjenovnim promjenama. No, s tim se podacima a priori ne raspolaže. Alternativno, varijanca se može umanjiti stratificiranjem skupine artikala u dvije podskupine, jedne ( $A_H$ ) s visokim izdacima u baznom razdoblju koja se promatra u cjelini i druge ( $A_L$ ) s niskim izdacima iz koje se bira slučajni uzorak  $\hat{A}_{L^t}$ . Novi procjenitelj indeksa za skupinu artikala jest vagani prosjek izdataka od  $\hat{P}_B^t(L)$ , koji je indeks cijena Laspeyresova tipa za podskupinu s nižim izdacima procijenjen u skladu s izrazom (2), i  $P^t(H)$ . Njegova je varijanca

$$(1 - \tau_H)^2 \text{Var} \left[ \hat{P}_B^t(L) \right]$$

gdje  $\tau_H$

označuje udio izdataka podskupine  $A_H$  u okviru skupine  $A$ . Dok ova metoda nužno ne umanjuje varijancu procijenjenog indeksa cijena, ona vjerojatno dovodi do toga kada se ukupna veličina uzorka povećava.

Izbor  $\tau_H$

i time veličine  $N_H$  stratuma  $A_H$  prilično je veliki problem. Bilo bi bolje kada bi se raspolagalo nekim kriterijem optimalnosti da bi se minimizirala varijanca. Ali budući da se ne raspolaže a priori poznavanjem cjenovnih promjena artikala i budući da prošli trendovi ne predviđaju točno buduće cjenovne promjene, optimalna veličina za  $A_H$  teško se može odrediti u praksi.

### “Cut-off” uzorak

Ako se  $A_L$  uopće ne promatra tako da je  $n=N_H$ , radi se o posebnom obliku “cut-off” uzorka. Indeks cijena skupine artikala procjenjuje se jednostavno sa.

$$\hat{P}_C^t = P^t(H) \quad \text{Svi } g \in A_H$$

sada imaju vjerojatnost uključivanja 1, a svi  $g \in A_L$

imaju vjerojatnost uključivanja 0. Budući da točno znamo koji će artikli biti odabrani, nije uključena randomizacija i varijanca za  $\hat{P}_C^t$

po definiciji je 0. Pristranost je jednaka stvarnoj grešci, tj. razlici između procijenjene vrijednosti i vrijednosti populacije:

$$\hat{P}_C^t - P^t = (1 - \tau_H) [P^t(H) - P^t(L)].$$

S veoma nejednolikom distribucijom izdataka, čak će i mala veličina uzorka dovesti će do velike vrijednosti  $\tau_H$

U tom se slučaju "cut-off" procjena može pokazati boljom od procjene stratifikacijom, uspoređujući prosječnu kvadratnu grešku. Može se ili fiksirati "cut-off" stopa  $\tau_H$  tako da je veličina uzorka određena sa  $\tau_H$ , ili fiksirati veličina uzorka u kojem slučaju  $\tau_H$  ovisi o izabranom  $n$ .

Korištenje "cut-off" postupaka može se opravdati sljedećim: 1) troškovi onemogućuju formiranje pouzdanog okvira izbora za čitavu populaciju i 2) pristranost se smatra zanemarivom. Naravno, pretpostavka 2) ne može se dokazati. Namjerno isključivanje dijela populacije iz uzorka može unatoč tome dati zadovoljavajuće rezultate kada se provedu prikkladne korekcije. No, "cut-off" postupci za odabir artikala nisu ispravni s obzirom na isključene artikle. Uz važnost troškova provođenja, ova se metoda ponekada brani pretpostavkom da se barem na dugi rok cjenovne promjene za manje značajne artikle neće mnogo razlikovati od onih za tržišne lidere u okviru iste skupine zbog slične strukture troškova proizvodnje.

### **Greška uzorka u indeksima cijena**

Jedan od glavnih ciljeva istraživanja na osnovi uzorka jest dobiti procjene karakteristika populacije. Takve procjene nikada nisu sasvim jednake karakteristikama populacije. Uvijek postoji određena greška. Ta greška može imati više uzroka.

Mogu se razlikovati dvije opće vrste: greške uzorka i neuzoračke greške. Greške uzorka nastaju zbog toga što se procjene zasnivaju na uzorcima, a ne na cjelokupnoj populaciji, a nestale bi ako bi se promatrala čitava populacija.

Greška uzorka može se raščlaniti na grešku odabira i grešku procjene. Greška odabira javlja se kada se koristi pogrešnim vjerojatnostima odabira. Vjerojatnost izbora elementa iz okvira izbora mora se podudarati s vjerojatnošću odabira koja je specificirana u planu uzorka. Greška procjene odnosi se na efekt koji uzrokuje korištenje uzorka koji se zasniva na proceduri slučajnoga odabira. Svaki novi odabir uzorka rezultirat će različitim odabirom elemenata, i odatle različitim vrijednostima procjena.

### **Prosječna kvadratna greška procjenitelja indeksa cijena**

Kvaliteta procjenitelja indeksa cijena može se razmatrati imajući u vidu točnost i relevantnost. Vrijednost indeksa  $\hat{P}$  koji se izračunava korištenjem podataka iz

uzorka, uzima se kao procjena indeksa cijena populacije  $P$  (egzaktni indeks). Točnost se odnosi na razliku između procijenjenog i egzaktnog indeksa. Ta se razlika može promatrati kao zbroj dviju komponenti: varijance procijenjenog indeksa i pristranosti procijenjenog indeksa. Egzaktni indeks samo je aproksimacija idealnog indeksa  $I$ . Relevantnost se odnosi na razliku između egzaktnog indeksa i idealnog indeksa. Simbolički, prosječna kvadratna greška procijenjenog indeksa u odnosu na idealni indeks može se izraziti kao<sup>5</sup>:

$$E(\hat{P} - I)^2 = E(\hat{P} - P)^2 + E(\hat{P} - E\hat{P})^2 + (P - I)^2 + 2(P - I)E(\hat{P} - P) \quad (5)$$

gdje je

$E(\hat{P} - P)^2$	kvadrat pristranosti procijenjenog indeksa prema egzaktnom indeksu
$E(\hat{P} - E\hat{P})^2$	varijanca procijenjenog indeksa
$(P - I)^2$	kvadrat razlike između egzaktnog indeksa i idealnog indeksa (relevantnost)
$2(P - I)E(\hat{P} - P)$	uzajamno djelovanje pristranosti i razlike između egzaktnog indeksa i idealnog indeksa

## Procjena varijance

Procjenitelj varijance ovisi o izabranom procjenitelju za indeks potrošačkih cijena i o planu uzorka. Samo četiri nacionalna statistička ureda u Europskoj uniji koriste se nekim oblicima “probability” metoda za odabir prodavaonica, a samo se statistički ured Švedske koristi “probability” metodama za odabir artikala. U odsutnosti “probability” metoda primjenjuju se tzv. metode prosudbe ili “cut-off” metode.

Zbog složenosti planova uzorka pri formiranju indeksa potrošačkih cijena, integralni je pristup procjenjivanju varijance problematičan. Što će reći, teško je prikazati uz pomoć jedne formule mjeru varijance indeksa u kojoj su obuhvaćeni svi izvori grešaka uzorka. Obično se razvijaju parcijalne (ili uvjetne) mjere, u kojima je kvantificiran samo efekt jednoga izvora varijabilnosti. Npr., Balk i Kersten (1986)<sup>6</sup> izračunali su varijancu za indeks potrošačkih cijena zbog varijabilnosti uzorka za

<sup>5</sup> vidjeti u Boon M., Opperdoes E., Schut C.; Effects of outlet sampling on consumer price indices: a case study using scanner data; Joint ECE/ILO Meeting on Consumer Price Indices; Geneva; 1997.; Internet, [www.unece.org/stats/documents/1997.11.ac.49.htm](http://www.unece.org/stats/documents/1997.11.ac.49.htm)

<sup>6</sup> Balk B.M., Kersten H.M.P.; On the precision of consumer price indices caused by sampling variability of budget surveys; Journal of Economic and Social Measurement 14, 1986.

istraživanja o potrošnji kućanstava. Te se varijance zasnivaju na pretpostavci da su parcijalni indeksi cijena poznati. Uvjetne greške uzorka objedinjene su u jedinstven sustav. S takvim se sustavom može procijeniti relativan značaj različitih izvora greške uzorka. Uz prilično ograničavajuće pretpostavke Balk (1989)<sup>7</sup> je izveo integralni sustav za grešku uzorka u indeksu potrošačkih cijena.

Postoje različiti tipovi postupaka za procjenjivanje varijanci uzorka za indekse potrošačkih cijena. Procjenitelji varijance koji se zasnivaju na planu uzorka (tj. varijance Horvitz-Thompsonovih procjenitelja) mogu biti korišteni u kombinaciji s postupcima Taylorove linearizacije za greške uzorka koje proizlaze iz “probability” planova. Npr., uz pretpostavku unakrsno klasificiranog plana uzorka, u kojem se uzorci artikala i prodavaonica biraju neovisno iz dvodimenzionalne populacije, s vjerojatnošću proporcionalnoj veličini u obje dimenzije može se izvesti formula za varijancu koja se zasniva na planu. Dalén i Öhlsson (1995)<sup>8</sup> ustanovili su da je greška uzorka za švedski indeks potrošačkih cijena iznosila 0.1 – 0.2%.

Glavni je problem “nonprobability” uzorcima to što ne postoji teoretski prihvatljiv način na koji bi se utvrdilo odražava li disperzija u uzorku točno disperziju u čitavoj populaciji. U tom je slučaju procjena standardnih grešaka veoma teška. U praksi, za “nonprobability” uzorke općenito se primjenjuju aproksimativne tehnike za procjenu varijance. Jedna je kvazi-randomizacija, tj. pretpostavlja se da je dani skup podataka generiran slučajnom procedurom premda “probability” odabir nije korišten (Särndal, Swensson, Wretman, 1992).<sup>9</sup> Pa je tako uzeta pretpostavka o vjerojatnosti izbora za svaki artikl i prodavaonicu. Problem je s tom metodom to što je teško pronaći model vjerojatnosti koji je adekvatna aproksimacija metodi kojom se koristi za odabir prodavaonica i artikala u praksi. Druga je tehnika metoda replikacije. To je potpuno neparametarska metoda procjenjivanja sampling distribucija i standardnih grešaka. Metoda replikacije provodi se biranjem velikog broja poduzoraka s ponavljanjem iz raspoloživog uzorka. Iz svakog poduzorka procjenjuje se parametar od interesa. Distribucija rezultirajućih procjena aproksimira stvarnu sampling distribuciju procjenitelja.

---

<sup>7</sup> Balk B.M., On calculating the precision of consumer price indices, Contributed Papers 47<sup>th</sup> Session of the ISI, Paris, 1989.

<sup>8</sup> Dalén J., Öhlsson E.; Variance estimation in the Swedish consumer price index; Journal of Business and Economic Statistics, 13(3), 1995.

<sup>9</sup> Särndal C.E., Swensson B., Wretman J.; Model Assisted Survey Sampling; Springer-Verlag, New York, 1992.

## Greške cjenovnih relativna

Uzmimo da su  $p_0$  i  $p_t$  cijene artikla u tekućem i baznom razdoblju. Pretpostavlja se da je izabran slučajni uzorak od  $n$  prodavaonica iz beskonačne populacije i da su cijene prilagođene za kvalitativne promjene. Aritmetička sredina i standardna devijacija cijena u baznom razdoblju označuje se sa  $\bar{p}_0$  i  $s_0$ , a u tekućem sa  $\bar{p}_t$  i  $s_t$ . Najbolji procjenitelj cjenovnog relativna<sup>10</sup> za artikl jest<sup>11</sup>:

$$R_{0t} = \bar{p}_t / \bar{p}_0 \quad (6)$$

Za veliki  $n$  (recimo  $n > 100$ ) sampling distribucija za približno je normalna sa sampling varijancom koja se može dati sa:

$$\text{var } R_{0t} = \frac{R_{0t}^2}{n} \left( \frac{s_0^2}{\bar{p}_0^2} + \frac{s_t^2}{\bar{p}_t^2} - 2\rho \frac{s_0 s_t}{\bar{p}_0 \bar{p}_t} \right) \quad (7)$$

gdje je  $\rho$  koeficijent korelacije između  $p_0$  i  $p_t$  u prodavaonicama. Skupljanje cijena daje sve informacije potrebne za ove izraze, osim  $\rho$ . Kada ne postoje druge informacije, uzima se da je  $\rho = 1/2$  kada se ne bi umanjila varijanca. Tada se 95% interval pouzdanosti za približno normalnu distribuciju za  $R_{0t}$  može napisati kao:

$$R_{0t} \pm 1.96 SE \quad \text{gdje je } SE^2 = \text{var } R_{0t} \quad (8)$$

## Greške pondera

Uzmimo  $n$  opservacija na varijabli  $x$  koje daju aritmetičku sredinu  $\bar{x}$  i standardnu devijaciju  $s_x$ , i na pridruženom ponderu  $e$  koje daju aritmetičku sredinu  $\bar{e}$  i standardnu devijaciju  $s_e$ . Pretpostavlja se da svaki  $x$  dolazi iz nezavisnog uzorka iz posebne populacije, ali sa zajedničkom varijancom,  $\text{var } e$ ; i da ne postoji korelacija između  $x$  i  $e$ .

Tada je sampling distribucija vagane aritmetičke sredine približno normalna za veliki  $n$  i u određenim će okolnostima imati varijancu:

<sup>10</sup> Alternativni procjenitelj, aritmetička sredina omjera uzoraka tj. sredina  $(p_t/p_0)$  u  $n$  prodavaonica, pokazuje pristranost i često nije dosljedan u smislu da ne teži cjenovnom relativu populacije kada se veličina uzorka povećava.

<sup>11</sup> Formule u poglavlju 6.2 preuzete su iz Allen R.G.D., *Index Numbers in Theory and Practice*, The Macmillan Press Ltd., London and Basingstoke, 1975.

var  $y = A \text{ var } x + B \text{ var } e$ , gdje je

$$A = \frac{1}{n} \left( 1 + \frac{s_x}{\bar{x}^2} \right) \left( 1 + \frac{s_e^2}{\bar{e}^2} \right) \quad \text{i} \quad B = \frac{1}{n} \frac{s_x^2}{\bar{x}^2} \left( 1 + \frac{s_e^2}{\bar{e}^2} \right) \quad (9)$$

A je veće od B u svim slučajevima, tako da greške u ponderima imaju manji utjecaj na  $y$ , nego greške u cjenovnim relativima. Presudan faktor u var  $y$  jest koeficijent varijacije cjenovnih relativna. Ako kretanja cijena pokazuju veliku disperziju od baznog prema tekućem razdoblju, koeficijent je velik i oba su člana u var  $y$  značajna. Ako kretanja cijena ne pokazuju veliku disperziju, tada koeficijent može biti dovoljno mali za B, tako da bude neznatan u odnosu na A. U ovim okolnostima greške u ponderima mogu se zanemariti.

No, to sve ne vrijedi ako postoji korelacija između pondera i cjenovnih relativna. To se može dogoditi ako postoje jedan ili dva dominantna pondera, ili češće, ako stavke s velikim ponderima pokazuju značajne cjenovne promjene.

Prije zanemarivanja grešaka u ponderima potrebno je provjeriti dvije činjenice. Prvo, nepostojenje značajne korelacije između kretanja pondera i cijena. Drugo, da disperzija cijena od baznog razdoblja nije velika.

### Postupci za umanjivanje greške

Greška procjene može se kontrolirati preko plana uzorka. Npr., povećavanjem veličine uzorka, ili uzimanjem vjerojatnosti odabira proporcionalnih s nekom dobro odabranom pomoćnom varijablom, greška se može umanjiti. Izbor primjerenog plana uzorka izrazito je složen zadatak.

Uz uvjet da je plan uzorka poznat, varijanca uzorka procijenjenog indeksa potrošačkih cijena može se općenito umanjiti na ove načine:

- povećavanjem uzorka kućanstava potrebnog za procjenjivanje udjela izdataka
- formiranjem homogenih artikala, tj. skupina vrsta koje imaju slične cjenovne promjene, ili povećavanjem uzorka prodavaonica po artiklu
- formiranjem skupina artikala koje se sastoje od artikala koji imaju slične cjenovne promjene, ili povećavanjem uzorka za artikl.

Važno je optimalno alocirati raspoloživa sredstva između i unutar različitih uzoraka, i to zato što loše alocirani uzorci mogu dovesti do nepotrebno visokih grešaka uzorka. Rezultati procjene varijance u Švedskoj (Dalén i Öhlsson, 1995)

pokazali su da je greška zbog uzorka artikala relativno visoka u usporedbi s onom zbog uzorka prodavaonica. Dakle, bilo bi potrebno povećati veličinu uzorka artikala i smanjiti veličinu uzorka prodavaonica. Na osnovu ovoga može se zaključiti da sustavno analiza grešaka uzorka nudi mogućnosti za povećanja preciznosti i/ili smanjenja troškova. Problem optimalne alokacije uzorka obično se formulira kao određivanje veličina uzoraka artikala i prodavaonica i njihove distribucije, tako da se greška uzorka sveopćeg indeksa svede na najmanju moguću mjeru u okviru raspoloživog budžeta.

Poslovni registar obično nije primjeren okvir izbora za prodavaonice, zato što posjeduje izrazito veći obuhvat. Preporučuje se da se formira prikladan okvir izbora sastavljanjem popisa glavnih prodavaonica unutar svake općine izabrane u uzorak. Taj bi popis sadržao i skupine artikala koje pripadaju njihovom asortimanu. Statistički ured Ujedinjenog Kraljevstva jedini je nacionalni statistički ured u EU koji formira okvir izbora popisivanjem prodavaonica u glavnim trgovačkim područjima unutar mjesta izabranog u uzorak. Manje je skup način organiziranja takvog istraživanja prodavaonica tražiti od skupljača cijena da naprave popis prodavaonica u kojima kućanstva kupuju, a to zato što oni obično veoma dobro poznaju situaciju.

Populacije artikala (i vrsta) i prodavaonica kontinuirano se mijenjaju kroz vrijeme. Sastav najvećeg broja skupina artikala nije konstantan kroz vrijeme zbog toga što neki artikli nestaju sa tržišta, a drugi se pojavljuju. Isto se događa i s prodavaonicama. Neke se prodavaonice zatvaraju (privremeno ili stalno), a druge se otvaraju. Značaj nekih tipova prodavaonica može se smanjiti ili povećati. Uzorci artikala (i vrsta) i prodavaonica moraju se periodično revidirati i potrebno ih je ažurirati da bi se sačuvala njihova reprezentativnost.

### **Efekti neuzoračkih grešaka na procjene greške uzorka**

Mnogi povremeni korisnici statistike interpretiraju procjene greške uzorka kao procjene ukupne greške za pokazatelj, što svakako nije točno. Postavlja se ovakvo pitanje: ako se koristi klasičnom formulom za izračunavanje greške uzorka, a podaci su također podložni neuzoračkim greškama, kako to utječe na procjenu greške uzorka? Odražavaju li procjene greške uzorka i neke neuzoračke greške?

U veoma restriktivnim okolnostima, procjene greške uzorka odražavat će neke neuzoračke greške. Najjednostavnije rečeno, to će se javiti: 1) samo za neuzoračke greške varijable, 2) samo ako mehanizmi koji uzrokuju neuzoračku grešku koincidiraju s definicijama jedinice uzorka i 3) samo za neuzoračke greške koje ne ovise o stvarnim jedinicama promatranja. Istraživanja obuhvaćaju mnoge postupke

koji se moraju provesti da bi se dobili konačni rezultati. Svi su oni potencijalni izvor neuzoračke greške. Greške proizlaze iz procesa istraživanja bez obzira skupljaju li se podaci iz čitave populacije ili samo iz uzorka. Najznačajnije kategorije neuzoračke greške jesu greška obuhvata, greška nedobijenih podataka, greška dobijenih podataka, greška obrade i greška procjene.

### ***Greška obuhvata***

Proizlazi iz izostavljanja dijela populacije (manji obuhvat) ili uključivanja jedinica izvan populacije (veći obuhvat), tako se, npr., izostavljaju gradovi, kućanstva, prodavaonice iz relevantnog okvira ili se uključuju u okvir kada to nije potrebno. Potencijalni izvor greške obuhvata jest vremenski razmak između koncipiranja istraživanja i skupljanja cijena za artikle. Zbog ovog se vremenskog razmaka artikli koje prodavaonice nude u vrijeme skupljanja cijena ne moraju poklapati s onima koje kupci kupuju.

### ***Greška nedobijenih podataka***

Nastaje kada podaci nisu sakupljeni za neke jedinice iz uzorka zbog propusta da se intervjuiraju kućanstva ili prodavaonice. To se javlja kada se odabrana kućanstva ili prodavaonice ne mogu kontaktirati ili odbijaju sudjelovati u istraživanju. Nedobijeni podaci mogu utjecati na pristranost indeksa ako se stope promjena cijena u jedinicama koje nisu dale odgovor razlikuju od stopa promjene u jedinicama koje su uključene.

### ***Greška dobivenih podataka***

Proizlazi iz skupljanja i korištenja u procjenjivanju pogrešnih, nepostojanih ili nepotpunih podataka. Nastaje zbog prikupljanja podataka od neprikladnih ispitanika, pamćenja, namjernog davanja pogrešnog odgovora, utjecaja anketara, pogrešnog bilježenja odgovora, prikupljanja cijena za pogrešne artikle, nesporazuma, pogrešne primjene postupaka skupljanja ili pogrešnog shvaćanja potreba istraživanja.



### ***Greška obrade***

Proizlazi iz pogrešnog editiranja, kodiranja i prenošenja podataka. Greške također mogu nastati zbog softverskih problema u računalskoj obradi koji uzrokuju gubitak korektnih podataka. Kompjutorsko provjeravanje i stručno pregledavanje podataka osigurava kontrolu točnosti obrade.

### ***Greška procjene***

Nastaje kada proces istraživanja ne mjeri točno ono što je bila namjera. Takve greške mogu biti konceptualne ili proceduralne po prirodi, a nastaju zbog pogrešnog razumijevanja ili pogrešne primjene pravila i postupaka. Supstitucija i prilagođivanje za kvalitativne promjene mogući su izvor greške procjene zbog proceduralnih poteškoća.

### **Korištenje skeniranih podataka u izračunavanju indeksa cijena**

Ne tako davno, empirijska su istraživanja indeksa cijena bila ograničena nepostojanjem potrebno raščlanjenih podataka o izdacima i količinama ili u najboljem slučaju raspoloživih u formi uzoraka malog opsega. Danas su neke od agencija za istraživanje tržišta uspjele formirati ogromne baze podataka, koristeći se skeniranim podacima.<sup>12</sup> U Europi artikli koji se mogu skenirati imaju jedinstveni linijski kod (European Article Number - EAN). Proizvođači dodjeljuju jedan i samo jedan EAN svakoj vrsti, veličini, tipu pakiranja, itd. za svaki artikl. To ima dvije posljedice: 1) neki se linijski kodovi mijenjaju veoma brzo, što otežava praćenje određenog proizvoda u vremenu, 2) neki linijski kodovi imaju beznačajne izdatke, pa se čini da je sustav klasifikacije previše potreban.

Postoji nekoliko načina da se poboljša indeks potrošačkih cijena koristeći se skeniranim podacima. Prvo, raspoloživost tekućih podataka o potrošnji omogućuje izračunavanje superlativnih indeksa, kojih su asimptotske vrijednosti bliže idealnom indeksu troškova života. Drugo, uzorci prodavaonica i artikala veoma su veliki, a to umanjuje grešku uzorka i nelinearnu pristranost. Treće, ti podaci pomažu da se identificiraju i ugrade u indeks novi proizvodi.

---

<sup>12</sup> Ovi podaci osiguravaju ažurne i točne informacije o: broju prodaja za vrstu artikla u izabranom razdoblju, ukupnu vrijednost tih prodaja i prosječnu transakcijsku "cijenu", popis individualnih karakteristika za promatranu vrstu i geografske i druge karakteristike prodavaonice.

Premda skenirani podaci imaju neke nedostatke, oni omogućuju provođenje empirijskih istraživanja različitih problema izbora uzorka. Istraživanja koja su provedena za izbor uzoraka artikala<sup>13</sup> pokazuju da se, unatoč svemu, za artikle koji su promatrani (kava, dječje pelene i toaletni papir) može preporučiti jednostavan slučajni uzorak. Vjeruje se da se ta preporuka može proširiti na sve skupine artikala kod kojih je distribucija izdataka veoma asimetrična. Ako statistički uredi žele primijeniti "probability" uzorak, postići će bolje rezultate koristeći se uzorakom s vjerojatnostima proporcionalnim izdacima. No, "cut-off" odabir može biti dobra (ili čak bolja) alternativa za one skupine artikala kod kojih cjenovne promjene nisu odveć nepostojeane. Agregiranjem skeniranih podataka može se ukazati na potrebnu "cut-off" stopu.

Takva su istraživanja provedena i za izbor uzoraka prodavaonica.<sup>14</sup> Ona su pokazala da ne postoji favorizirani plan uzorka koji daje najbolje rezultate (kao kriterij uzeta je prosječna kvadratna greška). Izbor ovisi o specifičnom artiklu. Također je promatrana kvaliteta procjenitelja varijance. Kod jednostavnog slučajnog uzorka procjenitelj varijance bio je približno točan, a kod sustavnog je  $\pi\pi\pi$  uzorka kvaliteta procjenitelja varijance manje dobra.

Procesiranje velikih baza skeniranih podataka posao je koji zahtijeva mnogo vremena, što će kočiti njihovu primjenu pri formiranju indeksa potrošačkih cijena u bližoj budućnosti.

### **Minimalni standardi za izbor uzorka u Europskoj uniji**

Statistička teorija predlaže korištenje slučajnog uzorka da bi se izbjegla pristranost kod pokazatelja. No, to nije jednostavno ostvariti, jer najveći broj zemalja primjenjuje namjerne ili reprezentativne postupke odabira, tako da odluka o tome koje će se cijene sakupljati ovisi o stupnju kooperativnosti prodavača ili o sklonostima pojedinih skupljača cijena. Ta praksa dovodi do neusporedivosti indeksa.

Samo nekoliko zemalja članica pokušava izračunati greške uzorka za njihove indekse potrošačkih cijena, a nijedna nema mjeru pristranosti. Istraživanja koja je naložio Eurostat pokazala su da je problem usporedivosti posebno naglašen kod izbora uzorka. Ona ukazuju da reprezentativan izbor u odnosu na "probability"

---

<sup>13</sup> Haan J., Opperdoes E., Schut C.; Item Sampling in the Consumer Price Index: A Case Study Using Scanner Data; Joint ECE/ILO Meeting on consumer price indices; Geneva; 1997.

<sup>14</sup> Boon M., Opperdoes E., Schut C.; Effects of outlet sampling on consumer price indices: a case study using scanner data; Joint ECE/ILO Meeting on consumer price indices; Geneva; 1997.

izbor može dati veće razlike za skupine artikala, premda one nisu značajne za prosjek. Ona su također pokazala da dio elementarnih agregata može imati kratkoročne efekte, a regionalne razlike ne predstavljaju problem.

Cilj članka 8. Uredbe Komisije 1749/96 jest poboljšati pouzdanost i usporedivost harmoniziranih indeksa potrošačkih cijena<sup>15</sup>, umanjujući greške koje proizlaze iz različitih planova uzorka i postupaka. On zahtijeva da zemlje članice provjeravaju uzorke cijena i do prilagode postupke izbora uzorka ako prosude da je to potrebno. Predlažu se istraživanja kojima bi se procijenila greška uzorka i pristranost indeksa u svrhu postavljanja prikladnih granica za takve greške.

Raširena primjena namjernog uzorka znači da ne postoji primjeren teoretski okvir kojim bi se prosudila pouzdanost (reprezentativnost i točnost) indeksa. Standardne greške izračunavaju se za neke indekse, a postoji nada da će se postupci kojima se koriste za formiranje indeksa prilagoditi da bi se dobio pokazatelj varijabilnosti kao potpora za efikasniji plan uzorka. Mnogo se toga još mora napraviti da bi se razumjela i kontrolirala pristranost. Prijeko je potrebno ustanoviti koja će kombinacija broja i specifikacija elementarnih agregata i broja cijena potrebnih za svaki elementarni agregat osigurati dovoljnu pouzdanost indeksa.

### **Plan uzorka za indeks potrošačkih cijena u SAD**

Plan uzorka za indeks potrošačkih cijena u SAD uključuje četveroetapni potpuni "probability" proces odabira. Takav uzorak luksuz je koji si najveći broj zemalja sebi ne može dopustiti.

Odabire se 88 geografskih primarnih jedinica izbora diljem cijele zemlje u kojima se skupljaju cijene. Dio jedinica izabire se sa sigurnošću, a ostatak se izabire s vjerojatnošću proporcionalnom veličini populacije. Svaka je jedinica područje ili skupina područja i može se kretati u rasponu populacije od nekoliko milijuna u velikim urbanim područjima do 20 000 u manjim urbanim područjima. Seoska su područja isključena.

Odabir prodavaonica zasniva se na tzv. "point-of-purchase" istraživanju koje se provodi za svaku odabranu jedinicu. To je istraživanje nastavak istraživanja potrošnje kućanstava kojim se utvrđuju prodavaonice u kojima su artikli kupljeni i iznos potrošen u svakoj prodavaonici za svaki artikl. Nakon toga, odabire se uzorak prodavaonica s vjerojatnošću proporcionalnom izdacima. Cijene se skupljaju u

---

<sup>15</sup> Svaka zemlja članica EU računa mjesečni harmonizirani indeks potrošačkih cijena. Ti indeksi moraju zadovoljiti zahtjeve koji se pred njih postavljaju. Svi se oni koriste istim nomenklaturama, koje su podijeljene u skupine i podskupine, i izvedene iz COICOP-a. Svi imaju isto područje, definirano iz COICOP-a kao podskup individualne potrošnje kućanstva.

približno 21 000 prodavaonica. Da bi indeks odražavao tekuće stanje, uzorak se prodavaonica svake godine rotira u 20% od primarnih jedinica izbora izabranih u uzorak.

Svi se izdaci iz istraživanja potrošnje kućanstava klasificiraju u isključive i iscrpne skupine artikala. Za svaku primarnu jedinicu izbora izabranu u uzorak artikli se odabiru za svaku skupinu artikala s vjerojatnostima proporcionalnim procijenjenim godišnjim izdacima kućanstva u danj primarnoj jedinici izbora. Tekući indeks obuhvaća 370 artikala. Uzorak artikala rotira se u oko 20% primarnih jedinica izbora izabranih u uzorak svake godine.

Unutar prodavaonica izabranih u uzorak, bira se jedna specifična vrsta za svaki artikl, koristeći se vjerojatnošću proporcionalnom prodajama artikla. Odabir vrste u svakoj prodavaonici može uključivati izbor u etapama. Za indeks koji obuhvaća sve artikle skuplja se 80 000 cijena svakog mjeseca.

Da bi se provele prilagodbe za nedobijene podatke, zabilježene se cijene preponderiraju, da bi se dobile procjene cjenovnih promjena artikla.

Broj cijena koji se skuplja za svaki artikl određuje se tako da dovodi do najmanje moguće greške uzorka za sveobuhvatni indeks u okviru ukupno raspoloživog budžeta. Ta alokacija zahtijeva pristup nelinearnog programiranja koji uzima u obzir relativnu važnost svakog artikla s obzirom na izdatke, jediničnu varijancu, komponente varijance, jedinične troškove skupljanja podataka i obrade. Artikli s većim udjelima izdataka zahtijevaju veću veličinu uzorka od artikala s manjim udjelima. Slično, artikli s većim jediničnim varijancama također zahtijevaju veću veličinu uzorka da bi se umanjio njihov doprinos varijanci sveobuhvatnog indeksa.

### **Novi pristupi ‘probability’ uzorku u SAD**

‘Probability’ uzorak izuzetno je skup u odnosu na namjerni uzorak. To je glavni razlog što najveći broj zemalja ne primjenjuje potpuni ‘probability’ uzorak. U američkom indeksu potrošačkih cijena broj geografskih područja izvodi se na osnovi broja stanovnika i veličine dohotka u urbanim područjima. No, ponovno biranje uzorka područja u cjelini nakon svakog popisa stanovništva izaziva velike dodatne troškove. Uvođenje novih područja kada je to potrebno i ažuriranje pondera za stanovništvo razboritije je rješenje.

Armknrecht (1997) smatra da bi biranje uzoraka prodavaonica i artikala iz izvora koji već postoje omogućilo postizanje ušteta. Provođenje redundantnih istraživanja i održavanje redundantnih fajlova čini statistički sustav neefikasnim. Takve podatke daju popisi stanovništva i baze podataka poduzeća, pa bi se statističkim uredima morao omogućiti pristup tim podacima u svrhe izbora uzorka.

Podacima popisa moguće je koristiti se za izbor uzoraka artikala i prodavaonica sličnih onima koji su trenutno raspoloživi (možda čak i bolji) za naciju, regije i gradove. Specifični artikli i vrste mogu se odabrati za nacionalnu komponentu (artikli tako uobičajeni da se mogu pronaći bilo gdje) i regionalne i lokalne komponente (dodatni artikli, gdje postoje regionalne i lokalne varijacije). Na lokalnoj razini, gdje su važne specifične lokalne vrste može se primijeniti odabir iz baza podataka poduzeća i drugih sekundarnih izvora. Kao dio ovoga procesa, potrebno je staviti naglasak na standardizaciju nacionalnih i regionalnih specifikacija za artikle da bi se omogućile prostorne usporedbe među gradovima i među državama.

## **Zaključak**

Mnogo je problema koji su samo djelomično riješeni u teoriji i praksi indeksa. Jedan od njih je utjecaj greške uzorka i drugih grešaka. Svaka procjena podložna je greškama i propustima koji se nikada neće moći u potpunosti izbjeći. Stoga je potrebno odbaciti koncept idealnog indeksa i težiti što točnijem indeksu. Kao i kod drugih statističkih pokazatelja, točnost indeksa izražava se njegovom varijancom.

Kvaliteta indeksa potrošačkih cijena uvjetovana je kvalitetom podataka, i posebno reprezentativnošću uzorka prodavaonica kojima se koristi za snimanje cijena i izborom artikala. U izboru uzoraka praksa u Republici Hrvatskoj ne razlikuje se u većoj mjeri od one u Europskoj uniji, jer se u zemljama članicama EU također uglavnom koristi namjernim izborom, koji se približava nekoj vrsti "cut-off" odabira u kojem se veliki dio populacije (obično artikli s najnižim izdacima) ne promatra. Jedine zemlje u EU koje nastoje uvesti potpuni "probability" uzorak jesu Ujedinjeno Kraljevstvo i Švedska. Glavni je problem s "nonprobability" uzorkom što ne omogućuje precizno procjenjivanje greške uzorka. Poznavanje greške uzorka nije značajno za korisnike već također i za statističare. Optimizacija veličine uzorka jedino je moguća u slučaju "probability" uzorka. Zemlje koje optimiziraju alokaciju uzorka koriste se sljedećim kriterijima: varijancom indeksa cijena, troškovima skupljanja cijena, troškovima putovanja i ponderima izdataka.

Potpuni "probability" uzorak zasad se jedino primjenjuje u Sjedinjenim Američkim Državama. Ovaj se plan može smatrati krajnjim ciljem za svaki statistički ured. Najteži problem u odabiru potpunog "probability" uzorka za indekse potrošačkih cijena jest nedostatak prikladnih okvira izbora za artikle.

Različiti planovi uzorka imaju za posljedicu različitu grešku. Izbor plana s najmanjom greškom osnovni je cilj. Kod namjernog izbora kojim se uglavnom koristi u RH varijancu je nemoguće procijeniti izravno iz uzorka. Jednom, kada se

bude koristilo "probability" uzorkom dobar će dio greške u izračunavanju indeksa doći pod kontrolu, pa će mjera preciznosti biti dana uz indeks.

## LITERATURA

1. Allard-Saulnier M., Report on the CPI Sample Diversification Project, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, 1999., Internet, [www.unece.org/stats/documents/1999.11.cpi.htm](http://www.unece.org/stats/documents/1999.11.cpi.htm)
2. Allen R.G.D., Index Numbers in Theory and Practice, The Macmillan Press Ltd., London and Basingstoke, 1975.
3. Armknecht P.A., Improving the Efficiency of the U.S. CPI, Third Meeting of the Ottawa Group on Price Indices, Voorburg, 1997., Internet, [www4.statcan.ca/secure/english/ottawagroup/meet.htm](http://www4.statcan.ca/secure/english/ottawagroup/meet.htm)
4. Astin J.A., Sellwood D.J.; Harmonization in the European Union: A Review of some Technical Issues; Third Meeting of the Ottawa Group on Price Indices; Voorburg; 1997; Internet, [www4.statcan.ca/secure/english/ottawagroup/meet.htm](http://www4.statcan.ca/secure/english/ottawagroup/meet.htm)
5. Balk B.M., On calculating the precision of consumer price indices, Contributed Papers 47th Session of the ISI, Paris, 1989.
6. Balk B.M., Kersten H.M.P.; On the precision of consumer price indices caused by the sampling variability of budget surveys; Journal of Economic and Social Measurement, 14; 1986.
7. Boon M., Quality assurance in a consumer price index: a conceptual framework, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, 1999., Internet, [www.unece.org/stats/documents/1999.11.cpi.htm](http://www.unece.org/stats/documents/1999.11.cpi.htm)
8. Boon M., Sampling and non-sampling errors in a consumer price index, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, 1999., Internet, [www.unece.org/stats/documents/1999.11.cpi.htm](http://www.unece.org/stats/documents/1999.11.cpi.htm)
9. Boon M., Sampling designs in constructing consumer price indices: current practices at statistical offices, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, 1997., Internet, [www.unece.org/stats/documents/1997.11.ac.49.htm](http://www.unece.org/stats/documents/1997.11.ac.49.htm)
10. Boon M., Opperdoes E., Schut C.; Effects of outlet sampling on consumer price indices: a case study using scanner data; Joint ECE/ILO Meeting on Consumer Price Indices; Geneva; 1997.; Internet, [www.unece.org/stats/documents/1997.11.ac.49.htm](http://www.unece.org/stats/documents/1997.11.ac.49.htm)

11. Bradley R., Cook B., Leaver S.G., Moulton B.R.; An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI; Third Meeting of the Ottawa Group on Price Indices; Voorburg; 1997.; Internet, [www4.statcan.ca/secure/english/ottawa\\_group/meet.htm](http://www4.statcan.ca/secure/english/ottawa_group/meet.htm)
12. Dalén J., Öhlsson E.; Variance estimation in the Swedish consumer price index; *Journal of Business and Economic Statistics*, 13(3), 1995.
13. Dondes-Connors I., Graf B.; Analysis and construction of price indexes; U.S. Department of Labor; Bureau of Labor Statistics; 1995.
14. Državni zavod za statistiku RH, Metodologija izračunavanja indeksa cijena na malo i indeksa troškova života, interni metodološki materijal
15. European Commission, Commission Regulation (EC) No 1749/96 on initial implementing measures for Council Regulation (EC) No 2494/95 concerning harmonized indices of consumer prices, 1996.
16. Fenwick D., The impact of sample design on the performance of the sample geometric mean and related issues, Fourth Meeting of the Ottawa Group on Price Indices, Washington, 1998., Internet, [www4.statcan.ca/secure/english/ottawagroup/meet.htm](http://www4.statcan.ca/secure/english/ottawagroup/meet.htm)
17. Fenwick D., Ball A.; Sampling in Consumer Price Indices: what role for scanner data?; Sixth Meeting of the Ottawa Group on Price Indices; Reykjavik; 1999., Internet, [www4.statcan.ca/secure/english/ottawagroup/meet.htm](http://www4.statcan.ca/secure/english/ottawagroup/meet.htm)
18. Fixler D., The Consumer Price Index: Underlying Concepts and Caveats, *Monthly Labor Review*, 12(December), 1993.
19. Groves R.M., Survey errors and survey costs, John Wiley & Sons, New York..., 1989.
20. Haan J., Opperdoes E., Schut C.; Item sampling in the Consumer Price Index: A Case Study Using Scanner Data; Joint ECE/ILO Meeting on Consumer Price Indices; Geneva; 1997.; Internet, [www.unece.org/stats/documents/1997.11.ac.49.htm](http://www.unece.org/stats/documents/1997.11.ac.49.htm)
21. Haan J., Item sampling in the CPI: The success of cut-off selection methods, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, 1999., Internet, [www.unece.org/stats/documents/1999.11.cpi.htm](http://www.unece.org/stats/documents/1999.11.cpi.htm)
22. Household budget surveys in the EU, Methodology and recommendations for harmonisation, Eurostat, Luxembourg, 1997.
23. Kish L., Survey Sampling, John Wiley & Sons, New York..., 1965.
24. Okamoto M., Empirical study of outlet sampling using scanner data, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, 1999., Internet, [www.unece.org/stats/documents/1999.11.cpi.htm](http://www.unece.org/stats/documents/1999.11.cpi.htm)

25. Porter-Hudak S., Hayes K.; A numerical methods approach to calculating cost-of-living indices; *Journal of Econometrics* 50; 1991.
26. Progress report on harmonisation of consumer price indices in the European Union, Eurostat, Luxembourg, 1996.
27. Remp J.M., *Community Price Indices*, Eurostat, Luxembourg, 1995.
28. Särndal C.E., Swensson B., Wretman J.; *Model Assisted Survey Sampling* (Springer-Verlag, New York)
29. Schmidt M.L., Effects of Updating the CPI Market Basket, *Monthly Labor Review*, 116 (December), 1993.
30. Sellwood D., Constraining unwarranted variations in sampling and quality adjustment practices in the construction of CPI elementary aggregates, Fourth Meeting of the Ottawa Group on Price Indices, Washington, 1998., Internet, [www4.statcan.ca/secure/english/ottawagroup/meet.htm](http://www4.statcan.ca/secure/english/ottawagroup/meet.htm)
31. Stoevska V., CPI methodologies in transition countries, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, 1999., Internet, [www.unece.org/stats/documents/1999.11.cpi.htm](http://www.unece.org/stats/documents/1999.11.cpi.htm)
32. Thompson S.K.; *Sampling*; John Wiley & Sons, Inc; New York...; 1992.
33. Tomohiko S., Outlet Sampling Method on the 1997 National Survey of Prices, Fourth Meeting of the Ottawa Group on Price Indices, Washington, 1998., Internet, [www4.statcan.ca/secure/english/ottawagroup/meet.htm](http://www4.statcan.ca/secure/english/ottawagroup/meet.htm)
34. Tryfos P.; *Sampling methods for applied research*; John Wiley & Sons, Inc; New York...; 1996.
35. Turvey R., Three kinds of monthly CPI, Second Meeting of the Ottawa Group on Price Indices, Stockholm, 1995., Internet, [www4.statcan.ca/secure/english/ottawagroup/meet.htm](http://www4.statcan.ca/secure/english/ottawagroup/meet.htm)



## SAMPLE DESIGNS IN CONSTRUCTION OF CONSUMER PRICE INDICES AND THEIR IMPACT ON SAMPLING ERROR

### Summary

All member countries of the European Union have a consumer price index in their statistical systems. It is used as inflation rate indicator and for price stability comparison among member countries. As cost of living index methodologically better corresponds to consumer price index than retail price index, the Croatian Bureau of Statics has intention in 2004 to transform methodologically the term into consumer price index.

The quality of consumer price index depends on the quality of data, and in particular on representation of both the sample of retail outlets used to monitor prices and the choice of items priced. As regards the sample choice practice in the Republic of Croatia it does not differ too much from the European Union practice. In member countries of the European Union are also mainly used judgmental sampling methods which are close to some kind of cut-off selection, in which a large part of population (usually the items with the lowest expenditures) is left unobserved. The only EU countries that appear to make serious efforts at full "probability" sampling are the United Kingdom and Sweden. The main problem with "non-probability" sampling is that it does not permit the precise statistical estimation of sampling errors. Knowing the sampling error is important not only for the users but for the statisticians as well. Optimization of the sample size is only possible in the case of a "probability" sample. Countries which optimize their sample allocation use the following criteria: price indices variance, cost of the price collection, travel costs and expenditure weights.

Full "probability" sample has been used only in the United States of America so far. This design can be considered as the ultimate goal for each statistical office. The most difficult problem in selecting full "probability" samples for CPI is the lack of appropriate sampling frames of items.