EESTI
STATISTIKA

Skannerandmete kasutamine tarbijahinnaindeksis

Metoodiline käsitlus ja aegridade ühendamine

Sissejuhatus

Tarbijahinnaindeksi (THI) peamine eesmärk on mõõta hinnataseme muutust ajas nii, et see kajastaks võimalikult täpselt leibkondade tegelikku hinnataju. Indeksi kasutusala ulatub makromajanduslikust analüüsist ja prognoosist indekseerimisotsuste ja poliitikameetmeteni. Seetõttu on THI puhul kriitilise tähtsusega nii mõõtmise täpsus kui ka aegridade järjepidevus ja ajaline võrreldavus.

Toiduainete ja mittealkohoolsete jookide hinna mõõtmine on viimastel aastatel muutunud eriti keerukaks. Jaekaubanduses on toodete arv kasvanud, sortiment vahetub kiiresti, kampaaniate osakaal on suur ning tarbijad reageerivad hinnamuutustele aktiivselt, valides kallimate toodete asemel odavamad. Traditsiooniline valimipõhine hinnakogumine ei ole enam piisav, et eespool nimetatud protsesse indeksis usaldusväärselt kajastada.

Skannerandmete kasutuselevõtt võimaldab hinnamõõtmist, mis põhineb tegelikel tehingutel koos müügimahtudega. Käesolev raport selgitab, kuidas kasutatakse skannerandmeid tarbijahinnaindeksis, millist indeksimetoodikat rakendatakse ning kuidas uue metoodikaga saadud andmed ühendatakse varasematega nii, et säilib inflatsiooninäitajate ajaline võrreldavus.

Alates 2026. aasta jaanuarist rakendab statistikaamet tarbijahinnaindeksi (THI) ja tarbijahindade harmoneeritud indeksi (THHI) koostamisel poekettide müügiandmeid ECOICOPi jaotises 01 ehk toiduainete ja mittealkohoolsete jookide puhul ning jaotises 02 ehk alkohoolsete jookide ja tubakatoodete puhul.

Selle ülevaate eesmärk ei ole valitud tehnilist lahendust ainult kirjeldada, vaid ka selgitada, millist probleemi metoodikamuudatusega lahendatakse, milliseid alternatiive kaaluti, millistele kriteeriumidele lahendus pidi vastama ning mis põhjustel valiti sobivaimaks meetodiks just skannerandmetel põhinev GEKS-Törnqvisti meetod koos aegridade ühendamisega. Eraldi käsitletakse ka seda, miks senist aegrida tagasiulatuvalt ümber ei arvutata.

Traditsioonilise metoodika järgi koguvad hinnaregistraatorid valitud toodete hinnad kauplustes käsitsi. Müügi- ehk skannerandmete meetod põhineb jaemüügi kassasüsteemidest pärineval detailsel tehingupõhisel infol kõigi müüdud toodete kohta. See võimaldab järgmist:

- hõlmata suuremat tootevalikut (tuhandeid tooteid valimi asemel);
- arvestada tegelikke tarbimismustreid ja müügimahte;
- kajastada toodete asendumist (nt kui tarbijad vahetavad kallimad tooted odavamate vastu);
- paremini käsitleda toodete pidevat turule tulekut ja sealt lahkumist.

Metoodikamuudatuse lähtekoht ja valikukriteeriumid

Skannerandmete kasutuselevõtt ei tähenda üksnes andmeallika ajakohastamist, vaid ka hinnakogumise kvaliteedi parandamist olukorras, kus traditsioonilise valimipõhise hinnakogumisega ei saa kiiresti muutuvas jaekaubanduskeskkonnas enam tegelikku hinnadünaamikat piisava täpsusega kajastada. Hindade täpsus on eriti oluline tooterühmades, kus sortiment uueneb kiiresti, kampaaniate osakaal on suur ning tarbijad kohandavad oma ostukäitumist vastavalt hinnamuutustele.

Metoodika valikul lähtuti neljast põhimõttest. Esiteks pidi uus meetod kajastama võimalikult hästi tegelikke tehinguid ja tarbimiskaale. Teiseks pidi see olema sobiv suure ja ajas kiiresti muutuva tootevaliku jaoks. Kolmandaks pidi lahendus olema kooskõlas THHI õigusraamistikuga ja sobima indeksite koostamiseks. Neljandaks peab uus meetod tagama avaldatud aegrea järjepidevuse, et metoodikamuudatus ise ei tekitaks inflatsiooninäitajates kunstlikku murdepunkti.

Skannerandmed ja hinnamõõtmise kontseptuaalne muutus

Skannerandmed põhinevad jaekettide kassasüsteemide tehinguandmetel ning hõlmavad triipkoodide põhjal koondatud hindu ja müügimahte. See tähendab, et hinnamõõtmise aluseks ei ole enam piiratud arv hinnavaatlusi, vaid kogu müügivoog. Indeksis lähevad arvesse sooduspakkumised, kampaaniad, uued ja müügilt kadunud tooted ning nende tegelik osatähtsus tarbijate kulutustes.

Majanduslikult on see suur muutus. Hinnamuutuse mõju tarbijale ei sõltu ainult sellest, kui palju hind muutub, vaid ka sellest, kui suur osa kulutustest langeb muutunud hinnaga tootele. Skannerandmed võimaldavad hinnamuutusi kaaluda tegelike kulutusosakaaludega ning kajastada tarbijate tooteasendusi ilma lisamodelleerimiseta.

Samal ajal toob selline andmestik kaasa metodoloogilised probleemid. Toodete arv on väga suur, sortiment muutub kiiresti ja ühe kindla baasperioodi valimine hinnavõrdluseks ei ole enam mõistlik. See eeldab teistsugust indeksit kui traditsiooniline Laspeyres'i tüüpi lähenemine.

Skannerandmed saadakse Eestis tegutsevatelt jaemüügikettidelt, mis katavad ligikaudu 60% Eesti toidukaupade jaemüügiturust. Andmed hõlmavad igakuiseid tehinguid ning võimaldavad hinnata toodete hindu ja müügimahte üksiktoote (EANi ehk triipkoodi) tasemel. Indeksi arvutamisse kaasatakse ainult ECOICOPi jaotistesse 01 ja 02 klassifitseeritud tooted.

GEKS-Törnqvisti indeksi põhimõtted

Hindade mõõtmisel kasutatakse multilateraalset GEKS-Törnqvisti indeksit, mis põhineb bilateraalsetel Törnqvisti indeksitel. Törnqvisti indeks on rahvusvaheliselt tunnustatud superlatiivne indeksi arvutamise meetod, mis arvestab toodete asendumist tarbijate ostukäitumises.

Multilateraalses lähenemises ei võrrelda vaatluskuu hindu ainult vahetult eelneva või ühe kindla baaskuuga, vaid korraga kõikide kuudega kindlas ajavahemikus. Indeks arvutatakse libiseva 14-kuulise ajavahemiku põhjal, kus iga kuu hinnasuhted määratakse kõigi teiste sama ajavahemiku kuude suhtes.

See erineb bilateraalsest aheldatud lähenemisest, kus indeks seotakse kuude kaupa järjestikku. Aheldamisega võivad aja jooksul tekkida väikesed juhuslikud hinnanihked või struktuursed muutused toodete valikus. Nihete kuhjumisel võib tekkida pikaajaline ühesuunaline ahelnihe (*chain drift*), mis ei kajasta tegelikku hinnadünaamikat.

Multilateraalne GEKS-meetod vähendab seda riski, kuna ükski kuu ei ole eelistatud baaskuu ja hinnasuhted määratakse sümmeetriliselt kogu ajavahemiku ulatuses. Selle tulemusena on indeksi taseme ja hinnamuutuste mõõtmine ajas stabiilsem ning vähem tundlik juhuslike või kumulatiivsete moonutuste suhtes.

Selle meetodiga saab:

- vähendada pikaajalisi ahelnihkeid võrreldes bilateraalsete aheldatud indeksitega (indeksi sidumisel eelneva perioodiga võib tekkida pikaajaline ühesuunaline ahelnihe (*chain drift*)). Multilateraalse meetodiga on selle mõju indeksile märksa väiksem;
- mõõta hinnamuutusi sümmeetriliselt ja stabiilselt;
- käsitleda toodete tavapärast turule tulekut ja sealt väljumist.

GEKS-indeksi oluline omadus on, et see määrab hinnasuhted, mitte absoluutse hinnataseme. Indeksi tase on defineeritud ainult kindla arvutusperioodi kontekstis. Kui periood nihkub ajas edasi, muutuvad korraga tootevalik, kaalud ja hinnasuhted ning kogu indeks skaleeritakse ümber.

Selle dünaamika katsetamiseks koostatud aknapõhised aegread kinnitavad seda. Sama kuu hinnadünaamika on eri akendes peaaegu identne, kuid indeksi tase erineb süstemaatiliselt. Tasemeerinevused ei tulene hinnamuutustest ega tarbijate käitumisest, vaid on multilateraalse arvutusmehhanismi paratamatu kaasnähtus.

Kui sellised aknapõhised indeksid lihtsalt järjestada, tekivad igal aknavahetusel väikesed, kuid korduvad tasemenihked. Makromajanduslikus analüüsis kanduksid need nihked inflatsioonimääradesse ja sealt edasi reaalnäitajatesse, kuigi tegelikult ei ole selliseid hinnamuutusi toimunud. Seetõttu tuleb multilateraalse indeksi puhul siduda kattuvad ajavahemikud ühte aegritta. Selleks kasutatakse kõigil kattuvatel perioodidel põhinevat indeksiridade ühendamist (*mean splice*), mille ühendustegur arvutatakse geomeetrilise keskmisena. Meetod põhineb juba avaldatud indeksitel ja tagab aegrea järjepidevuse.

Skannerandmete puhul kasutatakse Törnqvisti hinnaindeksit, mis kuulub nn superlatiivsete indeksite hulka. Superlatiivne indeks annab elukalliduse muutusele teist järku lähendi ning arvestab tarbijate tooteasendusi.

Perioodide *t* ja *s* vaheline Törnqvisti hinnaindeks, kus perioodi *t* võrreldakse perioodiga *s*, on defineeritud järgmiselt:

$$\ln P_{t,s}^{T} = \sum_{i} \frac{1}{2}\left(w_{i,t}+w_{i,s}\right)\ln\left(\frac{p_{i,t}}{p_{i,s}}\right)$$

kus $p_{i,t}$ on toote i hind perioodil t ning $w_{i,t}$on sama toote kulutusosakaal perioodil t.

Valemi majanduslik loogika on, et hinnamuutusi kaalutakse kahe perioodi keskmiste kulutusosakaaludega. Kui tarbijad valivad hinnatõusu korral odavamad tooted, väheneb kallima toote osakaal ja indeksi kasv pidurdub. Kui tarbimine koondub kallimatele toodetele, kajastub ka nende hinnamuutus indeksis tugevamalt. Seega ei eelda indeks fikseeritud tarbijakorvi püsimist, vaid kohandub tegeliku tarbimiskäitumisega.

Törnqvisti indeks on oma olemuselt kaheperioodiline. Skannerandmete puhul aga teeks ühe kindla baasperioodi valimine indeksi juhuslikuks ja ebastabiilseks, sest toodete valik ja tarbimismustrid

muutuvad pidevalt. Probleemi lahendamiseks kasutatakse GEKS-raamistikku (Gini–Eltetö–Köves–Szulc).

GEKS-meetod määrab perioodi t hinnataseme akna võrdlusperioodi suhtes, kasutades selleks kõiki aknasse kuuluvaid kaheperioodilisi Törnqvisti indekseid. Kui K tähistab kasutatavat arvutusakent, b akna võrdlusperioodi ning r kõiki aknasse kuuluvaid perioode, siis GEKS-Törnqvisti indeks perioodil t on järgmisel kujul:

$$G_t^K = \left(\prod_{r \in K} \frac{P_{r,b}^T}{P_{r,t}^T} \right)^{1/|K|}$$

See tähendab, et kuigi aknapõhine indeks esitatakse tehniliselt võrdlusperioodi b suhtes, ei põhine hinnavõrdlus ainult ühel baaskuul. Indeks kajastab akna kõigi kuude keskmist hinnasuhet ning on seetõttu stabiilsem ja vähem tundlik üksikute äärmuslike hinnaliikumiste suhtes. Praktikas kasutatakse libisevat 14-kuulist ajavahemikku, mis tagab piisava kattuvuse ja stabiilsuse.

Indeksiridade ühendamine

Indeksiridade ühendamine (*splicing*) on meetod, mille eesmärk on siduda eri arvutusakendes saadud indeksid üheks järjepidevaks aegreaks. Meetodiga saab eemaldada märkimisväärse osa arvutusakna vahetumisega kaasnevast tasemekõikumisest, säilitades samal ajal hinnamuutuste dünaamika. See ei ole eelneva hinnataseme korrigeerimine ega varasemate andmete revideerimine, vaid aegridade tehniline normaliseerimine.

Skannerandmete puhul võeti kasutusele kõigil kattuvatel perioodidel põhinev ühendamine (*mean splice*), mille aluseks on kahe järjestikuse arvutusakna kattuvad kuud. Olgu $I_t^{(k)}$ indeksi väärtus kuul t, arvutatud aknas k, ning O kahe akna kattuvate kuude hulk. Skaleerimistegur λ arvutatakse järgmiselt:

$$\lambda = \left(\prod_{s \in O} \frac{I_s^{(k)}}{I_s^{(k+1)}} \right)^{1/|O|}$$

See tegur on kattuvate kuude indeksite suhtarvude geomeetriline keskmine. Uue akna indeks korrutatakse seejärel teguriga λ, nii et see sobitub olemasolevasse aegritta.

Kõigil kattuvatel perioodidel põhineva ühendamise oluline omadus on, et see säilitab uue arvutusakna sisemised suhtelised hinnamuutused, kuid joondab kogu akna taseme kattuvate kuude põhjal varem avaldatud aegreaga. Inflatsioonidünaamika ei muutu, kuid eemaldatakse tehniline tasemenihe, mis muidu tekiks akna igakuisel vahetumisel. Seetõttu tuleb indeksiridade ühendamist käsitleda tehnilise ühildamisena, mitte sisulise ümberhindamisena.

Uue ja vana metoodika ühendamine

Skannerandmete kasutuselevõtt ei tähenda, et varasema hinnamõõtmise tulemused oleksid olnud valed. Varasem indeks kajastas tollal kättesaadavaid andmeid ja põhines parimal võimalikul metoodikal. Seetõttu aegrida tagasiulatuvalt ei korrigeerita.

Praktikas tähendab see, et alates kindlast hetkest hakatakse kasutama skannerandmetel põhinevat GEKS-Törnqvisti indeksit, kuid selle tulemused joondatakse varasema aegreaga nii, et

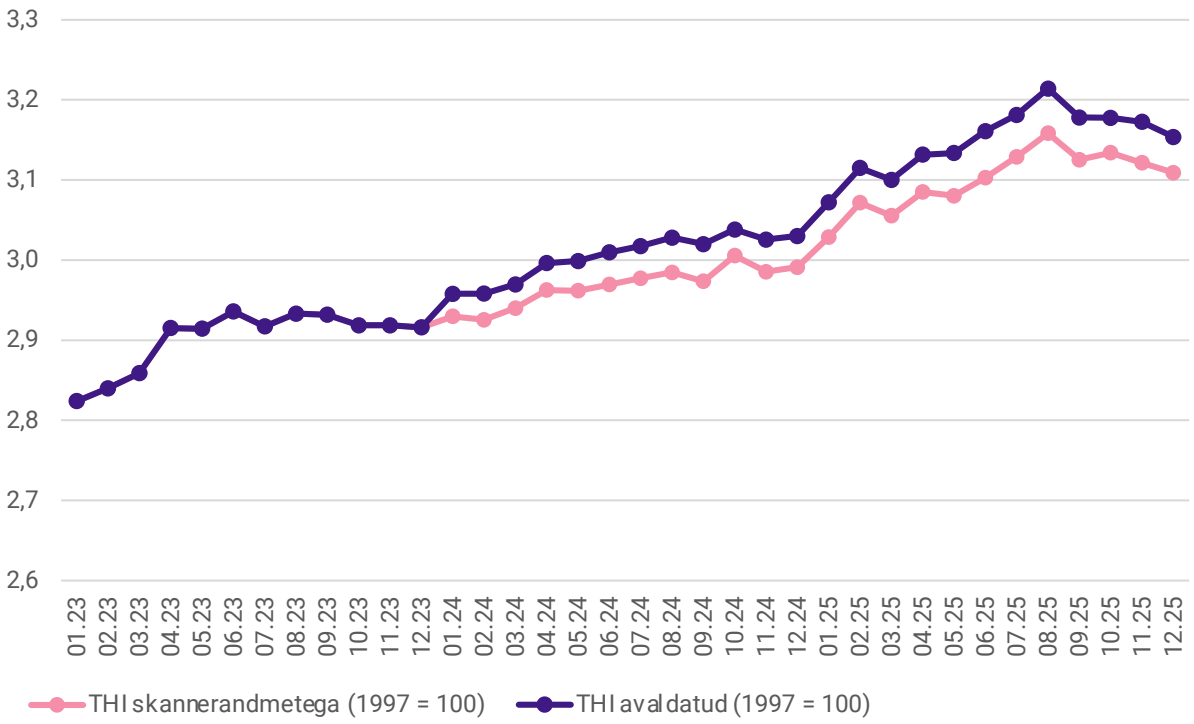
indeksi tase ei katke. Makromajanduslikus mõttes tähendab see, et ehkki edaspidised hinnamuutused põhinevad täpsemal ja tänapäevasemal metoodikal, siis varasemaid inflatsiooninäitajaid ei muudeta.

Üleminek skannerandmetel põhinevale metoodikale tehakse nii, et THI ja THHI aegridade järjepidevus ei katke. Selleks kasutatakse aegridade ühildamist ehk uue metoodika alusel arvutatud indeksid joondatakse varasema THHI tasemega.

Euroopa Parlamendi ja nõukogu määruse (EL) nr 2016/792 artiklis 4 sätestatud võrreldavusnõude ning seda täpsustavate rakendusmääruste kohaselt tuleb olulise metoodikamuudatuse korral vältida metoodikast tulenevat indeksi püsivat tasemenihet. Kui metoodikamuudatuse hinnanguline mõju THHI koguindeksi aastamuutusele ületab 0,1 protsendipunkti, rakendatakse indeksi korrektsioonifaktorit kõige detailsemal võimalikul tasemel, millel vana ja uut metoodikat saab omavahel ühendada.

Kui metoodikat oleks muudetud 2024. aasta jaanuaris ilma korrektsioonifaktorita (vt joonist), oleks metoodikast tulenev aastamuutuse erinevus olnud 0,3 protsendipunkti. Seega rakendatakse metoodika vahetamise kuust (jaanuar 2026) kuni kalendriaasta lõpuni indeksis ECOICOPi 5. taseme korrektsioonifaktorit[1]. Alates 2027. aasta jaanuarist korrektsioonifaktorit enam ei rakendata, kuna uue metoodika järgi koostatud indeks on selleks ajaks olemasoleva aegreaga täielikult ühendatud ja aegrida jätkub tavapäraselt.

Tarbijahinnaindeks (1997 = 100): metoodikavahetus ilma korrektsioonifaktorita, jaanuar 2023 – detsember 2025



[1] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A32024R1720

Alternatiivsed meetodid

Metoodikavalikul katsetati lisaks multilateraalsele GEKS-Törnqvistile ka bilateraalseid Törnqvisti indekseid, kasutades baasperioodina eelnevat kuud, vaatlusaastale eelneva aasta detsembrit ja kogu eelnevat aastat. Eesmärk oli hinnata, kas kaheperioodiline lähenemine võimaldaks skannerandmeid kasutada piisava täpsuse ja stabiilsusega ka ilma multilateraalse raamistikuta.

Katsetest ilmnes, et bilateraalsete indeksite rakendamisel tekkisid mitmed piirangud, mis tulenevad peamiselt tootevaliku kiiretest muutustest. Eelneva kuuga aheldamisel avaldus ahelnihe, mille korral kuude kaupa järjestikku seotud indeksisse kanduvad ajas väikesed struktuursed erinevused. Eelneva aasta detsembri või kogu eelneva aasta baasperioodina kasutamisel osutus peamiseks probleemiks see, et aasta edenedes otseselt võrreldavate toodete hulk vähenes. Selle põhjus on toodete turule tulek ja sealt kadumine ning hooajaliste toodete erinevus olenevalt aastaajast. Mida aasta lõpu poole, seda väiksemal toodete ühisosal võrdlus põhineb ning seda väiksem on ka indeksi representatiivsus ja stabiilsus.

Kaaluti ka akna keskperioodi kaudu ühendamist (*half splice*). Selle meetodi korral kasutatakse ühendusperioodina üht kindlat akna keskosas asuvat kuud, samal ajal kui kõigil kattuvatel perioodidel põhineval ühendamisel (*mean splice*) kasutatakse kõigi kattuvate kuude andmeid. Seetõttu oleneb keskperioodi kaudu ühendamisel saadud tulemus rohkem valitud sidumiskohast, kõigi kattuvate perioodide ühendamise tulemus aga hajutab mõju kattuva perioodi peale laiemalt. Käesoleval juhul peeti sobivamaks lahenduseks kõigil kattuvatel perioodidel põhinevat ühendamist (*mean splice*), kuna see meetod vähendab üksikust sidumisperioodist tulenevat volatiilsust ja kasutab ühendamisel kogu kattuva perioodi infot.

Kokkuvõte

Skannerandmete kasutuselevõtt tarbijahinnaindeksis on metoodiline muudatus, mille eesmärk on parandada hinnamõõtmise täpsust tooterühmades, kus traditsiooniline valimipõhine hinnakogumine ei kirjelda tegelikku hinnadünaamikat enam piisavalt hästi. Täpsus on eriti oluline kiiresti muutuva sortimendi ja suure kampaaniate osakaaluga kaubagruppides. Skannerandmetega saab kasutada tegelikke tehinguid koos müügimahtudega ja seetõttu kajastada hinnamuutusi tarbijate tegeliku ostukäitumise kontekstis.

Metoodika valikul lähtuti eesmärgist leida lahendus, mis oleks ühtaegu sisuliselt põhjendatud, indeksi koostamise protsessis rakendatav ja kooskõlas THHI nõuetega. Proovitud alternatiivide põhjal osutus sobivaimaks multilateraalne GEKS-Törnqvisti meetod. Bilateraalsed Törnqvisti indeksid osutusid skannerandmete puhul vähem sobivaks, kuna eelneva kuuga aheldamisel ilmnes ahelnihe ja pikema baasperioodi kasutamisel võrreldavate toodete hulk aasta jooksul vähenes. See puudutas eriti hooajalisi tooteid ja kiiresti muutuvat sortimenti, mille korral muutus võrdlusperioodide ühine tooteosa ajas järjest väiksemaks.

GEKS-Törnqvisti meetodi kasutamine võimaldab vähendada ahelnihet ning tagab hinnamuutuste sümmeetrilisema ja stabiilsema mõõtmise olukorras, kus tootevalik pidevalt muutub. Samal ajal vähendab multilateraalne lähenemine vajadust ühendada eri arvutusakendes saadud tulemused üheks järjepidevaks aegreaks. Selleks et eemaldada aknavahetusest tulenev tehniline tasemekõikumine ja tagada indeksi järjepidevus, kasutatakse kõigil kattuvatel perioodidel põhinevat ühendamist (*mean splice*).

Uue metoodika kasutuselevõtul ei arvutata aegrida tagasiulatuvalt ümber. Varasemad indeksi väärtused jäävad kehtima, kuna need põhinevad tolle aja andmetel ja tollal rakendatud parimal

võimalikul metoodikal. Alates 2026. aasta jaanuarist joondatakse skannerandmetel põhinev indeks olemasoleva aegreaga nii, et säilib THI ja THHI järjepidevus ning metoodikamuudatus ei tekita kunstlikke murdepunkte.

Kokkuvõttes võimaldab kirjeldatud metoodika parandada hinnamõõtmise kvaliteeti, säilitades samal ajal avaldatud statistika järjepidevuse ja ajalise võrreldavuse. Tegemist on lahendusega, mis ühendab detailsemad skannerandmed ja tänapäevasema metoodika hinnastatistika põhieesmärgiga mõõta hinnamuutusi võimalikult täpselt ja metoodiliselt järjepidevalt.