

Valtiontalouden tarkastusvirasto, finanssipolitiikan valvonta

Tarkastuskertomus 7/2018 Finanssipolitiikan valvonnan arvio valtiovarainministeriön makroennusteiden luotettavuudesta – Tarkastelussa BKT-kasvun, työttömyysasteen ja inflaation suhdanne-ennusteet vuosille 1976–2016

## Yhteenveto raporttiluonnoksesta annetuista lausunnoista

Valtiontalouden tarkastusvirasto on saanut viitekohdassa mainitusta raporttiluonnoksesta palautetta seuraavasti:

- Valtiovarainministeriö 19.3.2018 (VM/511/00.05.00/2018)
- Elinkeinoelämän tutkimuslaitos Etlä 22.3.2018
- Palkansaajien tutkimuslaitos PT 23.3.2018. (Lopullinen täydennetty lausunto 29.3.2018)

Lausuntojen määräaika oli 23.3.2018. Pellervon taloustutkimus (PTT) ei toimittanut lausuntoa.

Lausunnonantajilla on ollut tilaisuus

- lausua näkemyksensä raporttiluonnoksessa esitetyistä tarkastusviraston kannanotoista sekä
- esittää raporttiluonnoksen tekstiin korjaus- ja täsmennysehdotuksia.

Ennen lopullisen kertomuksen antamista tarkastusvirasto on kertomusta viimeisteltäessä ottanut huomioon palautteen seuraavasti:

### 1. Lausunnonantajien näkemykset kannanotoista

Palautteen antajien lausunnoissa on pääosin yhdytty raporttiluonnoksen kannanottoihin. Pääkannanotot olivat: ”Valtiovarainministeriön BKT:n, työttömyysasteen ja inflaation ennusteissa ei havaittu sellaisia piirteitä (vinoumia), jotka julkisen talouden suunnitelmaa (JTS) koskevan asetuksen mukaan edellyttäisivät toimenpiteitä asiantilan korjaamiseksi” sekä ”Valtiovarainministeriön ennusteiden luotettavuus on ollut vertailujoukon parhaimmista”. Lausunnot eivät ole antaneet aihetta muutoksiin lopullisen kertomuksen kannanotoissa.

### 2. Teknisluonteiset täsmennys- ja korjausehdotukset

Elinkeinoelämän tutkimuslaitos Etlä esittää lausunnossaan, että olisi itse esittänyt ensin osuvuusvertailun (ennustevirheiden) itseisarvojen keskiarvoa (tai mediaania) käyttäen ja edennyt sen jälkeen harhattomuutta mittaavan (ennustevirheiden) keskiarvoon. VTV toteaa, että puhtaasti matemaattisesti voisi ajatella tuotakin järjestystä. Sen sijaan tilastotieteessä tarkastellaan yleensä aina ensin otospohjaisen estimaattorin harhattomuutta (tässä otoksen 1976–2016 ennusteiden poikkeavuutta keskimäärin) populaatioarvosta tai todellisesta arvosta (tässä toteutuneesta tilastohavainnosta). Toiseksi tarkastellaan, onko estimaattori tarkentuva (tässä onko ennuste ollut tarkka verrattuna muihin). Kuten Etlä toteaa lausunnossaan, molemmat on selitetty raportissa huolellisesti, joten järjestyksen valinnalla ei voida katsoa olevan suurta merkitystä. Yllä esitetty syy on johtanut tässä raportissa valittuun järjestykseen.

Etlä esittää lisäksi lausunnossaan ”Selvyyden vuoksi raportissa olisi hyödyllistä tuoda eri kohdissa esille, mille periodille tarkastelu on kulloinkin kunkin laitoksen osalta tehty”. VTV toteaa, että tämä informaatio on tuotu näkyviin kunkin ennustajan omassa tuloslitteessä yksityiskohtaisten tulosten aluksi.

Palkansaajien tutkimuslaitos (PT) pohtii lausunnossaan kuvan täydentämistä raportissa käytettävien tilastollisten testien läpipääsyn helppoudesta. PT:n mukaan ”Raportissa suoritettuja, eri ennustelaitosten vertailuun käytettyjä testejä olisi ollut hyvä täydentää luomalla kvalitatiivinen kuva testeistä ”läpipääsemisen helppoudesta”. Tällainen kuva olisi voitu muodostaa suorittamalla (informaation kattavuuden testiä lukuun ottamatta) tarkastellut testit myös ns. ”naiiville ennusteelle” (jossa suurelle annetaan tarkasteluvuonna sama arvo kuin edellisenä vuonna).” VTV katsoo, että ehdotusta voidaan harkita myöhemmissä vastaavissa tarkasteluissa, tosin kuten PT toteaa, informaatiotehokkuuden testissä (testi 5) tämä ei toimi. VTV toteaa lisäksi, että testissä 4 testataan jo, onko kukin ennuste kattanut naiivin ennusteen informaation.

VTV haluaa täsmentää, että finanssipolitiikan valvonnan lakisäätöihin tehtäviin kuuluu virallisten, budjetin pohjana käytettävien ennusteiden luotettavuuden arviointi. Suomessa tämä tarkoittaa valtiovarainministeriön ennusteiden luotettavuuden arviointia. Alustavissa testauksissa nähtiin, että ministeriö läpäisee suuren osan luotettavuuden testeistä. Kuten arvioreportin luvussa 1.2 todetaan ”Alustavan tarkastelun jälkeen lopulliseen ennusteaineistoon poimittiin VM:n syksyn ennusteiden lisäksi sellaisten ennustajien seuraavan vuoden ( $t+1$ ) ja kuluvan vuoden ( $t+0$ ) ennusteet, joille Etlan tietokannasta löytyi riittävästi vuosiennusteita taaksepäin tämän arvion tilastollisten testien toteuttamiseksi.” Toisin sanoen VM:n ennusteiden luotettavuuden arvioissa käytettyjen tilastollisten testien läpipääsyn helppoutta on kuvattu VTV:n raportissa tekemällä samat testit niille neljälle Suomen talouden pääennustajista, joille Etlan tietokannasta löytyi testien edellyttämässä määrin riittävästi julkaistuja ennustehavaintoja taaksepäin.

PT pohtii edelleen lausunnossaan valittua BKT-tilaston ajankohtaa: ”Tarkasteltuja ennustearvoja vastaavina suureiden todellisina arvoina on raportissa käytetty kunkin tarkasteluvuoden kohdalla seuraavan vuoden heinäkuussa julkistettuja, ensimmäisen kerran tarkennettuja kansantalouden tilinpidon tietoja. Raportin mukaan ”heinäkuussa julkaistavan 1. kerran tarkennetun vuosi-BKT:n katsotaan kuvaavan riittävän tarkasti edellisen vuoden kehitystä”. Valintaa perustellaan myös esittämällä, että ko. julkistus on ”selvästi ennustehetkeä ja tällöin käytössä ollut informaatiota lähempänä” (s. 8). Valinnan perustelua on vaikea hyväksyä; kuuluuhan ennustelaitosten tehtäviin myös käytettävissä olevan informaation koherenssin ja uskottavuuden tarkastelu sekä Tilastokeskuksen julkistamien tietojen todennäköisimmän korjautumissuunnan arviointi. Myös näkemys, jonka mukaan Tilastokeskuksen ensimmäinen tarkastettu julkistus olisi riittävän tarkka, näyttää vaikeasti perusteltavalta. Esimerkiksi TK:n julkistama vuoden 2016 bruttokansantuotteen volyymikasvun arvo on raportissa käytetyn julkistuksen jälkeen kasvanut 0,2 prosenttiyksikköä. Tällaista tietojen myöhempää tarkentumista voidaan pitää – vertailtavien ennustelaitosten tasaväkisyyden huomioon ottaen – merkittävänä.”

VTV:llä ei sinänsä ole mitään sitä vastaan, että vuosien takaisen ennusteen vertailukohtana käytettäisiin vaikkapa viimeisimmän tilastojulkaisun mukaisia tietoja. Aineiston kokoamisen kannalta tämä olisi itse asiassa helpompaa. VTV on kuitenkin keskustellut ennusteiden vertailukohtana käytettävästä valittavasta tilastojulkistuksesta myös muiden ennustajien kanssa. Koska VTV on saanut näiltä näkemyksiä, jonka mukaan tilastoajankohdaksi tulisi valita viimeisimmän tilastojulkaisun sijaan lähempänä ennustejankohtaa oleva tilastojulkaisu, ennustajien ei voi sanoa olevan yksimielisiä asiasta. VTV on päättänyt tässä raportissa (samoin kuin vuoden 2016 tarkastuskertomuksessa) vuodesta 1997 eteenpäin käyttämään ennusteiden vertailukohtana Tilastokeskuksen ensimmäisen kerran vuosiaineistoihin perustuvaa BKT-tilastoa, joka on toistaiseksi julkaistu heinäkuussa edellistä vuotta koskien. (Vuosille 1976–1996 on käytetty aineiston kokoamisen aikaan elokuussa 2017 viimeistä julkaistua BKT-kasvun lukua.) Tilastokeskuksen web-sivuston mukaan vuosina 2006–2015 vuosi-BKT:n revisioiden – ensimmäisestä julkaisusta vuoden 2016 viimeisimpään julkaisuun – itseisarvojen keskiarvo on n. 0,6 prosenttiyksikköä. Samana aikana vuosina 2006–2015 tulevan vuoden ( $t+1$ ) BKT-kasvun kaikkien raportin viiden ennustajan ennustevirheiden itseisarvojen keskiarvo seuraavana vuonna heinäkuussa julkistusta BKT-kasvun tilastoluvusta oli 2,6 prosenttiyksikköä. Tilastojen tarkentuminen on toki yksi ennusteen osumista hankaloittava tekijä, mutta tästäkin yksinkertaisesta vertailusta päätelleen pääasiassa ennusteen osumista vaikeuttanevat muut tekijät.

PT toteaa lausunnossaan ”Työväen taloudellinen tutkimuslaitos julkisti vuosittaisen suhdanne-ennusteensa yleensä keväisin, ja raportin mukaan siinä käytetyistä TTT/PT:n ennusteista 17 on laadittu keväällä tai kesäkuussa. Tarkastelluista OECD:n ennusteista 34 on jo toukokuussa tai kesäkuussa laadittuja (s. 9). Vertailtavista ennusteista TTT/PT:n ja OECD:n ennusteet on siis laadittu muita ennusteita vähäisemmän ja epätarkemman informaation pohjalta.” VTV on tästä asiasta samaa mieltä. Juuri tästä syystä VTV on raportoinut luvussa 1.2 Tutkimusaineisto huolellisesti kunkin ennustajan kutakin vuotta koskeneen ennusteen julkaisuajankohdat, jonka perusteella tämä asia on tuotu esiin. Lisäksi VM:n tuloksia ja VM:n tilastollisten testien läpäisyä verrattaessa toisiinsa ennustajiin, raportissa muistutetaan usein OECD:n ja PT:n alkupään ennusteiden olevan jo keväällä tuotettu, mikä on todennäköisesti vaikuttanut näiden tuloksiin.

PT laajentaa lausunnossaan kuvaa edeltäjänsä Työväen Taloudellisen Tutkimuslaitoksen (TTT) ennuste-toiminnasta. ”PT:n ennusteella” tarkoitetaan vuosien 1976-1993 osalta syksyllä 1993 nimensä Palkansaajien tutkimuslaitokseksi muuttaneen Työväen taloudellisen tutkimuslaitoksen (TTT:n) suhdanne-ennustetta. ”Lisäksi TTT:n harjoittaman ennustetoiminnan tavoitteet olivat VM:n ja nykyisen PT:n ennusteiden tavoitteiden rinnalla vaatimattomia, eikä TTT ole aina edes pyrkinyt kilpailemaan ennustetarkkuudessa VM:n kanssa. Raportin graafisista esityksistä (s. 16) on helppoa havaita, että BKT:n kasvuennusteen ja työttömyysaste-ennusteen kohdalla TTT/PT:n menestys on ollut muihin ennustelaitoksiin verrattuna heikoimmillaan 1970- ja 1980-luvuilla, mutta TTT:n resurssit ja akateeminen tavoitteenasettelu olivat tuolloin nykyiseen PT:een verrattuna vaatimattomia. Tämä motivoisi vuosien 1976-2016 jakamista kahteen tai useampaan osaväliin ja vertailun suorittamista eri osaväleille erikseen. Tosin tällöin pienentynyt otoskoko saattaisi tehdä raportin tilastollisista testeistä käyttökelvottomia.” VTV ymmärtää tämän huolen TTT/PT:n ennusteiden osalta, varsinkin jos päätavoitteena olisi TTT/PT:n ennusteiden luotettavuuden arviointi (olkoonkin, että TTT on toki aikanaan päättänyt itse julkaista ennusteensa ja on ollut tietoinen, että sen ennusteita tullaan vertaamaan toisiin julkisiin ennusteisiin). Sen sijaan raportin varsinaisen kohteen, VM:n ennusteiden luotettavuuden verifiointissa tilastollisin testein, ei ole syytä lähteä pienentämään otosta ainakaan tämän raportin muuttujien kohdalla.

Euroopan komission omien ennusteidensa arvioinnin mukaisesti Valtiontalouden tarkastusvirasto (VTV) on tietoisesti valinnut tilastollisen testauksen arvioidessaan valtiovarainministeriön (VM) suhdanne-ennusteiden luotettavuutta. Tällä pyritään siihen, että luotettavuusarvion kannanotot perustuvat systemaattiseen tilastolliseen testaukseen eikä esimerkiksi subjektiivisiin näkemyksiin. Tilastolliset testit, kuten raportissa käytetty *t*-testi sekä toinen regressioparametrien testauksessa käytetty *F*-testi, edellyttävät riittävän suuren määrän havaintoja, jotta testin voima (power) olisi riittävä johtopäätösten tekemiseen. Mikäli havaintoja on vähän, testin voima (power) on alhainen, ja todennäköisyys H<sub>0</sub>-hypoteesin väärästä hylkäämisestä (tyypin 1 virhe) tai väärästä hyväksymisestä (tyypin 2 virhe) on selvästi suurentunut. Tämän vuoksi tässä tarkasteltavassa arvioreportissa on kerätty VM:n ennusteita vuodesta 1976 lähtien aina vuoteen 2016, yhteensä 41 kpl.<sup>1</sup> VM toteaaikin itse lausunnossaan ”Valtiovarainministeriön ennusteiden harhattomuutta on tutkittu usein, mutta VTV:n arvion kattavuus ylittää selvästi aikaisempien arvioiden kattavuuden käytettävän aineiston laajuuden sekä tilastollisten menetelmien soveltamisen osalta”.

PT:n lausunnon lisäksi teknisluonteista keskustelua käytiin sähköpostitse PT:n ja VTV:n asiantuntijoiden kesken. Tämä liittyi TTT/PT:n inflaatioennusteiden (*t+0*) karkeuteen jakson alkupäässä. PT:n asiantuntijat osoittivat, että TTT on ennustanut inflaatiota tarkastelujakson alkuvuosina vain yhden tai puolen prosentin tarkkuudella. Tämän pelättiin painottavan PT:n inflaation(*t+0*) keskimääräistä ennustevirhettä sekä ennustevirheen itseisarvojen keskiarvoa sekä vioittavan niiden luottamusvälejä kuvioissa

<sup>1</sup> Otoksoon varsin suuresta merkityksestä *t*-testiä käytettäessä tilastorevision (tai ennustevirheen) harhattomuuden testauksessa löytyy viimeaikainen esitys artikkelista Kokkinen, Arto ja Wouters, Hans (2016). EA and EU GDP flash estimates at 30 days, *Eurostat*, 1/2016, Annex B, s. 102–104, ks. erityisesti Figure 2 sivulla 104, haettavissa osoitteesta <http://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-GP-16-001>.

9 ja 10. Alkupään havaintojen poikkeavan suuret ennustevirheet PT:n inflaation ( $t+0$ ) ennusteissa näkyivät myös kuvion 8 ennustevirheissä siten, että alkupään ennustevirheiden vaihtelu (varianssi) oli selvästi suurempaa. Tämän keskustelun seurauksena VTV poisti PT:n inflaation ( $t+0$ ) ennusteista havainnot 1976–1981, minkä jälkeen ennustevirheiden vaihtelu tasaantui samaan kokoluokkaan koko loppujaksolla. Kaikki PT:n inflaation ( $t+0$ ) tilastolliset testit tehtiin uudelleen. Varmuuden vuoksi niissäkin tämän PT:n ennusteen tilastollisissa testeissä, joissa ei vielä oltu käytetty HAC-keskivirheitä, käytettiin nyt heteroskedastisuus-autokorrelaatio-konsistentteja (HAC) keskivirheitä.

Yllä esitetyn PT:n ja VTV:n asiantuntijoiden keskustelun seurauksena, VTV laski kuvioden 4 ja 5 sekä 9 ja 10 kaikkien ennustajien kaikkien muuttujien luottamusvälit vielä varmuuden vuoksi uudelleen käyttäen aina heteroskedastisuus-autokorrelaatio-konsistentteja (HAC) keskivirheitä. (Testeissä 3-5 vastaavia keskivirheitä olikin jo käytetty aina kun regression jäännös sisälsi piirteitä, jotka eivät kuulu riippumattomiin identtisen jakauman (i.i.d) jäännöksiin.)

Yllä kahden viimeisen kohdan teknisluontoiset keskustelut otettiin siten täysin huomioon lopullisessa arvioreportissa. Kahta viimeistä kohtaa edeltäneet teknisluontoiset kommentit eivät johtaneet yllä esitetyn perusteluihin muutoksiin lopullisessa arvioreportissa.

Lisätietoja: Vanhempi ekonomisti Arto Kokkinen, p. 09 432 5847