



BoF Online

5 • 2010

**Tasosiirtymät
taloudellisissa aikasarjoissa**

Heikki Hella

Tässä julkaisussa esitetyt mielipiteet ovat kirjoittajan omia eivätkä välttämättä edusta Suomen Pankin kantaa.



Suomen Pankki

Rahapolitiikka- ja tutkimusosasto

11.5.2010

Sisällys

1	Johdanto	3
2	Aikasarjojen perusominaisuudet ja kausitasoituksen pääperiaatteet	4
2.1	Makrotalouden suhdannesarjojen rakenne	4
2.2	Kausipuhdistusanalyysin pääperiaatteet	5
2.2.1	Historiaa lyhyesti	5
2.2.2	Kausipuhdistus tällä hetkellä	6
2.2.3	Mallipohjaisen kausipuhdistuksen pääperiaatteet	6
3	Tasosiirtymät Tilastokeskuksen julkaisemissa suhdannesarjoissa	8
4	Tilastokeskuksen tekemien ratkaisujen arviointi	11
4.1	Tasosiirtymätulkinnan haasteita ja näkökohtia	11
4.2	Suhdanneteoreettinen ”Macroeconomics”	11
4.3	Käytettävissä olevat vieras havainto -tyypit	11
4.4	Historiallisesti poikkeuksellinen tuotannon pudotus	12
4.5	Aikasarjojen esipuhdistuksen kovenevat vaatimukset	13
5	Tasosiirtymien vaikutus suhdanneaikasarjojen myöhempään kehitykseen	14
6	Yhteenveto ja johtopäätökset	16
	Kirjallisuutta ja internet-linkkejä	17
	Liite	20

Kuvioluettelo

Kuvio 1. BKT:n, viennin, teollisuustuotannon ja metalliteollisuuden tuotannon neljännesvuositainen aikasarja (2000Q1–2010Q1) sekä kausipuhdistetut arvot (sarjat perustuvat vuoden 2000 hintatasoon)	10
--	----

Taulukkoluetelo

Taulukko 1. ARIMA-mallit ja vieraat havainnot 2008Q4 ja 2009Q1, t-testiarvoineen	9
--	---

Kiitän Samu Kurria ja Veli-Matti Mattilaa (Suomen Pankki) monista hyödyllisistä kommenteista, korjauksista ja lisäyksistä sekä Faiz Alsuhaalia ja Samu Hakalaa (Tilastokeskus) saamastani informaatiosta koskien suhdannesarjojen kausipuhdistusprosessia. Mahdollisista virheistä vastaan luonnollisesti itse.
Heikki Hella

BoF Online

Päätoimittaja

Mika Pösö

ISSN

1796-9123 (online)

Postiosoite
PL 160
00101 HELSINKI

Käyntiosoite Snellmanin aukio
Puhelin 010 8311
Faksi (09) 174 872

Sähköposti
etunimi.sukunimi@bof.fi
www.suomenpankki.fi

Swift SPFB FI HH
Y-tunnus 0202248-1
Kotipaikka Helsinki

1 Johdanto

Selvityksessä tarkastellaan aikasarjoissa esiintyvien poikkeavien havaintojen (outlier) mallintamista suhdannesarjojen kausipuhdistusprosessissa. Tarkastelun erityiskohteena on Tilastokeskuksen kansantalouden neljännesvuositilinpäivityksessä vuoden 2008 viimeiseen ja 2009 ensimmäiseen kausipuhdistettuun havaintoon sovittamien tasomuutosten (tasosiirtymien) arviointi. Tasomuutosten seurauksena vuoden 2009 sisäinen kasvuprofiili muuttui merkittävästi.

Vieraiden havaintojen mallintaminen (modelling) on keskeinen osa aikasarjan esikäsittelyä ja -puhdistusta. Sen tavoitteena on turvata trendi/suhdanne- ja kausivaihtelukomponentin stokastisen estimaatin korkea tilastollinen laatu ja aikasarjan ennustetehokkuus. Tasosiirtymien ja muiden vieraiden havaintojen mallintamiseksi on tilastotieteen ja ekonometrian kirjallisuudessa esitetty useita vaihtoehtoisia laadintamenetelmiä ja niiden kehittämistyö jatkuu vilkkaana.

Tilastokeskus näyttää soveltaneen TRAMO/SEATS -kausipuhdistusohjelmaa asianmukaisesti myös tasosiirtymien mallintamisessa. Vaihtoehtoisista menettelytavoista voisi kuitenkin olla hyödyllistä kokeilla sarjojen ARIMA-mallintamisen (identifiointi, estimointi ja diagnostiset tarkistukset) päivityskertojen lisäämistä vuoden sisällä. Kausikomponentin ja sarjojen muiden komponenttien laatua saattaisi parantaa etenkin prosessin ennustamisosan tehostaminen. Vieraiden havaintojen mallintamisen yhteydessä niiden tulkinnan tarkastelu on tärkeä osa käyttäjäryhmien informointia. Toinen informoinnin väylä on internetissä julkaistava kausipuhdistusprosessin menetelmäkuvaus, jota voisi selkeyttää ja täsmentää. Kansainvälinen yhteistyö on alalla jatkunut intensiivisenä, minkä tuloksena uusi käsittelyohjelma on tulossa vielä kuluvana vuonna.

Selvityksessä tarkastellaan lyhyesti myös tilastotuotannon käytännön toimintaan liittyviä asioita. Tilastotuotannon yleinen laatuluokitus kehitettiin viime vuosikymmenellä. Nykyinen talouskriisi on paljastanut ongelmia ja puutteita koskien kausi- ja kalenteripuhdistuksen käytännön menettelyjä, suositusten luokitteluja ja sarjojen revisiointimenettelyjä. Näitä kaikkia tulisi edelleen yhdenmukaistaa ja selkeyttää koskemaan koko EU-alueen virallista tilastotuotantoa.

Asiasanat: tasosiirtymä, vieraat havainnot, kausipuhdistusprosessi, suhdannesarjan komponentit, robustit tilastolliset menetelmät

2 Aikasarjojen perusominaisuudet ja kausitasoituksen pääperiaatteet

2.1 Makrotalouden suhdannesarjojen rakenne

Kansantalouden suhdannesarjoissa, muuttujissa, erotetaan perinteisesti seuraavat neljä komponenttia:

- trendi eli pitkän ajan kehityssuunta (systemaattinen komponentti), (T)
- suhdannevaihtelu (cyclical), joka tapahtuu useamman vuoden syklinä, (C)
- kausivaihtelu, joka tapahtuu vuoden sisäisenä heilahteluna (seasonal fluctuations), (S) sekä
- epäsäännöllinen (irregular) komponentti, (I)

$$Y_t = T_t + C_t + S_t + I_t \quad \text{additiivisena} \quad (1)$$

ja

$$Y_t = T_t \times C_t \times S_t \times I_t \quad \text{multiplikatiivisena} \quad (1^*)$$

Epäsäännöllisen osaan luetaan aikasarjassa esiintyvät poikkeavat ilmiöt, kuten rakennekatkokset/siirtymät ja ns. vieraat havainnot (outlier).

Käytännössä nämä komponentit ovat suoraan havaitsemattomia (unobservable) stokastisia muuttujia, ja ne estimoidaan aikasarja-analyysin ja aikasarjaekonometrian menetelmin ja mallein suhdanteiden seurantaan, ennustamiseen, analysointiin ja tutkimusta varten. Komponenteista trendi/suhdannekomponentti mallinnetaan usein yhtenä kokonaisuutena. Näin tehdään esimerkiksi nykyisessä TRAMO/SEATS -kausipuhdistusmenetelmässä¹.

Kansantalouden tilinpidossa nämä erilaiset suhdannemuuttujat muodostavat systemaattisen järjestelmän, jossa on yksittäisten sarjojen ohella niistä summattuja eri tasojen aggregaattisarjoja. Esimerkkejä tilinpidon sarjoista ovat kokonaiskysyntä ja kokonaistarjonta, bkt, kulutus, investoinnit, tavaravienti ja tavaratuonti.

¹ TRAMO = "Time Series Regression with Arima Noise, Missing Observations, and Outliers"; SEATS = "Signal Extraction in Arima Time Series" (ks. Gómez and Maravall, 1996).

Tässä selvityksessä keskitytään suhdannesarjan kausikomponentin estimointiin tavoitteena ns. kausipuhdistettu sarja, eli alkuperäinen aikasarja miinus siitä poistettu kausikomponentti. Kausipuhdistettua sarjaa käytetään laajasti juuri lyhyen ajan suhdanneseurannan ja analyysin apuna. Tämän takia ekonomistit, tutkijat ja tiedotusvälineet ovat sen suurimmat käyttäjäryhmät. Tästä myös johtuu se, että kausipuhdistuksen tutkimusvälineet ja menetelmät ovat olleet tärkeä kvantitatiivisen taloudellisen analyysin kehittämiskohde lähes vuosisadan ajan. Nykyisenkaltaisen kausivaihteluanalyysin teoreettiset perusteet luotiin Yhdysvalloissa jo 1920-luvun lopulta alkaen.

2.2 Kausipuhdistusanalyysin pääperiaatteet

2.2.1 Historiaa lyhyesti

Kausipuhdistuksen teoreettinen ja empiirinen kehittämistyö alkoi Yhdysvalloissa 1920-luvun lopulla ja jatkui 1930-luvulla. Varsin nopeassa tahdissa se puettiin ajan kvantitatiivisten menetelmien soveltamisen myötä käytännön proseduuriksi. Korrelaatioanalyysin ja liukuvien keskiarvojen välineet muodostuivat tuolloin keskeisiksi. Alusta lähtien yleinen mielenkiinto suhdannesarjojen kausikäyttäytymistä kohtaan levisi tutkijapiirejä laajemmalle.

Kehitys jatkui kiivaana toisen maailmansodan jälkeen ja 1960–70-luvuilla U.S. Census Bureau toimi kehittämistyön keskuspaikkana. X-11 Census -menetelmäperhe ja myöhemmin (semiparametrinen) X-11 ARIMA² muodostuivat vähitellen legendaarisiksi käsittelymenetelmiksi tänä ei-parametrisiin liukuviin tasoitusfunktioihin perustuvien kausipuhdistusproseduurien aikakautena. Aikasarja-analyysin ja ekonometrian teorian ja tietokoneiden kehittymisen myötä kehitettiin yhä tehokkaampia laskentaproseduureja ja valmiudet ns. mallipohjaisten kausipuhdistusmenetelmien kehittämiseen olivat käytettävissä. Tämä tapahtui jo 1970-luvulla mutta pääosin 1980-luvulla, edelleen pääpaikkana U.S. Census Bureau, ja rinnalla näkyvästi myös Statistics Canada. Euroopassa kehittämistyön pääpaikka oli tuolloin Englannin keskuspankki.

Mallipohjaisen kausipuhdistuksen kehittämistyö eteni 1990-luvun alkupuolella, jolloin keskuspaikaksi muodostui Espanjan keskuspankki. Usean vuoden työn tuloksena muodostui nykyinen TRAMO/SEATS -mallipohjainen kausipuhdistusmenetelmä, jonka vastinparina oli Amerikassa (USA ja Kanada) yhteistyönä vuosien varrella kehitetty X-12-ARIMA – menetelmäpaketti. Uutta näissä menetelmissä oli erilaisten vieraiden havaintojen mallinnus

² Auto Regressive Integrated Moving Average time series model.

(testeineen) ja muut käsittely- ja diagnostiikan algoritmit, jotka olivat pääosin puuttuneet vanhemmista, ei-parametrisista proseduureista.

2.2.2 Kausipuhdistus tällä hetkellä

Tilastovirastojen käyttämistä mallipohjaisista ohjelmapaketeista kolme tärkeintä ovat TRAMO/SEATS, X-12-ARIMA ja STAMP³. Viimeksi mainittu on ns. rakenteellisten aikasarjamallien (STM) ohjelma. Demetra-käyttöliittymä sisältää kaksi ensin mainittua ohjelmaa. STAMP on tilavektorimalliin (sis. Kalman-suotimen) perustuva ohjelma.

Tällä hetkellä on loppuvaiheessaan usean vuoden kestänyt yhteistyöprojekti, jonka tavoitteena on uusi, entistä tehokkaampi kausi- ja kalenterivaihtelun analyysin menetelmäkokonaisuus. Uudistustyö keskittyy aikasarjan esipuhdistuksen (pre-treatment) tehostamiseen ja kausikomponentin mallin käyttökelpoisuuden tehokkaampaan testaamiseen ja diagnostisointiin⁴. Projektissa ovat mukana EKP, Eurostat, U.S. Census Bureau, Espanjan keskuspankki ja Belgian keskuspankki. Projektin tuloksena tulee olemaan Demetra+ - ohjelmapaketti, jonka ensimmäisen version odotetaan valmistuvan kuluvan vuoden loppupuoliskolla. Eurostat on lisäksi yhdessä amerikkalaisten kanssa kehittämässä menetelmällisesti vielä monipuolisempaa ja tehokkaampaa X-13-ARIMA-SEATS -ohjelmaa. Tämän ohjelman arvioitua valmistumisajankohtaa ei ole ilmoitettu. Siinä tulee olemaan mm. vieraita havaintoja koskevia diagnostiikkaosia testeineen (ks. Monsell, 2009b).

2.2.3 Mallipohjaisen kausipuhdistuksen pääperiaatteet

Aikasarjan kausipuhdistuksen⁵ (kausitasoituksen) prosessin pääosat ovat:

1. Sarjan esitarkastus ja -puhdistus, joka sisältää log-transformaation tarpeellisuuden testaamisen, alustavan ARIMA-mallin ja outlierien sekä rakennemurrosten etsinnän, estimoinnin ja testauksen, kalenteri- ja työpäiväkorjaukset, korjaukset koskien muita juhla-

³ Structural Time Series Analyzer, Modeller and Predictor.

⁴ Ohjelmassa, joka perustuu uusimpaan informaatioteknologiaan, tulee olemaan uutta graafista analyysiä, kausivaihtelun testejä ja spektrianalyysiin perustuvia mallien diagnostiikkatestejä sekä välineitä tutkia aikasarjan revisiohistoriaa. Ohjelman käyttäjällä on mahdollisuus valita kolmesta eri kausipuhdistusmenetelmästä: TRAMO/SEATS, X-12-ARIMA ja STAMP. Ohjelma tulee automaattisesti täyttämään Eurostatin vaatiman metatietolomakkeen (metadata template).

⁵ Termejä kausitasoitus ja kausipuhdistus käytetään kirjallisuudessa tarkoittamaan samaa asiaa. Koska kysymyksessä on viime vaiheessa stokastisen kausikomponentin poistamisoperaatio aikasarjasta, sana puhdistus lienee osuvampi.

päiviä ja kansallisia erikoisjuhlapäiviä. Vieraiden havaintojen (outlierien)⁶ tyypit ovat additiivinen vieras havainto eli AO, innovatiivinen vieras havainto eli IO, aikasarjan tasosiirtymää (step in mean level) ilmaiseva vieras havainto, LS, ja vaimenevan muutoksen vieras havainto, TC. Liitteessä on esitetty näiden matemaattiset mallit. Vieraiden havaintojen tilastollinen mallintaminen ja estimointi tehdään useiden iterointikierrosten avulla ja ARIMA-malli muodostetaan esipuhdistun aikasarjan perusteella (Tsay, 1988; Peña et al., 2001)

2. Kohdan 1. vaikutusten eliminointi ja korjausten vieminen aikasarjaan.
3. Esikäsitellyn aikasarjan komponenttien (ks. kaavat (1) tai (1*)) lopullinen ARIMA-mallinnus ja komponenttien *simultaaninen* estimointi (kanoninen dekomponointi). Saadut komponentit ovat yli ajan muuttuvia, stokastisia muuttujia.
4. Kausitasoitettu aikasarja saadaan vähentämällä alkuperäisestä sarjasta kohdassa 3 esitimoitu kausikomponentti. Kausitasoitettuun sarjaan jäävät muut osat, siis trendi/suhdannekomponentti ja epäsäännöllinen komponentti.
5. Trendi/suhdannekomponentti jää jäljelle, kun sarjasta vielä vähennetään kohdassa 3 esitimoitu epäsäännöllinen komponentti.

Operaatiot tehdään tässä järjestyksessä ja keskeisenä ideana on, että kohdassa 3 voidaan mallintaa mahdollisimman tehokas ARIMA-malli. *Huomattakoon, että kohdan 4 muodostamisessa esim. outlierien lopulliset estimaatit on lisätty takaisin kohdan 3 esipuhdistettuun aikasarjaan*⁷. Käytännön kysymys on tällöin, miten ”hyvin” aikasarjan ARIMA-malli (kohdat 1 ja 3) ja siitä riippuvat vieras havainto -mallit vaikutuksineen on esitimoitu ja miten laadukkaita estimaattiarvoja on silloin käytettävissä. Mallinnustulosten laatu on myös ratkaiseva sarjan ennusteiden ja suhdanneseurannassa käännepeisteiden arvioinnin kannalta.

Outlierien esiintyminen suosittaa vallitsevan teoreettisen tietämyksen perusteella ns. robustien mallintamismenetelmien soveltamista koko iteratiivisen mallintamisketjun (identifiointi/spesifiointi, parametrien estimointi ja diagnostiset tarkistukset) osalta. Robustien menetelmien käyttö yleensä vähentää parametriestimaattien harhaisuutta, mikä puolestaan terävöittää vieraiden havaintojen erottumista ARIMA-mallien residuaalisarjoissa. Tehokas vieras

⁶ Outlieria on syytä kutsua poikkeavaksi tai vieraaksi (outlying, aberrant) havainnoksi eikä äärihavainnoksi (extreme observations). Vieraan havainnon ei välttämättä tarvitse olla äärihavainto eikä äärihavainnon tarvitse välttämättä olla vieras havainto. Tilastotieteessä ja matematiikassa on erikseen muuttujien ääriarvohavaintoja käyttävä mallinnusala (Extreme Value Modelling), jolla on oma metodologia, sovellutusohjelmat ja soveltamisalat (mm. riskiteoria).

⁷ Tällöin AO ja TC lisätään epäsäännölliseen komponenttiin (I) ja LS lisätään trendikomponenttiin (T).

havainto -mallinnus johtaa yleensä sarjan ennusteiden keskivirheiden (MSE) pienenemiseen (esim. Chen and Tiao, 1990).

3 Tasosiirtymät Tilastokeskuksen julkaisemissa suhdannesarjoissa

Tasosiirtymien esiintyminen makrotalouden sekä reaali- että rahatalouden aikasarjoissa ei ole mitenkään harvinaista (Balke & Fomby, 1991 ja 1994; Maravall, 2007). Harvinaista on se, että nykyisen talouskriisin aiheuttamat tasosiirtymät tapahtuvat eri maissa niin samanaikaisesti ja ovat lisäksi poikkeuksellisen jyrkkiä.

Tässä selvityksessä tarkastellaan tasosiirtymää aikasarja-analyysin yhtenä poikkeavien (vieraiden) havaintojen, outlierien tyyppinä. Poikkeavien havaintojen mallintaminen kuuluu kunkin alkuperäisen sarjan esikäsittelyyn ennen kausikomponentin ja sarjan muiden teoreettisten komponenttien mallintamista. Sovellutus koskee Tilastokeskuksen 8.12.2009 - tiedotteessa mainittujen Suomen suhdannesarjojen (ks. liite) tasosiirtymiä. Sarjoissa on eriasteisia aggregaatteja. Kysymyksessä on kansantalouden tuotannon vapaan pudotuksen tilastollinen kuvaus, jossa poikkeavan havainnon mallintamisella on turvattu aikasarjan kausi- ja trendikomponentin ja niiden ennusteiden laadun pysyminen. Jos käytettävissä olisi ollut vain vanhakantainen, ei-parametrinen, liukuviin keskiarvoihin perustuva kausipuhdistusproseduuri, näin mittava talousromahdus olisi aiheuttanut esipuhdistuksen ja siten trendi/suhdanne- ja kausiestimoinnin epäonnistumisen lähes täysin. Tämä olisi luonnollisesti vaikuttanut suhdannesarjojen aikasarjaennusteiden laatuun.

Seuraavassa taulukossa (Taulukko 1) on esitetty muutaman keskeisen suhdannesarjan kausipuhdistusprosessin tuloksia koskien vuosineljänneksiä 2008Q4 ja 2009Q1. Taulukosta nähdään sarjan puhdistuksen lopullinen malli, estimoidun vieraan havainnon tyyppi ja sen tilastollinen merkitsevyys. Kaikki vieraat havainnot, LS, ovat t-testin perusteella tilastollisesti erittäin merkitseviä.

Taulukko 1. ARIMA-mallit ja vieraat havainnot 2008Q4 ja 2009Q1, t-testiarvoineen

Kerrannainen kausivaihtelumalli $ARIMA(p,d,q) \times (P,D,Q)_4$

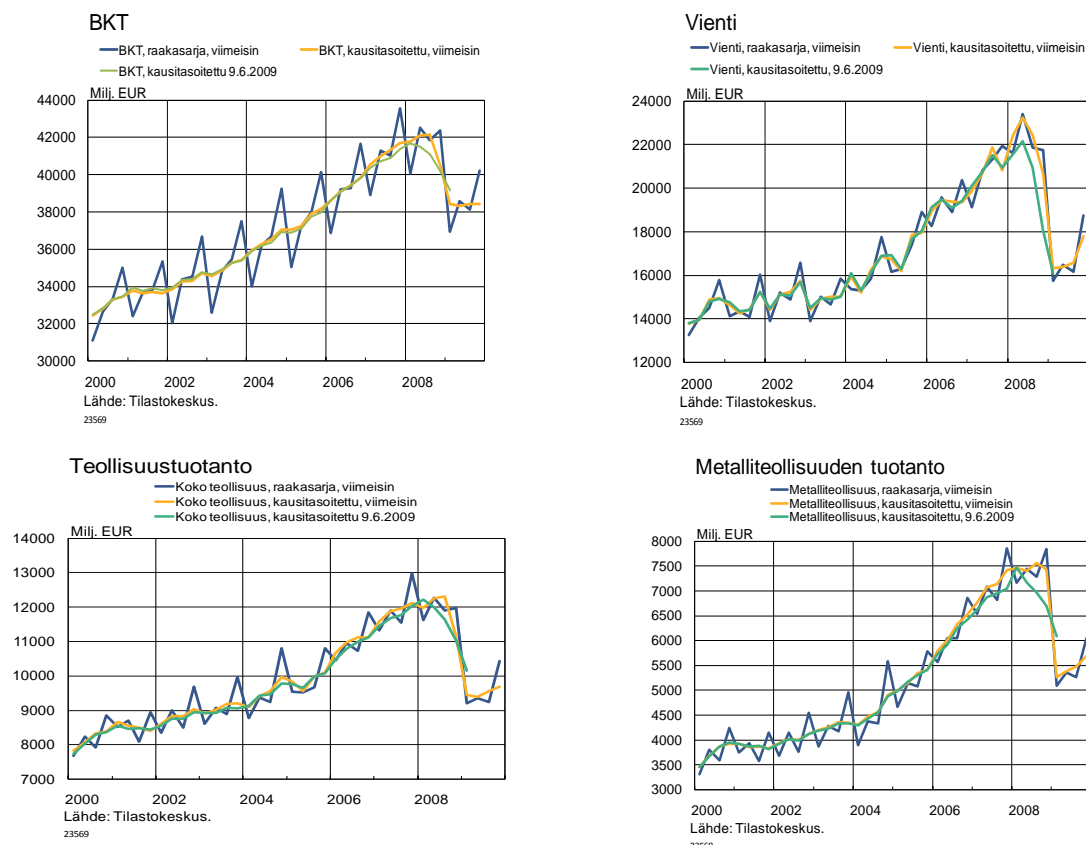
Suhdannesarja	ARIMA-malli	2008Q4	t-arvo	2009Q1	t-arvo
BKT	(0,1,1)(0,1,1)	LS	-4,53	LS	-5,17
Vienti yhteensä	(0,1,0)(0,1,1)	--	--	LS	-5,29
Koko teollisuus	(0,1,1)(0,1,1)	LS	-3,55	LS	-5,75
Metalliteollisuus	(1,0,0)(0,1,0)	--	--	LS	-8,52

Lähde: Tilastokeskus.

BKT:n ARIMA-malli on pysynyt samana kesästä 2008 lähtien. Sarjojen koko teollisuus, kauppa yhteensä ja arvonlisäys ARIMA-malli on koko ajan ollut sama (0,1,1)(0,1,1). Viennin ARIMA-mallina on aiemmin käytetty mallia (0,1,1)(0,1,1), mutta kesäkuusta 2009 lähtien mallia (0,1,0)(0,1,1). Tuonnin kausipuhdistusmallina on ollut koko ajan (0,1,0)(0,1,1).

Kuviossa 1 on esitetty neljän keskeisen suhdanneaggregaatin alkuperäinen, neljännesvuosittainen aikasarja 2000Q1–2010Q1 sekä kausipuhdistetut arvot kahdelta eri ajankohdalta. Kuten kuvioista nähdään, kausipuhdistus muuttuu uusien havaintojen myötä. Sarjojen revisiot eivät näy kuvioissa.

Kuvio 1. BKT:n, viennin, teollisuustuotannon ja metalliteollisuuden tuotannon neljännesvuosittainen aikasarja (2000Q1–2010Q1) sekä kausipuhdistetut arvot (sarjat perustuvat vuoden 2000 hintatasoon)



Tasosiirtymien koko ja jyrkkyys (ks. kuviot) osoittavat selkeästi talouskriisin poikkeuksellisen voimakkuuden. Kuten Tilastokeskuksen neljännesvuositilinpäätöksen julkaisusta ilmenee, kokonaisyntymän siirtymä vuosien 2008 ja 2009 vaihteessa oli erittäin voimakas samoin kuin BKT:n sekä koko viennin ja tuonnin. Koko teollisuuden, metalliteollisuuden sekä jalostuselinkeinojen arvonlisäsarjan tasosiirtymät olivat likimain samaa suuruusluokkaa. Kauppa yhteensä supistui selvästi vähemmän.

4 Tilastokeskuksen tekemien ratkaisujen arviointi

4.1 Tasosiirtymätulkinnan haasteita ja näkökohtia

Todetut tasosiirtymät ovat niin suuria, että ehkä useampi kerrannainen ARIMA-kausivaihtelumalli olisi testannut murtumakohdan oikein tasosiirtymäksi. Tasosiirtymän estimoinnin mahdollisti nyt se, että nykyisissä kausipuhdistusohjelmissa on sarjan esipuhdistuksessa käytettävissä vieraat havainnot mallintava osa. Peruskysymykseksi kuitenkin jää, miten *laadukas* LS-havainto on nyt pystytty estimoimaan ja miten tehokas ARIMA-malli (alustava ja lopullinen) on koko prosessissa aikasarjalle kyetty konstruoimaan.

4.2 Suhdanneteoreettinen ”Macroeconomics”

LS_t-vieraan havainnon teoreettinen tulkinta on, että muuttujan volyyymistä tai sisällöstä on ajankohtana $t = T$ tapahtuneen rakenne- tai häiriötekijän (-joiden, disturbance(s)) vaikutuksesta siirtynyt osia pysyvästi joko pois tai tullut lisää, jolloin taso alenee tai nousee pysyvästi. Voidaan kysyä esimerkiksi BKT-aggregaatin tapauksessa, onko joidenkin toimialojen tai sektorien osuus kokonaistuotannosta pysyvästi laskemassa (globaaliyritysten tai niiden osien myynnit, metsä-, metalli- ja elektroniikkateollisuuden jotkin osat jne.). Jatkossa havaittu aikasarja-aggregaatti tulee sisältämään em. tasomuutokset ja BKT:n muiden osien painotetun muutoksen (netto) määrättyllä aikavälillä mitattuna. Yksittäisen aikasarjan tapauksessa tällaisia LS-siirtymiä voi tapahtua esimerkiksi kaupan jonkin toimialan merkittävistä rakennemuutoksista, lainsäädännöstä, teknisistä ja tilastollisista luokittelumuutoksista johtuen. Tällaisia tasosiirtymätekijöitä on hankalampi soveltaa kansantalouden suuriin tilastoagregaatteihin.

4.3 Käytettävissä olevat vieras havainto -tyypit

Outlierien mallintaminen myös kausipuhdistuksen yhteydessä perustuu Tsayn artikkeliin (Tsay, 1988). Tässä Tsayn proseduuriksi kutsutussa lähestymistavassa määritellään testillä vieraan havainnon tyyppi (ks. edellä), sekä estimoidaan ja testataan sen koko ja paikka aikasarjassa. Tsayn proseduurista on kehitetty eri versioita. Syy näyttäisi olevan se, että vierai-

den havaintojen, IO:n ja LS:n, keskinäisessä erottamisessa on koettu sekä käytännön ongelmia että teoreettisia ongelmia ARIMA-mallien käytössä (Chen and Balke, 1990; Balke, 1993; Chen & Liu, 1993a, 1993b; Balke & Fomby 1991 ja 1994; Trivez, 1995 ja Vaage 2000). Todettakoon lisäksi, että vieraan havainnon paikka aikasarjassa määrää tällaisen havainnon vaikutusvoiman (influence) aikasarjan lyhyen aikavälin ennusteeseen. Mitä lähempänä vieras havainto on aikasarjan tuoreimpia havaintoja, sitä suurempi vaikutus sillä on laadittaviin ennustelukuihin.

TRAMO/SEATS -ohjelmalla voidaan mallintaa vieraat havainnot: AO, LS ja TC. Valikoima on suppea (vrt. liite). Esimerkiksi IO, joka on juuri ulkoisen shokki-vaikutuksen (vaikutus sarjassa eteenpäin, $t+1, \dots$) vieras havaintotyyppi, ei ilmeisesti edellä mainituista teoreettisista ja käytännön identifiointisyistä ole valikoimassa.

Yhtälö vaikeutuu lisää, kun suhdannesarjat ovat ei-stationaarisia $I(1)$ -sarjoja (eli niissä on trendi, TRAMO/SEATSissa on ARIMA(0,1,1)(0,1,1)-malli "Airline Model" oletusmallina), jolloin yksikköjuuren testaus saattaa antaa pahastikin harhaisia tuloksia, kun aikasarja sisältää vieraita havaintoja (ks. Balke & Fomby, 1991). Lisäksi kun ei-stationaarinen aikasarja, joka sisältää vieraita havaintoja, muunnetaan mallintamisessa tarvittavaan, stationaariseen muotoon $I(0)$, tuloksena voi olla entistä hankalammin käsiteltävä aineisto. Syynä on se, että differensointi yleensä lisää poikkeavien havaintojen lukumäärää aikasarjassa.⁸

4.4 Historiallisesti poikkeuksellinen tuotannon pudotus

Talouden vapaa pudotus tapahtui muutamassa kuukaudessa ja pohjalla on pysytty varsin pitkään shokkien laajetessa kansantalouksissa. Vieraiden havaintojen osalta romahduksessa oli ensin AO:n ja IO:n piirteitä ja pitkittyessään myös LS:n aineksia. Näyttää siltä, että käytettävissä oleva kausipuhdistusmenetelmä ei ole tarpeeksi tehokas ja monipuolinen tämän romahduksen kvantifioimiseksi selkeäksi vieraiden havaintojen rakenteeksi tai että tarvitaan lisää havaintoja niiden luonteen testaamiseksi. Teoreettisessa kirjallisuudessa viitataan eri vieras havaintotyyppien yhdistelmien soveltamiseen. Nykyisellään tämä kombinointi ei ole vielä mallinnettavissa, vaan se jää ensin tutkimuksen tehtäväksi (kehittää vieras havainto - kombinaatio teoreettisesti aikasarjamalleihin), sitten kokeiltavaksi ja lopulta puettavaksi tilastollisiksi ohjelmapaketeiksi.

⁸ Jos meillä on trendin sisältävä aikasarja $I(1)$ ja siinä AO-havainto, sarjan trendin poistaminen ottamalla 1. differenssit aiheuttaa muunnettuun sarjaan uuden, samankokoisen, mutta vastakkaismerkkisen AO-havainnon. Jos alkuperäinen $I(1)$ -sarja sisältää tasosiirtymän, sarjan 1. differensointi aiheuttaa AO-havainnon tasosiirtymän alkamisajankohtaan.

4.5 Aikasarjojen esipuhdistuksen kovenevat vaatimukset

Koko kausipuhdistusprosessin tuloksen laatu riippuu siitä, miten perusteellisesti ja tarkasti aikasarjan esipuhdistus kyetään toteuttamaan. Esimerkiksi vieraiden havaintojen vaikutusten arviointi on yleensä erittäin vaikeaa. Jos alkuvaiheen ARIMA-mallin muoto on virheellinen, voidaan sen myötä estimoida vääryyppisiä vieraita havaintoja. Tämä puolestaan heikentää esipuhdistuksen laatua, mikä johtaa keskivirheiden ja ennustevirheiden kasvuun (parametrin keskivirheiden ja jäännöstermien varianssien kasvu, t-testien merkitsevyysasteen aleneminen jne.).

Suomen neljännesvuositilinpidon ja muiden suhdannesarjojen kausipuhdistuksen kannalta tässä tilanteessa näyttäisi olevan kolme pääongelmaa. Ensinnäkin nykyinen TRAMO/SEATS on rakenteeltaan vanhentunut ja puutteellinen. Siitä mm. puuttuu kausikomponentin merkitsevyyttä ilmaiseva testi, joka on esimerkiksi X-12-ARIMAssa sekä STAMPissa. Toinen on Tilastokeskuksen varsin jäykältä tuntuva vuoden sisäinen kausipuhdistusmenettely, jossa kunkin suhdannesarjan ARIMA-malli konstruoidaan (identifioidaan, estimoidaan ja diagnostioidaan) vain kerran vuodessa.⁹ ARIMA-malli siis pidetään vuosi kiinteänä (fixed), ja vain parametrien estimaatit lasketaan uudestaan kullakin havaintokierroksella. Tosin aivan poikkeustapauksessa pääaggregaattisarjojen ARIMA-mallia voidaan säätää kesken vuotta. Kolmas seikka on kullakin kierroksella laskettava sarjakohtainen ARIMA-ennuste, jonka laatu voi kärsiä kohtien yksi ja kaksi ongelmista. Aikasarjan tuoreimpien havaintojen kausitasoituksessa käytetään näitä ARIMA-ennusteita.

Kaikki mainitut kohdat ovat omiaan vähentämään kausipuhdistuksen muutenkin rajoitettuja mahdollisuuksia lähisuhdanteen ja erityisesti suhdannekäänteiden tehokkaaseen arviointiin. Tarvittaisiin ilmeisesti lisäpanostusta (esim. interventio/regressiomalleja, robusteja suotimia ja tasointusproseduureja) ennusteiden laadun turvaamiseksi.

Edellä mainittuja puutteita lukuun ottamatta Tilastokeskus näyttää soveltavan TRAMO/SEATS -menetelmää huolellisesti. Log-muunnokset ja erilaiset kalenterikorjaukset on toteutettu asianmukaisesti. Käypähintaisille sarjoille on tehty ns. indirect-kausipuhdistus, eli alasarjat on ensin kausipuhdistettu ja sitten aggregoitu, kun taas kiinteähintaiset aggregaattisarjat on kausipuhdistettu suoraan (direct). Ongelmana näyttää olevan joustavuuden

⁹ Kyselytutkimuksen mukaan valtaosassa (80 %) maiden tilastovirastoista menetellään näin. Osalla virastoja estimoinnin aikaväli on muu kuin vuosi, esim. 6 kuukautta. Vajaa kolmasosa virastoista päivittää *tärkeimpien* sarjojen mallirakenteen aina uuden havainnon jälkeen. Noin 2/3 maista päivittää mallien parametriarvot aina uuden havainnon jälkeen (ks. tämän selvityksen kirjallisuusosa: julkaisu: Hungarian Central Statistical Office: SA Methods and Practices, 2007).

lisääminen varsinkin poikkeavien havaintojen tilanteissa. Tämä koskee lähinnä ARIMA-mallien rakenteen päivittämistä ja parametrien estimointeja ennusteineen kesken vuotta. ARIMA-mallit eivät tunnetusti ole kovin joustavia suhdannekäänteiden arvioinnissa ja ennustamisessa. Ilmeisesti tarvittaisiin joustavampia tilastollisia apuvälineitä ja menettelyjä tulosten laadun varmistamiseksi.

Tilastokeskus näyttää pyrkineen resurssiensa puitteissa seuraamaan alansa kehitystä myös aineistojen käsittelymenetelmien osalta. Tätä osoittavat ajantasaiset tilastojen menetelmäkuvaukset (myös internetissä) ja laaja osallistuminen kansainväliseen yhteistyöhön.

5 Tasosiirtymien vaikutus suhdanneaikasarjojen myöhempään kehitykseen

Tasosiirtymä on tässä poikkeava havainto ja sen vaikutus ulottuu sarjassa eteenpäin. Tämä estimoitu siirtymä näkyy lähiajan ennusteissa ja siten osaltaan pienentää ennustevirhettä. Aikasarjassa voi kuitenkin jatkossa olla erityyppisiä vieraita havaintoja, esimerkiksi uusia tasosiirtymiä ja/tai AO-havaintoja¹⁰. Tiedetään, että vieraat havainnot (erityisesti tasosiirtymät, LS ja TC, Balke & Fomby, 1991, s. 72–73) vaikeuttavat aina ARIMA-mallien konstruointia. Kausipuhdistuksen estimaatteihin vaikuttaa (ainakin vuoden sisällä) myös Tilastokeskuksen omaksuma teknis-hallinnollinen kausitasoituskäytäntö (concurrent, current adjustment)¹¹. Tässä kriisitilanteessa korostuu myös tilastojen revisioiden rooli, ja miten revisiot kytketään sarjojen kausipuhdistukseen. Kausikomponenttien muuttuminen ajassa on siksi tavanomais-

¹⁰ Tuoreessa EU-aikasarjojen analyysissä (Maravall, 2007) LS- ja TC-vieraat havainnot esiintyivät lähes yhtä usein (suurin frekvenssi oli AO-havainnoilla). Kun vieraiden havaintojen paikka aikasarjassa tiedetään, erityyppisten vieraiden havaintojen kombinaation yhteisvaikutusta voitaisiin kokeilla erillisten interventiomallien avulla.

¹¹ Näistä tilastovirastojen erilaisista, vuoden sisäisistä menettelyistä keskusteltiin viime marraskuussa Luxemburgin Workshopissa, ks. loppuraportti, 26.1.2010, tämän esityksen liitteen kirjallisuusosa. Esiin tuli joukko ehdotuksia erilaisista menettelykäytännöistä ("concurrent versus current adjustment", "controlled current adjustment"). Sarjojen (ainakin keskeiset sarjat) ARIMA-mallinnus aina uuden havainnon jälkeen saattaisi enemmän tukea suhdanteiden käännepeisteiden määrittystä ja luultavasti myös nostaisi aikasarjojen ennusteiden laatua. Nykyisen informaatioteknologian välineistön aikana teknisen työn määrä tuskin lisääntyisi kovin merkittävästi. Asiantuntijatyö sitä vastoin kasvaisi tuntuvasti. Verrattuna esimerkiksi nykyisten tilastollisten rekisterien tietokantasovellusten kokoon ja suorituspäämiin suhdanneseurannan tilastot ovat varsin pieniä kokonaisuuksia.

ta työläämpää kvantifioida ja hallita ja myös tulkita talusteoreettisesti. Tilastovirastojen kausipuhdistuksen teknis-hallinnollisessa menettelyssä tulisi saada aikaan suositukset mahdollisimman yhtenäisiksi säännöiksi. Tämä puolestaan nostaisi virallisen, kansainvälisen tilastoinnin laatua. Nykyinen talouskriisi näyttää paljastaneen puutteita kansainvälisen ja kansallisen tilastotuotannon menettelytapojen koordinoinnissa.

On myös tärkeää tutkia, onko talouden vapaa pudotus aiheuttanut keskeisten suhdannesarjojen perinteisen kausikäyttäytymisen särkymisen ja sysännyt kausikomponenteissa liikkeelle normaalia suuremman muuttumisprosessin. Nyt on tärkeää seurata sarjakohtaisesti kausikomponentin muutoksia, koska suuret muutokset voivat haitata a) sarjan esipuhdistusta ml. vieraat havainnot ja b) sarjan muiden komponenttien estimointia.

Ns. kausaalisen ekonometrisen kausitutkimuksen (esim. klassiset Engle, 1978 and Granger, 1978) tarve on lisääntymässä täydentämään puhtaasti tilastollista kausitutkimusta. Tämä kausaalisuuden tutkimus koskee yleensä yksityisen tulonmuodostuksen, säästämisen, kulutuksen ja investointien eri osa-alueiden sarjojen kausivaihtelua. Kun aineistoa (havaintoja) on kertynyt tarpeeksi lisää, voidaan soveltaa sarjojen kausikomponenttien tavanomaisia merkitsevyystestejä ja stabiilisuustestejä.

Myös kalenterivaihtelun kuten erilaisten kauppapäivä- ja työpäiväkorjausten tutkimuksen tarve on voimistunut (mm. kaupan sunnuntai-aukioloajan laajentuessa). Nämä erilaiset korjaukset ja sopeutukset kuuluvat sarjojen esipuhdistukseen ennen lopullisten kausikomponenttien estimointia. Sivutuotteena kertyvä historiatieto keskeisten suhdannesarjojen vieraisista havainnoista (mahdollisine tulkintoineen) voisi etenkin suhdannekäänteiden ja rakennemuutosten osalta olla todella arvokasta sekä tutkijoille että suhdanneseurannalle. USA:ssa tällainen seuranta on ollut käytössä pitkään (ks. Balke & Fomby, 1994).

Suhdanneanalyysin uusissa haasteissa edessä on lisätöitä ja vähän opiskeluakin tilastoasiantuntijoille, ekonomisteille ja muille tutkijoille – ja lisäperehtymistä medialle sekä tutkimuslaitoksille ja päättäjille, siis tilastojen käyttäjäryhmille. Vieraiden havaintojen tulkintaongelmat tulevat entistä useammin esille, joten ekonomistien on hyödyllistä ja tarpeellista perehtyä tähän aihepiiriin.

6 Yhteenveto ja johtopäätökset

Yleinen talouskriisi ja tuotannon romahdus ovat tuoneet näkyvästi esiin vieraat havainnot (outlierit) suhdannesarjoissa. Vuodenvaihteeseen 2008–2009 Suomessa ja monessa muusakin maassa tapahtunut kansantalouden vapaa pudotus näkyi ennen muuta tasosiirtymätyyppisenä vieraana havaintona.

Suhdannesarjojen kausitasoituksessa vieraiden havaintojen mallinnuksesta on tullut rutinia, mikä on mahdollista nykyaikaisten, mallipohjaisten kausipuhdistusohjelmien ansiosta. Kun esipuhdistuksen tehokkuus kasvaa, saadaan aiempaa tarkempia aikasarjan komponenttien estimaatteja ja niiden tarkempia ennusteita ja siten pienempiä keskivirheitä ja residuaalivariansseja.

Suhdanneseuranta on myös tullut nopeasykkeisemmäksi (mm. monet kuukausitason suhdannekuvaajat ja -indeksit), minkä tulisi näkyä kausipuhdistusprosessien menetelmissä, vuoden sisäisissä mallinnuskerroissa, toimintatavoissa ja menettelyissä. Vieraiden havaintojen, etenkin tasosiirtymien, vaikutusten eliminointi sarjojen esikäsittelyssä onkin edelleen hyvin haastava tehtävä.

Tilastokeskus on kaikesta päätellen soveltanut asianmukaisesti TRAMO/SEATS – kausipuhdistusohjelmaa. Tällä hetkellä jo hieman vanhentuneella ohjelmalla on estimoitu tilastollisesti erittäin merkitsevät tasosiirtymät joukolle suhdanneagregaatteja. Kausi-, trendi/suhdanne- ja irregulaarisen komponentin stokastisen estimaatin laatua olisi kuitenkin syytä edelleen parantaa, etenkin jos sarjassa on erityyppisiä vieraita havaintoja. Yhtenä keinona laadun nostamiseksi voisi kokeilla vuoden sisäisten mallintamispäivitysten frekvenssien lisäämistä koko estimointiprosessissa.

Tilastotuotannon laadunvalvonta on yleisesti parantunut ja avoimuus sekä läpinäkyvyys ovat lisääntyneet, mikä on hyvä tilastojen käyttäjien kannalta. Laadun jatkuva parantaminen edellyttää kuitenkin tilastotuotannon prosessien, käsittelymenetelmien, käytäntöjen ja toimintatapojen yhtenäistämistä EU-maiden tilastovirastoissa. Tarvitaan yhteisesti laaditut, selkeät menettelysuositukset ja niiden luokitukset. Revisioiden (havaintorevisiot, mallirevisiot) koon ja lukumäärän odotettavissa oleva kasvu vaatii lisää asiantuntijatyötä. Tilastokeskus näyttää kiinteästi seuranneen uusien kausipuhdistusohjelmien kansainvälistä kehittämistyötä. Alan sidosryhmillä ml. tilastojen käyttäjät on edessään perehtymistä uusiin käsitteisiin ja määrittelyihin.

Kirjallisuutta ja internet-linkkejä

Balke, N.S. (1993), Detecting Level shifts in time series. *Journal of Business and Economics*, 11, 81–92.

Balke, N.S. and Fomby, T.B. (1991), Shifting trends, segmented trends, and infrequent permanent shocks. *Journal of Monetary Economics*, 28, 61–85.

Balke, N.S. and Fomby, T.B. (1994), Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics*, 9, 181–200.

Bianco, A.M., Ben, M.G., Martinez, E.J. and Yohai, V.J. (2001), Outlier detection in regression models with ARIMA errors using robust estimates. *Journal of Forecasting*, 20, 565–579.

Chen, C. and Liu, L-M. (1993a), Forecasting Time Series with Outliers. *Journal of Forecasting*, 12, 13–35.

Chen, C. and Liu, L-M. (1993b), Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88, 284–297.

Chen, C. and Tiao, G.C. (1990), Random Level-Shift Time Series Models, ARIMA Approximations, and Level-Shift Detection. *Journal of Business and Economic Statistics*, 8, 83–97.

Engle, R. F. (1978), Estimating structural models of seasonality. In *Seasonal Analysis of Economic Time Series*, in: A. Zellner (ed.), 281-308 (with discussion). U.S. Department of Commerce, Washington.

The ESS Guidelines on Seasonal Adjustment: Workshop on SA, Eurostat, 24 November 2009, 55 pages.

Gómez, V. and Maravall, A. (1996), *Programs TRAMO and SEATS, Instructions for the User* (with some updates) Working Paper 9628, Servicio de Estudios, Banco de España.

Granger, C. W. J. (1978), Seasonality: causation, interpretation, and implications. In *Seasonal Analysis of Economic Time Series*, in: A. Zellner (ed.), 33–55 (with discussion). U.S. Department of Commerce, Washington.

Hella, H. (1996), On outliers in time series data. Keskustelunaiheita – Discussion papers No. 572, ETLA, Elinkeinoelämän Tutkimuslaitos, Helsinki.

Hungarian Central Statistical Office: Seasonal Adjustment, Methods and Practices, 131 pages. Budapest 2007. Internet-linkki: http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/SEASONAL_ADJUSTMENT_METHODS_PRACTICES.pdf.

Introduction to Seasonal Adjustment (2009). Artur Andrysiak (UN, ESCAP) Power-Point –esitys, internet-linkki: <http://unstats.un.org/unsd/nationalaccount/workshops/2009/india/AC186-12.PPT>.

Maravall, A. (2007), Automatic model identification in model-based seasonal adjustment with program TSW, a case study. International Statistical Institute (ISI), biannual conference, Lissabon.

Monsell, B. (2009a) (U.S., Census Bureau) Power-Point -esitys kausitasoitusanalyysistä, internet-linkki: <http://www.stat.fsu.edu/50th/Monsell.ppt>.

Monsell, B. (2009b) The X-13A-S Seasonal Adjustment Program. Linkki <http://www.census.gov/srd/www/x12a/>.

Report of Workshop on the implementation of the ESS Guidelines on seasonal adjustment, 24 November 2009, Eurostat, Luxembourg, (sent on 26 January 2010), 12 pages.

STAMP-ohjelman linkkiosoite: <http://www.stamp-software.com/index.html?content=/stampup.html>.

Suhdannekäännö ja taloudelliset aikasarjat, Tilastokeskus: Tutkimuksia 210, Erkki Pahkinen ja Veijo Ritola (toim.), Helsinki 1994.

A Course in Time Series Analysis, Peña, D., Tiao, G.C. and Tsay, R.S. (Eds.). John Wiley, New York, 2001.

Trivez, F.J. (1995), Level Shifts, Temporary Changes and Forecasting. *Journal of Forecasting*, 14, 543-550.

Trivez, F.J. & Nievas, J. (1998), Analyzing the effects of level shifts and temporary changes on the identification of ARIMA models. *Journal of Applied Statistics*, 25, 409–424.

Tsay, R.S. (1988), Outliers, Level Shifts, and Variance Changes in Time Series. *Journal of Forecasting*, 7, 1–20.

Vaage, K. (2000), Detection of Outliers and Level Shifts in Time Series: An Evaluation of Two Alternative Procedures. *Journal of Forecasting*, 19, 23–37.

Liite

A. Tilastokeskuksen suhdannesarjat, joihin on tehty tasomuutos kausipuhdistuksen yhteydessä (tiedote 8.12.2009)

Bruttokansantuote, arvonlisäys yhteensä, jalostus, koko teollisuus, tehdasteollisuus, puu- ja paperiteollisuus, metalliteollisuus, muu tehdasteollisuus, kauppa, tuonti yhteensä, tavaroiden tuonti, palveluiden tuonti, vienti yhteensä, tavaroiden vienti, palveluiden vienti, kokonaiskysyntä ja kokonaistarjonta. Muutokset koskevat volyymisarjojen ohella myös käypähintaisia aikasarjoja (paitsi puu- ja paperiteollisuus). Tasomuutokset ajoittuvat 2008N4 ja /tai 2009N1 neljänneksille.

Tasomuutoksia on lisäksi seuraavissa tilastoissa:

- tuotannon suhdannekuvaaja (kuukausisarja)
- teollisuustuotannon volyymi-indeksi
- teollisuuden uudet tilaukset
- teollisuuden liikevaihtokuvaaja
- rakentamisen liikevaihtokuvaaja
- kaupan liikevaihtokuvaaja
- palveluiden liikevaihtokuvaaja.

Lähde: Tilastokeskus.

B. Aikasarjojen tavallisimmat vieras havainto -tyypit¹²

Vieraiden havaintojen aiheuttajina voivat olla poikkeukselliset sääolot, lakot, tiedossa oleva muutosajankohta lainsäädännössä, sodat, luonnonmullistukset, tekninen virhe tiedon tallennusmekanismeissa, pysyvä muutos jossakin talouden käsitteessä, määritelmässä tai tilastoluokittelussa jne. Näiden havaintojen vaikutusten aikajänne voi olla hyvin eripituinen.

Vieraita havaintoja esiintyy sekä tavallisissa ARIMA-malleissa että kerrannaisissa kausivaihtelun ARIMA-malleissa. Lyhyiden vuoksi seuraavassa tarkastel-

¹² Aikasarja-analyysissä ja regressioanalyysissä ("i.i.d.-muuttuja-avaruus") on kummallakin oma vieras havainto -tyypittelynsä.

laan vain tavallisia ARIMA-malleja, joissa lisäksi ei esitetä vakiotermiä. Olkoon $\{Y_t\}$ aikasarja, jonka ARIMA-malli on muotoa

$$Y_t = \frac{\theta(B)}{\alpha(B)\phi(B)} a_t \quad t = 1, \dots, n,$$

(1)

missä n on havaintojen lukumäärä; viivepolynomien $\phi(B)$ ja $\theta(B)$ juurien oletetaan olevan yksikköympyrän ulkopuolella eli mallin autoregressiivisen osan (AR) oletetaan olevan stationaarinen ja liukuvan keskiarvon (MA) osan ns. kääntyvä prosessi; polynomien $\alpha(B)$ kaikkien juurien oletetaan olevan yksikköympyrällä; mallin (1) kohinaprosessin oletetaan olevan normaalijakaumaa noudattavaa valkoista kohinaa $a_t \approx N(0, \sigma_a^2)$. B on viiveoperaattori $Bx_t = x_{t-1}$.

Aikasarjoissa esiintyvät yleisimmät outlier-tyypit ovat seuraavat neljä.

1. Additiivisen outlierin malli (Additive Outlier, AO)

Additiivinen outlier on aikasarjan yksittäinen poikkeava havainto, esimerkkinä voidaan mainita havaintoarvo, jossa on desimaalipistevirhe. AO on riippumaton aikasarjan (realisaation) ARIMA-mallista. Jos oletetaan, että AO on aikasarjassa ajankohdassa $t = T$, AO –malli on muotoa

$$Z_t = Y_t + o_A P_t^{(T)},$$

(2)

missä $P_t^{(T)}$ on indikoiva (pulse, step) funktio ($P_t^{(T)} = 1$ kun $t = T$, muulloin 0) ja o_A on AO:n numeraalinen suuruus (magnitude). Mallissa (2) Z_t on todellinen havainto (siis sisältää outlierit) ja Y_t is on "puhdas" (ei-havaittu) prosessi ajankohdassa $t = T$.

2. Innovaatio-outlierin malli (Innovation Outlier, IO)

Tämän outlierin shokkimainen jälkivaikutus (subsequent) aikasarjaan on geneeroivan ARIMA-mallin viiverakenteiden mukainen. Merkinnät Z_t ja Y_t ovat samat kuin edellä. IO-malli on tällöin

$$Z_t = Y_t + \alpha(B)^{-1} \frac{\theta(B)}{\phi(B)} o_I P_t^{(T)}, \quad \text{or} \quad Z_t = \frac{\theta(B)}{\alpha(B)\phi(B)} (a_t + o_I P_t^{(T)}) \quad (3)$$

missä $P_t^{(T)}$ ja o_I ilmaisevat indikoivaa funktiota ja outlierin kokoa kuin vastaavassa mallissa edellä ajankohtana $t = T$.

3. Tasomuutoksen malli (Level Shift, LS)

$$Z_t = Y_t + \frac{1}{1-B} o_L P_t^{(T)},$$

(4)

missä o_L on sarjan pysyvä (permanent mean level) tasomuutos ajankohdasta $t = T$ alkaen.

4. Vaimenevan muutoksen malli (Transient Change, TC)

$$Z_t = Y_t + \frac{1}{1-\delta B} o_C P_t^{(T)} \quad 0 < \delta < 1,$$

(5)

missä o_C alkujankohdan vaikutuksen (muutoksen) suuruus, kun $t = T$. Muutos vaimenee tasaisesti (eksponentiaalisesti) kertoimen δ numeroarvon¹³ mukaisesti. Kun $\delta = 1$, kysymyksessä on LS ja kun $\delta = 0$, kysymyksessä on AO. Jos alkujankohda t on tunnettu, voidaan estimoida interventiomalli kuvaamaan vaimenevaa vaikutusta.

¹³ Makrotalouden aikasarjoissa tämän vaimenemiskertoimen arvo voi olla 0.7 (on oletusarvo X-12-ARIMAn manuaalissa; ja Maravall, 2007) tai 0.8. Vaimenemismuutoksia ei yleensä ilmoiteta aikasarjan havaintojen lukumääränä, vaikka se voidaan laskea kussakin tapauksessa.

Jotkut tutkijat nimittävät malleja (4) ja (5) *pysyvän tasomuutoksen malliksi* ja *vaimenevan tasomuutoksen malliksi* (Vaage, 2000).

5. Muita aikasarjoissa esiintyviä vieras havainto -tyyppejä ovat:

Reallocation (**RO**), kausivaihtelun outlieri (**SLS**), sekä ramp shift -outlieri (**RP**).

Vieraiden havaintojen tilastollinen estimointi ja testaus tapahtuvat iteratiivisena prosessina. Prosessi on hyvin tietokoneintensiivinen, koska laskenta testeineen tehdään aikasarjan kunkin havainnon Y_t kohdalla. Tsay (1988) on esittänyt vieras havainto -tyyppien estimaattien, varianssien, ennusteiden ja testimuuttujien kaavarakenteet. Kunkin outlierin uskottavuusosamäärätestin perusteella määritellään outlierin tyyppi, samalla estimoidaan outlierin koko (magnitute) ja tilastollinen merkitsevyys (Tsay, 1988; Chen and Liu, 1993a ja 1993b; Pena et al., 2001). Paikka aikasarjassa määräytyy samalla kertaa. LS- ja TC-estimoinnissa ajankohta t ilmaisee milloin estimoitu tasosiirtymä alkaa ja TC:n osalta milloin alkanut muutos on suurimmillaan, mutta alkaa seuraavana ajankohtana vaimeta tasaisesti (eksponentiaalisesti). Tasosiirtymän estimoinnista on kirjallisuudessa esitetty erilaisia versioita. Iteratiivisen estimoinnin rakennetta on myös kritisoitu. IO:n ja LS:n keskenään ongelmallista estimointia on tutkittu ja ehdotettu vaihtoehtoista mallinnusta (Balke, 1993; Balke & Fomby, 1994; Chen & Tiao, 1990; Chen & Liu, 1993a; Bianco et al. 2001, and Vaage, 2000).

Vieraan havainnon vaikutukset aikasarjoissa on osoitettu riippuvan ko. havainnon tyylistä, suhteellisesta koosta, paikasta aikasarjassa yleensä ja suhteessa muihin vieraisiin havaintoihin sekä aikasarjan generoiveen ARIMA-mallin viive-rakenteista ja kerroinparametrien suuruusluokasta ja etumerkistä. Kysymyksessä on siis varsin monisäikeinen teoreettinen kokonaisuus (ks. esim. Chen & Liu, 1993a ja 1993b; Peña et al., 2001). Vieraiden havaintojen teoriassa on paljon lisätutkimuksen tarvetta. Käytännössä näiden havaintojen käsittelyyn tarvitaan *robustin tilastotieteen* estimointimenetelmiä, diagnostiikkoja ja estimaattoreita. Näiden välineiden avulla laadittujen aikasarjaennusteiden tarkkuutta on useimmiten kyetty lisäämään verrattuna perinteisiin, klassisiin mallintamisen menetelmiin.

Masking-, swamping- ja smearing-vaikutus

Kun aikasarjassa on useita poikkeavia havaintoja, voi osoittautua, että ne peittävät (mask) toistensa vaikutukset aikasarjaan estäen tilastollisia testejä löytämästä lainkaan poikkeavia havaintoja. Joskus aikasarjan havainto voidaan virheellisesti tulkita vieraaksi havainnoksi. Tällöin kysymyksessä on swamping-ilmio. Smearing-vaikutus on vain aikasarja-analyyseissä esiintyvä ilmiö. Siinä vieraan havainnon vaikutus etenee aikasarjan peräkkäisten havaintojen kesken johtuen havaintojen korreloituneisuudesta.

Robusteilla suotimilla/tasoittimilla (filters/smoothers, esim. 4253H) voidaan tutkia sarjaa, jos on epäselvyyttä vieraiden havaintojen esiintymisestä¹⁴. Graafiset tarkastelut voivat tällöin paljastaa poikkeavat havainnot. Jo nykyisenkin kausipuhdistusohjelman yhteydessä voitaisiin joustavasti kokeilla näiden robustien tilastollisten suotimien käyttöä (log-transformaation tapaan) sarjojen esikäsittelyssä. Jos aikasarjan tuorein havainto on selvästi muista poikkeava, sen vieras havainto -tyyppiä ei voida määrittää, vaan aikasarjaan tarvitaan lisähavaintoja.

¹⁴ Robusti filteri 4253H on STATISTICA-ohjelmapaketissa, joka on käytössä myös Suomen Pankissa. Jos tällä mediaanien kombinaatioon perustuvalla, ei-parametrisella suotimella kuvataan BKT-neljännesvuosisarja ja kuukausipohjainen BKT-kuvaaja (Tilastokeskus), saadaan selväpiirteisesti esiin vuosien 2008 ja 2009 vaihteen jyrkkämuotoinen tasosiirtymä. Jatkossa tätä yksinkertaista filteriä kannattaa käyttää arvioitaessa, miten tasosiirtymä etenee. Huomattakoon, että tällöin tasoitettu sarja saadaan myös tuoreimmille havainnoille.