

<http://www.jyu.fi/library/tutkielmat/395/>

Riku Salonen

**MUUTOKSEN JA TASON ESTIMOINTI ROTATOIVASSA  
PANEELIAINEISTOSSA ERI ESTIMAATTOREIDEN AVULLA  
SOVELLUS TYÖVOIMATUTKIMUKSEN AINEISTOON**

Tilastotieteen  
pro gradu -tutkielma  
12.06.1997

Jyväskylän yliopisto  
Tilastotieteen laitos  
Informaatioteknologian maisteriohjelmat  
Tilastotoimen menetelmät

## **Tilastotoimen menetelmien maisteriohjelma**

(<http://www.stat.jyu.fi/>)

(Päivitetty 16.5.1997)

Tilastotoimen menetelmien maisteriohjelman tavoite on kouluttaa opiskelija tilastotiedon keruun, jatkojalostuksen ja käytön asiantuntijaksi nykyaikaisessa tilastojärjestelmäympäristössä, jossa datat ovat survey-, koeasetelma- tai rekisteriperusteisia. Koulutus järjestetään yhteistyönä Jyväskylän yliopiston tilastotieteen laitoksen kanssa. Tilastotieteelliseltä kannalta kyse on survey-menetelmiin, biostatistiikkaan ja ekonometriaan erikoistuneiden tilastoasiantuntijoiden koulutuksesta. Ohjelman kesto päätoimisella opiskelijalla on kaksi lukuvuotta.

Tilastotoimen ytimenä ovat tilastotieteen, erityisesti tilastotoimen teorian ja tietojenkäsittelyn kurssit. Ohjelmaan voidaan sisällyttää myös koulutustavoitteita tukevia opintojaksoja lähitieteistä (esim. talous-, yhteiskunta- ja viestintätieteet) sekä yritystoimintaan perehdyttäviä kursseja. Tarjottava opetus koostuu osittain tilastotieteen laitoksen opetusohjelmasta, erityisesti suunnitelluista tilastotoimen teorian kursseista ja ostopalveluina hankituista muista erikoiskursseista. Opettajista pääosa on kotimaasta. Vierailijoina on myös ulkomaalaisia asiantuntijoita, joten opetus on osin englanninkielistä.

Ohjelman tärkeä osa on yliopiston ulkopuolisessa yhteistyötoimipaikassa suoritettu pro gradu -tutkielma ja siihen liittyvä harjoittelu. Yhteistyötoimipaikat ovat virallisesta tilastotoimesta, suuryrityksistä tai tutkimuslaitoksista. Ne osallistuvat osaltaan harjoittelun kustannuksiin. Perusajatus on, että pro gradu -tutkielma tai osa siitä toisi tutkimustulostensa osalta lisäarvoa yhteistyötoimipaikalle. Tätä tukenee se, että harjoittelu jaetaan kahteen osaan, joista ensimmäinen eli orientoiva osa on ensimmäisen opintovuoden lopussa. Sen jälkeen opiskelija palaa yliopisto-opiskeluihin yhden lukukauden ajaksi syventääkseen tietojensa siinä tilastotieteen osa-alueessa, joka on tarpeen aiotussa pro gradu -tutkielmassa. Maisteriohjelman viimeinen lukukausi muodostaa yhtenäisen tutkimusjakson, jonka aikana tehdään pro gradu -tutkielma.

## **Master's Programme in Statistical Systems**

(<http://www.stat.jyu.fi/>)

(Updated 16.5.1997)

The main target of the Program is to educate high qualified professionals in the collection analysis, managing and dissemination of large data sets. The Program has been built mainly on the regular curriculum of the Department of Statistics in the University of Jyväskylä. In point of view of statistical sciences, the Program in Statistical Systems concentrate in the knowledge of survey methodology, biometry and econometry. Full-time students can graduate in two Academic years. The core courses cover the advanced theory in mathematical statistics, of statical systems especially the survey methodology and some courses in information technology. Still the program is flexible that a moderate amount of different kinds of studies in the neighboring sciences can be included as studies in economics, social sciences and communication. Most of the instructors come from Finland but visiting professors from outboard are eventual using English.

An important part of the Program crows from the cooperational research work created between the Department of Statistics and the research units located outside the University of Jyväskylä. Those units are research and development departments in the bodies of official statistics, big business firms and research institutes. They share the costs of practices. Basic idea is that the MS.c. Thesis will be written from the topics given by the cooperational research units and so the results could be of some contributed value to them. As an appropriate mode of policy may be here thus that the time of practice skills for the Thesis. This is proceeded in 6 final months of the Program, in collaboration with the same cooperating research unit.

## Tilastotoimen menetelmien maisteriohjelman pro gradu -tutkielma sarja

1. Salmikuukka, J. (1997) Aikasarjojen perusrakennemalleista ja niiden soveltaminen Jyväskylän kaukolämmön kulutuksen analysointiin ja ennustamiseen. (76 s., 1 liite) Jyväskylän Energia Oy, Jyväskylä
2. Yrjölä, T. (1997) Lasten päivähoiton tuottavuusvertailu suurissa kaupungeissa DEA-menetelmällä. (72 s., 2 liitettä) Jyväskylän kaupungin terveystoimi, Jyväskylä
3. Ainiala, N. (1997) Helsingin osa-alueiden työvoimatilastojen estimointi pienaluetekniikalla valtakunnallisesta työvoimatutkimuksesta. (73 s., 3 liitettä) Helsingin kaupungin tietokeskus, Helsinki
4. Puhakka, E. (1997) Kiintiöpöiminnan tilastolliset ominaisuudet pk-yritysbarometritutkimuksessa. Sovelluksena 2/1996 aineisto (82 s., 2 liitettä) (salainen) Tietoykkönen, Jyväskylä
5. Salonen, R. (1997) Muutoksen ja tason estimointi ratatoivassa paneeliaineistossa eri estimaattoreiden avulla. Sovellus työvoimatutkimuksen aineistoon. (39 s, 4 liitettä) Tilastokeskus, Helsinki
6. Kunttu, S. (1997) Alueellisen teollisuustuotannon volyyymi-indeksin estimointi Etelä-Pohjanmaalle. (103 s.) Tilastokeskus, Seinäjoki

## Muutoksen ja tason estimointi rotatoivassa paneeliaineistossa eri estimaattoreiden avulla

Sovellus työvoimatutkimuksen aineistoon

Riku Salonen

Tilastotiede. 12.06.1997. Jyväskylän yliopisto.

Sivuja 39, liitteitä 4.

Tutkielmassa tarkastellaan toistuviin mittauksiin perustuvan tulosmuuttujan keskivirheen mittaamista. Tarkastelun kohteena on sekä tulosmuuttujan tasoarvion että sen muutosarvion tilastollinen tarkkuus. Tarkastelussa tullaan käymään läpi sekä estimoinnin teoreettinen osuus että sovellus Tilastokeskuksen kuukausittaisen työvoimatutkimuksen vuosien 1994 ja 1995 kuukausi- ja vuosiaineistoihin.

Kuukausittaisen työvoimatutkimuksen nykyinen laskentatapa ei ota huomioon muutosta ja tasoa estimoitaessa kaikkia rotaatioasetelman riippuvuuksia. Riippuvuuksia on olemassa niin peräkkäisien neljännesvuosien kuin vuosienkin välillä. Tutkielman keskeisenä tavoitteena onkin ollut sellaisten estimaattoreiden muodostaminen, jotka ottaisivat työvoimatutkimuksen rotaatorakenteen huomioon. Tutkimuksessa käytetty rotaatorakenne voidaan ottaa muutosta ja tasoa estimoitaessa huomioon jakamalla paneeli päällekkäisiin ja riippumattomiin osiin. Työvoimatutkimuksen rotaatorakenteen hyödyntävä estimaattori saadaan paneelin päällekkäisten ja riippumattomien osien yhdistelmäestimaattorina.

Keskeinen tulos on, että kuukausittaisen työvoimatutkimuksen rotatoiva paneeliasetelma huomioimalla voidaan tehostaa henkilön työmarkkina-asemaan liittyviä estimaatteja sekä muutoksen että tason estimoinnin tapauksessa. Estimoinnin tehostuminen ilmenee perusjoukon parametrien kapeampina luottamusväleinä.

Avainsanat: päällekkyyys, rotatoiva paneeliasetelma, yhdistelmäestimaattori

# The estimation of level and change in a rotating panel design

## Application to Finnish Labour Force Survey

Riku Salonen

Statistics. 12.06.1997. University of Jyväskylä.  
39 pages, 4 appendices.

The design of the Finnish monthly Labour Force Survey is a rotating panel. Repeated measurement of labour market-related issues often involve a positive correlation between successive answers to the same questions. When estimating a change between successive quarters or years and an average of year, the rotating panel structure must be taken into account.

A method of estimation, that takes in account the structure of the rotating panel is to divide the sample in an overlapping part and a non-overlapping part. Then we can estimate total number of persons separately within these parts of sample. The estimator in the overlapping part of panel is based on answers from the same persons in different points in time, whereas the estimator in the non-overlapping part of the panel is based on cross-sectional answers from random sample. We can estimate the change or the average by using the composite estimator in the overlapping and the non-overlapping parts of the panel.

We can compare the variance of the composite estimator to the variance of the simple estimator, which do not take into account the structure of the rotating panel. The composite estimators are more efficient estimators of change and level than the simple estimators.

Keywords: composite estimator, partial overlapping, rotating panel

# Sisällysluettelo

<b>1 Johdanto</b>	<b>2</b>
1.1 Tutkimuksen tavoite	2
1.2 Kuukausittainen työvoimatutkimus	3
<b>2 Rotatoiva paneeliasetus ja siihen liittyvät estimaattorit</b>	<b>8</b>
2.1 Tutkimusasetelma	8
2.2 Rotaatiomallin kuvaus	10
2.3 Estimaattorin kuvaus tutkittavana ajankohtana	12
2.4 Muutoksen estimointi	13
2.4.1 Nykyestimaattori muutokselle	13
2.4.2 Vaihtoehtoinen estimaattori muutokselle	14
2.5 Tason estimointi	20
2.5.1 Nykyestimaattori vuositasoa estimoitaessa	20
2.5.2 Vaihtoehtoiset estimaattorit vuositasoa estimoitaessa	21
<b>3 Tuloksia</b>	<b>29</b>
3.1 Estimointituloksia	29
3.2 Tehokkuusvertailu	31
<b>4 Yhteenveto</b>	<b>34</b>
<b>Lähteet</b>	<b>38</b>
Liite 1. Synteettisen aineiston kuvaus	40
Liite 2. Neljännesvuosimuutoksen estimoinnissa käytetty kuukausittaisen työvoimatutkimuksen aineisto	42
Liite 3. Vuosimuutoksen estimoinnissa käytetty kuukausittaisen työvoimatutkimuksen aineisto	45
Liite 4. Vuositason estimoinnissa käytetty kuukausittaisen työvoimatutkimuksen aineisto	48

# Luku 1

## Johdanto

### 1.1 Tutkimuksen tavoite

Tutkielmassa tarkastellaan muutoksen ja tason estimointia rotatoivassa paneeliaineistossa eri estimaattoreiden avulla. Tarkoituksena on tutkia vaihtoehtoisia tapoja tulosmuuttujan muutoksen keskivirheen mittaamiseen neljännesvuosi- ja vuositasolla sekä tulosmuuttujan tason eli keskiarvon keskivirheen mittaamiseen vuositasolla, kun paneeliaspekti on mukana tutkimuksessa. Nykyestimaattoriksi kutsutaan tässä tapauksessa estimaattoria, joka ei huomioi rotatoivan paneeliaineiston rakennetta. Vaihtoehtoiseksi estimaattoriksi nimitetään estimaattoria, joka pyrkii ottamaan huomioon rotatoivan paneeliaineiston rakenteen. Tarkastelussa käydään läpi sekä estimoinnin teoreettinen osuus että sovellus Tilastokeskuksen kuukausittaisen työvoimatutkimuksen aineistoon. Monimutkaisesta tutkimusasetelmasta johtuen työvoimatutkimuksessa ei ole tähän asti laskettu muutosten eikä vuositason keskivirheitä.

Teoreettinen osuus käsittelee yleisesti rotatoivan paneeliasetelman hyväksikäyttöä muutoksen ja tason estimoinnin yhteydessä. Estimointitekniikat on johdettu pitkälti Kishin (1965, 1978 ja 1995) ja Cochranin (1977) mukaan. Indikaattorimuuttujien käytön sekä muutoksen että tason päällekkäisen osan varianssiestimaattoreita muodostettaessa on työpaperissaan esittänyt Madsen (Eurostat-workshop, Avignon, 13-15. July 1996).

Tutkielman empiirisessä osassa tarkastellaan nykyestimaattoreiden ja muodostettavien vaihtoehtoisten estimaattoreiden toimintaa sekä synteettisellä aineistolla että todellisilla Tilastokeskuksen kuukausittaisen työvoimatutkimuksen aineistoilla. Tutkielmassa esitettävissä estimointiproseduureissa ei ole mukana työvoima-

tutkimuksessa käytettävää jälkiositusta. Niinpä tutkielman empiirisessä osassa esitettävät estimointitulokset eivät ole täysin vertailtavissa työvoimatutkimuksesta saatuihin tuloksiin. Synteettinen aineisto on nykyestimaattoreiden ja vaihtoehtoisten estimaattoreiden vertailua varten rakennettu keinotekoinen aineisto, joka pyrkii mukailemaan todellisia työvoimatutkimuksen kuukausiaineistoja.

## 1.2 Kuukausittainen työvoimatutkimus

Kuukausittainen työvoimatutkimus on Labour Force Survey -tyyppinen haastattelututkimus. Suomessa tutkimusta - jonka nimenä oli ensin työvoimatiedustelu, sittemmin työvoimatutkimus - teki vuodesta 1959 vuoteen 1970 kulkulaitosten ja yleisten töiden ministeriö, sittemmin työvoimaministeriön tutkimustoimisto. Vuodesta 1971 lähtien tutkimus on tehty Tilastokeskuksessa. (Tilastokeskus, 1993, 3.)

Työvoimatutkimuksessa työikäinen väestö jaetaan työmarkkina-aseman mukaisiin ryhmiin. Perusluokittelu on jako työllisiin, työttömiin ja työvoimaan kuulumattomiin.

- **Työllinen** on henkilö, joka tutkimusviikkona teki yhtenäkin päivänä työtä palkkaa tai voittoa saadakseen tai työskenteli yrittäjäperheen jäsenenä vähintään kolmanneksen alan normaalista työajasta<sup>1</sup> tai oli työpaikastaan tilapäisesti poissa.
- **Työtön** on henkilö, joka oli koko tutkimusviikon työtä vailla, työhön käytettävissä sekä etsi työtä. Myös henkilö, joka oli työpaikastaan lomautettu<sup>2</sup> tai odotti sovituksen alkamista kuuluu työttömiin. Työttömäksi luetaan myös henkilö, joka oli osan tutkimusviikkoa työttömänä ja muun osan viikkoa muualla kuin työssä esimerkiksi opiskelijana. Työttömyyseläkkeellä olevista vain työtä etsivät luetaan työttömiin.
- **Työvoimaan kuulumaton** väestö muodostuu henkilöistä, jotka tutkimusviikolla eivät olleet työssä tai työttöminä. Työvoimaan kuulumattomiin luokitellaan opiskelijat, asevelvollisuuttaan suorittavat, omaa kotitalouttaan hoitavat, van-

---

<sup>1</sup> Tämä käytäntö muuttui vuoden 1997 alusta. Yrittäjäperheen jäseniin sovelletaan nykyisin samaa työnteon rajaa kuin muihinkin työllisiin. (Tilastokeskus, 1997, 48.)

<sup>2</sup> Myös tämä käytäntö muuttui vuoden 1997 alusta. Määräajaksi lomautetut luetaan työllisiksi, muut lomautetut työttömiksi tai työvoimaan kuulumattomiksi samoin ehdoin kuin muukin väestö. (Tilastokeskus, 1997, 47.)



huuseläkkeellä olevat, työkyvyttömät, korko- ja pääomatuloilla elävät sekä eräät muut edellisiin luokkiin kuulumattomat.

Tärkeimmät laskettavat tunnusluvut ovat työvoimaosuus ja työttömyysaste. Lisäksi työvoimatutkimus antaa tietoa myös työajasta, ammattijakaumasta ja työttömyyden laadusta. Tutkimustuloksia käytetään ensisijaisesti talouden suhdanteiden analysointiin ja seurantaan. Muita tärkeitä käyttöalueita ovat kansantulolaskelmat ja kansainväliset vertailut. (Tilastokeskus, 1995.)

Työvoimatutkimus perustuu satunnaisotokseen. Tutkimusasetelmana on rotatoiva paneeliasetelma ja perusjoukkona vakinaisesti Suomessa asuva 15-74-vuotias väestö. Otoskehikkona on väestön keskusrekisteri, joka on lajiteltu yksikäsitteisen kotipaikkatunnuksen, iän ja sukupuolen mukaan. Otos poimitaan systemaattisesti 14-74-vuotta täyttäneestä väestöstä. Otoskehikon lajittelujärjestys tuottaa maantieteellisesti sekä ikä- ja sukupuolirakenteen mukaan itsepainottuvan otoksen. Koska Ahvenanmaan maakunnan asukkaista poimitaan pieni lisäotos, palautuu otanta-asetelma ositetuksi otannaksi, jossa ositteet on muodostettu alueellisesti manner-Suomesta ja Ahvenanmaan maakunnasta. Kiintiöintiperiaate on lähinnä suhteellinen.

Noin 18 000 henkilön suuruiset satunnaisotokset poimitaan kahdesti vuodessa. Tilastokeskuksessa ne tahdistetaan sellaisen rekisterin kanssa, jossa on aikaisemmissa tutkimuksissa mukana olleiden henkilöiden tunnistetiedot. Viiden viimeisen vuoden aikana mukana olleet henkilöt vapautetaan otoksesta ja jäljelle jääneet henkilöt arvotaan kuukausiotoksiin. Vapautettujen henkilöiden osuus on ollut noin seitsemän prosenttia. (Djerf, 1994, 23-24.)

Tutkimuksen tiedot kerätään yhdeltä tutkimusviikolta haastattelemalla kuukausittain noin 12 000 työkäistä suomalaista. Tutkimusviikkona on yleensä kuukauden 15. päivän sisältävä viikko, joka edustaa kyseistä kuukautta. Haastatteluista 95 prosenttia tehdään puhelinhaastatteluna ja loput käyntihaastatteluna. Paneelin rotaatorakenne on sellainen, että otokseen kuuluva henkilö on tutkimuksessa mukana viisi kertaa puolentoista vuoden aikana. Otos vaihtuu asteittain siten, että kolmena peräkkäisenä kuukautena vastaamisvuorossa ovat eri henkilöt ja peräkkäisten kuukausien otokset ovat aina toisistaan riippumattomia. Kolmen peräkkäisen kuukauden haastatteluaineistot yhdistämällä saadaan neljännesvuosiaineisto. Pe-

räkkäisinä vuosineljänneksinä 60 prosenttia vastaajista on samoja. Vuoden kuluttua otosten päällekkäisyys on 40 prosenttia. (Tilastokeskus, 1995)

Työvoimatutkimuksen tulokset estimoidaan jälkiosittamalla otos asuinläänin, viisivuotisikäryhmän ja sukupuolen mukaan. Kun pääkaupunkiseutua käsitellään kuten lääninä, ositteita saadaan 312. Tulosten laskennassa käytetään ositteita kuvaavia väestötilastoja. Ositteittaisten väkilukujen kautta lasketaan koko väestöä koskevat tulokset. Keskivirheitä laskettaessa oletetaan systemaattisen otantamennettelyn palautuvan yksinkertaiseen satunnaisotantaan, sillä poimintajärjestyksen ja tutkittavien muuttujien oletetaan olevan riippumattomia. (Djerf, 1994, 24.)

Tilastokeskuksessa tutkittiin mahdollisuutta käyttää estimoinnin tehostamiseen väestötietojen ohella myös työnhakijarekisterin tietoja (THR-tieto). Saatujen tulosten perusteella estimoinnin tehostamiskokeilu onnistui eli otantavirhe hieman pieneni. Toisaalta hieman ongelmia aiheutti työnhakijarekisterin käytöstä seuraava estimaatin tason muutos ja aikasarjojen käyttäytymisen vaikeampi tulkinta. Ongelmat olivat kuitenkin kokonaisuuden kannalta niin vähäisiä, että Tilastokeskus alkoi käyttää kuukausitutkimuksen tulosten estimoinnissa työnhakijarekisterin tietoja vuoden 1997 alusta. Työnhakijarekisterin tiedot julkaisee työministeriö työvälitystilastossaan, jossa ilmenee tiedot työttömistä työnhakijoista kuukauden viimeiseltä arkipäivältä. Tilastokeskuksen ja työministeriön työttömyysluvut eivät ole vertailukelpoisia. Ne muuttuvat yleensä kuitenkin samansuuntaisesti siten, että työministeriön luvut ovat olleet muutaman kymmenen tuhannen verran suurempia.

Työttömyyslukujen systemaattisen eron pääasiallinen syy on se, että tilastoissa sovellettavat työttömyyden kriteerit poikkeavat toisistaan. Työvälitystilastossa työttömän työnhakijan määrittely on seuraava: Työttömänä pidetään työnhakijaa, joka on ilman työtä ja on käytettävissä työhön, jossa työaika on vähintään puolet alan normaalista työajasta, tai odottaa sovitun työn alkamista. Työnhakijaa, joka voi ottaa tarjotun työn vastaan vasta määräajan kuluttua tai joka hakee vain työtä, jossa työaika on lyhyempi kuin puolet alan normaalista työajasta, ei merkitä työttömäksi. Työministeriön tilastossa henkilö lasketaan edellisten määrittelyjen mukaisesti työttömäksi, kun hänellä samalla on voimassa oleva työnhaku työvoimatoimiston asiakasrekisterissä.

Lisäksi eroa aiheuttavat osaltaan työnhakijarekisteriin liittyvä päivittymisviive, joka johtuu tiedonsaannin hitaudesta tai kiireistä asiakastyössä. Se on nostanut

rekisteröidyn työttömyyden tasoa. Päivittymisviiveen takia osa työttömistä työnhakijoista on työvoimatutkimuksen mukaan ollut työssä, vaikka tieto työllistymisestä ei työministeriön tilastoon ole vielä ehtinytkään<sup>3</sup>. Työvoimatutkimukseen liittyy vastauskato, joka puolestaan saattaa jonkin verran alimitoittaa Tilastokeskuksen työttömyystietoa. Myös tilastoinnin ajankohdat poikkeavat toisistaan. Tilastokeskuksen tutkimushan kuvaa yleensä kuukauden 15. päivän sisältävää viikkoa. (Tilastokeskus, 1997.)

Satunnaisotokseen perustuvat tilastot sisältävät aina satunnaisuudesta johtuvaa otantavirhettä. Otantavirheen vaikutusta työvoimatutkimuksen lukuihin arvioidaan laskemalla kuukausi- ja neljännesvuositietojen *keskivirheitä* (standard error). Taulukossa 1.1 on esitetty neljännesvuositainen ja kuukausittainen keskivirhe (Tilastokeskus, 1995, 32). Keskivirheen ohella laatuvaatimuksena käytetään myös suhteellista mittaria, joka ilmoittaa keskivirheen suhteen estimaattiinsa eli prosenttista *variaatiokerrointa* (coefficient of variation).

Taulukko 1.1: Kuukausittaisen työvoimatutkimuksen kuukausi- ja neljännesvuositietojen keskivirheitä vuodelta 1994.

<b>1994</b>	<b>Estimaatti</b>	<b>Keskivirhe</b>	<b>%-variaatio- kerroin</b>
<b>1. neljännes</b>			
Työvoima	2 439 000	7 000	0,3
Työttömät	487 000	7 000	1,4
Työttömyysaste %	20,0	0,3	1,3
<b>Tammikuu</b>			
Työvoima	2 410 000	12 000	0,5
Työttömät	497 000	12 000	2,3
Työttömyysaste %	20,6	0,5	2,3

Vuositietojen tarkkuuden arviointi keskivirheen avulla on monimutkaista. Vuositiedot sisältävät kuitenkin vähemmän satunnaisvaihtelua kuin neljännesvuositiedot, jotka taas ovat tarkempia kuin kuukausittaiset tiedot. Vuosiestimaattien keskivirhe on pienempi, koska otoskoko vuodessa on suurempi kuin neljännesvu-

<sup>3</sup> Vuoden 1997 alussa tapahtunut työttömyysturvalainsäädännön muutos muutti ilmoittautumiskäytäntöjä ja pienensi päivittymisviivettä. (Tilastokeskus, 1997, 34.)

dessa ja kuukaudessa. Arvioitaessa kahden ajankohdan välillä tapahtuneiden muutosten satunnaisvaihtelua, on muutosten keskivirheelle laskettu yläraja-arvio. Se on pitkälti riippuvainen otospäällekkäisyydestä. Tilastokeskuksessa on laskettu, että esimerkiksi koko maan työvoimaluvussa jonkin neljänneksen ja edellisen vuoden vastaavan neljänneksen erotuksen on oltava yli 20 000 ja vastaavasti työttömyysluvussa yli 9 000, jotta muutokset olisivat tilastollisesti merkitseviä 90 prosentin luotettavuustasolla (1995, 31-33.). Laskelma perustuu ajatukseen riippumattomista satunnaisotoksista ja niiden luottamusväleistä.

EU-kriteerien mukaisessa kuukausittaisessa työvoimatutkimuksessa ollaan kiinnostuneita myös vuosittaisien estimaattien tason ja muutoksen tarkkuudesta sekä neljännesvuosiestimaattien muutoksen tarkkuudesta. Pelkät keskivirheiden yläraja-arviot eivät enää riitä tutkimuksen tarkkuuden arvioimiseksi. On tarpeen kehittää laskentamenetelmiä, joiden avulla muutoksen ja tason keskivirheet voidaan mitata. Laskentamenetelmien on oltava sellaisia, että ne ottavat tutkimuksen rotaatorakenteen huomioon.

## Luku 2

# Rotatoiva paneeliasetelma ja siihen liittyvät estimaattorit

### 2.1 Tutkimusasetelma

Rotatoivan paneelitutkimuksen tapauksessa sama henkilö on otoksessa mukana useampia kertoja. Kun henkilö jättää lopullisesti paneelin, hänen tilalleen otetaan uusi haastateltava. Tällöin otos vaihtuu asteittain siten, että peräkkäisinä kuukausina, vuosineljänneksinä ja vuosina voi olla päällekkyyttä eli osa vastaajista on samoja (Kish, 1995, 8). Duncanin ja Kaltonin mukaan paneelitutkimuksessa, jossa samoja henkilöitä haastatellaan useita kertoja, saattaa ongelmia aiheuttaa paneelikato. Samoin vastaustaakka koetaan usein ongelmana paneelitutkimuksen tapauksessa. Rotatoivassa paneeliasetelmassa paneelikadon ja vastausrasitteen aiheuttamat ongelmat ovat huomattavasti vähäisempiä, sillä se on sekoitus uusia haastateltavia ja jo aiemmin haastateltuja. Rotatoivan paneeliasetelman etuna on myös ajantasaisuus, joka saavutetaan käyttämällä riittävä määrä uusia haastateltavia kunakin tutkittavana ajankohtana (Duncan&Kalton, 1987, 103). Myös Kish on kiinnittänyt huomiota samoihin seikkoihin (1987, 153).

Duncanin ja Kaltonin mielestä tutkimusasetelman etuna voidaan pitää edullisuutta, sillä uusintahaastattelu on halvempi kuin alkuperäinen. Lisäksi päällekkyyden kannattaa maksimoida niillä tutkimusajankohdilla tai -jaksoilla, joilla halutaan parantaa sekä poikkileikkaustyyppistä että muutoksen estimointia. Tällöin rotatoiva paneeliasetelma tuottaa tarkemmat estimaatit kuin saman suuruiset riippumattomat otokset (1987, 103-104). Estimoinnin tehostuminen etenkin muutosta estimoidaessa perustuu peräkkäisten vastauskertojen tietojen hyödyntämiseen. Pakki-

sen ja Lehtosen mukaan apuinformaatiolla eli tässä tapauksessa ottamalla huomioon edellinen vastauskerta voidaan tehostaa valmiiksi poimitusta otoksesta suoritettavaa estimointia (1989, 88). Myös Särndalin mielestä on selvää, että edellisen haastattelukerran tietojen hyödyntäminen parantaa estimointia (1992, 370). Edellytyksenä on se, että tutkittavat muuttajat ovat suhteellisen vakaita, s.o. että peräkkäisten vastausten välillä on selvä (positiivinen) korrelaatio.

Työvoimatutkimuksessa poimitaan samanlaisina toistettavia otoksia kehikkopetusjoukosta, johon lukeutuu kunakin tutkimusajankohtana Suomessa asuva 15-74-vuotias väestö. Se jaetaan työmarkkina-aseman mukaisesti ryhmiin eli perusluokittelun mukaan työllisiin, työttömiin ja työvoimaan kuulumattomiin. Perusjoukko muuttuu yleensä tutkimusajankohtien välillä sekä lukumäärältään että rakenteeltaan. Uusina jäseninä siihen tulevat 15 vuotta täyttäneet henkilöt ja toisaalta 75 vuotta täyttäneet henkilöt poistuvat siitä. Muita syitä perusjoukon muuttumiseen ovat esimerkiksi kehikkoon kuuluvan henkilön kuolema, maahanmuutto tai maastamuutto. (Tilastokeskus, 1995.)

Yleisesti ottaen perusjoukon parametrien muutokset eri ajankohtina ovat seurausta siitä, että 1) perusjoukkoon kuuluva henkilö muuttaa **työmarkkina-asemaansa** tai 2) perusjoukko muuttuu **kooltaan** tai **rakenteeltaan**. (Kish, 1995, 8.)

### **Muutoksen ja tason estimointi rotatoivassa paneelissa**

Estimoitaessa perusjoukon parametrien muutosta eri ajankohtien välillä voidaan tutkimuksen paneelirakennetta hyödyntää laskettaessa varianssiestimaattoria. Saman henkilön eri ajankohtina antamat vastaukset korreloivat työvoimatutkimuksen tapauksessa selvästi. Seuraavassa tarkastellaan havaintoyksiköihin  $i$  ja  $j$  liittyvää muutoksen varianssia kahdessa eri tapauksessa, jotka ovat saman havaintoyksikön toistuva mittaus ja riippumattomien havaintojen peräkkäinen mittaus perusjoukossa  $U$ :

$$\begin{aligned} i \in U: \quad \text{var}(y_{it} - y_{i,t-1}) &= \text{var}(y_{it}) + \text{var}(y_{i,t-1}) - 2\text{kov}(y_{it}, y_{i,t-1}) \\ i \neq j \in U: \quad \text{var}(y_{it} - y_{j,t-1}) &= \text{var}(y_{it}) + \text{var}(y_{j,t-1}) \end{aligned} \quad (2.1)$$

Jos paneelin päällekkäiseen osaan kuuluvan henkilön eri haastatteluissa antamat vastaukset korreloivat positiivisesti, pienenee muutoksen varianssiestimaatin arvo.

Negatiivisen korrelaation tapauksessa muutoksen varianssiestimaatin arvo puolestaan kasvaa. Riippumattomien havaintojen peräkkäisten mittausten välillä ei luonnollisestikaan ole korrelaatiota, joten muutoksen varianssiestimaattori ei myöskään sisällä kovarianssitermiä.

Tasoa estimoitaessa saman henkilön eri vastauksertoina antamien vastausten välinen positiivinen korrelaatio aiheuttaa merkittävän tehokkuuden menetyksen suhteessa riippumattomiin otoksiin. Seuraavassa tarkastellaan havaintoyksiköiden  $i$  ja  $j$  tasoestimaattien varianssia paneeliasetelman ja riippumattomien otosten tapauksissa perusjoukossa  $U$ :

$$\begin{aligned} i \in U: \quad \text{var}\left(\frac{y_{it} - y_{i,t-1}}{2}\right) &= \frac{\text{var}(y_{it}) + \text{var}(y_{i,t-1}) + 2\text{kov}(y_{it}, y_{i,t-1})}{4} \\ i \neq j \in U: \quad \text{var}\left(\frac{y_{it} - y_{j,t-1}}{2}\right) &= \frac{\text{var}(y_{it}) + \text{var}(y_{j,t-1})}{4} \end{aligned} \quad (2.2)$$

Mikäli paneelin päällekkäiseen osaan kuuluvan henkilön eri vastauksertoina antamat vastaukset korreloivat positiivisesti, suurentaa positiivinen kovarianssitermi tason varianssiestimaatin arvoa. Negatiivinen korrelaatio puolestaan lisää tason estimoinnin tehokkuutta suhteessa riippumattomiin otoksiin, sillä negatiivinen kovarianssitermi pienentää tason varianssiestimaatin arvoa. Paneelin riippumattomassa osassa tason varianssiestimaattori ei sisällä kovarianssitermiä, koska eri henkilöiden vastaukset ovat toisistaan riippumattomia. (Cochran, 1977, 345.)

## 2.2 Rotaatiomallin kuvaus

Tilastokeskuksen kuukausittaisessa työvoimatutkimuksessa otos on jaettu kuukausittain viiteen eri rotaatioryhmään taulukossa 2.1 esitetyllä tavalla:

1. Tutkimuksessa ensimmäistä kertaa olevat henkilöt.
2. Toista kertaa tutkimuksessa olevien henkilöiden muodostama rotaatioryhmä. He olivat ensimmäisessä haastattelussa edellisellä vuosineljänneksellä.
3. Kolmatta kertaa haastateltavat henkilöt. Heidän toinen haastattelukertansa oli edellisellä vuosineljänneksellä.

4. Tutkimuksessa neljättä kertaa olevista henkilöistä koostuva rotaatioryhmä. Heidän ensimmäinen haastattelukertansa oli vuotta aiemmin ja edellinen haastattelukertansa kuusi kuukautta aiemmin.
5. Viidettä kertaa tutkimuksessa olevien henkilöiden muodostama rotaatioryhmä. He olivat neljännessä haastattelussa edellisellä vuosineljänneksellä.

Taulukko 2.1: Tilastokeskuksen kuukausittaisen työvoimatutkimuksen rotaatioryhmät tutkimuskerran mukaan kuukausittain vuosina 1994 ja 1995.

<i>Rotaatioryhmät vuonna 1994</i>												
Tutkimus- kerta	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
5.	241	242	243	311	312	313	321	322	323	331	332	333
4.	311	312	313	321	322	323	331	332	333	341	342	343
3.	331	332	333	341	342	343	<b>411</b>	412	413	421	422	423
2.	341	342	343	<b>411</b>	412	413	421	422	423	431	432	433
1.	<b>411</b>	412	413	421	422	423	431	432	433	441	442	443

<i>Rotaatioryhmät vuonna 1995</i>												
Tutkimus- kerta	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
5.	341	342	343	<b>411</b>	412	413	421	422	423	431	432	433
4.	<b>411</b>	412	413	421	422	423	431	432	433	441	442	443
3.	431	432	433	441	442	443	511	512	513	521	522	523
2.	441	442	443	511	512	513	521	522	523	531	532	533
1.	511	512	513	521	522	523	531	532	533	541	542	543

Tummennus osoittaa esimerkkinä yhden rotaatioryhmän (**411**) mukanaolon tutkimuksessa. Rotaatioryhmän identifioiva koodi muodostuu tiedoista, jotka ilmaisevat rotaatioryhmän tutkimukseen mukaantulo ajankohdan. Koodissa sadat ilmaisevat vuoden (1994), kymmenet vuosineljänneksen (**1.**) ja ykköset kuukauden (**1.**) kyseisen vuosineljänneksen sisällä. Rotaatioryhmät neljännesvuosittain on saatu yhdistämällä neljännesvuoden sisältämät kuukausittaiset rotaatioryhmät tutkimuskerran mukaan taulukossa 2.2 esitetyllä tavalla uusiksi rotaatioryhmiksi (the pooled samples). Yhdistäminen voidaan tehdä, koska peräkkäiset kuukaudet ovat toisistaan riippumattomia (Djerf, 1994, 24). Näin esimerkiksi vuoden 1994 tammi-maaliskuun rotaatioryhmät (411, 412 ja 413), jotka olivat ensimmäistä kertaa tut-



kimuksessa on yhdistetty ja ne ovat saaneet koodikseen **41** eli kyseessä on neljännesvuositason identifikaatio.

Taulukko 2.2: Tilastokeskuksen kuukausittaisen työvoimatutkimuksen yhdistetyt rotaatioryhmät tutkimuskerran mukaan neljännesvuosittain vuosina 1994 ja 1995.

Tutkimuskerta	1994				1995			
	I	II	III	IV	I	II	III	IV
5.	24	31	32	33	34	<b>41</b>	42	43
4.	31	32	33	34	<b>41</b>	42	43	44
3.	33	34	<b>41</b>	42	43	44	51	52
2.	34	<b>41</b>	42	43	44	51	52	53
1.	<b>41</b>	42	43	44	51	52	53	54

Tummennetut rotaatioryhmät osoittavat esimerkkinä yhden yhdistetyn rotaatioryhmän (**41**) mukanaolon tutkimuksessa. Taulukosta 2.2 voidaan havaita, että peräkkäisinä vuosineljänneksinä **3/5** vastaajista on samoja. Jos tarkastellaan ensimmäistä ja toista vuosineljänneistä vuonna 1994, niin kumpanakin vuosineljänneksenä esiintyvät rotaatioryhmät 31, 34 ja 41. Vuoden kuluttua otosten päällekkyyks on **2/5**. Tarkasteltaessa ensimmäisiä vuosineljänneksiä vuosina 1994 ja 1995 huomataan, että kumpanakin vuosineljänneksenä esiintyvät rotaatioryhmät 34 ja 41.

## 2.3 Estimaattorin kuvaus tutkittavana ajankohtana

Työvoimatutkimuksessa pyritään selvittämään taloudellisesti aktiivisen väestön jakaumista taloudellisen aktiviteetin mukaan. Niinpä onkin tärkeää estimoida työmarkkina-aseman mukaisten perusluokiteltujen ryhmien eli työllisten, työttömien ja työvoiman ulkopuolisten kokonaismääriä tutkittavana ajankohtana. Jos estimoidaan esimerkiksi työttömien kokonaismäärää vuosineljänneksellä  $t$ , merkitään muuttujalle  $y_{jt}$  arvo

- 1, jos henkilö  $j$  on työtön vuosineljänneksellä  $t$ .
- 0, muulloin.

Työttömien kokonaismääräksi vuosineljänneksellä  $t$  saadaan:

$$Y_t = \sum_{j=1}^{N_t} y_{jt} \quad (2.3)$$

jossa  $N_t$  on tutkimuksen perusjoukkoon kuuluvat 15-74 vuotiaat Suomessa asuvat. (Cochran, 1977, 50-51.)

Työttömien kokonaismäärää estimoitaessa voidaan käyttää seuraavaa Horvitz-Thompson -tyyppistä estimaattoria (**HT-estimaattori**) edellyttäen, että systemaattinen otantamenettely palautuu työvoimatutkimuksen tapauksessa yksinkertaiseen satunnaisotantaan palauttamatta (**SRSWOR**). Samoin otoksen jälkiosittaminen sukupuolen, viisivuotiskäryhmän ja asuinläänin väkilukutietojen suhteen jätetään huomioimatta, koska jälkiositusmenettelyn vaikutus kohdistuu etupäässä katoharhan pienentämiseen. (Djerf, 1994, 24.)

$$\hat{Y}_t = \frac{N_t}{n_t} \sum_{j=1}^{n_t} y_{jt} \quad (2.4)$$

Työttömien kokonaismäärän estimaatti paneelin vuosineljänneksellä  $t$  saadaan siis tietyn henkilön sisällymistodennäköisyyden käänteisluvun ja otoksessa olevien työttömien kokonaismäärän tulona.

## 2.4 Muutoksen estimointi

### 2.4.1 Nykyestimaattori muutokselle

Neljännesvuosimuutosta estimoitaessa ollaan kiinnostuneita tutkittavien perusjoukon parametrien muutoksesta kahden peräkkäisen vuosineljänneksen välillä. Nykyestimaattori, joka olettaa tutkimuksen kohteena olevat vuosineljännekset toisistaan riippumattomiksi, ei hyödynnä rotatoivan paneelitutkimuksen rakennetta optimaalisella tavalla eikä siten huomioi tutkimukseen kuuluvien henkilöiden edellisillä haastattelukerroilla antamia tietoja.

Nykyestimaattori neljännesvuosimuutokselle eli kahden peräkkäisen vuosineljänneksen väliselle muutokselle  $D_t = Y_t - Y_{t-1}$  on muotoa (vrt. Kish, 1987, 162):

$$\hat{D}_t = \hat{Y}_t - \hat{Y}_{t-1} \quad (2.5)$$

Neljännesvuosimuutoksen varianssiestimaattori saadaan, kuten kaavan (2.1) riippumattomien otosten tapauksessa:

$$\text{var}(\hat{D}_t) = \text{var}(\hat{Y}_t) + \text{var}(\hat{Y}_{t-1}) \quad (2.6)$$

Vuosimuutosta estimoitaessa ollaan kiinnostuneita perusjoukon parametrien muutoksesta tutkittavan vuosineljänneksen ja edeltävän kalenterivuoden vastaavan vuosineljänneksen välillä, joten nykyestimaattoriksi vuosimuutokselle  $D_t = Y_t - Y_{t-4}$  saadaan:

$$\hat{D}_t = \hat{Y}_t - \hat{Y}_{t-4} \quad (2.7)$$

Neljännesvuosimuutoksen varianssin estimaattori saadaan myös, kuten kaavan (2.1) riippumattomien vuosineljännesten tapauksessa:

$$\text{var}(\hat{D}_t) = \text{var}(\hat{Y}_t) + \text{var}(\hat{Y}_{t-4}) \quad (2.8)$$

## 2.4.2 Vaihtoehtoinen estimaattori muutokselle

Muutosta estimoitaessa kuukausittaisen työvoimatutkimuksen nykyinen laskentatapa ei ota rotaatioasetelman kaikkia riippuvuuksia huomioon. Riippuvuuksia on olemassa niin peräkkäisten neljännesvuosien kuin vuosienkin välillä. Peräkkäiset tutkimuskuukaudet ovat sitävastoin toisistaan riippumattomia. Muutosta estimoitaessa tutkimuksessa käytetty rotaatorakenne voidaan ottaa huomioon jakamalla paneeli tarkasteltavien vuosineljänneksien osalta sekä **päällekkäisiin (p)** että **riippumattomiin (q)** osiin (Cochran, 1977, 346).

Tutkittavan osajoukon kokonaismäärä voidaan siten estimoida erikseen sekä paneelin päällekkäisessä että riippumattomassa osassa (Madsen, 1996, 3):

$$\hat{Y}_{tp} = \frac{N_t}{n_{tp}} \sum_{j=1}^{n_{tp}} y_{jtp} \quad (2.9)$$

$$\hat{Y}_{tq} = \frac{N_t}{n_{tq}} \sum_{j=1}^{n_{tq}} y_{jtq}$$

Kokonaismäärän estimaattoreista (2.9) muodostettua yhdistelmäestimaattoria eli päällekkäisten ja riippumattomien osan **linearikombinaatiota** voidaan käyttää  $Y_t$ :n estimaattorina (vrt. Cochran, 1977, 346):

$$\hat{Y}_t = \phi \hat{Y}_{tp} + (1 - \phi) \hat{Y}_{tq}, \quad 0 \leq \phi \leq 1 \quad (2.10)$$

Vaihtoehtoinen estimaattori neljännesvuosimuutokselle, joka estimoitaessa kahden vuosineljänneksen  $t-1$  ja  $t$  välistä muutosta käyttää hyväkseen muutoksen lineaarikombinaatiota sekä paneelin päällekkäisissä että riippumattomissa osissa. (Madsen, 1996, 3):

$$\begin{aligned} \hat{D}_t(\phi) &= \left( \phi \hat{Y}_{tp} + (1 - \phi) \hat{Y}_{tq} \right) - \left( \phi \hat{Y}_{t-1,p} + (1 - \phi) \hat{Y}_{t-1,q} \right) \\ &= \phi (\hat{Y}_{tp} - \hat{Y}_{t-1,p}) + (1 - \phi) (\hat{Y}_{tq} - \hat{Y}_{t-1,q}) \\ &= \phi \hat{D}_{tp} + (1 - \phi) \hat{D}_{tq} \end{aligned} \quad (2.11)$$

Työvoimatutkimuksen rotaatorakenteen hyödyntävä neljännesvuosimuutoksen estimaattori saadaan siis paneelin päällekkäisten ja riippumattomien osien muutosten **yhdistelmäestimaattorina** (composite estimator). Muutoksen estimaattori paneelin päällekkäisessä osassa perustuu samojen henkilöiden antamiin vastauksiin, kun taas muutoksen estimaattori paneelin riippumattomassa osassa perustuu eri henkilöiltä saatuihin vastauksiin. Vaihtoehtoinen estimaattori vuosimuutokselle saadaan korvaamalla vuosineljännes  $t-1$  vuosineljänneksellä  $t-4$ .

Madsenin mukaan neljännesvuosimuutoksen varianssiestimaattori voidaan esittää seuraavassa muodossa (1996, 3):

$$\text{var}\left(\hat{D}_t(\phi)\right) = \phi^2 \text{var}\left(\hat{D}_{tp}\right) + (1-\phi)^2 \text{var}\left(\hat{D}_{tq}\right) \quad (2.12)$$

jossa

$$\text{var}\left(\hat{D}_{tp}\right) = \text{var}\left(\hat{Y}_{tp}\right) + \text{var}\left(\hat{Y}_{t-1,p}\right) - 2\text{kov}\left(\hat{Y}_{tp}, \hat{Y}_{t-1,p}\right)$$

$$\text{var}\left(\hat{D}_{tq}\right) = \text{var}\left(\hat{Y}_{tq}\right) + \text{var}\left(\hat{Y}_{t-1,q}\right)$$

Paneelin päällekkäisen osan muutoksen varianssiestimaattoria laskettaessa on huomioitava paneelin päällekkäiseen osaan kuuluvien samoilta henkilöiltä saatujen vastausten välillä yleensä oleva positiivinen korrelaatio. Korrelaatioiden huomioiminen tapahtuu kovarianssitermin avulla, kuten kaavassa (2.1) havaittiin. Riippumattoman osan muutoksen varianssiestimaattoria laskettaessa ei kovarianssitermiä sen sijaan tarvita, koska siinä vastaukset on saatu eri henkilöiltä.

Neljännesvuosimuutoksen tapauksessa kertoimet  $\phi$  ja  $1-\phi$  saadaan **minimoimalla** varianssiestimaattori (vrt. Cochran, 1977, 346). Kun **derivoidaan varianssiestimaattori  $\phi$ :n suhteen ja asetetaan derivaatta nolaksi**, saadaan  $\phi$ :n ja  $1-\phi$ :n optimaaliset arvot:

$$\frac{\partial \text{var}\left(\hat{D}_t(\phi)\right)}{\partial \phi} = 0 \quad \Rightarrow \quad 2\phi \text{var}\left(\hat{D}_{tp}\right) - 2(1-\phi) \text{var}\left(\hat{D}_{tq}\right) = 0 \quad \Rightarrow$$

$$\phi = \frac{1}{1 + \frac{\text{var}\left(\hat{D}_{tp}\right)}{\text{var}\left(\hat{D}_{tq}\right)}} \quad \text{ja} \quad 1-\phi = \frac{1}{1 + \frac{\text{var}\left(\hat{D}_{tq}\right)}{\text{var}\left(\hat{D}_{tp}\right)}} \quad (2.13)$$

Päällekkäisten ja riippumattomien osien muutoksen varianssiestimaattoreiden osalta oletetaan, että populaation koko ja otoskoot ovat vakioita kaikilla tutkimuskerroilla. Samoin oletetaan, että otoskoko on vakio myös kaikilla rotaatioryhmillä. Lisäksi paneelin päällekkäisiä (p) ja riippumattomia (q) osia kuvaavien lukujen oletetaan olevan vakioita. (Madsen, 1996, 3-4.)

Kuukausittaisessa työvoimatutkimuksessa poimitaan samanlaisina toistettavia otoksia kehikkoperusjoukosta, jonka muodostaa Suomessa asuva 15-74-vuotias väestö. Populaation koko kasvaa hieman tutkimusajankohtien välillä, koska uusi- ja jäseninä tutkimukseen tulevia 15 vuotta täyttäneitä henkilöitä on enemmän kuin 75 vuotta täyttäneitä tutkimuksesta poistuvia henkilöitä. Populaation koko kasvoi esimerkiksi vuosina 1994-95 keskimäärin noin 1 400 henkilöllä kuukaudessa. Työvoimatutkimuksen otoskoko on kaikkina tutkimuskuukausina noin 12 000 henkilöä, joten kunkin vuosineljänneksen otokseen kuuluu suunnilleen 36 000 työkäistä suomalaista. Otoksessa haastateltavat jakautuvat tasaisesti kaikkiin viiteen eri rotaatioryhmään eli kunkin rotaatioryhmän otoskoko on noin 2 500 henkeä. Paneelin päällekkäisiä (p) ja riippumattomia (q) osia kuvaavat luvut ovat vakioita niin neljännesvuosimuutoksen ( $p=3/5$ ,  $q=2/5$ ) kuin vuosimuutoksenkin ( $p=2/5$ ,  $q=3/5$ ) tapauksessa.

Paneelin eri osien muutoksen varianssiestimaattoreiden lausekkeet voidaan edellä esitetyt oletukset huomioiden kirjoittaa myös seuraavassa muodossa (Madsen, 1996, 4):

$$\begin{aligned}\text{var}(\hat{D}_p) &= \text{var}(\hat{Y}_p) + \text{var}(\hat{Y}_{t-1,p}) - 2\text{kov}(\hat{Y}_p, \hat{Y}_{t-1,p}) \\ &= 2 \text{var}(\hat{Y}_p) - 2\text{kov}(\hat{Y}_p, \hat{Y}_{t-1,p})\end{aligned}\tag{2.14}$$

$$\begin{aligned}\text{var}(\hat{D}_{tq}) &= \text{var}(\hat{Y}_{tq}) + \text{var}(\hat{Y}_{t-1,q}) \\ &= 2 \text{var}(\hat{Y}_q)\end{aligned}$$

Tällöin oletetaan lisäksi, että paneelin päällekkäisten ja riippumattomien osien varianssiestimaattoreiden varianssit ovat vakioita kaikilla tutkimuskerroilla:

$$\text{var}(\hat{Y}_p) = \text{var}(\hat{Y}_{t-1,p}) = \text{var}(\hat{Y}_p)\tag{2.15}$$

$$\text{var}(\hat{Y}_{tq}) = \text{var}(\hat{Y}_{t-1,q}) = \text{var}(\hat{Y}_q)$$

Paneelin eri osien varianssiestimaattoreiksi saadaan kaikkien mainittujen oletusten ollessa voimassa (vrt. Cochran, 1977, 52):

$$\text{var}(\hat{Y}_{\text{tp}}) = \text{var}(\hat{Y}_{t-1,p}) = \text{var}(\hat{Y}_p) = N^2 \frac{N - n_p}{n_p N} s^2 \quad (2.16)$$

$$\text{var}(\hat{Y}_{\text{tq}}) = \text{var}(\hat{Y}_{t-1,q}) = \text{var}(\hat{Y}_q) = N^2 \frac{N - n_q}{n_q N} s^2$$

Riippumattoman osan muutoksen varianssiestimaattori voidaan kaikki oletukset huomioiden kirjoittaa seuraavassa muodossa:

$$\text{var}(\hat{D}_{\text{tq}}) = 2N^2 \frac{N - n_q}{n_q N} s^2 \quad (2.17)$$

jossa varianssitermi määräytyy otantasuhteiden perusteella (Cochran, 1977, 52):

$$s^2 = \frac{n_q}{n_q - 1} \left( \frac{\sum_{j=1}^{n_q} y_j}{n_q} \right) \left( 1 - \frac{\sum_{j=1}^{n_q} y_j}{n_q} \right) \quad (2.18)$$

Muutoksen estimaattorin varianssiestimaattori päällekkäiselle osalle muodostetaan eri tavalla johtuen kovarianssitermistä. Kovarianssihan aiheutuu siitä, että paneelin päällekkäisessä osassa ovat kyseessä samat henkilöt molemmilla tarkasteltavilla vuosineljänneksillä.

$$\begin{aligned} \hat{D}_{\text{tp}} &= \hat{Y}_{\text{tp}} - \hat{Y}_{t-1,p} \\ &= \frac{N_t}{n_{\text{tp}}} \sum_{j=1}^{n_p} y_{\text{tpj}} - \frac{N_{t-1}}{n_{t-1,p}} \sum_{j=1}^{n_p} y_{t-1,pj} \\ &\cong \frac{N}{n_{\text{tp}}} \sum_{j=1}^{n_p} (y_{\text{tpj}} - y_{t-1,pj}) \\ &= \frac{N}{n_{\text{tp}}} \sum_{j=1}^{n_p} d_{\text{tpj}} \end{aligned} \quad (2.19)$$

jossa **indikaattorimuuttuja d** osoittaa muutosta vuosineljännesten t-1 ja t välillä päällekkäisissä osissa p. Vuosimuutoksen tapauksessa indikaattorimuuttuja d osoittaa muutosta vuosineljännesten t-4 ja t välillä. Muuttuja d saa arvon

- -1, jos henkilö j jättää osajoukon.
- 0, jos henkilön j status ei muutu.
- 1, jos henkilö j tulee osajoukkoon.

Kun sovelletaan vaihtoehtoista estimaattoria kuukausittaisen työvoimatutkimuksen rotaatorakenteeseen, saadaan päällekkäisen osan varianssiestimaattoriksi:

$$\text{var}(\hat{D}_p) = N^2 \frac{N - n_p}{n_p N} s_{d_p}^2 \quad (2.20)$$

jossa varianssitermi kuvaa indikaattorimuuttuja d:n vaihtelua otoksessa (vrt. Pahkinen&Lehtonen, 1989, 25):

$$s_{d_p}^2 = \frac{1}{n_p - 1} \sum_{j=1}^{n_p} (d_{pj} - \bar{d}_p)^2 \quad (2.21)$$

Indikaattorimuuttajaa d käytetään siis laskettaessa muutoksen varianssiestimaattoria paneelin päällekkäisessä osassa, jossa saman henkilön eri ajankohtina antamien vastausten välillä on yleensä selvä positiivinen korrelaatio. Indikaattorimuuttujan käytön perustelu voidaan johtaa lähtien kaavan (2.1) yleisestä tilanteesta, jossa kovarianssitermit ovat mukana pienentämässä muutoksen varianssiestimaattoria. Jos paneelin päällekkäisessä osassa saman henkilön eri ajankohtina antamien vastausten välillä on negatiivista korrelaatiota, niin kaavan kovarianssitermin etumerkki vaihtuu negatiivisesta positiiviseksi. Siinä tapauksessa kovarianssitermit siis suurettavat muutoksen varianssiestimaattoria.

Seuraavassa indikaattorimuuttujan d käyttö perustellaan vain neljännesvuosimuutoksen estimoinnin tapauksessa, mutta indikaattorimuuttujan käyttö voidaan perustella vastaavalla tavalla soveltuvaksi myös vuosimuutoksen estimointiin. Vuosimuutosta estimoidessa vuosineljännes t-1 korvataan vuosineljänneksellä t-4, mutta muilta osin seuraavana oleva perustelu pysyy ennallaan.



**Todistus.** Indikaattorimuuttaja  $d$ :n käyttö laskettaessa muutoksen varianssiestimaattoria paneelin päällekkäisessä osassa:

$$\begin{aligned}
\text{var}(\hat{D}_{\text{tp}}) &= \text{var}\left(\hat{Y}_{\text{tp}} - \hat{Y}_{t-1,p}\right) \\
&= \text{var}(\hat{Y}_{\text{tp}}) + \text{var}(\hat{Y}_{t-1,p}) - 2\text{kov}(\hat{Y}_{\text{tp}}, \hat{Y}_{t-1,p}) \\
&= N^2 \frac{N - n_p}{n_p N} \frac{1}{n_p - 1} \sum_{j=1}^n \left[ (y_{\text{tpj}} - \bar{y}_{\text{tp}})^2 + (y_{t-1,pj} - \bar{y}_{t-1,p})^2 - 2(y_{\text{tpj}} - \bar{y}_{\text{tp}})(y_{t-1,pj} - \bar{y}_{t-1,p}) \right] \\
&= N^2 \frac{N - n_p}{n_p N} \frac{1}{n_p - 1} \sum_{j=1}^n \left[ (y_{\text{tpj}} - y_{t-1,pj})^2 - 2(y_{\text{tpj}} - y_{t-1,pj})(\bar{y}_{\text{tp}} - \bar{y}_{t-1,p}) + (\bar{y}_{\text{tp}} - \bar{y}_{t-1,p})^2 \right] \\
&= N^2 \frac{N - n_p}{n_p N} \frac{1}{n_p - 1} \sum_{j=1}^n (d_{\text{pj}}^2 - 2d_{\text{pj}} \cdot \bar{d}_p + \bar{d}_p^2) \\
&= N^2 \frac{N - n_p}{n_p N} \frac{1}{n_p - 1} \sum_{j=1}^n (d_{\text{pj}} - \bar{d}_p)^2 \\
&= N^2 \frac{N - n_p}{n_p N} S_{d_p}^2
\end{aligned}$$

## 2.5 Tason estimointi

### 2.5.1 Nykyestimaattori vuositasoa estimoidessa

Mielenkiinnon kohteena on estimoida tutkittavan osajoukon kalenterivuoden ajalta saatujen vastausten keskiarvoa:

$$\bar{Y} = \frac{Y_1 + \dots + Y_{20}}{20} = \frac{1}{20} \sum_{t=1}^{20} Y_t \quad (2.22)$$

Tällöin voidaan estimoida esimerkiksi työttömien osajoukon rotaatioryhmittäisten kokonaismäärien keskiarvo. Yhden kalenterivuoden aikana Tilastokeskuksen kuukausittaisessa työvoimatutkimuksessa on mukana 20 rotaatioryhmää. Nykyestimaattori ei hyödynnä rotaativan paneelin rakennetta optimaalisella tavalla, koska se olettaa kaikkien rotaatioryhmien olevan toisistaan riippumattomia. Tasoa estimoidessa saman henkilön eri vastauksena antamien vastausten välillä oleva

positiivinen korrelaatio suurentaa varianssiestimaattoria, joka on suurempi kuin toisistaan riippumattomien rotaatioryhmien tapauksessa. Vastausten välinen negatiivinen korrelaatio puolestaan pienentää varianssin estimaattoria. (Madsen, 1996, 1.)

Nykyestimaattori, joka olettaa rotaatioryhmät toisistaan riippumattomiksi, voidaan ilmaista seuraavalla lausekkeella. Siinä tutkittavan osajoukon kokonaismäärät on estimoitu erikseen kaikissa 20 rotaatioryhmässä ja vuositason estimaattori muodostuu kyseisten ryhmien keskiarvosta:

$$\hat{\bar{Y}} = \frac{\hat{Y}_1 + \dots + \hat{Y}_{20}}{20} = \frac{1}{20} \sum_{t=1}^{20} \hat{Y}_t \quad (2.23)$$

Nykyestimaattorin vuositason varianssiestimaattori saadaan, kuten kaavan (2.1) riippumattomien otosten tapauksessa:

$$\text{var}(\hat{\bar{Y}}) = \frac{\text{var}(\hat{Y}_1) + \dots + \text{var}(\hat{Y}_{20})}{20^2} \quad (2.24)$$

### 2.5.2 Vaihtoehtoiset estimaattorit vuositasoa estimoitaessa

Tarkastelussa on ollut mukana kahdenlaisia vaihtoehtoisia estimaattoreita, jotka kumpikin pyrkivät hyödyntämään kuukausittaisen työvoimatutkimuksen paneelirakenteen. Ensimmäinen tarkasteltavista vaihtoehtoisista estimaattoreista on pitkälti nykyestimaattorin kaltainen eikä siis varsinainen yhdistelmäestimaattori. Se poikkeaa nykyestimaattorista ainoastaan siten, että varianssiestimaattoria laskettaessa otetaan riippuvuudet eli kaikki mahdolliset kovarianssitermit huomioon. Vaihtoehtoisen **ei-yhdistelmäestimaattorin** tuottamat vuositason estimaatit eivät poikkeaa nykyestimaattorin tuottamista estimaateista, mutta sen varianssiestimaattori on huomattavasti suurempi johtuen kovarianssitermien huomioimisesta. Vuositasolla 90% rotaatioryhmistä on sellaisia, joissa on kovarianssitermi mukana.

Toinen tarkasteltavista vaihtoehtoisista estimaattoreista painottaa rotaatioryhmiä erilaisten kertoimien avulla eli kyseessä on **yhdistelmäestimaattori**. Paneelin

päällekkäiseen osaan kuuluville rotaatioryhmille annetaan eri kertoimet kuin riippumattomaan osaan kuuluville rotaatioryhmille. Tästä seuraa, että yhdistelmäestimaattorin varianssiestimaattori tulee pienemmäksi kuin kertoimia hyödyntämättömän vaihtoehdoisen ei-yhdistelmäestimaattorin varianssin estimaattori.

Tilastokeskuksen kuukausittaisen työvoimatutkimuksen rotatoivassa paneelissa on kalenterivuoden aikana kaikkiaan 20 rotaatioryhmää, kuten taulukosta 2.3 havaitaan.

Taulukko 2.3: Tilastokeskuksen kuukausittaisen työvoimatutkimuksen neljännesvuositasolle identifioidut rotaatioryhmät vuosineljänneksittäin vuonna 1995.

1. neljännes	34	43	41	44	51									
2. neljännes			41	44	51	42	52							
3. neljännes					51	42	52	43	53					
4. neljännes							52	43	53	44	54			

Taulukko 2.4: Tilastokeskuksen kuukausittaisen työvoimatutkimuksen neljännesvuositasolle identifioidut rotaatioryhmät tutkimuskerran mukaan vuonna 1995.

1. haastattelu	2. haastattelu	3. haastattelu	4. haastattelu	5. haastattelu
54				
53	53			
51	51	51		
52	52	52		
	44	44	44	
		43	43	43
			41	41
			42	42
				34

Edellä olevasta taulukosta 2.4 huomataan, että päällekkäisiä rotaatioryhmiä joko peräkkäisinä vuosineljänneksinä tai pidemmällä aikavälillä on kaikkiaan 18. Riip-

pumattomia rotaatioryhmiä on kaksi eli rotaatioryhmät 34 ja 54, jotka kuuluvat vuositasoa estimoitaessa paneelin riippumattomaan osaan. Kaikkien vuositasolla tarkasteltavien rotaatioryhmien estimaattorit ovat taulukossa 2.5. Niitä tarvitaan nykyestimaattoria ja vaihtoehtoisia estimaattoreita konstruoitaessa.

Taulukko 2.5: Tilastokeskuksen kuukausittaisen työvoimatutkimuksen neljännesvuositasolle identifioidut rotaatioryhmien estimaattorit tutkimuskerran mukaan vuonna 1995.

1. haastattelu	2. haastattelu	3. haastattelu	4. haastattelu	5. haastattelu
$\hat{Y}_{14q}^{(54)}$				
$\hat{Y}_{13p}^{(53)}$	$\hat{Y}_{24p}^{(53)}$			
$\hat{Y}_{11p}^{(51)}$	$\hat{Y}_{22p}^{(51)}$	$\hat{Y}_{33p}^{(51)}$		
$\hat{Y}_{12p}^{(52)}$	$\hat{Y}_{23p}^{(52)}$	$\hat{Y}_{34p}^{(52)}$		
	$\hat{Y}_{21p}^{(44)}$	$\hat{Y}_{32p}^{(44)}$	$\hat{Y}_{44p}^{(44)}$	
		$\hat{Y}_{31p}^{(43)}$	$\hat{Y}_{43p}^{(43)}$	$\hat{Y}_{54p}^{(43)}$
			$\hat{Y}_{41p}^{(41)}$	$\hat{Y}_{52p}^{(41)}$
			$\hat{Y}_{42p}^{(42)}$	$\hat{Y}_{53p}^{(42)}$
				$\hat{Y}_{51q}^{(34)}$

Rotaatioryhmien estimaattorit yksilöidään erilaisten ylä- ja alaindeksien avulla. Estimaattorin yläindeksinä oleva luku on kyseisen rotaatioryhmän koodi, kun rotaatioryhmä on neljännesvuositasolle identifioitu. Alaindeksi puolestaan kertoo, mikä haastattelukerta ja vuosineljännes on kyseessä. Esimerkiksi rotaatioryhmä 54 on vuoden 1995 aikana tutkimuksessa mukana yhden kerran ja se lasketaan vuositasoa estimoitaessa kuuluvaksi riippumattomiin rotaatioryhmiin. Kyseessä on tällöin ensimmäinen haastattelukerta ja tutkimusajankohtana neljäs vuosineljännes. Rotaatioryhmä 52 puolestaan on tutkimuksessa mukana kolme kertaa vuoden 1995 aikana ja se lasketaan vuositasoa estimoitaessa kuuluvaksi päällekkäisiin rotaatioryhmiin. Ensimmäinen haastattelukerta on toisella vuosineljänneksellä, toinen haastattelu kolmannella vuosineljänneksellä ja kolmas haastattelu neljännellä vuosineljänneksellä.

Paneelin kalenterivuoden aikana sisältämien päällekkäisten ja riippumattomien rotaatioryhmien lukumäärät on otettava huomioon laskettaessa vuositason yhdistelmäestimaattoria. Yhdistelmäestimaattori muodostetaan samaan tapaan kuin muutoksen estimoinnissakin eli kyseessä on paneelin päällekkäisen ja riippumattoman osan yhdistävä lineaarikombinaatio. (vrt. Madsen, 1996, 5.)

$$\begin{aligned}\hat{\bar{Y}}(\phi) &= \frac{1}{18} \left[ \phi \left( \hat{Y}_{13p}^{(53)} + \hat{Y}_{24p}^{(53)} + \hat{Y}_{11p}^{(51)} + \hat{Y}_{22p}^{(51)} + \hat{Y}_{33p}^{(51)} + \hat{Y}_{12p}^{(52)} + \hat{Y}_{23p}^{(52)} + \hat{Y}_{34p}^{(52)} + \right. \right. \\ &\quad \hat{Y}_{21p}^{(44)} + \hat{Y}_{32p}^{(44)} + \hat{Y}_{44p}^{(44)} + \hat{Y}_{31p}^{(43)} + \hat{Y}_{43p}^{(43)} + \hat{Y}_{54p}^{(43)} + \\ &\quad \left. \left. \hat{Y}_{41p}^{(41)} + \hat{Y}_{52p}^{(41)} + \hat{Y}_{42p}^{(42)} + \hat{Y}_{53p}^{(42)} \right) \right] + \\ &\quad \frac{1}{2} \left[ (1-\phi) \left( \hat{Y}_{14q}^{(54)} + \hat{Y}_{51q}^{(34)} \right) \right] \\ &= \frac{1}{18} \phi \sum \hat{Y}_{\wp} + \frac{1}{2} (1-\phi) \sum \hat{Y}_{\text{tq}}\end{aligned}\tag{2.25}$$

Yhdistelmäestimaattorissa päällekkäisten rotaatioryhmien kokonaismäärän estimaattoreiden summa siis jaetaan päällekkäisten rotaatioryhmien lukumäärällä ja riippumattomien rotaatioryhmien kokonaismäärän estimaattoreiden summa riippumattomien rotaatioryhmien lukumäärällä.

Myös vuositason yhdistelmäestimaattorin varianssiestimaattoria laskettaessa on huomioitava päällekkäisten ja riippumattomien rotaatioryhmien lukumäärät (vrt. Madsen, 1996, 6):

$$\text{var}(\hat{\bar{Y}}(\phi)) = \frac{1}{18^2} \phi^2 \text{var}\left(\sum \hat{Y}_{\wp}\right) + \frac{1}{2^2} (1-\phi)^2 \text{var}\left(\sum \hat{Y}_{\text{tq}}\right)\tag{2.26}$$

Päällekkäisten rotaatioryhmien kokonaismäärän varianssiestimaattoreiden summa jaetaan neliöidyllä päällekkäisten rotaatioryhmien lukumäärällä. Vastaavalla tavalla riippumattomien rotaatioryhmien kokonaismäärän varianssiestimaattoreiden summa jaetaan neliöidyllä riippumattomien rotaatioryhmien lukumäärällä.

Kun vuositason yhdistelmäestimaattorin varianssiestimaattoria muodostettaessa käytettäviä paneelin päällekkäisen osan kokonaismäärän varianssiestimaattoreita lasketaan kaavan 2.27 mukaisesti, on huomioitava paneelin päällekkäiseen osaan kuuluvien samoilta henkilöiltä saatujen vastausten välillä yleensä oleva positiivinen korrelaatio. Päällekkäiseen osaan kuuluu neljä sellaista neljännesvuositasolle

identifioitua rotaatioryhmää (51, 52, 44 ja 43), jotka ovat tutkimuksessa mukana kolme kertaa vuoden 1995 aikana. Lisäksi päällekkäiseen osaan kuuluu kolme rotaatioryhmää (53, 41 ja 42), joita on haastateltu kahdesti vuonna 1995. Korrelaatioiden huomioiminen tapahtuu kovarianssitermien avulla, kuten kaavassa (2.2) havaittiin. Jos henkilöiden antamat vastaukset korreloivat positiivisesti, niin kovarianssitermin etumerkki on positiivinen. Negatiivisen korrelaation tapauksessa kovarianssitermin etumerkki vaihtuu negatiiviseksi.

Riippumattoman osan kokonaismäärän varianssiestimaattoreita laskettaessa ei kovarianssitermiä sen sijaan tarvita, koska siinä eri henkilöiltä saatujen vastausten välillä ei ole korreloituneisuutta. Riippumattomaan osaan kuuluvat rotaatioryhmät (34 ja 54) ovat yhden kerran mukana tutkimuksessa vuoden 1995 aikana.

$$\begin{aligned}
& \text{var}\left(\sum \hat{Y}_{tp}\right) = \\
& \text{var}(\hat{Y}_{13p}^{(53)}) + \text{var}(\hat{Y}_{24p}^{(53)}) + 2\text{kov}(\hat{Y}_{13p}^{(53)}, \hat{Y}_{24p}^{(53)}) + \\
& \text{var}(\hat{Y}_{41p}^{(41)}) + \text{var}(\hat{Y}_{52p}^{(41)}) + 2\text{kov}(\hat{Y}_{41p}^{(41)}, \hat{Y}_{52p}^{(41)}) + \\
& \text{var}(\hat{Y}_{42p}^{(42)}) + \text{var}(\hat{Y}_{53p}^{(42)}) + 2\text{kov}(\hat{Y}_{42p}^{(42)}, \hat{Y}_{53p}^{(42)}) + \\
& \text{var}(\hat{Y}_{11p}^{(51)}) + \text{var}(\hat{Y}_{22p}^{(51)}) + \text{var}(\hat{Y}_{33p}^{(51)}) + \\
& 2\text{kov}(\hat{Y}_{11p}^{(51)}, \hat{Y}_{22p}^{(51)}) + 2\text{kov}(\hat{Y}_{11p}^{(51)}, \hat{Y}_{33p}^{(51)}) + 2\text{kov}(\hat{Y}_{22p}^{(51)}, \hat{Y}_{33p}^{(51)}) + \\
& \text{var}(\hat{Y}_{12p}^{(52)}) + \text{var}(\hat{Y}_{23p}^{(52)}) + \text{var}(\hat{Y}_{34p}^{(52)}) + \\
& 2\text{kov}(\hat{Y}_{12p}^{(52)}, \hat{Y}_{23p}^{(52)}) + 2\text{kov}(\hat{Y}_{12p}^{(52)}, \hat{Y}_{34p}^{(52)}) + 2\text{kov}(\hat{Y}_{23p}^{(52)}, \hat{Y}_{34p}^{(52)}) + \\
& \text{var}(\hat{Y}_{21p}^{(44)}) + \text{var}(\hat{Y}_{32p}^{(44)}) + \text{var}(\hat{Y}_{44p}^{(44)}) + \\
& 2\text{kov}(\hat{Y}_{21p}^{(44)}, \hat{Y}_{32p}^{(44)}) + 2\text{kov}(\hat{Y}_{21p}^{(44)}, \hat{Y}_{44p}^{(44)}) + 2\text{kov}(\hat{Y}_{32p}^{(44)}, \hat{Y}_{44p}^{(44)}) + \\
& \text{var}(\hat{Y}_{31p}^{(43)}) + \text{var}(\hat{Y}_{43p}^{(43)}) + \text{var}(\hat{Y}_{54p}^{(43)}) + \\
& 2\text{kov}(\hat{Y}_{31p}^{(43)}, \hat{Y}_{43p}^{(43)}) + 2\text{kov}(\hat{Y}_{31p}^{(43)}, \hat{Y}_{54p}^{(43)}) + 2\text{kov}(\hat{Y}_{43p}^{(43)}, \hat{Y}_{54p}^{(43)})
\end{aligned}
\tag{2.27}$$

$$\text{var}\left(\sum \hat{Y}_{tq}\right) = \text{var}(\hat{Y}_{14q}^{(54)}) + \text{var}(\hat{Y}_{51q}^{(34)})$$

Vuositason yhdistelmäestimaattorissa ja sen varianssiestimaattorissa käytetyt kertoimet  $\phi$  ja  $1-\phi$  saadaan **minimoimalla** vuositason varianssiestimaattori (vrt. Madsen, 1996, 6). Kun derivoidaan varianssiestimaattori  $\phi$ :n suhteen ja asetetaan derivaatta nolllaksi, saadaan  $\phi$ :n ja  $1-\phi$ :n optimaaliset arvot:

$$\frac{\partial \text{var}(\hat{Y}(\phi))}{\partial \phi} = 0 \quad \Rightarrow \quad \frac{1}{18^2} \phi \text{var}(\sum \hat{Y}_p) - \frac{1}{2^2} (1-\phi) \text{var}(\sum \hat{Y}_{tq}) = 0 \quad \Rightarrow$$

$$\phi = \frac{\frac{1}{4} \text{var}(\sum \hat{Y}_{tq})}{\frac{1}{18^2} \text{var}(\sum \hat{Y}_p) + \frac{1}{4} \text{var}(\sum \hat{Y}_{tq})}$$

ja (2.28)

$$1-\phi = \frac{\frac{1}{18^2} \text{var}(\sum \hat{Y}_p)}{\frac{1}{18^2} \text{var}(\sum \hat{Y}_p) + \frac{1}{4} \text{var}(\sum \hat{Y}_{tq})}$$

Riippumattomien rotaatioryhmien kokonaismäärän varianssiestimaattoreiden summan lauseke voidaan kirjoittaa kaavan (2.27) perusteella seuraavassa muodossa (vrt. Madsen, 1996, 6):

$$\text{var}\left(\sum \hat{Y}_{tq}\right) = 2N^2 \frac{N-n_q}{n_q N} s^2 \quad (2.29)$$

Vuositasoakin estimoitaessa oletetaan päällekkäisten ja riippumattomien osien kokonaismäärän varianssiestimaattoreiden osalta, että populaation koko ja otoskoot ovat vakioita kaikilla tutkimuskerroilla. Samoin oletetaan, että otoskoko on vakio myös kaikilla rotaatioryhmillä. Mainitut oletukset toteutuvat vuositasoa estimoitaessa samaan tapaan kuin muutosta estimoitaessa jo mainittiin. Lisäksi paneelin päällekkäisiä (p) ja riippumattomia (q) osia kuvaavien lukujen oletetaan olevan vakioita. Kalenterivuoden aikana esiintyvistä rotaatioryhmistä on 18 aina päällekkäisiä ja kaksi riippumattomia. Kaavassa (2.29) on lisäoletetuksena myös, että riippumattomien rotaatioryhmien varianssit ovat vakioita kaikilla tutkimuskerroilla.

Koska päällekkäisten rotaatioryhmien kokonaismäärän estimaattorit perustuvat samojen henkilöiden antamiin vastauksiin, voidaan niiden summa kirjoittaa seuraavaan muotoon (Madsen, 1996, 6):

$$\begin{aligned}
\sum \hat{Y}_{tp} &= \hat{Y}_{t-2,p} + \hat{Y}_{t-1,p} + \hat{Y}_{tp} \\
&= \frac{N_{t-2}}{n_{t-2,p}} \sum_{j=1}^{n_p} y_{t-2,pj} + \frac{N_{t-1}}{n_{t-1,p}} \sum_{j=1}^{n_p} y_{t-1,pj} + \frac{N_t}{n_{tp}} \sum_{j=1}^{n_p} y_{tpj} \\
&\equiv \frac{N}{n_{tp}} \sum_{j=1}^{n_p} (y_{t-2,pj} + y_{t-1,pj} + y_{tpj}) \\
&= \frac{N}{n_{tp}} \sum_{j=1}^{n_p} I_{tpj}
\end{aligned} \tag{2.30}$$

jossa **indikaattorimuuttuja I** saa arvon

- 0, jos henkilö j ei kuulu osajoukkoon kertaakaan tutkittavan kalenterivuoden aikana.
- 1, jos henkilö j kuuluu osajoukkoon yhden kerran tutkittavan kalenterivuoden aikana.
- 2, jos henkilö j kuuluu osajoukkoon kaksi kertaa tutkittavan kalenterivuoden aikana.
- 3, jos henkilö j kuuluu osajoukkoon kaikilla mahdollisilla haastattelukerroilla tutkittavan kalenterivuoden aikana.

Paneelin päällekkäisen osan kokonaismäärän varianssiestimaattori voidaan kuukausittaisen työvoimatutkimuksen tapauksessa määritellä seuraavasti:

$$\text{var}(\sum \hat{Y}_{tp}) = N^2 \frac{N - n_p}{n_p N} s_{I_p}^2 \tag{2.31}$$

jossa varianssitermi kuvaa indikaattorimuuttuja I:n vaihtelua otoksessa (vrt. Pahkinen&Lehtonen, 1989, 25):

$$s_{I_p}^2 = \frac{1}{n_p - 1} \sum_{j=1}^{n_p} (I_{pj} - \bar{I}_p)^2 \tag{2.32}$$



Indikaattorimuuttajaa I käytetään siis laskettaessa vuositason varianssiestimaattoria paneelin päällekkäisessä osassa, jossa saman henkilön eri ajankohtina antamien vastausten välillä on yleensä selvää positiivista korrelaatiota. Indikaattorimuuttujan käytön perustelu voidaan johtaa lähtien kaavan (2.2) yleisestä tilanteesta, jossa kovarianssitermit ovat mukana huomioimassa paneeliasetelman riippuvuuksia.

**Todistus.** Indikaattorimuuttaja I:n käyttö laskettaessa vuositason varianssiestimaattoria paneelin päällekkäisessä osassa:

$$\begin{aligned}
\text{var}\left(\sum \hat{Y}_{tp}\right) &= \text{var}\left(\hat{Y}_{t-2,p} + \hat{Y}_{t-1,p} + \hat{Y}_{tp}\right) \\
&= \text{var}(\hat{Y}_{t-2,p}) + \text{var}(\hat{Y}_{t-1,p}) + \text{var}(\hat{Y}_{tp}) + \\
&\quad 2\text{kov}(\hat{Y}_{t-2,p}, \hat{Y}_{t-1,p}) + 2\text{kov}(\hat{Y}_{t-2,p}, \hat{Y}_{tp}) + 2\text{kov}(\hat{Y}_{t-1,p}, \hat{Y}_{tp}) \\
&= N^2 \frac{N - n_p}{n_p N} \frac{1}{n_p - 1} \sum_{j=1}^n \left[ (y_{t-2,pj} - \bar{y}_{t-2,p})^2 + (y_{t-1,pj} - \bar{y}_{t-1,p})^2 + (y_{tpj} - \bar{y}_{tp})^2 + \right. \\
&\quad 2(y_{t-2,pj} - \bar{y}_{t-2,p})(y_{t-1,pj} - \bar{y}_{t-1,p}) + 2(y_{t-2,pj} - \bar{y}_{t-2,p})(y_{tpj} - \bar{y}_{tp}) + \\
&\quad \left. 2(y_{t-1,pj} - \bar{y}_{t-1,p})(y_{tpj} - \bar{y}_{tp}) \right] \\
&= N^2 \frac{N - n_p}{n_p N} \frac{1}{n_p - 1} \sum_{j=1}^n \left[ (y_{t-2,pj} + y_{t-1,pj} + y_{tpj})^2 - \right. \\
&\quad \left. 2(y_{t-2,pj} + y_{t-1,pj} + y_{tpj})(\bar{y}_{t-2,p} + \bar{y}_{t-1,p} + \bar{y}_{tp}) + (\bar{y}_{t-2,p} + \bar{y}_{t-1,p} + \bar{y}_{tp})^2 \right] \\
&= N^2 \frac{N - n_p}{n_p N} \frac{1}{n_p - 1} \sum_{j=1}^n \left( I_{pj}^2 - 2I_{pj} \cdot \bar{I}_p + \bar{I}_p^2 \right) \\
&= N^2 \frac{N - n_p}{n_p N} \frac{1}{n_p - 1} \sum_{j=1}^n \left( I_{pj} - \bar{I}_p \right)^2 \\
&= N^2 \frac{N - n_p}{n_p N} S_{I_p}^2
\end{aligned}$$

# Luku 3

## Tuloksia

### 3.1 Estimointituloksia

Estimointituloksia laskettaessa on ollut käytössä Tilastokeskuksen kuukausittaisen työvoimatutkimuksen vuosien 1994 ja 1995 kuukausiaineistot. Alustavien estimointitulosten laskennassa on käytetty synteettistä aineistoa (liite 1).

#### Neljännesvuosimuutoksen estimointi

Tapausesimerkinä neljännesvuosimuutoksesta on kolmannen ja neljännen vuosineljänneksen välinen muutos vuonna 1995 työmarkkina-aseman mukaan tarkasteltuna. Taulukossa 3.1 vaihtoehtoinen tutkimuksen rotaatorakenteen huomioiva neljännesvuosimuutoksen estimaattori ja sen keskivirhe. Taulukossa 3.2 on puolestaan nykyestimaattori neljännesvuosimuutokselle. Nykyestimaattori olettaa, että tarkasteltavien vuosineljännesten välillä ei ole riippuvuutta. Neljännesvuosimuutoksen estimoinnissa käytetty kuukausittaisen työvoimatutkimuksen aineisto on liitteessä 2. Synteettisestä aineistosta lasketut alustavat tulokset olivat keskivirheidensä osalta aika oikean suuntaisia, sillä vaihtoehtoisen estimaattorin keskivirheet olivat työllisten osalta noin 50%, työttömien noin 30% ja työvoimaan kuulumattomien noin 55% pienempiä kuin nykyestimaattorin vastaavien ryhmien keskivirheet.

Taulukko 3.1: Vaihtoehtoinen yhdistelmäestimaattori neljännesvuosimuutokselle.

<b>Työmarkkina-asema</b>	$\phi$	$1 - \phi$	$\hat{D}_t(\phi)$	s.e. $\hat{D}_t(\phi)$
Työlliset	0,89	0,11	-80 388	7 461
Työttömät	0,79	0,21	-12 414	6 660
Työvoimaan kuulumattomat	0,93	0,07	92 802	6 055

Taulukko 3.2: Nykyestimaattori neljännesvuosimuutokselle.

<b>Työmarkkina-asema</b>	$\hat{Y}_t$	$\hat{Y}_{t-1}$	$\hat{D}_t$	s.e. $\hat{D}_t$
Työlliset	2 018 933	2 091 904	-72 972	14 490
Työttömät	404 560	418 559	-13 999	8 987
Työvoimaan kuulumattomat	1 418 172	1 331 202	86 970	13 925

### Vuosimuutoksen estimointi

Vuosimuutoksen yhteydessä tapausesimerkkinä käytetään vuosien 1994 ja 1995 neljänsien vuosineljännesten välistä muutosta työmarkkina-aseman mukaan tarkasteltuna. Taulukossa 3.3 nähdään vaihtoehtoinen tutkimuksen rotaatorakenteen huomioiva vuosimuutoksen estimaattori ja sen keskivirhe. Taulukossa 3.4 on puolestaan nykyestimaattori vuosimuutokselle. Nykyestimaattori olettaa, että tarkasteltavien vuosineljännesten välillä ei ole riippuvuutta. Vuosimuutoksen estimoinnissa käytetty kuukausittaisen työvoimatutkimuksen aineisto on liitteessä 3. Synteettisestä aineistosta lasketut alustavat tulokset olivat keskivirheiden osalta aivan kuten neljännesvuosimuutoksenkin tapauksessa hyvin suuntaa antavia.

Taulukko 3.3: Vaihtoehtoinen yhdistelmäestimaattori vuosimuutokselle.

<b>Työmarkkina-asema</b>	$\phi$	$1 - \phi$	$\hat{D}_t(\phi)$	s.e. $\hat{D}_t(\phi)$
Työlliset	0,73	0,27	27 429	9 743
Työttömät	0,58	0,42	-17 110	7 872
Työvoimaan kuulumattomat	0,76	0,24	-10 319	8 791

Taulukko 3.4: Nykyestimaattori vuosimuutokselle.

<b>Työmarkkina-asema</b>	$\hat{Y}_t$	$\hat{Y}_{t-4}$	$\hat{D}_t$	s.e. $\hat{D}_t$
Työlliset	2 016 831	1 991 126	25 704	14 424
Työttömät	404 139	418 955	-14 816	8 935
Työvoimaan kuulumattomat	1 416 695	1 427 584	-10 888	13 947

## Vuositason estimointi

Vuositasoa estimoitaessa on tapausesimerkkinä käytetty vuotta 1995 työmarkkina-aseman mukaan tarkasteltuna. Taulukossa 3.5 nähdään vaihtoehtoinen tutkimuksen rotaatorakenteen huomioiva yhdistelmäestimaattori vuositasolle ja sen keskivirhe. Taulukossa 3.6 on puolestaan sekä nykyestimaattori että vaihtoehtoinen ei-yhdistelmäestimaattori vuositasolle. Nykyestimaattori olettaa keskivirhettä laskettaessa, että kaikki tutkimuksessa vuoden aikana esiintyvät rotaatioryhmät ovat toisistaan riippumattomia. Ei-yhdistelmäestimaattori puolestaan ottaa riippuvuudet huomioon keskivirhettä laskettaessa. Nykyestimaattorin ja vaihtoehtoisen ei-yhdistelmäestimaattorin keskivirheet siis eroavat toisistaan, mutta vuositason piste-estimaatit ovat samat. Vuositason estimoinnissa käytetty kuukausittaisen työvoimatutkimuksen aineisto on liitteessä 4. Synteettisestä aineistosta lasketut alustavat tulokset eivät eronneet keskivirheiden prosentuaalisten muutosten osalta paljoakaan todellisista aineistoista lasketuista tuloksista.

Taulukko 3.5: Vaihtoehtoinen yhdistelmäestimaattori vuositasolle.

Työmarkkina-asema	$\phi$	$1 - \phi$	$\hat{Y}(\phi)$	s.e. $\hat{Y}(\phi)$
Työlliset	0,79	0,21	2 035 880	7 487
Työttömät	0,80	0,20	420 483	4 389
Työvoimaan kuulumattomat	0,78	0,22	1 381 910	7 258

Taulukko 3.6: Vaihtoehtoinen ei-yhdistelmäestimaattori ( $\hat{Y}_{ALT}$ ) ja nykyestimaattori ( $\hat{Y}$ ) vuositasolle.

Työmarkkina-asema	$\hat{Y}_{ALT}$ ja $\hat{Y}$	s.e. $\hat{Y}_{ALT}$	s.e. $\hat{Y}$
Työlliset	2 037 738	7 775	5 140
Työttömät	422 562	4 515	3 225
Työvoimaan kuulumattomat	1 378 280	7 543	4 940

## 3.2 Tehokkuusvertailu

Estimaattoreiden keskimääräistä suorituskkyä mitataan asetelmakertoimen estimaattorin **deff** yleistyksen eli strategiakertoimen estimaattorin **seff** avulla (Lehtonen&Pahkinen, 1995, 14-15). Strategiakertoimen estimaattori määrittellään muutoksen estimoinnin tapauksessa, kuten kaavassa (3.1) vaihtoehtoisen estimaattorin varianssiestimaattorin ja nykyestimaattorin varianssiestimaattorin suhteena.

$$\text{seff}_{\text{srs}}(\hat{D}_t(\phi)) = \frac{\text{var}_{\text{srs}}(\hat{D}_t(\phi))}{\text{var}_{\text{srs}}(\hat{D}_t)} \quad (3.1)$$

Vuositasoa estimoitaessa vertailu suoritetaan vaihtoehtoisen yhdistelmäestimaattorin ja vaihtoehtoisen ei-yhdistelmäestimaattorin varianssiestimaattoreiden kesken.

$$\text{seff}_{\text{srs}}(\hat{Y}(\phi)) = \frac{\text{var}_{\text{srs}}(\hat{Y}(\phi))}{\text{var}_{\text{srs}}(\hat{Y}_{\text{ALT}})} \quad (3.2)$$

Vuositason estimointia koskevassa tehokkuusvertailussa olisi epäjohdonmukaista käyttää vaihtoehtoisen yhdistelmäestimaattorin ja nykyestimaattorin varianssiestimaattoreiden suhdetta, koska nykyestimaattori osoittautuisi vaihtoehtoista estimaattoria tehokkaammaksi rotaatioasetelman puutteellisen huomioimisen takia. Nykyestimaattorihan on laskettu olettaen kaikki tutkittavan kalenterivuoden aikana tutkimuksessa mukana olevat rotaatioryhmät toisistaan riippumattomiksi.

Taulukko 3.7: Vaihtoehtoisen estimaattorin ja nykyestimaattorin tehokkuusvertailu neljännesvuosimuutosta estimoitaessa.

<b>Työmarkkina-asema</b>	$\text{var}(\hat{D}_t(\phi))$	$\text{var}(\hat{D}_t)$	seff
Työlliset	55 659 599	209 958 364	0,27
Työttömät	44 351 394	80 760 415	0,55
Työvoimaan kuulumattomat	36 667 441	193 891 921	0,19

Kuten taulukoista 3.7. ja 3.8 huomataan, vaihtoehtoinen estimaattori osoittautuu muutosta estimoitaessa tehokkaammaksi kuin nykyestimaattori. Neljännesvuosi-muutos (60%) sisältää enemmän päällekkyyttä kuin vuosimuutos (40%), mikä selittää neljännesvuosimuutoksen keskivirheiden suuremman vähenemisen. Työttömien osajoukko erosi työllisistä ja työvoiman ulkopuolisista siten, että siinä keskivirheen väheneminen oli pienintä. Eroa selittää paljolti se, että työttömien osajoukossa henkilö vaihtaa työmarkkina-asemaansa paljon useammin kuin työllisten tai työvoiman ulkopuolisten osajoukoissa. Työlliset ja työvoiman ulkopuolisista esimerkiksi eläkeläiset eivät siis vaihda statustaan siirryttäessä vuosineljänneksestä toiseen yhtä usein kuin työttömät.

Taulukko 3.8: Nykyestimaattorin ja vaihtoehtoisen estimaattorin tehokkuusvertailu vuosimuutosta estimoitaessa.

Työmarkkina-asema	$\text{var}(\hat{D}_t(\phi))$	$\text{var}(\hat{D}_t)$	seff
Työlliset	94 933 824	208 062 992	0,46
Työttömät	61 963 775	79 827 627	0,78
Työvoimaan kuulumattomat	77 287 400	194 510 630	0,40

Taulukossa 3.9 oleva vuositason vaihtoehtoisten estimaattoreiden keskinäinen tehokkuusvertailu ei ole kovinkaan mielekästä eikä yhdistelmäestimaattorin käyttö tuokaan mukanaan huomattavia estimaattorin suorituskyvyn parannuksia. Vuositasoa estimoitaessa pääpaino onkin siinä, että on käytettävissä kuukausittaisen työvoimatutkimuksen rotaatorakenteen huomioivia estimaattoreita. Vuositason keskivirheitähän ei ole kuukausittaisessa työvoimatutkimuksessa aiemmin estimoitu.

Taulukko 3.9: Vaihtoehtoisen yhdistelmäestimaattorin ja vaihtoehtoisen ei-yhdistelmäestimaattorin tehokkuusvertailu vuositasoa estimoitaessa.

Työmarkkina-asema	$\text{var}(\hat{Y}(\phi))$	$\text{var}(\hat{Y}_{ALT})$	seff
Työlliset	56 054 863	60 443 805	0,93
Työttömät	19 259 273	20 383 904	0,94
Työvoimaan kuulumattomat	52 680 173	56 895 955	0,93

## Luku 4

### Yhteenveto

Tilastokeskuksen kuukausittainen työvoimatutkimus on suurkysely, jossa sama henkilö on tutkimuksessa mukana kaksi kertaa tai useammin, mutta ei välttämättä aina peräkkäisinä haastattelukertoina. Kun henkilö on ollut riittävän ajan tutkimuksessa, hänet vapautetaan ja tilalle valitaan uusi haastateltava. Tällaista asetelmaa kutsutaan rotatoivaksi eli kiertäväksi paneeliksi. Siinä otos vaihtuu siis asteittain siten, että peräkkäisinä kuukausina, vuosineljänneksinä ja vuosina osa vastaajista on samoja. Koska vastaajien työmarkkina-asema ei aina muutu tutkimuskerrojen välillä, korreloivat peräkkäisissä haastatteluissa annetut vastaukset keskenään. Tutkittavat otokset eivät siis aina ole toisistaan riippumattomia, vaikka riippumattomuutta pidetäänkin työvoimatutkimuksessa oletuksena estimoitaessa muutosta ja tasoa.

Rotatoivan paneelin käyttöä suurkyselyssä voidaan perustella monin eri tavoin. Tutkimuksen käytännön järjestelyn kannalta rotatoiva paneeli on edullinen, sillä uusintahaastattelu on aina alkuperäistä halvempi. Paneelikadon ja vastausrasitteen aiheuttamat ongelmat ovat rotatoivassa paneelissa vähäisempiä kuin puhtaassa paneeliasetelmassa, koska rotatoiva paneeli on sekoitus uusia ja jo aiemmin haastateltuja vastaajia. Lisäksi etuna on otoksen pysyminen ajan tasalla. Tämä johtuu siitä, että tutkittavaan joukkoon lisätään uusia haastateltavia kunakin tutkittavana ajankohtana.

Rotatoivassa paneelissa päällekkyyks kannattaa yleensä maksimoida niillä tutkimusajankohdilla tai -jaksoilla, joilla halutaan parantaa poikkileikkauksen tai muutoksen estimointia. Tällöin estimointi tehostuu kapeamman luottamusvälin ansiosta. Luottamusvälin kapeneminen perustuu edellisten vastauskertojen tietojen hyväksikäyttöön siten, että keskivirhettä eli varianssiestimaattorin neliöjuurta voi-

daan pienentää ns. paneelihistorian antamalla informaatiolla. Keskivirheellä on suuri merkitys laskettaessa luottamusväliä perusjoukon parametrille, sillä keskivirheen kasvaessa myös luottamusväli kasvaa.

Tilastokeskuksen kuukausittaisessa työvoimatutkimuksessa poimitaan samanlaisina toistettavia 12 000 hengen otoksia kehikkoperusjoukosta, johon lukeutuu kunakin tutkimusajankohtana Suomessa asuva 15 - 74-vuotias väestö. Tutkimuksessa annettujen vastausten perusteella otos jaetaan työmarkkina-aseman mukaisesti ryhmiin eli osajoukkoihin, jotka ovat työlliset, työttömät ja työvoiman ulkopuoliset. Perusjoukko muuttuu yleensä tutkimusajankohtien välillä sekä lukumäärältään että rakenteeltaan. Yleisesti ottaen perusjoukon parametrien muutokset eri ajankohtina ovat seurausta siitä, että perusjoukkoon kuuluva henkilö muuttaa työmarkkina-asemaansa tai perusjoukko muuttuu joko kooltaan tai rakenteeltaan. Kuukausittaisessa työvoimatutkimuksessa otos on jaettu siten, että kuukausittain haastatellaan viittä eri rotaatioryhmää. Otokseen kuuluva henkilö on tutkimuksessa mukana puolentoista vuoden ajan. Otos vaihtuu asteittain siten, että kolmena peräkkäisenä kuukautena vastausvuorossa ovat eri henkilöt. Peräkkäiset tutkimuskuukaudet ovat siis toisistaan riippumattomia. Peräkkäisinä vuosineljänneksinä 3/5 vastaajista on samoja ja vuoden kuluttua otosten päällekkyyks on 2/5.

Muutosta ja tasoa estimoitaessa työvoimatutkimuksen nykyinen laskentatapa ei ota rotaatioasetelman kaikkia riippuvuuksia huomioon. Muutosta ja tasoa estimoitaessa oletetaan, että kuukausien, neljännesvuosien ja vuosien välillä ei ole riippuvuuksia. Neljännesvuosimuutosta eli perusjoukon parametrien muutosta ja sen keskivirhettä kahden peräkkäisen vuosineljänneksen välillä ei ole siis estimoitu riippuvuudet huomioiden, kuten ei myöskään vuosimuutosta. Kyseisten muutosten keskivirheitä arvioidaan kylläkin laskemalla niille yläraja-arvio. Vuosimuutosta estimoitaessa ollaan kiinnostuneita tutkittavan vuosineljänneksen ja edeltävän kalenterivuoden vastaavan vuosineljänneksen välisestä muutoksesta ja sen keskivirheestä. Vuositasoa eli esimerkiksi työttömien osajoukon vuosineljänneksittäisten kokonaismäärien keskiarvoa ja sen keskivirhettä ei työvoimatutkimuksessa ole myöskään estimoitu riippuvuudet huomioiden, vaan keskivirhettä on arvioitu neljännesvuosikeskivirheiden avulla.

Nykyistä tarkempi estimaattien tehokkuuden arviointi edellyttää työvoimatutkimuksen paneelirakenteen huomioon ottamisen. Neljännesvuosi- ja vuosimuutosta estimoitaessa käytetty rotaatorakenne otetaan huomioon jakamalla paneeli vuosi-



neljänneksittäin sekä päällekkäisiin että riippumattomiin osiin. Jos paneelin päällekkäiseen osaan kuuluvien henkilöiden eri vastaukseton välillä on positiivista korrelaatiota eli heidän työmarkkina-asemansa ei ole paljoakaan muuttunut tutkimuskertojon välillä, pienenee muutoksen keskivirhe. Negatiivisen korrelaation tapauksessa vaikutus on päinvastainen eli muutoksen keskivirhe suurenee. Päällekkäisen ja riippumattoman osan yhdistävää yhdistelmäestimaattoria voidaan käyttää tutkittavan muutoksen estimaattorina. Yhdistelmäestimaattori painottaa päällekkäistä osaa eri kertoimen avulla kuin riippumatonta osaa.

Tilastokeskuksen kuukausittaisen työvoimatutkimuksen vuosien 1994 ja 1995 kuukausiaineistoista laskettujen estimointitulosten mukaan riippuvuus huomioimalla saadaan muutoksen keskivirhe ja sitä kautta tietysti myös estimoitavan parametrin luottamusväli pienemään huomattavasti. Tarkasteltavasta osajoukosta riippuen vähennys keskivirheessä on neljännesvuosimuutoksen tapauksessa 26 - 57 % ja vuosimuutoksen osalta 12 - 37 %. Neljännesvuosimuutos sisältää enemmän riippuvuutta kuin vuosimuutos, mikä selittää neljännesvuosimuutoksen keskivirheen suuremman vähenemisen. Työttömien osajoukko erosi työllisistä ja työvoiman ulkopuolisista siten, että siinä keskivirheen väheneminen oli pienintä. Eroa selittää paljolti se, että työttömien osajoukossa henkilö vaihtaa työmarkkina-asemaansa paljon useammin kuin työllisten tai työvoiman ulkopuolisten osajoukoissa. Työlliset ja työvoiman ulkopuolisista esimerkiksi eläkeläiset eivät siis vaihda statustaan siirryttäessä vuosineljänneksestä toiseen yhtä usein kuin työttömät.

Kuukausittaisen työvoimatutkimuksen rotaatorakenteen huomioiva vuositason estimaattori perustuu myös yhdistelmäestimaattorin käyttöön, joten siinäkin riippumattomat rotaatorioryhmät saavat eri kertoimen kuin päällekkäiset ryhmät. Tasoa estimoidaessa saman henkilön eri vastaukseton välisestä positiivisesta korrelaatiosta on merkittävää haittaa, sillä positiivinen korrelaatio suurentaa tason keskivirhettä. Negatiivinen korrelaatio paneelin päällekkäiseen osaan kuuluvan henkilön eri vastaukseton välillä vaikuttaa juuri päinvastoin eli tason keskivirhe pienenee. Työvoimatutkimuksessa on kalenterivuoden aikana kaksi riippumattomiksi tulkittavaa rotaatorioryhmää, sillä ne ovat yhden kerran tutkimuksessa mukana tarkasteltavan vuoden aikana. Useammin kuin kerran vuoden aikana esiintyvät ryhmät kuuluvat päällekkäisiin rotaatorioryhmiin, joista kolme ryhmää esiintyy kaksi kertaa vuoden aikana ja neljä ryhmää kolmasti.

Kaiken kaikkiaan vaikuttaa siltä, että riippuvuudet huomioimalla onnistutaan tehostamaan tuntuvasti sekä neljännesvuosi- että vuosimuutoksen estimointia. Vuositasoa estimoitaessa on puolestaan pääpaino sillä, että nyt mahdollistuu vuositason keskivirheen estimointi vieläpä tutkimuksen rotaatorakenne huomioiden.

# Lähteet

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. New York. Wiley & Sons.

Djerf, Kari (1994). Yleistetty Raking-menetelmä otantatutkimuksen estimoinnin tehostamismenetelmänä. Sovellus työvoimatutkimuksen aineistoon. Jyväskylä: Jyväskylän yliopiston tilastotieteen laitos. Tilastotieteen laudaturtyö.

Duncan, G.J. and Kalton, G. (1987). Issues of Design and Analysis of Surveys Across Time. *International Statistical Review*, Vol. 55, No. 1. 97-117.

Kish, L. (1995). Questions/Answers (1978-1994) from the Survey Statistician. IASS.

Kish, L. (1987). *Statistical Design for Research*. New York. Wiley & Sons.

Kish, L. (1965). *Survey Sampling*. New York. Wiley & Sons.

Lehtonen, R. and Pahkinen E.J. (1995). *Practical Methods for Design and Analysis of Complex Surveys*. New York. Wiley & Sons.

Madsen, B. (1996). Estimation Procedures in a Rotating Panel Survey. Paper presented at Eurostat-workshop "Weighting Practices of the Community LFS", Avignon, 13-15. July 1996.

Pahkinen E. ja Lehtonen, R.(1989). *Otanta-asetelmat ja tilastollinen analyysi*. Helsinki. Gaudeamus.

Särndal C.-E., Swensson B. and Wretman J. (1992). *Model Assisted Survey Sampling*. New York. Springer-Verlag.

Tilastokeskus (1997). Työttömyys ja työllisyys tilastoissa. Tilastokeskuksen ja Työministeriön tilastojen vertailua. Helsinki: Tilastokeskus.

Tilastokeskus (1993). Työvoimatilasto. Tietoja työllisyydestä, työttömyydestä ja työvoimasta vuosilta 1960-1991. Helsinki: Tilastokeskus.

Tilastokeskus (1995). Työvoimatilasto 1994. Työvoimatutkimuksen tuloksia vuosilta vuosilta 1980-1994. Helsinki: Tilastokeskus.

# Liite 1.

## Synteettisen aineiston kuvaus

Synteettinen aineisto sisältää 12 paneelia, jotka on kerätty työvoimatutkimuksen kuukausiaineistoista 10/1993 - 12/1995. Aineistossa, joka ei pidä sisällään haastattelussa mukana olleiden henkilöiden tunnistetietoja, on noin 30 000 havaintoyksikköä.

### **Muuttajat**

Synteettisessä aineistossa on mukana työvoimatutkimuksen muuttujista tutkimuskuukausi (TUTKK), kohdenumero (KONU), sukupuoli (SP), ikä (IKA), asuinkunta (ASKU), lääni (LAANI), toiminnan laatu (TOILAA), työtunnit (TYOTUNN) ja työnhakijarekisterin työvoimatieto (TYOTON). Muuttujat on sukupuolta ja kohdenumeroa lukuun ottamatta nimetty seuraavasti (esim. toiminnan laatu): ensimmäisellä haastattelukerralla TOILAA1, toisella haastattelukerralla TOILAA2 jne. Näin ollen aineistossa on kaikkiaan 37 muuttujaa.

Tutkimuskuukausi on tutkimuskuukauden numero, joka samalla ilmoittaa viiteajankohdan vuoden ja kuukauden (vvkk). Kohdenumero on muotoa xxx xxxx x, jossa kolme ensimmäistä merkkiä ilmoittaa kyseisen rotaatioryhmän, keskimmäiset neljä merkkiä koostuvat juoksevista numeroista ja viimeinen merkki on tarkistusmerkki. Sukupuoli on muotoa 1 = Mies ja 2 = Nainen. Ikä ilmoitetaan täysinä vuosina (15 - 74). Asuinkunta on KELAn kuntakoodi. Lääni on Tilastokeskuksen kuntarekisteristä asuinkunnan perusteella saatava tieto. Työnhakijarekisterin tieto on aineistoon jälkikäteen lisätty tieto, joka ilmoittaa onko henkilö THR:n mukaan 1 = työtön vai 2 = ei-työtön. Työtunnit ilmoittaa henkilön tekemät työtunnit tutkimusviikolla päätyössä mukaan lukien ylityötunnit. Työtunnit voivat olla välillä 1 - 168 tuntia/tutkimusviikko.

Toiminnan laatu -muuttujasta selviää oliko henkilö tutkimusviikolla työllinen, työtön vaiko työvoiman ulkopuolella. (Tilastokeskus, 1995.)

## Synteettisen aineiston rotaatorakenne

Tutkimuskerta	Rotaatioryhmät				
	t-4	t-3	t-2	t-1	t
5.	412	411	421	343	342
4.	411	421	343	342	341
3.	343	342	341	413	422
2.	342	341	413	422	423
1.	341	413	422	423	431

- Kussakin rotaatioryhmässä on 2 000 havaintoa ja kussakin vuosineljänneksessä näin ollen yhteensä 10 000 havaintoa.
- Katotapaukset eivät ole mukana aineistossa.

## Liite 2.

### Neljännesvuosimuutoksen estimoinnissa käytetty kuukausittaisen työvoimatutkimuksen aineisto

#### Riippumaton osa

Työmarkkina-asema Vuosineljännes	N	$n_q$	$\sum_{j=1}^{n_q} y_j$	$s^2$	$\hat{Y}_q$	$\text{var}(\hat{Y}_q)$
<b>Työlliset:</b>						
t	3 841 665	13 923	7 397	0,249	2 040 997	263 024 794
t - 1	3 841 665	13 817	7 549	0,248	2 098 916	263 804 763
				Muutos	<b>-57 920</b>	<b>526 829 557</b>
<b>Työttömät:</b>						
t	3 841 665	13 923	1 424	0,092	392 913	96 979 693
t - 1	3 841 665	13 817	1 475	0,095	410 108	101 493 989
				Muutos	<b>-17 194</b>	<b>198 473 682</b>
<b>Työvoimaan kuulu-</b>						
<b>mattomat:</b>						
t	3 841 665	13 923	5 102	0,230	1 407 755	243 200 689
t - 1	3 841 665	13 817	4 793	0,232	1 332 641	247 106 036
				Muutos	<b>75 114</b>	<b>490 306 724</b>

- Kuukausittaisen työvoimatutkimuksen perusjoukon kokona N on käytetty vuosineljännesten t ja t-1 perusjoukkojen keskiarvoa.
- Kaavan (2.15) oletusta riippumattomien osien varianssiestimaattoreiden varianssien yhtäsuuruudesta ei käytetä, vaan kummankin vuosineljänneksen riippumattoman osan varianssiestimaattorit on laskettu erikseen. Oletuksen käyttö ei sanottavasti vaikuta estimointituloksiin, sillä otoksista lasketut yhden havainnon varianssit eivät vuosineljänneksestä toiseen juurikaan poikkea toisistaan.

## Päällekkäinen osa

Työmarkkina-asema Vuosineljännes	N	$n_p$	$\sum_{j=1}^{n_p} y_j$	$S_{d_p}^2$	$\hat{Y}_p$	$\text{var}(\hat{Y}_p)$
Työlliset:	3 841 665	20 006		0,085		
t	3 841 665	20 832	10 868		2 004 187	
t - 1	3 841 665	20 721	11 258		2 087 229	
				Muutos	<b>-83 042</b>	<b>62 234 703</b>
Työttömät:	3 841 665	20 006		0,072		
t	3 841 665	20 832	2 236		412 345	
t - 1	3 841 665	20 721	2 288		424 194	
				Muutos	<b>-11 850</b>	<b>52 679 157</b>
Työvoimaan kuulu- mattomat:	3 841 665	20 006		0,053		
t	3 841 665	20 832	7 728		1 425 134	
t - 1	3 841 665	20 721	7 175		1 330 242	
				Muutos	<b>94 892</b>	<b>39 000 071</b>

- Kursivoidut luvut ilmoittavat ne päällekkäiseen osaan kuuluvat haastateltavat, joilta on saatu vastaus kummallakin vuosineljänneksellä. Heidän antamiensa vastausten perusteella on laskettu päällekkäisen osan varianssiestimaattorit. Jommalla kummalla tai kummallakin vuosineljänneksellä vastaamatta jättäneiden muodostamien katotapausten oletetaan jakautuvan työmarkkina-asemansa mukaan kuten vastanneidenkin.
- Päällekkäisen osan neljännesvuosimuutoksen estimaattoria laskettaessa hyväksytään kaikki vastaukset, jotka on saatu jommalta kummalta tai kummaltakin vuosineljännekseltä.



## Ei huomioida riippuvuutta

Työmarkkina-asema Vuosineljännes	N	n	$\sum_{j=1}^n y_j$	$s^2$	$\hat{Y}$	$\text{var}(\hat{Y})$
<b>Työlliset:</b>						
t	3 841 665	34 755	18 265	0,249	2 018 933	104 928 361
t - 1	3 841 665	34 538	18 807	0,248	2 091 904	105 030 003
				<b>Muutos</b>	<b>-72 972</b>	<b>209 958 364</b>
<b>Työttömät:</b>						
t	3 841 665	34 755	3 660	0,094	404 560	39 648 113
t - 1	3 841 665	34 538	3 763	0,097	418 559	41 112 303
				<b>Muutos</b>	<b>-13 999</b>	<b>80 760 415</b>
<b>Työvoimaan kuulu-</b>						
<b>mattomat:</b>						
t	3 841 665	34 755	12 830	0,233	1 418 172	97 998 220
t - 1	3 841 665	34 538	11 968	0,226	1 331 202	95 893 700
				<b>Muutos</b>	<b>86 970</b>	<b>193 891 921</b>

- Nykyestimaattorin, joka olettaa tarkasteltavat vuosineljännekset toisistaan riippumattomaksi, laskentaa varten.

## Liite 3.

### Vuosimuutoksen estimoinnissa käytetty kuukausittaisen työvoimatutkimuksen aineisto

#### Riippumaton osa

Työmarkkina-asema Vuosineljännes	N	$n_q$	$\sum_{j=1}^{n_q} y_j$	$s^2$	$\hat{Y}_q$	$\text{var}(\hat{Y}_q)$
<b>Työlliset:</b>						
t	3 837 665	20 766	10 891	0,249	2 012 714	175 932 351
t - 4	3 837 665	21 143	10 959	0,250	1 989 168	172 959 738
				<b>Muutos</b>	<b>23 546</b>	<b>348 892 090</b>
<b>Työttömät:</b>						
t	3 837 665	20 766	2 207	0,095	407 865	67 003 609
t - 4	3 837 665	21 143	2 313	0,097	419 833	67 496 804
				<b>Muutos</b>	<b>-11 967</b>	<b>134 500 413</b>
<b>Työvoimaan kuulumattomat:</b>						
t	3 837 665	20 766	7 668	0,233	1 417 086	164 296 356
t - 4	3 837 665	21 143	7 871	0,234	1 428 665	161 890 498
				<b>Muutos</b>	<b>-11 579</b>	<b>326 186 854</b>

- Kuukausittaisen työvoimatutkimuksen perusjoukon kokona N on käytetty vuosineljännesten t ja t-4 perusjoukkojen keskiarvoa.
- Kaavan (2.15) oletusta riippumattomien osien varianssiestimaattoreiden varianssien yhtäsuuruudesta ei käytetä, vaan kummankin vuosineljänneksen riippumattoman osan varianssiestimaattorit on laskettu erikseen. Oletuksen käyttö ei sanottavasti vaikuta estimointituloksiin, sillä otoksista lasketut yhden havainnon varianssit eivät vuosineljänneksestä toiseen juurikaan poikkea toisistaan.

## Päällekkäinen osa

Työmarkkina-asema Vuosineljännes	N	$n_p$	$\sum_{j=1}^{n_p} y_j$	$S_{d_p}^2$	$\hat{Y}_p$	$\text{var}(\hat{Y}_p)$
Työlliset:	3 837 665	13 512		0,120		
t	3 837 665	13 989	7 374		2 022 942	
t - 4	3 837 665	14 105	7 329		1 994 062	
				Muutos	<b>28 880</b>	<b>130 421 667</b>
Työttömät:	3 837 665	13 512		0,090		
t	3 837 665	13 989	1 453		398 608	
t - 4	3 837 665	14 105	1 535		417 640	
				Muutos	<b>-19 032</b>	<b>98 153 766</b>
Työvoimaan kuulu- mattomat:	3 837 665	13 512		0,092		
t	3 837 665	13 989	5 162		1 416 115	
t - 4	3 837 665	14 105	5 241		1 425 963	
				Muutos	<b>-9 848</b>	<b>100 289 102</b>

- Kursivoidut luvut ilmoittavat ne päällekkäiseen osaan kuuluvat haastateltavat, joilta on saatu vastaus kummallakin vuosineljänneksellä. Heidän antamiensa vastausten perusteella on laskettu päällekkäisen osan varianssiestimaattorit. Jommalla kummalla tai kummallakin vuosineljänneksellä vastaamatta jättäneiden muodostamien katotapausten oletetaan jakautuvan työmarkkina-asemansa mukaan kuten vastanneidenkin.
- Päällekkäisen osan vuosimuutoksen estimaattoria laskettaessa hyväksytään kaikki vastaukset, jotka on saatu jommalta kummalta tai kummaltakin vuosineljännekseltä.

## Ei huomioida riippuvuutta

Työmarkkina-asema Vuosineljännes	N	n	$\sum_{j=1}^n y_j$	$s^2$	$\hat{Y}$	$\text{var}(\hat{Y})$
<b>Työlliset:</b>						
t	3 837 665	34 755	18 265	0,249	2 016 831	104 708 972
t - 4	3 837 665	35 248	18 288	0,250	1 991 126	103 354 020
				Muutos	<b>25 704</b>	<b>208 062 992</b>
<b>Työttömät:</b>						
t	3 837 665	34 755	3 660	0,094	404 139	39 565 215
t - 4	3 837 665	35 248	3 848	0,097	418 955	40 262 413
				Muutos	<b>-14 816</b>	<b>79 827 627</b>
<b>Työvoimaan kuulu-</b>						
<b>mattomat:</b>						
t	3 837 665	34 755	12 830	0,233	1 416 695	97 793 322
t - 4	3 837 665	35 248	13 112	0,234	1 427 584	96 717 309
				Muutos	<b>-10 888</b>	<b>194 510 630</b>

- Nykyestimaattorin, joka olettaa tarkasteltavat vuosineljännekset toisistaan riippumattomaksi, laskentaa varten.

## Liite 4.

### Vuositason estimoinnissa käytetty kuukausittaisen työvoimatutkimuksen aineisto

#### Riippumaton osa

Työmarkkina-asema Rotaatioryhmä	N	$n_q$	$\sum_{j=1}^{n_q} y_j$	$s^2$	$\hat{Y}_q$	$\text{var}(\hat{Y}_q)$
Työlliset:						
54	3 838 580	6 898	3 690	0,249	2 053 401	530 534 247
34	3 838 580	7 163	3 719	0,250	1 992 975	512 618 730
				Yhteensä	<b>4 046 376</b>	<b>1 043 152 977</b>
Työttömät:						
54	3 838 580	6 898	709	0,092	394 542	196 661 112
34	3 838 580	7 163	768	0,096	411 564	196 564 827
				Yhteensä	<b>806 106</b>	<b>393 225 939</b>
Työvoimaan kuulu- mattomat:						
54	3 838 580	6 898	2 499	0,231	1 390 637	492 689 051
34	3 838 580	7 163	2 676	0,234	1 434 042	480 559 784
				Yhteensä	<b>2 824 678</b>	<b>973 248 835</b>

- Kuukausittaisen työvoimatutkimuksen perusjoukon kokona N on käytetty vuoden 1995 kaikkien tutkimuskuukausien perusjoukkojen keskiarvoa.
- Kaavan (2.15) oletusta riippumattomien osien varianssiestimaattoreiden varianssien yhtäsuuruudesta ei käytetä, vaan kummankin riippumattomaan osaan kuuluvan rotaatioryhmän varianssiestimaattorit on laskettu erikseen. Oletuksen käyttö ei sanottavasti vaikuta estimointituloksiin, sillä otoksista lasketut yhden havainnon varianssit eivät rotaatioryhmien välillä juurikaan poikkea toisistaan.
- Vuositasoa estimoitaessa paneelin päällekkäiseen osaan kuuluvat rotaatioryhmät työmarkkina-aseman mukaan eriteltynä sivuilla 49-51. Kursivoidut luvut ovat vaihtoehtoisten estimaattoreiden ja muut nykyestimaattoreiden laskentaa varten.

**Päällekkäinen osa**

Työlliset:

Rotaatioryhmä Vuosineljännes	N	$n_p$	$\sum_{j=1}^{n_p} y_j$	$s_{I_p}^2$	$\hat{Y}_p$	$\text{var}(\hat{Y}_p)$
53	3 838 580	6 670	7 108	0,913	4 090 649	2 014 210 722
t	3 838 580	6 895	3 571	0,250	1 988 045	532 685 353
t - 1	3 838 580	6 782	3 690	0,248	2 088 523	538 059 034
51	3 838 580	6 656	10 861	1,933	6 263 644	4 271 862 421
t - 1	3 838 580	6 991	3 846	0,248	2 111 741	520 743 691
t - 2	3 838 580	6 955	3 771	0,248	2 081 278	524 994 150
t - 3	3 838 580	6 898	3 568	0,250	1 985 511	532 504 440
52	3 838 580	6 589	10 707	1,959	6 237 620	4 372 778 317
t	3 838 580	6 973	3 630	0,250	1 998 286	526 498 756
t - 1	3 838 580	6 939	3 750	0,248	2 074 460	526 518 468
t - 2	3 838 580	6 784	3 694	0,248	2 090 170	537 817 738
44	3 838 580	6 766	11 016	1,945	6 249 748	4 227 780 055
t	3 838 580	7 025	3 707	0,249	2 025 568	521 876 334
t - 2	3 838 580	7 034	3 860	0,248	2 106 471	517 836 769
t - 3	3 838 580	7 096	3 667	0,250	1 983 663	517 650 286
43	3 838 580	6 747	10 846	1,934	6 170 630	4 216 208 028
t	3 838 580	6 964	3 667	0,249	2 021 263	526 584 836
t - 1	3 838 580	7 000	3 818	0,248	2 093 671	521 018 445
t - 3	3 838 580	7 048	3 605	0,250	1 963 405	521 493 173
41	3 838 580	6 632	6 912	0,923	4 000 643	2 047 061 630
t - 2	3 838 580	6 722	3 572	0,249	2 039 781	544 968 661
t - 3	3 838 580	6 789	3 400	0,250	1 922 400	541 712 857
42	3 838 580	6 698	7 266	0,904	4 164 097	1 984 467 910
t - 1	3 838 580	6 826	3 703	0,248	2 082 371	534 881 957
t - 2	3 838 580	6 823	3 647	0,249	2 051 781	536 441 173
				Yhteensä	<b>37 177 031</b>	<b>23 134 369 084</b>
				Yhteensä	<b>36 708 386</b>	<b>9 524 286 122</b>

**Päällekkäinen osa**

Työttömät:

Rotaatioryhmä Vuosineljännes	N	$n_p$	$\sum_{j=1}^{n_p} y_j$	$s_{I_p}^2$	$\hat{Y}_p$	$\text{var}(\hat{Y}_p)$
53	3 838 580	6 670	1 443	0,323	830 445	713 203 568
t	3 838 580	6 895	746	0,097	415 313	205 855 483
t - 1	3 838 580	6 782	755	0,099	427 326	214 592 065
51	3 838 580	6 656	2 144	0,637	1 236 466	1 408 425 118
t - 1	3 838 580	6 991	751	0,096	412 355	201 752 889
t - 2	3 838 580	6 955	766	0,098	422 768	207 288 681
t - 3	3 838 580	6 898	778	0,100	432 939	213 394 962
52	3 838 580	6 589	2 279	0,675	1 327 686	1 507 505 180
t	3 838 580	6 973	752	0,096	413 970	202 970 700
t - 1	3 838 580	6 939	780	0,100	431 488	211 510 770
t - 2	3 838 580	6 784	896	0,115	506 982	248 572 809
44	3 838 580	6 766	2 061	0,598	1 169 275	1 299 915 600
t	3 838 580	7 025	715	0,091	390 688	191 426 607
t - 2	3 838 580	7 034	731	0,093	398 920	194 744 195
t - 3	3 838 580	7 096	756	0,095	408 958	197 319 532
43	3 838 580	6 747	2 158	0,623	1 227 754	1 359 145 648
t	3 838 580	6 964	738	0,095	406 788	200 126 737
t - 1	3 838 580	7 000	753	0,096	412 922	201 736 749
t - 3	3 838 580	7 048	794	0,100	432 439	208 634 414
41	3 838 580	6 632	1 518	0,344	878 613	762 121 334
t - 2	3 838 580	6 722	774	0,102	441 991	222 976 935
t - 3	3 838 580	6 789	797	0,104	450 633	224 516 316
42	3 838 580	6 698	1 438	0,323	824 108	710 019 293
t - 1	3 838 580	6 826	724	0,095	407 139	204 335 160
t - 2	3 838 580	6 823	767	0,100	431 510	215 124 023
				Yhteensä	<b>7 494 348</b>	<b>7 760 335 741</b>
				Yhteensä	<b>7 645 129</b>	<b>3 766 879 026</b>

**Päällekkäinen osa**

Työvoimaan kuulumattomat:

Rotaatioryhmä Vuosineljännes	N	n <sub>p</sub>	$\sum_{j=1}^{n_p} y_j$	$s_{I_p}^2$	$\hat{Y}_p$	$\text{var}(\hat{Y}_p)$
53	3 838 580	6 670	4 789	0,867	2 756 066	1 912 133 586
t	3 838 580	6 895	2 578	0,234	1 435 223	499 441 981
t - 1	3 838 580	6 782	2 337	0,226	1 322 731	489 886 102
51	3 838 580	6 656	6 963	1,817	4 015 630	4 014 971 680
t - 1	3 838 580	6 991	2 394	0,225	1 314 484	473 796 772
t - 2	3 838 580	6 955	2 418	0,227	1 334 534	479 677 441
t - 3	3 838 580	6 898	2 552	0,233	1 420 130	497 077 209
52	3 838 580	6 589	6 781	1,799	3 950 434	4 015 642 739
t	3 838 580	6 973	2 591	0,234	1 426 325	492 600 700
t - 1	3 838 580	6 939	2 409	0,227	1 332 633	480 465 430
t - 2	3 838 580	6 784	2 194	0,219	1 241 428	474 492 900
44	3 838 580	6 766	7 221	1,847	4 096 717	4 015 690 508
t	3 838 580	7 025	2 603	0,233	1 422 324	488 384 581
t - 2	3 838 580	7 034	2 443	0,227	1 333 189	474 056 596
t - 3	3 838 580	7 096	2 673	0,235	1 445 959	486 714 052
43	3 838 580	6 747	7 237	1,857	4 117 356	4 047 413 294
t	3 838 580	6 964	2 559	0,232	1 410 529	490 970 189
t - 1	3 838 580	7 000	2 429	0,227	1 331 987	476 161 676
t - 3	3 838 580	7 048	2 649	0,235	1 442 735	489 600 878
41	3 838 580	6 632	4 834	0,870	2 797 903	1 930 115 196
t - 2	3 838 580	6 722	2 376	0,229	1 356 808	500 133 013
t - 3	3 838 580	6 789	2 592	0,236	1 465 547	511 437 635
42	3 838 580	6 698	4 692	0,842	2 688 955	1 849 166 324
t - 1	3 838 580	6 826	2 399	0,228	1 349 070	491 215 442
t - 2	3 838 580	6 823	2 409	0,228	1 355 289	492 464 400
				Yhteensä	<b>24 423 061</b>	<b>21 785 133 327</b>
				Yhteensä	<b>24 740 926</b>	<b>8 788 576 996</b>