

Suomen Pankin kirjasto



0000090790

IVA5a

Kirjasto: alaholvi

SUOMEN PANKKI D

Suomen Pankin investointikyselyn otantaan liittyvistä  
Suomen pankin taloustieteellisen tutkimuslaitoksen julk.  
24 1970

SUOMEN PANKIN INVESTOINTIKYSELYN OTANTAAN

LIITTYVISTÄ ONGELMISTA

Heikki Koskenkylä

Suomen Pankin taloustieteellinen  
tutkimuslaitos

Sarja D:24 Monistettuja tutkimuksia

Helmikuu 1970

Suomen Pankin taloustieteellisen tutkimuslaitoksen julkaisuja

Sarja:D. Monistettuja tutkimuksia

Bank of Finland Institute for Economic Research Publications

Series:D. Mimeographed Studies

1. PERTTI KUKKONEN On the Measurement of Seasonal Variations. 1963. 11 s.
2. The Index Clause System in the Finnish Money and Capital Markets. 1964. 15 s.
3. J.J. PAUNIO Adjustment of Prices to Wages. 1964. 15 s.
4. HEIKKI VALVANNE and JAAKKO LASSILA The Taxation of Business Enterprises and the Development of Financial Markets in Finland. 1965. 26 s.
5. MARKKU PUNTILA Likvidien varojen kysyntä ja yleisön likviditeetin kehitys Suomessa vuosina 1948 - 1962. 1965. s. 110.
6. J.J. PAUNIO Taloudellinen kasvu ja suhdannevaihtelut dynaamisen makrotarkastelun valossa. 1965. 117 s.
7. AHTI MOLANDER Kokonaistaloudelliseen hinta- ja palkkatasoon vaikuttavat tekijät Suomessa vuosina 1949 - 1962. 1965. 159 s.
8. ERKKI PIHKALA Keskinäisen taloudellisen avun neuvoston pysyvät komissiot työnjaon toteuttajina. 1965. 35 s.
9. KARI NARS Statens prispolitiska parametrar. 1965. 118 s.
10. HEIKKI VALVANNE The Framework of the Bank of Finland's Monetary Policy. 1965. 34 s.
11. JOUKO SIVANDER Ulkomaankaupan substituoitujoustojen teoriasta ja mittauksesta. 1965. 91 s.
12. TIMO HELELÄ - PAAVO GRÖNLUND - AHTI MOLANDER Muistio palkkaneuvotteluja varten. 1965. 56 s.
13. ERKKI LAATTO Suomen ulkomaisen tavarakaupan volyyymi-indeksit neljännesvuosittain vuosina 1949 - 1964 eräistä lyhytaikaisista vaihteluista puhdistettuna (English Summary). 1965. 24 s.
14. DOLAT PATEL The Share of the Developing Countries in Finnish Foreign Trade. 1966. 31 s.
15. PEKKA LAHIKAINEN Tuotoksen ja työpanoksen välisen suhteen vaihteluista. 1966. 25 s.
16. HEIKKI U. ELONEN Yrityksen rahoituspääomien kysynnästä ja tarjonnasta. 1966. 88 s.
17. TIMO HELELÄ and J.J. PAUNIO Memorandum on Incomes Policy. 1967. 10 s.
18. KARI NARS Undersökning av efterfrågetrycket. 1967. 119 s.
19. KARI PUUMANEN Indeksivaateet valintakohteina. 1968. 186 s.

SUOMEN PANKIN  
KIRJASTO

Tilastotieteen laudaturtyö,  
Helsinki 1969

# SISÄLLYSLUETTELO

|   | sivu |
|---|------|
| 1. Johdanto .....   | 1    |
| 2. Tietojenhankintajärjestelmän yleispiirteitä .....                      | 3    |
| 2.1. Yleiset puitteet .....   | 3    |
| 2.2. Tietojenhankintajärjestelmän joustavuus .....                        | 5    |
| 3. Estimointi- ja otantamenetelmien vertailua .....                       | 6    |
| 3.1. Estimaattien jakautuminen .....                                      | 6    |
| 3.2. Suhde-estimaattori .....   | 11   |
| 3.3. Empiirisiä tuloksia eri estimointi- ja otanta-<br>menetelmillä ..... | 12   |
| 3.4. Apumuuttujan valinta suhdemenetelmään .....                          | 16   |
| 3.4.1. Korrelaatioanalyysi .....  | 16   |
| 3.4.2. Regressioanalyysi .....  | 19   |
| 3.4.3. Eri apumuuttujilla saadut varianssit                               | 21   |
| 4. Optimaalinen allokoiminen ja otoksen koon estimoi-<br>minen .....      | 24   |
| 4.1. Optimaalinen allokoiminen varianssien perus-<br>teella .....         | 24   |
| 4.1.1. Optimaalisen allokoinnin kaava .....                               | 26   |
| 4.1.2. Optimaalisen allokoinnin tulokset ....                             | 28   |
| 4.1.3. Minimivarianssit .....   | 32   |
| 4.2. Otoksen koon estimoiminen .....                                      | 34   |
| 4.3. Konvergoiminen kohti optimaalista allokoointia                       | 36   |
| 4.4. Optimaalinen allokointi $\bar{X}_n$ :n perusteella ....              | 41   |
| 5. Optimaalisten ositerajojen määrääminen .....                           | 44   |
| 5.1. Teoreettinen kaava .....   | 44   |
| 5.2. Optimaalisen ositerajan määrääminen käytännös-<br>sä .....           | 46   |
| 5.2.1. Otokseen ottamisaste .....   | 46   |
| 5.2.2. Optimaalisen ositerajan kaava .....                                | 47   |
| 6. Käytännölliset johtopäätökset .....                                    | 51   |
| Lähdeluettelo .....   | 52   |
| Liitteitä .....   | 53   |

## 1. Johdanto

Suomen Pankin taloustieteellisessä tutkimuslaitoksessa on jo muutaman vuoden ajan toimeenpantu teollisuuden investointeja koskeva kysely kaksi kertaa vuodessa, keväällä ja syksyllä. Kyselystä saadun otanta-aineiston avulla pyritään estimoimaan teollisuuden toteutuneet investoinnit kyselyä edeltävälle vuodelle sekä investointisuunnitelmat kysely- ja kyselyä seuraavalle vuodelle.

Tässä tutkimuksessa pyritään selvittämään kyselyyn liittyvän otannan empiirisiä ja teoreettisia ongelmia. Tehdasteollisuus on valittu tutkimuksen kohteeksi, koska se muodostaa yhtenäisen kokonaisuuden ja sen investoinnit ovat noin 90 % koko teollisuuden investoinneista. Tutkimuksen ulkopuolelle jätettyjen kaivannaisteollisuuden sekä sähkö- ja kaasulaitosten investoinnit ovat niin pieni osa teollisuuden kokonaisinvestoinneista, ettei niiden jättäminen pois vähennä tutkimuksen arvoa koko teollisuuden kannalta katsottuna. Tähän vaikuttaa myös se, että kaivannaisteollisuudessa kysely kohdistuu koko populaatioon.

Populaation muodostavat ne tehdasteollisuuden yritykset, joiden työntekijöiden lukumäärä on suurempi kuin 20. Otokseen kuuluvat yritykset on poimittu tasavälisellä ositetulla otannalla Tilastollisen päätoimiston yrityskohtaisilta reikäkorteilta. Käytettyä otantamenetelmää voidaan pitää ositettuna umpimähkäisenä otantana (Yates s. 29). Tehdasteollisuus on vanhastaan jaettu kyselyssä ja otosta otettaessa kolmeen osaan: puu- ja paperiteollisuuteen, metalliteollisuuteen sekä muuhun tehdasteollisuuteen. Nämä kolme pääryhmää ovat otannan kannalta osapopulaatioita, joista jokaisesta valittiin erikseen otos em. ositetulla umpimähkäisellä otannalla ja joille otoksista estimoidaan erikseen halutut osapopulaation suureet. Kukin osapopulaatio on jaettu kolmeen ositteeseen, joissa käytetyt otantasuhteet ovat samat kaikissa osapopulaatioissa. Suurista yrityksistä on valittu otokseen kaikki, keskisuurista joka toinen ja pienistä joka viides. Suuret yritykset on määritelty siten, että niiden työntekijöiden lukumäärä on yli 500, keskisuurten yritysten 100-499 ja pienten yritysten 20-99. Suuret yritykset

muodostavat siis ositteen, jonka otantasuhde on 100 %. Siksi varsinainen otos muodostuukin tämän tutkimuksen kannalta pienistä ja keskisuurista yrityksistä. Osittamisella on ollut kahdenlainen tarkoitus. Toisaalta on pyritty saamaan mahdollisimman tarkkoja estimaatteja osapopulaatioiden suureista ja toisaalta on haluttu saada erikseen toteutuneiden ja suunniteltujen investointien estimaatit yrityksen koon mukaan, koska on ilmeistä, että yrityksen investointitoiminta vaihtelee yrityksen koosta riippuen. Suurten yritysten sekä toisaalta pienten ja keskisuurten yritysten investointitoiminnan erot johtuvat ennen kaikkea siitä, että suuret yritykset suunnittelevat investointinsa pitemmällä tähtäimellä kuin pienet ja keskisuuret yritykset, sekä siitä, että niillä on vientikysynnän osuus varsin suuri osa kokonaiskysynnästä. Jälkimmäinen tekijä vaikuttaa mm. rahoitusmahdollisuuksien erilaisuuteen pieniin ja keskisuuriin yrityksiin verrattuna. Ositerajoja alun perin määrättäessä ei suoritettu varsinaista tilastollista tutkimusta optimaalisten ositerajojen löytämiseksi, vaan ositerajat valittiin harkinnanvaraisesti. Otoksen ja ositteiden koot on myös määrätty a priori olettamusten perusteella ilman varsinaista tutkimusta. Tässä tutkimuksessa pyritäänkin a posteriori analysoimaan, kuinka hyvin käytetty otantamenetelmä, estimointimenetelmä sekä otoksen ja ositteiden koot vastaavat analyttisesti saatuja menetelmiä ja optimaalisia arvoja. Samalla pyritään selvittämään, onko aihetta muuttaa käytössä olevaa otantasysteemiä siten, että tulevissa kyselyissä saadaan tarkempia tuloksia. Korostettakoon kuitenkin, että tällä tutkimuksella ei pyritty koko otantasysteemin optimaaliseen järjestämiseen, vaan ainoastaan otantasysteemiin liittyvien tiettyjen yksityisongelmien optimaaliseen ratkaisemiseen. Esim. optimaalisia ositteiden kokoja määrättäessä oletetaan, että ositerajat ja otoksen koko säilyvät nykyisellään.

Tutkimuksen ensimmäisessä osassa kuvaillaan lyhyesti käytetyn tietojenhankintajärjestelmän yleispiirteitä, joilla on myös merkitystä seuraavissa luvuissa esitettävien optimointitehtävien kannalta. Tässä yhteydessä määritellään myös tietojenhankintajärjestelmän joustavuus.

Tutkimuksen toisessa osassa tarkastellaan nykyisin käytössä olevaa estimointimenetelmää, joka on suhdemenetelmä, ja verrataan sillä saatuja estimaatteja ja estimaattien variansseja keskiarvomenetelmällä saatuihin vastaaviin arvoihin. Lyhyesti tutkitaan myös ositetun umpimähkäisen otannan ja yksinkertaisen umpimähkäisen otannan keskinäistä tarkkuutta. - Kun otanta-aineistosta lasketaan suhde-estimaattoreita, tarvitaan ns. apumuuttujaa, jolla otoksesta saadut arvot muutetaan perusjoukkoa koskeviksi luvuiksi. Erityisesti tutkitaan toistaiseksi käytetyn apumuuttujan - työntekijöiden lukumäärä - vaikutusta estimaattien tarkkuuteen verrattuna muihin mahdollisiin apumuuttujiin, joita ovat tuotannon brutto- ja jalostusarvo. Suoritetulla analyysillä tullaan osoittamaan, että tehokkain apumuuttuja vaihtelee osapopulaatioittain ja riippuu ennen kaikkea siitä, onko osapopulaatio työvoima- vai pääomaintensiivinen teollisuusala.

Luvussa 4 esitetään tutkimuksen tärkeimmät tulokset. Tällöin estimoidaan otoksen koko ja suoritetaan ositteisiin allokoiminen optimaalisesti. Analyysi suoritetaan osapopulaatioittain. Tutkimustuloksissa osoitetaan, että muuttamalla nykyisin käytössä olevaa allokointia optimaalisen allokoinnin mukaiseksi voidaan saada huomattavasti tarkempia estimaatteja.

Viimeisessä luvussa osoitetaan, kuinka teoreettisesti ja empiirisesti voidaan määrätä optimaaliset ositteiden rajat.

## 2. Tietojenhankintajärjestelmän yleispiirteitä

### 2.1. Yleiset puitteet

Kyselylomakkeissa tiedustellaan yritysten investointien arvoa pääomahyödyketyypeittäin eli kone- ja laiteinvestointeja, rakennusinvestointeja sekä maa- ja vesirakennusinvestointeja. Tämän tutkimuksen kannalta katsottiin riittäväksi aggregoida investoinnit pääomatyypeittäin yhteen vuositasolla, jolloin es-timoitaviksi suureiksi tulivat kyselyä edeltävän vuoden toteutunut investointien arvo sekä kysely- ja kyselyä seuraavan vuoden investointisuunnitelmien arvo. Kyselystä estimoidaan SP:n

käyttöä varten myös disaggregoidut arvot kullekin kolmelle vuodelle, mutta tässä tutkimuksessa estimoitavat suureet rajoitettiin kolmeen, jottei mm. optimointia olisi jouduttu suorittamaan liian usean suureen suhteen.

Yrityskuoleman ja yritysten yhteensulautumisen johdosta joudutaan otosta aika ajoin uusimaan, koska otoksen vanhentumessa poisjääminen kasvaa em. syistä suuresti. Uusittaessa säilyy otoksesta kuitenkin noin 50 % ennallaan ennen kaikkea siksi, että suurten yritysten ositteessa otantasuhde on 100 %. Otosta uusittaessa on sen jatkuvuus siten turvattu. Koska keskisuurten yritysten otantasuhde on ollut noin 50 % ja pienten yritysten noin 20 %, tulee näistä ositteista otokseen sitä uusittaessa aina melkoisesti uusiakin yrityksiä.

Otokseen kuuluville yrityksille lähetettävät kyselylomakkeet on laadittu siten, että niihin vastaaminen on mahdollisimman helppoa ja että kysymykset samalla täyttävät ne vaatimukset, joita kyselyistä halutuille tiedoille asetetaan. Vastauksissa arvot annetaan markkamääräisinä. Kyselyyn vastaajille annetaan myös yksityiskohtaiset kirjalliset ohjeet.

Kyselyn aikataulu on laadittu siten, että sen avulla saadaan Suomen Pankin sisäiseen käyttöön tarkoitettuihin investointiennusteisiin mahdollisimman tuoretta aineistoa. Toinen kyselyistä suoritetaan keväällä toukokuussa ja toinen syksyllä marraskuussa. Kyselylomakkeet lähetetään vastaajille noin 2-3 viikkoa ennen kuin vastaukset pyydetään palauttamaan. Vaikka pääosa vastauksista palautetaankin varsin nopeasti Suomen Pankkiin, venyy viimeisten vastausten saapuminen jopa 2-3 kuukauteen kysymysten lähettämisestä. Tämä viivästyttää melkoisesti aineiston lopullista analyysia.

Aineistolle tehdään tietokoneella sekä alustavia käsitteilyitä sitä mukaa kuin vastauksia saapuu että lopullinen analyysi. Alustavissa ja lopullisessa käsittelyssä suoritetaan estimointi eri tavoin. Tässä tutkimuksessa analysoidaan vain lopullisessa käsittelyssä käytettyä estimointimenetelmää, jolloin suhde-estimaattorissa on apumuuttujana työntekijöiden lukumäärä, tuotannon bruttoarvo tai tuotannon jalostusarvo. Aineiston alustavissa estimoinneissa käytetään myös suhdemenetelmää, mutta



apumuuttujana on tällöin edellisestä kyselystä saatu investointien arvo. Tällöin ei tunneta apumuuttujan todellista populaation kokonaisarvoa, vaan vain sen estimaatti. Näiden kahden menetelmän antamissa estimaateissa ei ole huomattu kovin suuria eroja, mutta estimaattien varianssi ilmeisesti vaihtelee melkoisesti menetelmästä riippuen. Alustavan analyysin menetelmällä saadaan suhteellisen luotettavia estimaatteja ennen kaikkea investointien muutosprosentteille.

## 2.2. Tietojenhankintajärjestelmän joustavuus

Tietojenhankintajärjestelmän joustavuudella tarkoitetaan otantasysteemin käsittelyyn kuuluvaa kokonaisaikaa kysymysten postittamisesta lopulliseen analyysiin. Joustavuus on näin ollen riippuvainen eri osapopulaatioihin ja ositteisiin kuuluvien yritysten keskimääräisestä vastausajasta. Seuraavassa tutkitaan varianssianalyysillä, onko keskimääräinen vastausaika riippuvainen yrityksen koosta. Yksisuuntainen varianssianalyysi suoritetaan kahden kyselyn aineistolle. Taulukossa 1 esitetään keskimääräiset vastausajat kyselyittäin sekä taulukossa 2 saadut F-luvut ja yksisuuntaiselle testille 99 %:n luottamusarvo. Nollahypoteesina on, että ryhmien välinen varianssi ei ole ryhmien sisäistä varianssia suurempi. Jos nollahypoteesia ei hylätä, niin keskimääräiset vastausajat eivät poikkea toisistaan merkitsevästi valitulla konfidenssitasolla (Mills s. 553). F-luvussa on ryhmien välinen varianssi osoittajassa. Jos F on pienempi kuin yksi, niin oletetaan, että ainoastaan satunnaistekijät vaikuttavat ryhmien väliseen varianssiin. Vain jos F on merkitsevästi suurempi kuin yksi, hylätään nollahypoteesi, että todellinen  $F=1$  (Mills s. 553).

Taulukko 1. Keskimääräiset vastausajat eri kyselyissä, osapopulaatioissa ja ositteissa, vrk

S = suuret yritykset, K = keskisuuret yritykset, P = pienet yritykset

|                          | kevät 1967 |     |      | kevät 1968 |     |     |
|--------------------------|------------|-----|------|------------|-----|-----|
|                          | S          | K   | P    | S          | K   | P   |
| Puu- ja paperiteollisuus | 5.3        | 7.0 | 10.0 | 4.7        | 6.6 | 8.4 |
| Metalliteollisuus        | 5.7        | 5.0 | 7.3  | 4.6        | 8.4 | 4.9 |
| Muu tehdasteollisuus     | 4.8        | 4.3 | 5.5  | 5.0        | 5.9 | 8.8 |

Taulukko 2. Varianssianalyysin tulokset,  $F = v^2$

Puu- ja paperiteollisuus

| Aineisto | Kevät 1968                    | Kevät 1967                    |
|----------|-------------------------------|-------------------------------|
|          | $F(2,66) = 1.10$              | $F(2,74) = 1.59$              |
|          | $P(v^2(2,66) < 4.95) = 99 \%$ | $P(v^2(2,74) < 4.90) = 99 \%$ |

Metalliteollisuus

| Aineisto | Kevät 1968                     | Kevät 1967                     |
|----------|--------------------------------|--------------------------------|
|          | $F(2,135) = 3.33$              | $F(2,116) = 1.66$              |
|          | $P(v^2(2,135) < 4.77) = 99 \%$ | $P(v^2(2,110) < 4.80) = 99 \%$ |

Muu tehdasteollisuus

| Aineisto | Kevät 1968                     | Kevät 1967                     |
|----------|--------------------------------|--------------------------------|
|          | $F(2,237) = 2.95$              | $F(2,283) = 0.46$              |
|          | $P(v^2(2,237) < 4.70) = 99 \%$ | $P(v^2(2,283) < 4.68) = 99 \%$ |

Taulukosta voidaan havaita, että nollahypoteesia ei hylätä yhdessäkään tapauksessa. Näin on voitu osoittaa, että kyselystä ja osapopulaatiosta riippumatta keskimääräiset vastausajat eivät eroa merkitsevästi toisistaan eri kokoryhmissä. Kun taulukosta 1 voidaan vielä havaita, että keskimääräiset vastausajat eivät eroa paljoa toisistaan eri osapopulaatioissa, voidaan kokotosta pitää varsin homogeenisena keskimääräisen vastausajan suhteen.

### 3. Estimointi- ja otantamenetelmien vertailua

#### 3.1. Estimaattien jakautuminen

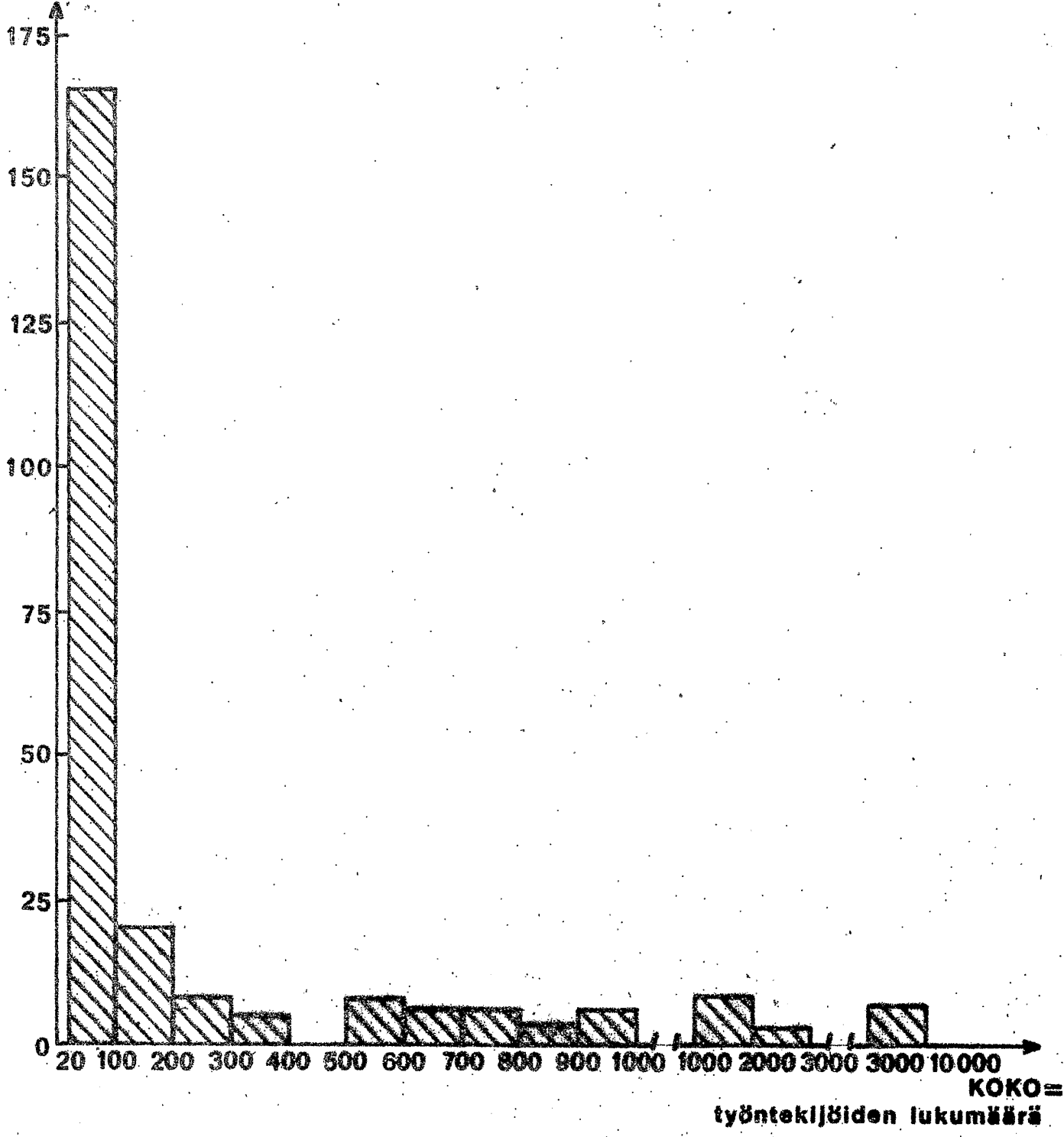
Oletetaan, että parametri  $c$  on jonkin otoksesta määrättävän suureen todellinen arvo ja  $\underline{t}$  saman suureen otoksesta saatu arvo. Otoksen antama tieto parametrissa  $c$  voidaan ilmoittaa nk. konfidenssivälillä, jolloin on tehtävä oletus  $\underline{t}$ :n jakautumasta konfidenssivälin laskemiseksi. Yleensä oletetaan, että  $\underline{t}$  on jakautumaltaan ainakin likimäärin normaalin (Elfvig

s. 125-126). Matemaattisessa tilastotieteessä on osoitettu, että äärellisen hajonnan omaavasta populaatiosta otetun otoksen keskiarvon jakautuma on asympotoottisesti normaalin havaintojen lukumäärän suhteen. Populaation edellytetään kuitenkin tällöin olevan ääretön. Madow on osoittanut saman tuloksen pätevän myös äärelliselle populaatiolle eräin otoksen kokoa koskevin ehdoin (Cochran s. 22). A priori on kuitenkin vaikea sanoa, kuinka suuri on otoksen koon oltava, jotta normaalin approksimaatio pätsi. Seuraavassa on graafisten tarkasteluiden avulla tutkittu kevään 1968 kyselyn aineistosta puu- ja paperiteollisuudessa, voidaanko normaalista approksimaatiota käyttää tässä otoksessa. Ensin esitetään kuitenkin kuviossa 1 frekvenssitaulukon muodossa yritysten kokojakautuma samasta aineistosta. Kuvioista voidaan havaita, että koon frekvenssijakautuma on positiivisesti vino. Usein positiivinen vinous taloudellisten yksiköiden koon jakautumassa heijastuu estimoitavan suureen jakautumaan. Kuviossa 2 esitetään vuodelle 1967 estimoidun investointien arvon otosjakautuma puu- ja paperiteollisuudessa. Otokset, joista estimaatit laskettiin, valittiin 30 yrityksen kokoisina aliotoksina varsinaisesta otoksesta, jossa oli 60 yritystä. Aliotoksia voidaan pitää umpimähkäisinä otoksina, koska ne valittiin satunnaislukujen avulla ja koska varsinaisen otos oli valittu ositetulla umpimähkäisellä otannalla, jolloin ositteessa 1 oli 49 ja ositteessa 2 19 yritystä. Aliotoksissa valittiin ositteeseen 1 20 yritystä ja ositteeseen 2 10 yritystä. Aliotoksia valittiin kaikkiaan 48 ja niiden estimaatit laskettiin suhdemenetelmällä (ratio-separate). Kuvioista 2 havaitaan estimoidun investointien kokonaisarvon jakautumassa myös lievää positiivista vinoutta. Normaalin approksimaatio vaikuttaa kuitenkin varsin hyvältä. Suurten yritysten poistaminen otannan kohteena olevasta populaatiosta sillä perusteella, että otantasuhde on niiden osalta 100 %, alentaa vinoutta ja parantaa normaalista approksimaatiota. Myös muissa osapopulaatioissa on yritysten kokojakautuma positiivisesti vino ja on oletettavissa, että tämä vinous heijastuu myös otosestimointien jakautumassa. Muita osapopulaatioita ei ole tässä tutkittu tarkemmin, koska ne ovat käyttäytymiseltään ilmeisen samankaltaisia puu- ja paperiteollisuuden kanssa.

Kuvio 1

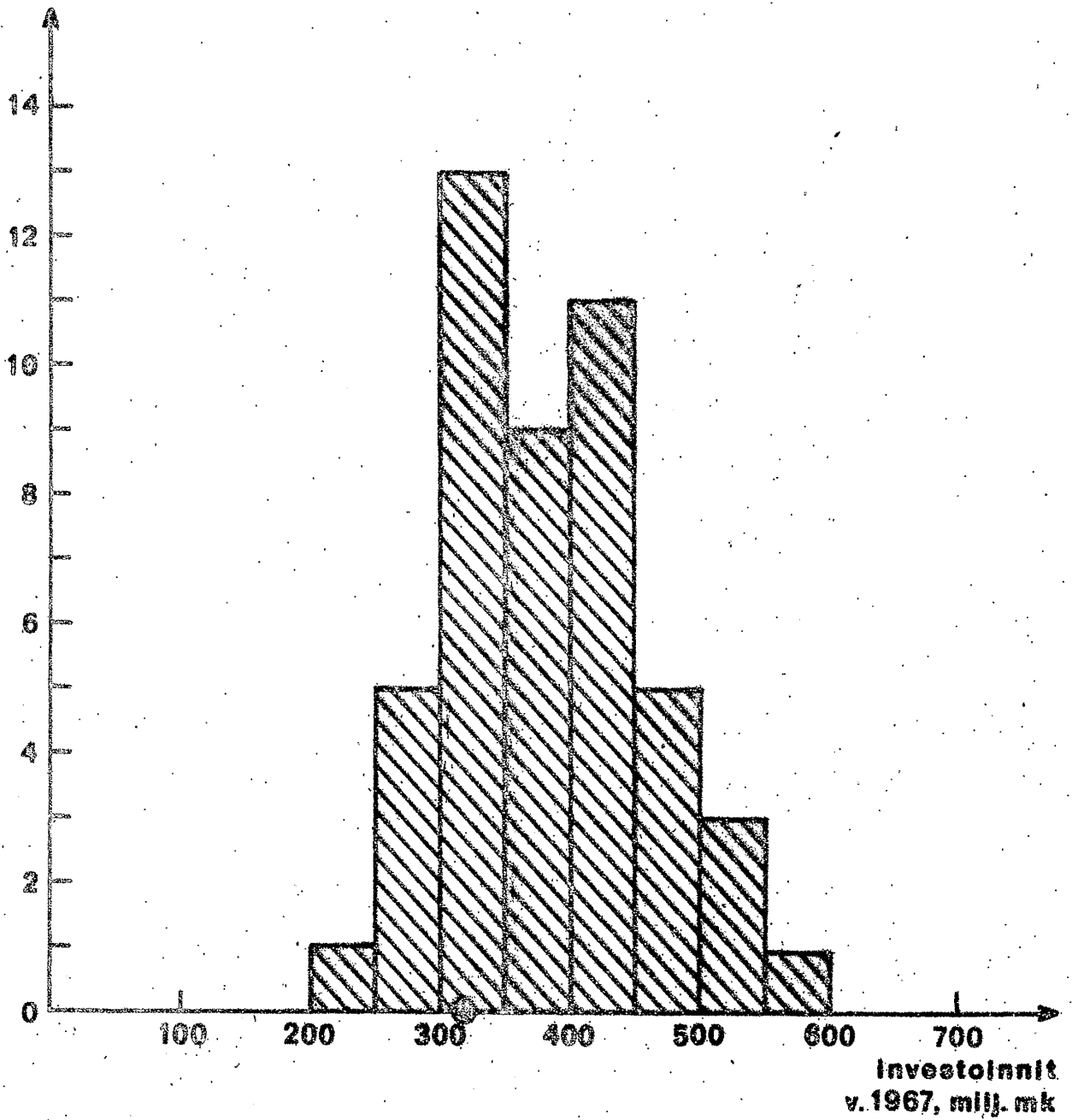
**Yritysten kokojakautuma puu-  
ja paperiteollisuudessa**

yritysten  
lukumäärä



Kuvio 2

Katimatin otosjakausten puu-  
ja paperiteollisuudessa



⊙ investointien todellisen populaation erro  
lähde 1967

Symbolitaulukko

- $x_i$  = apumuuttuja suhde-estimaattorissa,  $i$ :s yksikkö
- $x_h$  =  $x_i$ :den kokonaisarvo otoksen ositteessa  $h$
- $y_i$  = tarkasteltavan suureen  $i$ :nnen yksikön arvo
- $y_h$  =  $y_i$ :den kokonaissumma otoksen ositteessa  $h$
- $X_h$  =  $x_i$ :den kokonaissumma populaation ositteessa  $h$
- $Y_h$  =  $y_i$ :den        "-                        "-                        "-
- est  $Y_h = \hat{Y}_h$
- $r$  =  $R(x_i, y_i)$  =  $x_i$ :n ja  $y_i$ :n välinen korrelaatiokerroin
- $S_x$  =  $x_i$ :n hajonta koko populaatiossa
- $S_y$  =  $y_i$ :n        "-                        "-
- $\bar{X}$  =  $\frac{X}{N}$
- $\bar{Y}$  =  $\frac{Y}{n}$
- $N$  = yritysten lukumäärä koko populaatiossa
- $N_h$  =        "-                        "-        populaatiossa ositteessa  $h$
- $n$  = yritysten lukumäärä koko otoksessa
- $n_h$  =        "-                        "-        otoksen ositteessa  $h$
- $s_x$  = est  $S_x$
- $s_y$  = est  $S_y$
- $S_x / \bar{X}$  = coefficient of variation of  $x$  merk. CV
- $S_y / \bar{Y}$  =        "-                        "-         $y$  merk. CV

### 3.2. Suhde-estimaattori

Matemaattisessa tilastotieteessä esitetyt estimointimenetelmät edellyttävät, että tunnetaan otosaineiston frekvenssijakautuman funktionaalinen muoto. Tällöin jokin estimointimenetelmä soveltuu juuri tätä jakautumaa varten. Yleensä otanta-teoriassa on tehty vain hyvin väljiä oletuksia frekvenssijakautumasta. On tyydytty vain toteamaan, että jakautuma on esim. vino tai melkein symmetrinen eikä ole tarkemmin spesifioitu jakautuman tyyppiä. Siksi ei otantatutkimuksissa ole juuri sovellettu varsinaisia estimointimenetelmiä. Pääasiassa on käytetty kahta menetelmää, nimittäin suhde- ja regressiomenetelmää (Cochran s. 111). Tässä otantatutkimuksessa keskitytään ennen kaikkea suhdemenetelmän käyttöön ja tutkimiseen.

Suhdemenetelmässä käytetään hyväksi apumuuttujaa  $x_i$ , joka korreloi suureen  $y_i$  kanssa. Apumuuttujan arvo on tunnettava jokaiselle otosyksikölle  $i$ . Lisäksi on tunnettava  $x_i$ :den populaation kokonaisarvo  $X$ . Käytännössä  $x_i$  on usein  $y_i$ :n arvo jonakin aikaisempaan ajankohtana. Tässä tutkimuksessa  $x_i$  on työntekijöiden lukumäärä, tuotannon bruttoarvo tai tuotannon jalostusarvo. Suhdemenetelmän tarkoituksena on lisätä estimaattien tarkkuutta suhteessa keskiarvomenetelmään (simple expansion) käyttämällä hyväksi  $x_i$ :n ja  $y_i$ :n välistä riippuvuutta (Cochran s. 112).

Investointien kokonaisarvo ositteessa  $h$  saadaan estimoiduksi suhdemenetelmällä (ratio separate) seuraavasti:

$$\hat{Y}_h = \text{est } Y_h = \frac{y_h}{x_h} X_h \quad (\text{Cochran s. 129})$$

Investointien kokonaisarvon estimaatti koko populaatiolle on tällöin

$$\hat{Y}_{Rs} = \sum_h \frac{y_h}{x_h} X_h \quad (\text{Cochran s. 129})$$

Edellä oleva suhdemenetelmä on ns. ratio separate -menetelmä. Siinä edellytetään suhteen  $y_h / x_h$  laskemista erikseen kussakin ositteessa. Toinen suhdemenetelmä on ns. ratio combined-menetelmä. Sitä varten täytyy laskea arvot  $\hat{Y}_{st} = \sum_h N_h \bar{y}_h$  ja  $\hat{X}_{st} = \sum_h N_h \bar{x}_h$ , jotka ovat  $y_i$ :n ja  $x_i$ :n populaation kokonaisar-

von estimaatteja ja ne on laskettu keskiarvomenetelmällä ositetusta otoksesta (Cochran s. 131). Combined suhde-estimaatti  $\hat{Y}_{Rc}$  (c for combined) on tällöin

$$\hat{Y}_{Rc} = \frac{\hat{Y}_{st}}{\hat{X}_{st}} X = \frac{\bar{y}_{st}}{\bar{x}_{st}} X$$

jossa  $\bar{y}_{st} = \hat{Y}_{st} / N$ ,  $\bar{x}_{st} = \hat{X}_{st} / N$  ovat estimoituja populaation keskiarvoja ositetusta otoksesta (Cochran s. 131). Olennainen ero näiden kahden suhde-estimaattorin välillä on siinä, että estimaattorin  $\hat{Y}_{Rc}$  laskemisessa tarvitsee tuntea vain  $X$ , mutta ei  $X_h$ :ta. Jos  $R_h$ :t,  $R_h = \frac{Y_h}{X_h}$ , vaihtelevat suuresti ositteittain, antaa  $\hat{Y}_{Rs}$  tarkempia estimaatteja (Murthy s. 377). Jos sitä vastoin  $R_h$ :t ovat suunnilleen yhtäsuuria ositteittain, saadaan molemmilla menetelmillä suunnilleen yhtä tarkkoja estimaatteja (Sukthatme s. 170).

Apumuuttujaksi pyritään valitsemaan muuttuja, jonka korrelaatio  $y_i$ :n kanssa on mahdollisimman suuri (Murthy s. 370). Tarkat estimaatit edellyttävät myös, että  $y_i$ :n ja  $x_i$ :n välinen regressiosuora kulkee likimäärin origon kautta eli on muotoa  $y_i = ax_i$  ja että  $y_i$ :n varianssi regressiosuoran ympärillä kasvaa suhteessa  $x_i$ :hin (Cochran s. 140, Yates s. 156, Sukthatme s. 154). Jos kaikki nämä ehdot ovat edes likimäärin voimassa, on vaikea löytää suhdemenetelmää parempaa estimointimenetelmää (Cochran s. 124).

### 3.3. Empiirisiä tuloksia eri estimointi- ja otantamenetelmillä

Seuraavissa taulukoissa esitetään eri menetelmillä saatuja estimaatteja kevään 1968 aineistosta osapopulaatioittain. Taulukon yhteydessä on esitetty myös kaavat, jotka perustuvat ositetun umpimähkäisen otannan käyttöön.



Taulukko. Estimointimenetelmien vertailua niillä saatujen estimaattien valossa

Puu- ja paperiteollisuus

|     |  | Estimoitava suure, milj.mk |      |      |      |
|-----|--|----------------------------|------|------|------|
|     |  | investoinnit               | 1967 | 1968 | 1969 |
| (1) | $N \bar{y}_{st} = \sum_h N_h \bar{y}_h$ (Cochran s. 66)                  |                            | 20.1 | 30.9 | 16.0 |
| (2) | $\hat{Y}_{RS} = \sum_h \frac{y_h}{x_h} \cdot X_h$ ( -"- s. 129)          |                            | 21.2 | 31.7 | 16.6 |
| (3) | $\hat{Y}_{RC} = \frac{\hat{Y}_{st}}{\hat{Y}_{st}} \cdot X$ ( -"- s. 131) |                            | 22.2 | 32.3 | 17.6 |

Metalliteollisuus

| Estimointimenetelmä | inv. | 1967 | 1968  | 1969 |
|---------------------|------|------|-------|------|
| (1)                 |      | 92.8 | 101.5 | 87.1 |
| (2)                 |      | 94.9 | 114.4 | 94.7 |
| (3)                 |      | 95.3 | 104.3 | 89.5 |

Muu tehdasteollisuus

|     | inv. | 1967  | 1968  | 1969  |
|-----|------|-------|-------|-------|
| (1) |      | 239.5 | 195.0 | 148.9 |
| (2) |      | 257.2 | 207.8 | 163.8 |
| (3) |      | 251.5 | 204.8 | 156.6 |

Tuloksista voidaan havaita, että eri menetelmillä saadut estimaatit poikkeavat jonkin verran toisistaan. Suhde-estimaatit poikkeavat kuitenkin vähemmän toisistaan kuin keskiarvoestimaatista (1). Tuloksista voidaan tehdä se johtopäätös, että työntekijöiden lukumäärän käyttö ositusmuuttujana ja suhde-estimaatissa apumuuttujana vaikuttaa samalla tavalla estimaatin muodostumiseen.

Kullakin estimointimenetelmällä saadulle estimaatille on seuraavassa laskettu varianssit ja tutkittu niiden perusteella eri estimointimenetelmien keskinäistä tarkkuutta. Sen lisäksi on vertailtu myös ositettua ja yksinkertaista umpimähkäistä otantaa laskemalla niistä keskiarvomenetelmällä saatujen esti-

maattien varianssit. Vaikka käytetty otantamenetelmä onkin ositetty umpimähkäinen otanta, voidaan kuitenkin laskea samasta populaatiosta yksinkertaisella umpimähkäisellä otoksella saatavien estimaattien variansseja, kun estimointimenetelmänä on keskiarvomenetelmä (Cochran s. 97-99) (kaava 5.61 s. 99 kerrottuna  $N^2$ :lla, koska estimoidaan populaation kokonaisarvoa). Liitteessä 1 esitetään varianssien laskemisessa tarvittavat apusuureet ja käytetyt varianssien kaavat. Eri menetelmillä saatiin seuraavia variansseja kevään 1968 aineistosta:

Taulukko. Estimaattien varianssit

| <u>Puu- ja paperiteollisuus</u> |   | estimoitava suure |        |        |        |
|---------------------------------|---|-------------------|--------|--------|--------|
| otantamenetelmä                 | estimointimenetelmä                     | inv.              | 1967   | 1968   | 1969   |
| (1) simple random               | $N \bar{y}$                             |                   | 58.1   | 150.3  | 35.8   |
| (2) stratified random           | $N \bar{y}_{st} = \sum_h N_h \bar{y}_h$ |                   | 10.9   | 53.7   | 7.0    |
| (3) -"-                         | $\hat{Y}_{Rs}$                          |                   | 8.9    | 42.6   | 7.5    |
| (4) -"-                         | $\hat{Y}_{Rc}$                          |                   | 9.0    | 43.2   | 7.3    |
| <u>Metalliteollisuus</u>        |   | investoinnit      | 1967   | 1968   | 1969   |
| (1)                             |   |                   | 448.8  | 1195.0 | 1072.6 |
| (2)                             |   |                   | 118.8  | 400.0  | 146.7  |
| (3)                             |   |                   | 132.2  | 170.6  | 155.2  |
| (4)                             |   |                   | 135.6  | 186.6  | 170.3  |
| <u>Muu tehdasteollisuus</u>     |   | investoinnit      | 1967   | 1968   | 1969   |
| (1)                             |   |                   | 3160.0 | 1420.0 | 890.0  |
| (2)                             |   |                   | 494.4  | 285.0  | 154.4  |
| (3)                             |   |                   | 421.8  | 267.7  | 146.8  |
| (4)                             |   |                   | 435.5  | 285.6  | 150.6  |

Seuraavassa taulukossa on menetelmällä (1) saatua varianssia merkitty 100:lla. Muilla menetelmillä saadut varianssit on saatu kaavalla  $\left[ \frac{(k)}{(1)} \right] \cdot 100$ , jolloin  $k = (2), (3), (4)$ . Menetelmän tehokkuuden mittana on sillä saadun varianssin pienuus.

Taulukko. Estimaattien varianssit suhteessa toisiinsa

Puu- ja paperiteollisuus

|     | investoinnit | 1967  | 1968  | 1969  |
|-----|--------------|-------|-------|-------|
| (1) |              | 100.0 | 100.0 | 100.0 |
| (2) |              | 18.8  | 35.0  | 19.6  |
| (3) |              | 15.3  | 37.2  | 20.9  |
| (4) |              | 15.5  | 37.4  | 20.4  |

Metalliteollisuus

|     |  |       |       |       |
|-----|--|-------|-------|-------|
| (1) |  | 100.0 | 100.0 | 100.0 |
| (2) |  | 26.5  | 33.3  | 13.6  |
| (3) |  | 29.4  | 14.2  | 14.4  |
| (4) |  | 30.2  | 15.6  | 15.8  |

Muu tehdasteollisuus

|     |  |       |       |       |
|-----|--|-------|-------|-------|
| (1) |  | 100.0 | 100.0 | 100.0 |
| (2) |  | 15.6  | 20.1  | 17.3  |
| (3) |  | 13.3  | 18.9  | 16.5  |
| (4) |  | 13.8  | 20.1  | 16.9  |

Taulukosta havaitaan, että ositetusta umpimähkäisestä otoksesta sekä keskiarvomenetelmällä että suhdemenetelmällä saatujen estimaattien varianssit ovat huomattavasti pienempiä kuin yksinkertaisesta umpimähkäisestä otoksesta keskiarvomenetelmällä saadun estimaatin varianssi. Ositettu umpimähkäinen otanta on siis selvästi tarkempi kuin yksinkertainen umpimähkäinen otanta.

Työntekijöiden lukumäärän käyttö ositusmuuttujan lisäksi apumuuttujana suhdemenetelmässä alentaa varianssia noin 10 - 20 % osapopulaatiosta riippuen, joten suhdemenetelmä on ositetussa otannassa keskiarvomenetelmää jonkin verran tarkempi. Molemmilla suhdemenetelmillä (separate and combined) saatujen estimaattien varianssit eroavat sitä vastoin hyvin vähän toisistaan.

Yhteenvedona voidaan todeta, että jokaisessa osapopulaatiossa ja kaikkien estimoitavien suureiden suhteen on otantamenetelmistä ositettu umpimähkäinen otanta yksinkertaista umpimähkäistä otantaa selvästi tarkempi. Estimointimenetelmistä on suhdemenetelmä (separate) tarkin. Suhdemenetelmän ja regressiomene-

telmän keskinäistä tehokkuutta vertaillaan seuraavassa luvussa  $y_i$ :n ja apumuuttujan  $x_i$  välisen regressiosuoran kulun perusteella (Sukthatme s. 221).

### 3.4. Apumuuttujan valinta suhdemenetelmään

Kuten edellisessä luvussa (s.10-11) todettiin, riippuu apumuuttujan tehokkuus sen korrelaatiosta suureen  $y_i$  kanssa sekä apumuuttujan ja  $y_i$ :n välisen regressiosuoran kulutusta (Cochran s. 140, Yates s. 156). Suhdemenetelmässä on toistaiseksi käytetty apumuuttujana työntekijöiden lukumäärää. Sen lisäksi voidaan apumuuttujana käyttää tuotannon bruttoarvoa tai jalostusarvoa. Seuraavassa tutkitaan eri apumuuttujien paremmuutta suhdestimointimenetelmässä estimaattien tarkkuuden suhteen. Analyysi suoritetaan korrelaatiokertoimien, regressiosuorien kulun sekä estimaattien varianssien perusteella osapopulaatioittain. Tutkimusaineistona on ollut pääasiassa syksyn 1968 kysely. Korrelaatiokertoimet on laskettu myös kevään 1968 kyselyn aineistosta, mutta niitä ei ole esitetty, koska niistä voidaan tehdä samat johtopäätökset kuin syksyn 1968 kyselyn aineistosta.

#### 3.4.1. Korrelaatioanalyysi

Korrelaatiokertoimiksi saatiin osapopulaatioittain seuraavat:

Merkintä:  $M$  = aritmeettien keskiarvo

Taulukko. Puu- ja paperiteollisuus

I apumuuttuja = työntekijöiden lukumäärä

| estimoitava suure | osite 1      | osite 2      | koko otos    |
|-------------------|--------------|--------------|--------------|
| investoinnit 1967 | .6315        | -.4044       | .2511        |
| " 1968            | .1135        | .6346        | .6193        |
| " 1969            | <u>.1162</u> | <u>.0728</u> | <u>.4319</u> |
|                   | $M = .2871$  | $M = .1010$  | $M = .4341$  |

II apumuuttuja = tuotannon bruttoarvo

| estimoitava suure | osite 1      | osite 2      | koko otos    |
|-------------------|--------------|--------------|--------------|
| investoinnit 1967 | .6114        | .7784        | .6634        |
| " 1968            | .6532        | .7794        | .7742        |
| " 1969            | <u>.2966</u> | <u>.5265</u> | <u>.6656</u> |
|                   | M = .5204    | M = .6948    | M = .7011    |

III apumuuttuja = tuotannon jalostusarvo

| estimoitava suure | osite 1      | osite 2      | koko otos    |
|-------------------|--------------|--------------|--------------|
| investoinnit 1967 | .6239        | .6427        | .6097        |
| " 1968            | .5844        | .5104        | .6513        |
| " 1969            | <u>.3416</u> | <u>.3541</u> | <u>.5654</u> |
|                   | M = .5166    | M = .5024    | M = .6088    |

Taulukosta voidaan havaita, että korrelaatiokertoimien suuruudessa on apumuuttujittain suuria eroja. Kummassakin ositteessa ja siten myös koko otoksessa on tuotannon brutto- ja jalostusarvon korrelaatio jokaisen estimoitavan suureen kanssa noin 2-3 -kertainen työntekijöiden lukumäärällä saatuun korrelaatioon verrattuna. Havaittua korrelaatiokertoimien suuruuden eroa voidaan pitää varsin merkittävänä. Tämän perusteella pitäisi tuotannon brutto- ja jalostusarvolukujen käytön apumuuttujana antaa pienempiä estimaattien variansseja kuin työntekijöiden lukumäärän käytön. Tuotannon brutto- ja jalostusarvolukujen väliset erot korrelaatiokertoimien suuruudessa ovat sitä vastoin niin pienet, ettei korrelaatiokertoimien perusteella voida pitää toista tarkempana apumuuttujana kuin toista.

Metalliteollisuudessa saatiin seuraavia korrelaatiokertoimia:

Taulukko. Metalliteollisuus

I apumuuttuja = työntekijöiden lukumäärä

| estimoitava suure | osite 1      | osite 2      | koko otos    |
|-------------------|--------------|--------------|--------------|
| investoinnit 1967 | .0559        | .1701        | .3522        |
| investoinnit 1968 | .3148        | .4914        | .4786        |
| investoinnit 1969 | <u>.2004</u> | <u>.1098</u> | <u>.3548</u> |
|                   | M = .1904    | M = .2571    | M = .3952    |

II apumuuttuja = tuotannon bruttoarvo

| estimoitava suure | osite 1      | osite 2      | koko otos    |
|-------------------|--------------|--------------|--------------|
| investoinnit 1967 | .0016        | .0987        | .2175        |
| " 1968            | -.0459       | .1083        | .1711        |
| " 1969            | <u>.0807</u> | <u>.2265</u> | <u>.2045</u> |
|                   | M = .0102    | M = .1445    | M = .1977    |

III apumuuttuja = tuotannon jalostusarvo

| estimoitava suure | osite 1       | osite 2      | koko otos    |
|-------------------|---------------|--------------|--------------|
| investoinnit 1967 | .0468         | .0969        | .2360        |
| " 1968            | -.0195        | .1139        | .1903        |
| " 1969            | <u>-.0046</u> | <u>.1835</u> | <u>.1757</u> |
|                   | M = .0227     | M = .1314    | M = .2007    |

Taulukosta havaitaan, että työntekijöiden lukumäärällä saadut korrelaatiokertoimet ovat selvästi suuremmat kuin tuotannon brutto- ja jalostusarvoilla saadut korrelaatiokertoimet. Ositteessa yksi on työntekijöillä saatu korrelaatiokerroin kymmenkertainen ja ositteessa kaksi kaksinkertainen verrattuna brutto- ja jalostusarvoluvuilla saatuihin korrelaatiokertoimiin. Korrelaatiokertoimien erojen perusteella tulisi työntekijöiden lukumäärän antaa tarkempia estimaatteja kuin tuotannon brutto- ja jalostusarvolukujen.

Puunjalostusteollisuudessa saatiin päinvastainen tulos kuin metalliteollisuudessa. Tähän on syynä se, että puunjalostusteollisuudessa yritykset ovat kooltaan keskimäärin suurempia ja ennen kaikkea pääomavaltaisempia kuin metalliteollisuudessa. Yrityksen koon ja investointien kasvaessa puunjalostusteollisuudessa ei työntekijöiden lukumäärä kasva samassa suhteessa kuin investointien arvo, kun sitä vastoin tuotannon brutto- ja jalostusarvo kasvavat. Tämän vuoksi tuotannon brutto- ja jalostusarvo soveltuvat paremmin käytettäväksi apumuuttujana suhdemene- telmässä. Metalliteollisuus sitä vastoin on erittäin työvoima- intensiivinen ala, joten on odotettavissa, että yrityksen inves- tointien kasvaessa työntekijöiden lukumäärä kasvaa suunnilleen samassa suhteessa. Täten selittyy se, että työntekijöiden luku- määrä soveltuu metalliteollisuudessa paremmin apumuuttujaksi kuin puunjalostusteollisuudessa.

Muussa tehdasteollisuudessa saatiin seuraavia korrelaatio-kertoimia:

| Taulukko.                                |      | Muu tehdasteollisuus |              |              |
|--|------|----------------------|--------------|--------------|
| I apumuuttuja = työntekijöiden lukumäärä |      |                      |              |              |
| estimoitava suure                        |      | osite 1              | osite 2      | koko otos    |
| investoinnit 1967                        |      | .0984                | .4575        | .5413        |
| "  | 1968 | .1606                | .1768        | .3821        |
| "  | 1969 | <u>.1801</u>         | <u>.2496</u> | <u>.4404</u> |
|  |      | M = .1464            | M = .2946    | M = .4546    |

| II apumuuttuja = tuotannon bruttoarvo |      |              |               |              |
|---------------------------------------|------|--------------|---------------|--------------|
| estimoitava suure                     |      | osite 1      | osite 2       | koko otos    |
| investoinnit 1967                     |      | .1573        | .7937         | .7875        |
| "                                     | 1968 | .1744        | .2916         | .3782        |
| "                                     | 1969 | <u>.0764</u> | <u>-.0953</u> | <u>.0480</u> |
|                                       |      | M = .1361    | M = .3300     | M = .4046    |

| III apumuuttuja = tuotannon jalostusarvo |      |              |               |              |
|--|------|--------------|---------------|--------------|
| estimoitava suure                        |      | osite 1      | osite 2       | koko otos    |
| investoinnit 1967                        |      | .3323        | .7160         | .7472        |
| "  | 1968 | .3040        | .3292         | .4470        |
| "  | 1969 | <u>.1643</u> | <u>-.1342</u> | <u>.0717</u> |
|  |      | M = .2669    | M = .3037     | M = .4220    |

Muussa tehdasteollisuudessa ei voida havaita selviä paremmuuseroja eri apumuuttujien kesken. Tämä johtunee siitä, että muu tehdasteollisuus on varsin heterogeeninen ala, jossa on sekä pääoma- että työvoimaintensiivisiä sektoreita. Koska estimoitaessa koko muu tehdasteollisuus aggregoidaan yhteen, niin ilmeisesti apumuuttujien paremmuuserot eri sektoreilla keskimäärin kumoavat toisensa, jolloin koko aggregaatin tasolla ei saada selvää paremmuuseroa eri apumuuttujien välille.

### 3.4.2. Regressioanalyysi

Seuraavaksi tutkitaan eri apumuuttujien keskinäistä tehokkuutta estimoitavan suureen ja käytetyn apumuuttujan välisen regressiosuoran kulun perusteella osapopulaatioittain.

Liitteissä A1, B1 ja C1 on esitetty mainitut regressiosuo-

rat osapopulaatiossa puu- ja paperiteollisuus. Aineistona on syksyn 1968 kysely, kuten korrelaatioanalyysissakin. Tässä kyselyssä tiedusteltiin yritysten investointien arvoa vuonna 1967 ja ennustetta vuosille 1968 ja 1969, jotka siis olivat es- timoitavia suureita. Kuvioista voidaan selvästi havaita, että käytettäessä apumuuttujana tuotannon brutto- tai jalostusarvoa kulkevat regressiosuorat kaikkien estimoitavien suureiden suh- teen erittäin läheltä origoa. Käytettäessä sitä vastoin apu- muuttujana työntekijöiden lukumäärää kulkevat regressiosuorat selvästi origon ulkopuolelta. Regressiosuorien kulun perusteel- la päädytään samaan tulokseen kuin korrelaatioanalyysinkin poh- jalta. Siis tuotannon brutto- ja jalostusarvo ovat tehokkaam- pia apumuuttujia suhdemenetelmässä kuin työntekijöiden luku- määrä (Yates s. 156). Sitä vastoin tuotannon brutto- ja jalos- tusarvolukujen välille ei regressioanalyysin perusteella voida tehdä paremmuuseroa.

Liitteissä A2, B2 ja C2 on esitetty vastaavat regressio- suorat metalliteollisuudelle. Myös niistä saatavat tulokset ovat yhdenmukaisia korrelaatiokertoimien perusteella tehtyjen johtopäätösten kanssa. Kun apumuuttujana on työntekijöiden lu- kumäärä, kulkevat regressiosuorat selvästi lähempänä origoa, kuin jos apumuuttujana on tuotannon brutto- tai jalostusarvo. Työntekijöiden lukumäärän käyttö apumuuttujana suhdemenetelmäs- sä johtaisi siis tämän perusteella tarkempiin estimaatteihin kuin tuotannon brutto- tai jalostusarvon käyttö.

Muun tehdasteollisuuden regressiosuorat on esitetty liit- teissä A3, B3 ja C3. Niiden perusteella ei voida tehdä selviä johtopäätöksiä eri apumuuttujien keskinäisestä paremmuudesta. Samaan tulokseen päädyttiin myös korrelaatioanalyysin perusteel- la. Voidaan kuitenkin todeta, että kahden estimoitavan suureen suhteen (inv 69, 68) työntekijöiden lukumäärä (liite A3) näyt- täisi soveltuvan varsin hyvin apumuuttujaksi. Useimpien regres- siosuorien kulku poikkeaa kuitenkin niin paljon origosta, että mitään käytettyä apumuuttujaa ei voida pitää erityisen hyvänä. On ilmeistä, että muu tehdasteollisuus olisi tulevissa kyse- lyissä disagregoitava pääoma- ja työvoimavaltaisiin elinkeinoi- hin, jotta muun tehdasteollisuuden estimaattien tarkkuutta voi-



taisiin parantaa sekä eri apumuuttujien käytölle saataisiin selviä kriteerejä. Muussa tehdasteollisuudessa saataisiin regressiomenetelmällä tarkempia estimaatteja, koska mainitut regressiosuorat eivät kulje origon kautta (Sukthatme s. 221).

Regressiosuorien kulkua tutkittiin myös ositteittain jokaisessa osapopulaatiossa. Koska niiden perusteella voitiin tehdä samat johtopäätökset kuin edellä esitettyjen koko otoksesta laskettujen regressiosuorien kulun perusteella, ei tilan säästämissä ole esitetty regressiosuoria ositteittain.

### 3.4.3. Eri apumuuttujilla saadut varianssit

Lopuksi on eri apumuuttujien keskinäistä tehokkuutta tutkittu niillä saatujen estimaattien varianssien perusteella. Työntekijöiden lukumäärällä saatua varianssia on merkitty 100:lla ja muilla apumuuttujilla tuotetut varianssit on saatu jakamalla ko. apumuuttujalla saatu varianssi työntekijöiden lukumäärällä saadulla varianssilla ja kertomalla saatu osamäärä 100:lla. Todelliset varianssit on esitetty taulukossa w seuraavalla sivulla.

Taulukko. Eri apumuuttujilla saatujen varianssien suhteellinen paremmuus toisiinsa nähden

#### Puu- ja paperiteollisuus

| apumuuttuja →   | työntekijöiden lukum. | jalostusarvo | bruttoarvo  |
|-----------------|-----------------------|--------------|-------------|
| investoinnit 67 | 100                   | 80.6         | 89.6        |
| " 68            | 100                   | 80.3         | 96.8        |
| " 69            | 100                   | <u>84.9</u>  | <u>84.9</u> |
|                 |                       | M = 81.9     | M = 90.4    |

#### Metalliteollisuus

| apumuuttuja →   | työntekijöiden lukum. | jalostusarvo | bruttoarvo   |
|-----------------|-----------------------|--------------|--------------|
| investoinnit 67 | 100                   | 116.6        | 116.6        |
| " 68            | 100                   | 117.1        | 117.2        |
| " 69            | 100                   | <u>114.1</u> | <u>114.0</u> |
|                 |                       | M = 116.0    | M = 116.1    |

Muu tehdasteollisuus

| apumuuttuja →   | työntekijöiden lukum. |  | jalostusarvo | bruttoarvo   |
|-----------------|-----------------------|--|--------------|--------------|
| investoinnit 67 | 100                   |  | 100.9        | 201.6        |
| " 68            | 100                   |  | 133.5        | 133.6        |
| " 69            | 100                   |  | <u>102.1</u> | <u>102.1</u> |
|                 |                       |  | M = 112.2    | M = 145.7    |

Taulukko w. Estimointimenetelmä: ratio separate  
 Estimaattien varianssit: aineisto S-68  
 milj. mk

Puu- ja paperiteollisuus

| apumuuttuja →   | työntekijöiden lukum. |  | jalostusarvo | bruttoarvo  |
|-----------------|-----------------------|--|--------------|-------------|
| investoinnit 67 | 10.3                  |  | 8.3          | 9.5         |
| " 68            | 22.3                  |  | 17.9         | 21.6        |
| " 69            | <u>21.8</u>           |  | <u>18.5</u>  | <u>18.5</u> |
|                 | M = 18.1              |  | M = 14.9     | M = 16.5    |

Metalliteollisuus

| apumuuttuja →   | työntekijöiden lukum. |  | jalostusarvo  | bruttoarvo    |
|-----------------|-----------------------|--|---------------|---------------|
| investoinnit 67 | 204.4                 |  | 238.3         | 238.4         |
| " 68            | 267.3                 |  | 312.7         | 311.9         |
| " 69            | <u>966.0</u>          |  | <u>1104.6</u> | <u>1104.6</u> |
|                 | M = 479.2             |  | M = 551.8     | M = 551.9     |

Muu tehdasteollisuus

| apumuuttuja →   | työntekijöiden lukum. |  | jalostusarvo  | bruttoarvo    |
|-----------------|-----------------------|--|---------------|---------------|
| investoinnit 67 | 535.0                 |  | 539.6         | 1078.8        |
| " 68            | 627.0                 |  | 837.0         | 836.9         |
| " 69            | <u>1607.0</u>         |  | <u>1640.8</u> | <u>1640.9</u> |
|                 | M = 923.0             |  | M = 1006.0    | M = 1186.1    |

Investoinnit 68 ja 69 tarkoittavat siis vuosille 1968 ja 1969 tehtyjä investointisuunnitelmia. Taulukosta havaitaan, että jokaisessa osapopulaatiossa ovat investointisuunnitelmien varianssit huomattavasti suuremmat kuin vuonna 1967 toteutuneen investointien varianssi. Tämä johtuu ennen kaikkea siitä, että osa yrityksistä yliarvioi ja osa aliarvioi tulevat investointinsa.

Estimaattien varianssit tukevat täysin korrelaatio- ja regressioanalyysillä saatuja tuloksia. Voidaan todeta, että puu- ja paperiteollisuudessa jalostusarvon käyttö apumuuttujana alentaa estimaattien varianssia keskimäärin noin 18 % ja bruttoarvon käyttö keskimäärin noin 9 % työntekijöiden lukumäärällä saadusta varianssista. Metalliteollisuudessa sekä tuotannon brutto- että jalostusarvon käyttö nostavat varianssia keskimäärin noin 16 % työntekijöiden lukumäärällä saadusta varianssista. Kuten aikaisemmin todettiin, ei muussa tehdasteollisuudessa päästy korrelaatio- ja regressioanalyysin avulla selviin tuloksiin eri apumuuttujien paremmuudesta paitsi siinä suhteessa, että kahdelle estimoitavalle suurelle työntekijöiden lukumäärä näytti regressiosuorien perusteella soveltuvan parhaiten. Taulukosta w havaitaan, että bruttoarvolla saatu varianssi on keskimäärin noin 28 % ja jalostusarvolla saatu varianssi keskimäärin noin 9 % suurempi kuin työntekijöiden lukumäärällä saatu varianssi. Eri estimoitavien suureiden suhteen ovat erot kuitenkin edelleen varsin suuret, joten selvää johtopäätöstä eri apumuuttujien paremmuudesta ei voida tehdä.

Yhteenvetona voidaan todeta, että esitettyjen tulosten perusteella on puu- ja paperiteollisuudelle sekä metalliteollisuudelle löydetty parhaiten soveltuva apumuuttuja, joka puu- ja paperiteollisuudessa on tuotannon brutto- tai jalostusarvo sekä metalliteollisuudessa työntekijöiden lukumäärä. Sitä vastoin muussa tehdasteollisuudessa ei korrelaatio- ja regressioanalyysi antanut selvää kuvaa eri apumuuttujien paremmuudesta. Koska estimaatin varianssi on kuitenkin tärkein kriteeri apumuuttujan tehokkuutta tarkasteltaessa, näyttää työntekijöiden lukumäärä keskimäärin tehokkaimmalta apumuuttujalta muussa tehdasteollisuudessa. On kuitenkin todettava, että eräiden estimoitavien suureiden suhteen tuotannon brutto- ja jalostusarvo ovat yhtä tehokkaita apumuuttujia kuin työntekijöiden lukumääräkin muussa tehdasteollisuudessa.

#### 4. Optimaalinen allokoiminen ja otoksen koon estimoiminen

Allokoiminen tarkoittaa ositteiden koon  $n_h$  määräämistä otoksen kokonaiseen  $n$  ollessa annettu, Ositetussa otannassa optimaalinen allokoiminen suoritetaan erilaisten rajoittavien ehtojen vallitessa. Optimoiminen suoritetaan useimmiten joko kustannusfunktion tai otoksen koon ollessa rajoittavana ehtona. Optimaalinen ratkaisu saadaan minimoimalla estimoitavan suureen varianssi jollakin rajoittavalla ehdolla.

Optimaalinen allokoiminen tullaan seuraavassa suorittamaan erikseen jokaisessa osapopulaatiossa. Rajoittavana ehtona käytetään otoksen kokoa  $n$  kyseisessä osapopulaatiossa. Kustannusfunktiota ei käytetä rajoittavana ehtona, koska kustannukset eivät käytettävissä otantasysteemissä aseta estettä otoksen koon suurentamiselle jossakin ositteessa tai koko osapopulaatiossa niissä rajoissa, jotka tietojenhankintajärjestelmän joustavuus sallii. Myöskään otoksen kokoa estimoitaessa ei kustannusfunktio aseta estettä otoksen koon suurentamiselle. Otoksen kokoa estimoitaessa ylärajan koolle asettaa sitä vastoin tietojenhankintajärjestelmän joustavuus, jonka määräämäksi suhteellisen karkeaksi ylärajaksi saadaan noin 800 - 900 yritystä. Tätä rajaa huomattavasti suuremman yritysmäärän käyttö saattaisi viivästyttää aineiston lopullista analyysia liian paljon SP:n investointienustejien laatimisen kannalta. - Seuraavassa tarkoitetaan otoksen koolla  $n$  otoksen kokoa tietyssä osapopulaatiossa.

##### 4.1. Optimaalinen allokoiminen varianssien perusteella

Optimaalista allokoimintia suoritettaessa täytyy tuntea estimoitavan suureen varianssi kussakin ositteessa, joita siis on kaksi käytetyssä otantasysteemissä. Koska estimoitavan suureen todellista varianssia ei kuitenkaan tunneta etukäteen, on pyrittävä saamaan arviot varianssille kummassakin ositteessa, jotta optimaalinen allokoiminen voitaisiin suorittaa. Parhaat arviot varianssien suuruudelle saataneen laskemalla varianssien arvot aikaisemmista kyselyistä ja tekemällä sen jälkeen oletuksia varianssien muuttumisesta kyselyjen välillä (Cochran s. 81). Optimaalinen allokoiminen suoritetaan seuraavassa syksyn 1968 otokselle. Tästä otoksesta saatavat varianssit arvioidaan keväällä 1968 suoritettuna otoksen aineiston perusteella. Vuoden 1967 kyse-

lyn tietoja ei voitu käyttää hyväksi, koska silloin kysely suoritettiin devalvaatiosta johtuen ainoastaan suppeana puhelin-tiedusteluna. Tutkimusaineistoksi on valittu nimenomaan kahden peräkkäisen kyselyn tiedot, jotta päästäisiin selville siitä, kuinka hyvin edellisen kyselyn tiedot soveltuvat seuraavan kyselyn optimaalisen allokoinnin suorittamiseen. Edellisen kyselyn tietojen hyvyys optimiratkaisua etsittäessä riippuu siitä, kuinka paljon estimoitavien suureiden ositteiden varianssien suhteet muuttuvat peräkkäisten kyselyjen välillä. Tämä muuttuminen riippuu toisaalta yleisen taloudellisen aktiviteetin ja ennen kaikkea investointien kehittymisestä kyselyjen välillä eri ositteissa.

Edellä olevan perusteella vuoden 1968 aineistosta saatuja tuloksia voidaan pitää vain suuntaa antavina tulevien kyselyjen yhteydessä suoritettaville optimoinneille. Tämä johtuu siitä, että jossain muussa kuin vuonna 1968 vallinneessa suhdannetilanteessa saattaa estimaattien varianssien suhde muuttua toisin peräkkäisten kyselyjen välillä kuin vuonna 1968. Tulevaisuudessa otosta uusittaessa onkin siksi syytä verrata useampien peräkkäisten kyselyjen välillä tapahtuneita muutoksia ja vasta sitten tehdä oletukset vallitsevassa suhdannetilanteessa tapahtuneista varianssien suhteiden muutoksista. Tältä pohjalta päästään ilmeisesti tarkempiin optimaalisiin ratkaisuihin.

Optimaalista allokointia suoritettaessa syksyn 1968 otokselle kevään 1968 kyselystä saatujen varianssien suhteiden perusteella tehtiin oletus, että estimoitavien suureiden ositteiden varianssien suhde ei muutu kyselyjen välillä ja ratkaistiin optimaalinen allokointi tällä perusteella. Oletus tehtiin siksi, että tutkimuksen loppuvaiheessa saatiin syksyn 1968 kysely suoritetuksi loppuun ja näin ollen voitiin laskea syksyn 1968 otoksesta "todelliset" varianssit ja määrätä niiden perusteella "todellinen" optimaalinen allokointi. Täten saatiin syksyn 1968 otokselle kaksi erilaista optimaalista allokointiratkaisua, joiden erojen perusteella voitiin tehdä johtopäätöksiä siitä, kuinka paljon estimaattien varianssien suhteet saattavat muuttua kyselyjen välillä ja kuinka paljon tämä muuttuminen vaikuttaa optimaaliseen allokointiin.

#### 4.1.1. Optimaalisen allokoinnin kaava

Optimaalisen allokoinnin suorittamista varten laskettiin estimaattien varianssit seuraavalla kaavalla:

$$V(\hat{Y}_{RS}) = \sum_h \frac{N_h(N_h - n_h)}{n_h} (S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h r_h S_{yh} S_{xh}) \quad (\text{Cochran s. 129})$$

Varianssin estimaatti saatiin korvaamalla populaation arvot  $S_{yh}^2$ ,  $S_{xh}^2$  ja  $R_h$  otoksesta saaduilla estimaateilla (Cochran s.129 ja s. 132). Kaavan symboleista osa on jo aikaisemmin esiintynyt sivuilla 10 ja 11. Muiden merkitys on seuraava:

$S_{yh}$  =  $y_i$ :n hajonta populaation ositteessa h

est  $S_{yh} = s_{yh}$

$S_{xh}$  = apumuuttujan  $x_i$  hajonta populaation ositteessa h

est  $S_{xh} = s_{xh}$

$R_h = \frac{Y_h}{X_h}$ , est  $R_h = \hat{R}_h$

$r_h$  =  $x_i$ :n ja  $y_i$ :n korrelaatiokerroin ositteessa h

est  $V(\hat{Y}_{RS}) = v(\hat{Y}_{RS})$

Kaava siis olettaa, että osittotusta umpimähkäisestä otoksesta lasketaan suhdemenetelmällä (separate) estimaatit populaation suureille. Perustelut kaavan käytölle on esitetty jo sivulla 1.

Optimaalisen allokoinnin suorittamista varten otetaan käyttöön seuraava merkintätapa:

$$(1) \quad V'_h = s_{yh}^2 + R_h^2 s_{xh}^2 - 2 R_h r_h s_{yh} s_{xh}$$

Optimaalinen allokointi suoritetaan siten, että estimaatin varianssi minimoidaan, kun rajoittavana ehtona on otoksen koko  $n=n_1+n_2$ . Tehtävänä on siis minimoida lauseke

$$(2) \quad v(\hat{Y}_{RS}) = \sum_{h=1}^2 \frac{N_h(N_h - n_h)}{n_h} V'_h \quad (\text{Cochran s. 135})$$

rajoittavalla ehdolla  $n_1 + n_2 = n$ . Tehtävä suoritetaan siten, että valitaan  $n_h$  ja Lagrangen kerroin  $k$  siten, että summa (3) minimoituu.

$$(3) \quad v(\hat{Y}_{RS}) + k(n_1 + n_2 - n)$$

Derivoidaan lauseke ensin  $n_h$ :n suhteen ja saadaan

$$(4) \quad -\frac{N_h^2}{n_h^2} V_h' + k = 0 \quad (h = 1, 2)$$

Tästä saadaan

$$(5) \quad \sqrt{k} n_h = N_h \sqrt{V_h'}$$

Summaamalla yli  $h$ :n saadaan

$$(6) \quad \sqrt{k} n = \sum_h N_h \sqrt{V_h'}$$

Jakamalla (5) (6):lla saadaan edelleen (7) ja (8).

$$(7) \quad \frac{n_h}{n} = \frac{N_h \sqrt{V_h'}}{\sum_h N_h \sqrt{V_h'}} = B$$

$$(8) \quad n_h = n B$$

Kaavasta (7) saadaan

$$(9) \quad \frac{n_h}{N_h} = n \frac{\sqrt{V_h'}}{\sum_h N_h \sqrt{V_h'}}$$

Tuloksesta (9) voidaan päätellä, että otantasuhteen  $\frac{n_h}{N_h}$

tulisi olla suhteessa  $\sqrt{V_h'}$ :uun, joka on  $y_i$ :n residuaalivarianssi  $y_i$ :n ja  $x_i$ :n välisessä regressiossa ositteessa  $h$  (Cochran s.132, Sukthatme s. 170).

Optimaalinen allokointi suoritetaan myös toisen suhdemenetelmän suhteen (ratio combined). Näillä kahdella suhdemenetelmällä laskettujen estimaattien variansseissa on vain se ero, että

combined-menetelmässä  $\hat{R}_h$ :t korvataan kummassakin ositteessa koko otoksesta lasketulla  $\hat{R}$ :llä. Näiden kahden suhdemenetelmän eroa selostettiin sivuilla 11 ja 12.

#### 4.1.2. Optimaalisen allokoinnin tulokset

Optimaalisen allokoinnin määrittämisessä tarvittavat suuret

$N_h \sqrt{V'_h}$  ja  $\sum_h N_h \sqrt{V'_h}$  sekä  $V'_h$ :n laskemisessa käytetyt muuttujien arvot esitetään liitteissä 2-4 osapopulaatioittain ja ositteittain. Seuraavassa esitetään ensin kevään 1968 aineistosta saatujen varianssien perusteella suoritettu optimaalinen allokointi osapopulaatioittain. Käytetään seuraavia merkintöjä:  
 $h = 1, 2$

$n_h$  = syksyn 1968 kyselyssä käytetty ositteen  $h$  koko;  $n_h$ :n arvot saatiin siten, että pienistä yrityksistä (osite 1) valittiin joka viides ja keskisuurista yrityksistä (osite 2) joka toinen.

$n_{hopts}$  = estimointimenetelmää ratio - separate vastaava optimaalinen ositteen  $h$  koko

$n_{hopte}$  = estimointimenetelmää ratio-combined vastaava optimaalinen ositteen  $h$  koko

Koska kyselyssä tiedusteltiin yrityksiltä niiden investointien arvoa vuonna 1967 ja suunniteltuja investointeja vuosiksi 1968 ja 1969, optimoiminen on suoritettu erikseen kaikkien kolmen estimoitavan suureen suhteen. Täten kullekin ositteelle saatiin kolme optimiarvoa, joiden keskiarvo katsottiin ositteen lopulliseksi optimaaliseksi arvoksi, koska optimiarvot eri estimoitavien suureiden suhteen eivät poikenneet paljoa toisistaan. Tämä johtunee siitä, että estimoitavat suuret ovat suuresti riippuvia toisistaan yksityisen yrityksen kohdalla (Cochran s. 84 - 85, Hansen - Hurwitz - Madow s. 224-225). Keskiarvoa on merkitty  $M$ :llä. Seuraavissa taulukoissa esitetään saadut tulokset:



Taulukko.

Puu- ja paperiteollisuus

| osite | $N_h$  | $n_h$     |
|-------|--------|-----------|
| 1     | 207    | 41        |
| 2     | 39     | <u>19</u> |
|       | N = 60 |           |

| estimoitava suure | $n_1$ | $n_{1opts}$ | $n_{1optc}$ | $n_2$ | $n_{2opts}$ | $n_{2optc}$ |
|-------------------|-------|-------------|-------------|-------|-------------|-------------|
| investoinnit 1967 | 41    | 22.4        | 23.0        | 19    | 37.6        | 37.0        |
| " 1968            | 41    | 14.7        | 14.8        | 19    | 43.1        | 45.8        |
| " 1969            | 41    | 22.6        | 23.6        | 19    | 37.4        | 36.4        |
|                   | M= 41 | M=20.3      | M=21.0      | M=19  | M=39.7      | M=39.0      |

Metalliteollisuus

| osite | $N_h$ | $n_h$     |
|-------|-------|-----------|
| 1     | 461   | 91        |
| 2     | 74    | <u>36</u> |
|       | n=127 |           |

| estimoitava suure | $n_1$ | $n_{1opts}$ | $n_{1optc}$ | $n_2$ | $n_{2opts}$ | $n_{2optc}$ |
|-------------------|-------|-------------|-------------|-------|-------------|-------------|
| investoinnit 1967 | 91    | 79.0        | 79.6        | 36    | 48.0        | 47.4        |
| " 1968            | 91    | 82.1        | 82.5        | 36    | 44.9        | 44.5        |
| " 1969            | 91    | 52.4        | 54.3        | 36    | 74.9        | 72.7        |
|                   | M=91  | M=71.2      | M=72.0      | M=36  | M=55.8      | M=55.0      |

Muu tehdasteollisuus

| osite | $N_h$ | $n_h$      |
|-------|-------|------------|
| 1     | 874   | 172        |
| 2     | 208   | <u>103</u> |
|       | n=275 |            |

| estimoitava suure | $n_1$ | $n_{1\text{opts}}$ | $n_{1\text{optc}}$ | $n_2$ | $n_{2\text{opts}}$ | $n_{2\text{optc}}$ |
|-------------------|-------|--------------------|--------------------|-------|--------------------|--------------------|
| investoinnit 1967 | 172   | 81.4               | 84.2               | 103   | 193.8              | 190.8              |
| " 1968            | 172   | 111.2              | 111.2              | 103   | 163.8              | 162.1              |
| " 1969            | 172   | 103.9              | 113.7              | 103   | 171.1              | 161.3              |
|                   | M=172 | M=98.8             | M=103.0            | M=103 | M=176.1            | M=172.0            |

Saaduista tuloksista voidaan tehdä selvät johtopäätökset. Kaikissa osapopulaatioissa molemmilla estimointimenetelmillä saadut tulokset osoittavat, että käytetty allokointi poikkeaa huomattavasti optimaalisesta allokoinnista. Jokaisessa osapopulaatiossa optimaalinen allokointi edellyttää keskisuurten yritysten (osite 2) lukumäärän huomattavaa lisäämistä ja vastaavasti pienien yritysten lukumäärän supistamista. Puu- ja paperiteollisuudessa keskisuurten yritysten lukumäärää olisi lisättävä noin 100 %:lla, metalliteollisuudessa noin 55 %:lla ja muussa tehdasteollisuudessa noin 70 %:lla. Puu- ja paperiteollisuudessa ja muussa tehdasteollisuudessa optimaalinen allokointi edellyttäisi suunnilleen kaikkien keskisuurten populaatioon kuuluvien yritysten ottamista otokseen eli otantasuhteen nostamista 100 %:iin. Näissä osapopulaatioissa otos tulisi tällöin käsittämään vain yhden ositteen eli pienet yritykset, joista varsin pieni otos riittää antamaan luotettavat estimaatit. Metalliteollisuudessa optimaalinen allokointi edellyttäisi sitä vastoin huomattavasti vähäisempää ositteiden kokojen muuttamista nykyisestä. Arvojen  $n_{\text{hopts}}$  ja  $n_{\text{hoptc}}$  väliset erot olivat kaikissa osapopulaatioissa varsin pieniä.

Seuraavassa taulukossa esitetään "todellinen" optimaalinen allokointi osapopulaatioittain, jolloin allokointi perustuu syksyn 1968 kyselystä saatuihin "todellisiin" variansseihin. Käytetyt merkinnät ovat samat kuin edellä. Koska  $n_{\text{hopts}}$  ja  $n_{\text{hoptc}}$  arvot poikkesivat jälleen vain vähän toisistaan, on taulukossa esitetty vain  $n_{\text{hopts}}$  arvot.

Taulukko.

Puu- ja paperiteollisuus

| estimoitava suure | $n_1$ | $n_{1\text{opts}}$ | $n_2$ | $n_{2\text{opts}}$ |
|-------------------|-------|--------------------|-------|--------------------|
| investoinnit 1967 | 41    | 26.8               | 19    | 33.2               |
| " 1968            | 41    | 21.7               | 19    | 38.3               |
| " 1969            | 41    | 8.6                | 19    | 50.4               |
|                   | M=41  | M=18.2             | M=19  | M=42.8             |

Metalliteollisuus

| estimoitava suure | $n_1$ | $n_{1\text{opts}}$ | $n_2$ | $n_{2\text{opts}}$ |
|-------------------|-------|--------------------|-------|--------------------|
| investoinnit 1967 | 91    | 95.8               | 36    | 31.8               |
| " 1968            | 91    | 103.6              | 36    | 23.4               |
| " 1969            | 91    | 107.0              | 36    | 20.0               |
|                   | M=91  | M=101.9            | M=36  | M=25.1             |

Muu tehdasteollisuus

| estimoitava suure | $n_1$ | $n_{1\text{opts}}$ | $n_2$ | $n_{2\text{opts}}$ |
|-------------------|-------|--------------------|-------|--------------------|
| investoinnit 1967 | 172   | 140.4              | 103   | 134.6              |
| " 1968            | 172   | 111.1              | 103   | 163.1              |
| " 1969            | 172   | 150.6              | 103   | 124.4              |
|                   | M=172 | M=134.0            | M=103 | M=141.0            |

Verrattaessa "todellisen" optimaalisen allokoinnin tuloksia aikaisemman approksimatiivisen optimaalisen allokoinnin tuloksiin havaitaan, että tulokset käyvät puu- ja paperiteollisuuden ja muun tehdasteollisuuden osalta erittäin hyvin yksiin. Puu- ja paperiteollisuudessa myös "todellinen" optimaalinen allokoiminen edellyttäisi kaikkien keskisuurten yritysten ottamista otokseen. Muussa tehdasteollisuudessa keskisuurten yritysten lukumäärää olisi lisättävä noin 40 % "todellisessa" optimaalisessa allokoinnissa, kun vastaava tulos approksimatiivisessa allokoinnissa oli noin 70 %. Voidaan tehdä se johtopäätös, että näissä osapopulaatioissa ei kevään ja syksyn kyselyjen välillä ositteiden varianssien suhde ole paljoa muuttunut, joten edellisen kyselyn aineistosta saatuja variansseja voitiin sellaisenaan käyttää

tehokkaasti hyväksi optimaalista allokointia määrättäessä syksyn kyselylle. Sitä vastoin metalliteollisuudessa tapahtui kyselyjen välillä ositteiden varianssien suhteessa selvä muutos, koska "todellisen" optimaalisen allokoinnin perusteella keski-suurten yrittysten lukumäärää olisi hiukan pienennettävä, kun taas approksimatiivisen optimaalisen allokoinnin mukaan niiden lukumäärää olisi jonkin verran lisättävä.

Metalliteollisuudessa saadut tulokset viittaavat siihen, että edellisestä kyselystä saatuja variansseja ei tule sellaisenaan käyttää seuraavan kyselyn optimaalista allokointia suoritettaessa, vaan on pyrittävä tekemään oletuksia varianssien suhteen muuttumisesta kyselyjen välillä esim. analysoimalla kyseisen sektorin investointien kehittymistä kyselyjen välillä.

#### 4.1.3. Minimivarianssit

Estimaattien varianssien minimiarvot saadaan, kun optimaaliset  $n_h$  arvot sijoitetaan yleiseen varianssin kaavaan. Yleinen varianssin kaava oli

$$v(\hat{Y}_{Rs}) = \sum_h \frac{N_h^2}{n_h} V_h' - \sum_h N_h V_h'$$

Kun sijoitetaan tähän kaavaan  $n_{hopt} = n \frac{N_h \sqrt{V_h'}}{\sum_h N_h \sqrt{V_h'}}$

saadaan

$$v_{\min}(\hat{Y}_{Rs}) = \frac{1}{n} \left[ \sum_h N_h \sqrt{V_h'} \right]^2 - \sum_h N_h V_h'$$

Optimaaliseen allokointiin perustuvat minimivarianssien arvot ja saadut todelliset varianssit esitetään seuraavassa taulukossa osapopulaatioittain:

Taulukko.

Puu- ja paperiteollisuus

| estimoitava suure | milj. mk                |                          |
|-------------------|-------------------------|--------------------------|
|                   | saatu $v(\hat{Y}_{Rs})$ | $v_{\min}(\hat{Y}_{Rs})$ |
| investoinnit 1967 | 8.9                     | 4.9                      |
| " 1968            | 42.6                    | 2.4                      |
| " 1969            | <u>7.5</u>              | <u>3.4</u>               |
|                   | M= 19.7                 | M= 3.6                   |

Metalliteollisuus

| estimoitava suure | saatu v      | $v_{\min}$  |
|-------------------|--------------|-------------|
| investoinnit 1967 | 132.0        | 123.0       |
| " 1968            | 171.0        | 169.0       |
| " 1969            | <u>155.0</u> | <u>70.0</u> |
|                   | M=152.7      | M=121.0     |

Muu tehdasteollisuus

| estimoitava suure | saatu v      | $v_{\min}$  |
|-------------------|--------------|-------------|
| investoinnit 1967 | 422.0        | 184.0       |
| " 1968            | 268.0        | 189.0       |
| " 1969            | <u>147.0</u> | <u>91.0</u> |
|                   | M=279.0      | M=154.7     |

Taulukon arvoista havaitaan, että optimaalinen allokointi alentaisi varianssia puu- ja paperiteollisuudessa keskimäärin 83 %, metalliteollisuudessa 22 % sekä muussa tehdasteollisuudessa keskimäärin 45 %.

Seuraavassa taulukossa on käytetyllä allokoinnilla saatua varianssia merkitty 100:lla, jolloin optimaalisella allokoinnilla saatu varianssi on  $\left[ v_{\min} / \text{saatu } v \right] \cdot 100$ .

Taulukko.

Puu- ja paperiteollisuus

| estimoitava suure | käytetty allokointi | opt. allokointi |
|-------------------|---------------------|-----------------|
| investoinnit 1967 | 100                 | 54.8            |
| " 1968            | 100                 | 5.7             |
| " 1969            | 100                 | <u>45.6</u>     |
|                   |                     | M=35.0          |

Metalliteollisuus

| estimoitava suure | käytetty allokointi | opt. allokointi |
|-------------------|---------------------|-----------------|
| investoinnit 1967 | 100                 | 93.2            |
| " 1968            | 100                 | 99.0            |
| " 1969            | 100                 | <u>45.2</u>     |
|                   |                     | M=79.0          |

Muu tehdasteollisuus

| estimoitava suure | käytetty allokointi | opt. allokointi |
|-------------------|---------------------|-----------------|
| investoinnit 1967 | 100                 | 43.6            |
| " 1968            | 100                 | 70.5            |
| " 1969            | 100                 | <u>61.9</u>     |
|                   |                     | M=59.0          |

4.2. Otoksen koon estimoiminen

Otoksen koon estimoiminen käytettäessä optimaalista allokointiä voidaan suorittaa minimivarianssin kaavan perusteella kiinnittämällä varianssi. Kun kaavasta ratkaistaan n saadaan

$$n = \frac{\left( \sum_h N_h \sqrt{V_h'} \right)^2}{v_{\min} + \sum_h N_h V_h'}$$

Otoksen koko voidaan siis estimoida, kun kiinnitetään varianssin arvo. Edellä olevan kaavan perusteella saadaan tällöin ratkaistuksi se  $n:n$  arvo, jolla kiinnitetty varianssi saavutetaan, olettaen, että on käytetty optimaalista allokointia.

Seuraavassa varianssi on kiinnitetty jonkin verran minimiarvoa (kiinteälle  $n$ :lle) pienemmäksi, mikä merkitsee sitä, että otoksen kokoa on nostettava nykyisestä, jotta kiinnitetty varianssi saavutettaisiin. Puu- ja paperiteollisuudessa sekä metalliteollisuudessa, joissa otokseen kuuluvien yritysten investointien arvo on huomattavasti pienempi osa näiden sektoreiden investointien kokonaisarvosta kuin muussa tehdasteollisuudessa, on kiinnitetty varianssi suurempi suhteessa investointien arvoon kuin muussa tehdasteollisuudessa. Puu- ja paperiteollisuudessa sekä metalliteollisuudessa varianssi kiinnitettiin 10 %:iin investointien kokonaisarvosta ja muussa tehdasteollisuudessa 5 %:iin. Tämä merkitsee sitä, että ns. coefficient of variation (cv) on puu- ja paperiteollisuudessa sekä metalliteollisuudessa 0.10 ja muussa tehdasteollisuudessa 0.05 (Cochran s. 89).

Otoksen kooksi saatiin osapopulaatioittain seuraavat arvot:

Taulukko.

Puu- ja paperiteollisuus

| estimoitava suure | saatu v | $v_{\min}$ | kiinnitetty v | estimoi-<br>tu n | nykyi-<br>nen n |
|-------------------|---------|------------|---------------|------------------|-----------------|
| investoinnit 1967 | 9.0     | 4.9        | 4.0           | 61               | 60              |
| " 1968            | 42.6    | 2.4        | 9.0           | 52               | 60              |
| " 1969            | 7.5     | 3.4        | 3.0           | <u>64</u>        | <u>60</u>       |
|                   |         |            |               | M=59             | M=60            |

Metalliteollisuus

| estimoitava suure | saatu v | $v_{\min}$ | kiinnitetty v | estimoi-<br>tu n | nykyi-<br>nen n |
|-------------------|---------|------------|---------------|------------------|-----------------|
| investoinnit 1967 | 132.2   | 123.0      | 80.0          | 165              | 127             |
| " 1968            | 170.6   | 170.0      | 130.0         | 210              | 127             |
| " 1969            | 155.2   | 70.0       | 80.0          | <u>122</u>       | <u>127</u>      |
|                   |         |            |               | M=169            | M=127           |

Muu tehdasteollisuus

| estimoitava suure | saatu v | $v_{\min}$ | kiinnitetty v | estimoitu n | nykyinen n |
|-------------------|---------|------------|---------------|-------------|------------|
| investoinnit 1967 | 422.0   | 184.0      | 165.0         | 285         | 275        |
| " 1968            | 268.0   | 189.0      | 110.0         | 344         | 275        |
| " 1969            | 146.0   | 91.0       | 70.0          | <u>306</u>  | <u>275</u> |
|                   |         |            |               | M=310       | M=275      |

Tuloksista voidaan todeta, että puu- ja paperiteollisuudessa käytössä oleva otoksen koko riittää kiinnitetyn varianssin saavuttamiseksi, koska estimoitavalle suureelle, investoinnit 1968, varianssi kiinnitettiin  $v_{\min}$ :ä suuremmaksi. Metalliteollisuudessa otoksen kokoa olisi nostettava noin 40 yrityksellä ja muussa tehdasteollisuudessa noin 35 yrityksellä kiinnitetyn varianssin saavuttamiseksi. Jotta kiinnitetty varianssi saavutetaan, on siis koko ajan edellytetty, että estimoitu otoksen koko allokoidaan optimaalisesti ositteisiin 1 ja 2. Lisäksi on otettava huomioon, että eri estimoitavien suureiden suhteen estimoidut otoksen koot poikkeavat metalliteollisuudessa ja muussa tehdasteollisuudessa jonkin verran toisistaan. Kullekin osapopulaatiolle saatiin estimoitu n:n arvo keskiarvona eri estimoitaville suureille saaduista n:stä, joten saatu keskiarvo n ei ehkä anna aivan tarkalleen kiinnitettyä varianssia kullekin estimoitavalle suureelle. Aritmeettisella keskiarvolla saatu n osoittaa kuitenkin selvästi, että vaikka käytetäänkin optimaalista allokointia, niin on otoksen kokoa nostettava metalli- ja muussa tehdasteollisuudessa, jotta kiinnitetty varianssi saavutettaisiin likimäärin. Otoksen kokonaiskoon kasvu on kuitenkin sellaista suuruutta, ettei se aiheuta hankaluuksia tietojenhankintajärjestelmän joustavuudelle (vrt. s. 5 ja s. 24).

4.3. Konvergoiminen kohti optimaalista allokointia

Luvussa 4.2 estimoitiin otoksen koko käyttämällä optimaalista allokointia ja kiinnittämällä estimaatin varianssi halutuksi. Seuraavassa tutkitaan, kuinka poikkeaminen optimaalises-



ta allokoinnista vaikuttaa otoksen kokoon, kun varianssi oletetaan kiinnitetyksi samaksi kuin luvussa 4.2. Tutkimuksen kohteeksi on valittu osapopulaatiot puu- ja paperiteollisuus sekä muu tehdasteollisuus, joissa toistaiseksi käytetty allokointi poikkesi huomattavasti optimaalisesta allokoinnista.

Varianssin estimaatin kaava oli

$$v(\hat{Y}_{RS}) = \sum_h \frac{N_h^2}{n_h} V'_h - \sum_h N_h V'_h$$

Sijoitetaan kaavan  $n_h = w_h n$ , jossa  $w_h$  on ositteen  $h$  otantasuhte. Tällöin saadaan

$$v(\hat{Y}_{RS}) = \frac{1}{n} \sum_h \frac{N_h^2}{w_h} V'_h - \sum_h N_h V'_h$$

Ratkaisemalla kaavasta  $n$  saadaan

$$n = \frac{\sum_h \frac{N_h^2}{w_h} V'_h}{v(\hat{Y}_{RS}) + \sum_h N_h V'_h}$$

Sijoittamalla kaavaan  $v$ :n paikalle haluttu varianssi ja antamalla  $w_h$ :lle eri arvoja siten, että  $w_h \longrightarrow w_{hopt}$ , saadaan kuva siitä, kuinka  $n$ :n arvo muuttuu optimaalista allokoointia lähestyttäessä. Eri  $n$ :n arvojen määrittämisessä tarvittavat suureet  $V'_h$ ,  $N_h$  ja  $v_{hal}$  ovat esiintyneet jo aikaisempien lukujen yhteydessä, joten seuraavassa esitetään vain saadut tulokset. Koska estimoitavia suureita oli kolme, nimittäin investoinnit vuosina 1967, 1968, 1969, saatiin kullekin  $w_h$ :lle kolme  $n$ :n arvoa, joista kutakin  $w_h$ :ta vastaava lopullinen  $n$ :n arvo laskettiin keskiarvona. Seuraavissa taulukoissa esitetään saadut tulokset, jollain on voimassa yhtälö  $w_2 = 1 - w_1$ .

Taulukko. Puu- ja paperiteollisuus

käytetty allokointi  $w_2 = 0.317$   
 optimaalinen "  $w_2 = 0.702$

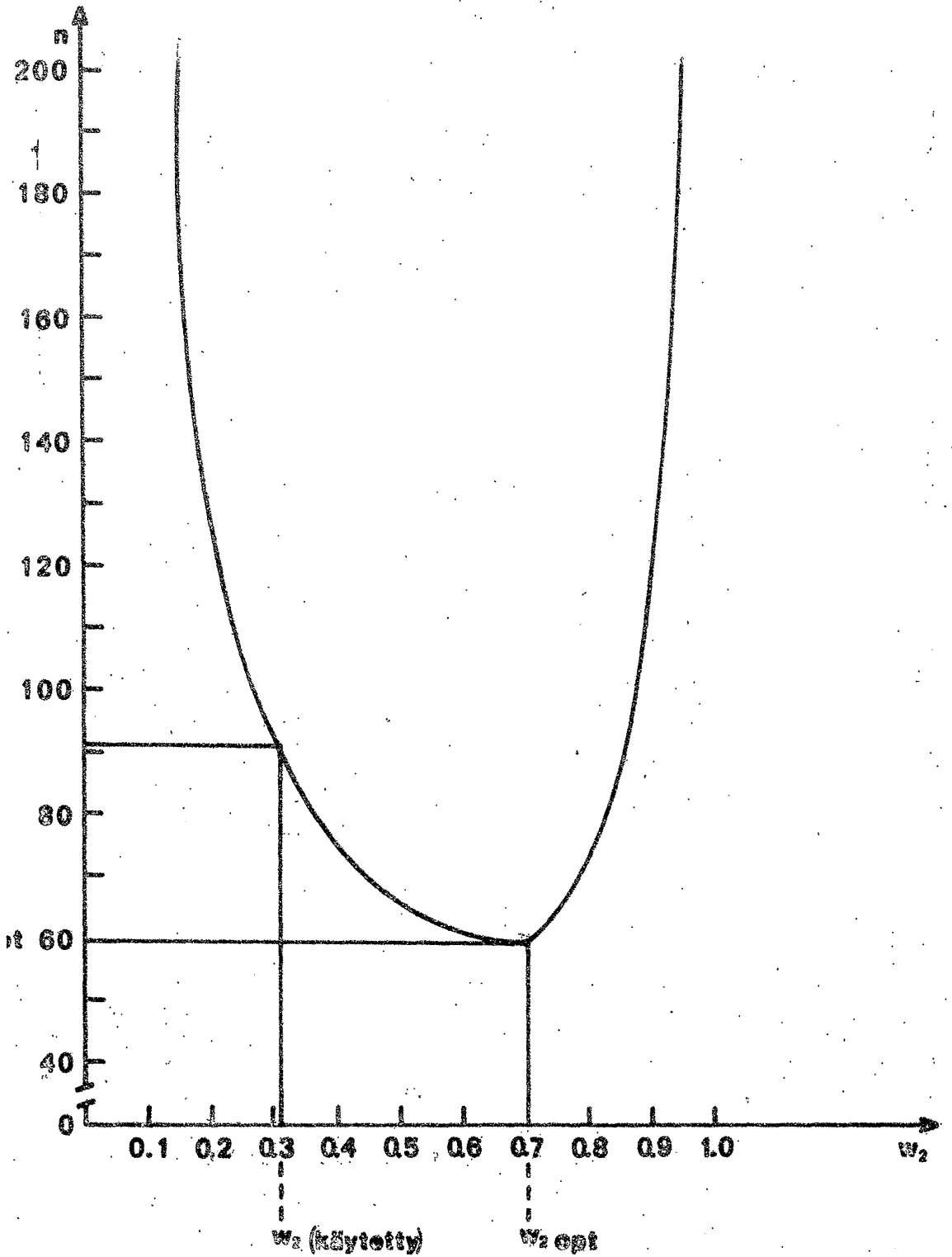
| $w_2$       | n         | $w_2$ | n   |
|-------------|-----------|-------|-----|
| .10         | 251       | .73   | 63  |
| .20         | 132       | .75   | 67  |
| .25         | 116       | .80   | 74  |
| <u>.317</u> | <u>92</u> | .85   | 87  |
| .35         | 82        | .90   | 114 |
| .40         | 75        | .95   | 196 |
| .45         | 69        | .99   | 900 |
| .50         | 66        |       |     |
| .55         | 63        |       |     |
| .60         | 62        |       |     |
| .65         | 60        |       |     |
| <u>.702</u> | <u>59</u> |       |     |

Taulukosta ja kuviosta 3 havaitaan, että otoksen koon muutos  $w_{2opt}$ :n suhteen ei ole symmetrinen. Jos  $w_2 < w_{2opt}$ , niin suuretkaan poikkeamat  $w_{2opt}$ :sta eivät nosta kuin muutamalla yrityksellä  $n:n$  arvoa, jotta haluttu varianssi saavutettaisiin. Käytetty  $w_2 = 0.317$  poikkeaa kuitenkin siksi paljon  $w_{2opt}$ :sta, että otoksen kokoa on lisättävä 32 yrityksellä eli yli 50 %, jotta haluttu varianssi saavutettaisiin. Kun sitävastoin  $w_2 > w_{2opt}$ , on  $n$  varsin herkkä poikkeamisille  $w_{2opt}$ :sta. Jos esim.  $w_2 = w_{2opt} + 0.2$ , on otoksen kokoa nostettava 90 %, jotta haluttu varianssi saavutettaisiin. Jos  $w_2 = w_2 - 0.2$ , niin otoksen kokoa on nostettava vain 10 %, jotta haluttu varianssi saavutettaisiin.

Kuvio 3

Puu- ja paperiteollisuus: funktion  $n = g_1(w_2)$

kuvaaja,  $w_2 = \frac{n_2}{n}$



Taulukko. Muu tehdasteollisuus

käytetty allokointi  $w_2 = 0.364$   
 optimaalinen "  $w_2 = 0.523$

| <u><math>w_2</math></u> | <u>n</u> |
|-------------------------|----------|
| .30                     | 680      |
| .35                     | 520      |
| .364                    | 460      |
| .40                     | 380      |
| .45                     | 320      |
| .475                    | 315      |
| .50                     | 312      |
| .523                    | 310      |
| .55                     | 315      |
| .575                    | 325      |
| .60                     | 340      |
| .65                     | 372      |
| .70                     | 480      |
| .75                     | 677      |

Kuviosta 4 ja taulukosta havaitaan, että otoksen koon muuttuminen  $w_{2opt}$ :sta tapahtuvien poikkeamien suhteen on varsin symmetrinen. Taulukosta havaitaan, että otoksen koko ei muutu paljoa, kun poikkeama  $w_{2opt}$ :sta on noin  $\pm 0.075$ . Kun poikkeama kasvaa tätä suuremmaksi, kasvaa myös halutun varianssin saavuttamiseksi tarvittava otoksen koko varsin nopeasti. Jotta nykyisin käytetyllä allokoinnilla saavutettaisiin haluttu varianssi, olisi otoksen kokoa lisättävä noin 175 yrityksellä eli noin 60 %.

Saadun empiirisen funktion  $n = g(w_2)$  kuvaaja on hyvin samankaltainen hyperbolisen cosinifunktion kuvaajan (ketjukäyrä) kanssa. Hyperbolinen cosinifunktio on  $y = \cosh x =$

$$\frac{e^x + e^{-x}}{2} .$$

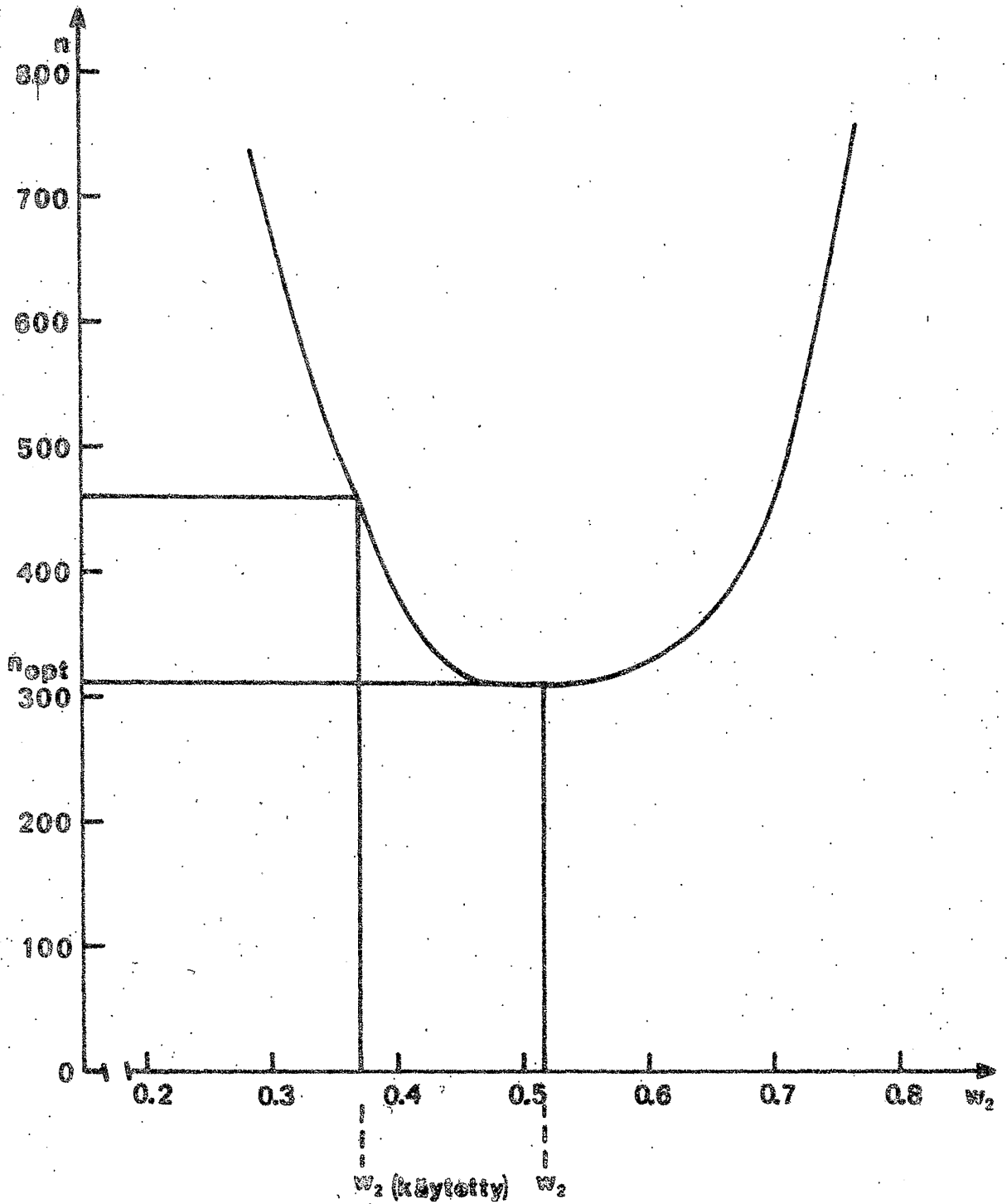
Matemaattinen funktionaalinen vastine saadulle em-

piiriselle käyrälle voidaan johtaa  $\cosh x$ :ää muuntamalla. Funktion minimi saadaan siirretyksi pisteeseen  $w_{2opt}$  muunnoksella

Kuvio 4

Käy tehdaesteollisuus: funktion  $n = G_2(w_2)$

Kuvassa,  $w_2 = \frac{D_2}{n}$



$$y = \cosh ( x - w_{2opt} ) = \frac{e^{x-w_{2opt}} + e^{-x+w_{2opt}}}{2}$$

Tämän funktion arvo (minimi) pisteessä  $w_{2opt}$  saadaan yhtäsuureksi empiirisen havainnon kanssa lisäämällä korjaustermi  $f(w_2)$ , jolloin  $f(w_{2opt}) = n_{opt} - 1$ , koska  $\cosh(w_{2opt} - w_{2opt}) = 1$ . Saadaan siis uusi funktio

$$\left\{ \begin{array}{l} z = \frac{e^{x-w_{2opt}} + e^{-x+w_{2opt}}}{2} + f(x) = h(x) \\ f(w_{2opt}) = n_{opt} - 1 = 310 - 1 = 309 \end{array} \right.$$

Jotta funktion  $h(x)$  kuvaaja vastaisi empiirisillä havainnoilla saatua käyrää, olisi korjausfunktion  $f(x)$  muoto spesifioitava, mitä ei kuitenkaan ole katsottu vaivan arvoiseksi. On mahdollista, että  $f(x)$  on kvadraattista muotoa.

Yhteenvetona voidaan todeta, että "kohtuulliset" poikkeamat optimaalisesta allokoinnista eivät useimmiten vaikuta paljoa otoksen kokoon, jolla kiinnitetty varianssi saavutetaan. Välillisesti voidaan tällöin myös päätellä, että estimaattien varianssi ei poikkea paljoa minimivarianssista, kun käytetty allokointi poikkeaa "kohtalaisesti" optimaalisesta allokoinnista.

#### 4.4. Optimaalinen allokointi $\bar{X}_h$ :n perusteella

Suoritettaessa optimaalista allokoointia ei tarvitse tuntea varianssien absoluuttisia arvoja eri ositteissa (Deming s. 230). Sitävastoin on olennaista tietää ositteiden varianssien suhde. Usein riittää lisäksi melko karkea arvio varianssien suhteellisesta koosta varsinkin, jos varianssit eri ositteissa poikkeavat huomattavasti toisistaan, kuten tämän tutkimuksen aineistossa (Hansen - Hurwitz - Madow s. 213). Aikaisemmin on luvussa 4.2 osoitettu, että optimaaliset  $n_h$ :n arvot johtavat huomattavaan varianssin laskuun verrattuna käytettyyn allokointiin. Varianssien suureneminen ei kuitenkaan ole ko-

vin huomattava poikkeamisen optimaalisesta allokoinnista ollessa "kohtuullinen". Tätä johtopäätöstä tukevat luvun 4.2 tulokset. Esitetyn käyrän minimialue on leveä eikä herkkä kohtuulliselle poikkeamiselle optimaalisuudesta. Näin ollen jos löydetään riittävän hyvät approksimaatiot varianssien suhteille tai jos löydetään muuttujat, jotka ovat suhteessa (proportional) variansseihin, niin näiden muuttujien avulla voidaan päästä lähes yhtä lähelle optimaalista allokoointia, kuin jos tunnettaiisiin "todelliset" varianssit.

Kun otoksen yksilöiden (sample unit) koko on suhteellisen stabiili ajan suhteen, mutta toisaalta koko vaihtelee suuresti yksiköittäin, voidaan saada hyvä optimaalisen allokoinnin approksimaatio olettamalla, että ositteen hajonta on suhteessa yksiköiden keskimääräiseen kokoon tässä ositteessa (Hansen - Hurwitz - Madow s. 215, merk. H - H - M s. 215).

Oheisessa tutkimuksessa yrityksen työntekijöiden lukumäärä täyttää varsin hyvin yksiköiden koon stabilisuudelle ajassa ja vaihtelevuudelle yksiköittäin asetetut vaatimukset. Tältä pohjalta etsitäänkin seuraavassa optimiratkaisut ja verrataan siten saatuja tuloksia aikaisemmin luvussa 4.2 saatuihin tuloksiin.

Käytetään merkintöjä

$X_h$  = työntekijöiden lukumäärä osapopulaation ositteessa  $h$

$X$  = koko osapopulaation työntekijämäärä

$\bar{X}_h = X_h/N_h$

Tehdään edellä olevan perusteella oletus, että  $\bar{X}_h$  on suhteessa  $\sqrt{V'_h}$ :uun eli  $\sqrt{V'_h} = K \bar{X}_h$ , jossa  $K =$  vakio. Optimaalisen allokoinnin lausekkeessa sijoitetaan  $\sqrt{V'_h}$ :n tilalle  $K \bar{X}_h$ , jolloin saadaan approksimatiivisen optimaalisen allokoinnin kaava

$$(1) n_h = n \frac{N_h \sqrt{V'_h}}{\sum_h N_h \sqrt{V'_h}} = n \frac{N_h \cdot K \cdot \frac{X_h}{N_h}}{\sum_h N_h \cdot K \cdot \frac{X_h}{N_h}} = n \frac{X_h}{X} \quad (h = 1, 2)$$

Tästä saadaan

$$(2) \frac{n_h}{N_h} = \bar{X}_h \frac{n}{X} \quad (\text{H - H - M s. 215})$$

joka on optimaalinen otantasuhde ositteessa  $h$  ja ilmaisee sen, että otantasuhde ositteessa  $h$  on suhteessa (proportionate) tässä ositteessa olevien yksiköiden keskimääräisen koon mittaan. Tässä allokoinnissa on yrityksen todennäköisyys joutua otokseen suunnilleen suhteessa yrityksen kokoon.

Tällainen optimaalisen allokoinnin approksimoiminen perustuu oletukseen, että cv:t (coefficient of variation esim.

$S_y/\bar{Y}$ ) ovat suunnilleen suhteessa  $\bar{Y}_h$ :hon tai  $Y_h$ :hon, joiden tarkkaa arvoa ei tiedetä, koska niitä juuri estimoidaan. Mutta jos  $Y_{hi}$ :t ovat karkeasti suhteessa koon mittaan  $X_{hi}$ , niin  $\bar{Y}_h$  on likimäärin suhteessa  $\bar{X}_h$ :hon, josta tulos seuraa (H-H-M s. 217). Esitettyjen oletusten paikkansa pitävyydestä päästään selville, kun verrataan kaavalla (1) saatuja optimaalisen allokoinnin tuloksia aikaisemmin luvussa 4.2 saatuihin tuloksiin.

Seuraavassa taulukossa esitetään kaavalla (1) saadut ositteiden optimikoot osapopulaatioittain. Kuten luvussa 4.2 on kutakin ositetta vastaava optimaalinen  $n_h$  laskettu erikseen kunkin estimoitavan suureen suhteen. Ositteen  $n_{hopt}$  on laskettu niistä keskiarvona, koska optimiratkaisut eri estimoitavien suureiden suhteen poikkesivat vain vähän toisistaan. Kaavalla (1) saatua optimiarvoa merkitään  $n'_{hopts}$ :lla ja  $n_{hopts}$  tarkoittaa luvussa 4.2 saatua optimiarvoa ja  $n_h$  käytettyä ositteen kokoa.

Taulukko. Puu- ja paperiteollisuus

| $n_1$ | $n_1$ opts | $n'_1$ opts | $n_2$ | $n_2$ opts | $n'_2$ opts |
|-------|------------|-------------|-------|------------|-------------|
| 41    | 18.2       | 19.6        | 19    | 42.8       | 40.4        |

Metalliteollisuus

| $n_1$ | $n_1$ opts | $n'_1$ opts | $n_2$ | $n_2$ opts | $n'_2$ opts |
|-------|------------|-------------|-------|------------|-------------|
| 91    | 101.9      | 107.3       | 36    | 25.1       | 19.7        |



Muu tehdasteollisuus

|       |            |             |       |            |             |
|-------|------------|-------------|-------|------------|-------------|
| $n_1$ | $n_1$ opts | $n_1'$ opts | $n_2$ | $n_2$ opts | $n_2'$ opts |
| 172   | 134        | 127.9       | 103   | 141.0      | 147.1       |

Saadut tulokset osoittavat, että optimaalinen allokointi voidaan varsin hyvin suorittaa  $\bar{X}_h$ :n perusteella. Variansseihin ja  $\bar{X}_h$ :hon perustuvien optimaalisten allokointien eroja ei voida missään osapopulaatiossa pitää merkitsevinä. Näin ollen ne oletukset, joihin  $\bar{X}_h$ :n käyttäminen optimaaliseen allokointiin perustui, pitävät ilmeisesti paikkansa. Saadut tulokset merkitsevät myös sitä, että tulevien kyselyjen yhteydessä optimaalinen allokointi voitaneen suorittaa tällä yksinkertaisella approksimatiivisella tavalla. Variansseihin perustuva optimaalinen allokointi vaatii moninkertaisen työmäärän tässä esitettyyn menetelmään verrattuna.

5. Optimaalisten ositerajojen määrääminen

5.1. Teoreettinen tulos (kaava)

Dalenius ja Gurney ovat johtaneet teoreettisen kaavan, jolla optimaalinen stratifiointi voidaan suhdemenetelmän tapauksessa suorittaa (Dalenius ja Gurney s. 136). Oletetaan, että

suhde-estimaatti on tyyppiä  $r = \left( \frac{\bar{y}}{\bar{x}} = \frac{\sum_h p_h \bar{y}_h}{\sum_h p_h \bar{x}_h} = \frac{\hat{Y}}{\hat{X}} \right)$

ja  $p_h = \frac{N_h}{N}$  ja  $y$  ja apumuuttuja  $x$

ovat riippuvia stratifiointimuuttujasta  $s$  eli

$x = x(s)$  ja  $y = y(s)$ . Otetaan käyttöön

muuttuja  $z_h = y_h(s) - R x_h(s)$ , jossa  $R = \frac{Y}{\bar{X}}$  on

populaation arvo, jota estimoidaan. Jos on käytetty optimaalista allokointia, on  $r$ :n varianssi  $V_N(r)$  seuraava:

$$V_N(r) = \frac{1}{n\bar{X}} \left( \sum_h p_h S_{hz} \right)^2$$

jossa  $\bar{X} = \frac{X}{N}$  ja  $S_{hz}$  on  $z$ :n hajonta populaation ositteessa  $h$  (D ja G s. 137). Soveltamalla Daleniuksen ja Gurneyn johtamaa yleistä optimaalisen stratifiointin kaavaa (D ja G s. 133) saadaan kahden ositteen tapaukselle approksimatiivinen ehto

$$\frac{S_{1z}^2 + (z - \bar{Z}_1)^2}{S_{1z}} = \frac{S_{2z}^2 + (z - \bar{Z}_2)^2}{S_{2z}}$$

jossa  $\bar{Z}_h$  on  $z$ :n populaation keskiarvo ositteessa  $h$ . Saadun tuloksen perusteella voitaisiin helposti johtaa vastaava kaava suhde-estimaatille  $\hat{Y}_R = \frac{\bar{Y}}{\bar{X}} X$ , koska  $V(\hat{Y}_R)$  saadaan kertomalla  $V_N(r) X^2$ :lla.

Daleniuksen ja Gurneyn johtamalla optimaalisen stratifiointin kaavalla on hankala määrätä käytännössä optimaalinen stratifiointikohta kahdelle ositteelle. Kaavat edellyttävät populaation arvojen kuten  $\bar{Z}_h$ :n tuntemista, jota käytännössä ei tunneta. Toisaalta voitaisiin ehkä käyttää populaation arvojen sijasta niille otoksesta laskettuja estimaatteja. Optimaalisen stratifiointipisteen löytäminen em. menetelmällä vaatisi kuitenkin niin paljon kokeellista (trial and error) työtä, ettei tässä tutkimuksessa katsottu aiheelliseksi ryhtyä siihen. Seuraavassa onkin johdettu yksinkertaisempi menetelmä optimaalisen stratifiointikohdan löytämiseksi.

## 5.2. Optimaalisen ositerajan määrääminen käytännössä

### 5.2.1. Otokseen ottamisaste

Kuten tutkimuksen alussa todettiin, on käytetyssä otantajärjestelmissä kolme ositetta, joista osite 3 käsittää suuret yritykset. Tässä ositteessa on otantasuhde 100 %. Suuriin yrityksiin kuuluivat ne, joiden työntekijöiden lukumäärä on yli 500. Tätä rajaa käytettiin kaikissa osapopulaatioissa. Kyseinen raja valittiin varsin subjektiivisin perustein. Käytännön tutkimuksissa on usein osoittautunut, että jos  $n$  on otoksen koko ja  $X$  populaation yksiköiden kokojen summa, niin karkea optimiraja niille yrityksille, jotka kaikki olisi otettava otokseen, on  $\frac{X}{2n}$ . Siis ne yritykset, joiden koko on suurempi kuin  $\frac{X}{2n}$ , muodostaisivat ositteen, jossa otantasuhde on 100 % (H-H-M s. 220). Saatua rajaa  $\frac{X}{2n}$  voitaisiin kutsua otokseen ottamisasteeksi. Seuraavassa on osapopulaatioittain esitetty arvot  $\frac{X}{2n}$ .

#### Taulukko

##### Puu- ja paperiteollisuus

###### Otokseen ottamisaste

|                |     |
|----------------|-----|
| Käytetty       | 500 |
| $\frac{X}{2n}$ | 660 |

##### Metalliteollisuus

###### Otokseen ottamisaste

|                |     |
|----------------|-----|
| Käytetty       | 500 |
| $\frac{X}{2n}$ | 420 |

##### Muu tehdasteollisuus

###### Otokseen ottamisaste

|                |     |
|----------------|-----|
| Käytetty       | 500 |
| $\frac{X}{2n}$ | 410 |

Taulukosta havaitaan, että käytetyt ja  $\frac{\bar{X}}{2n}$  :llä saadut otokseen ottamisasteet poikkeavat jonkin verran toisistaan.

### 5.2.2. Optimaalisen ositerajan kaava

Ne yritykset, joiden työntekijöiden lukumäärä oli alle 500, muodostivat tässä tutkimuksessa varsinaisen otoksen, joka jaettiin kahteen ositteeseen. Ositeraja oli aikaisempien kyselyjen yhteydessä valittu subjektiivisesti 100:n työntekijälukumäärän kohdalle. Seuraavassa pyritään tutkimaan analyyttisemmin, kuinka voitaisiin määrätä approksimatiivinen optimaalinen ositeraja. Karkea approksimatiivinen kaava voisi olla se, että kummassakin ositteessa yritysten työntekijöiden kokonaissumma olisi suunnilleen sama (H-H-M s. 219).

Seuraavassa johdetaan varsinainen kaava optimaalisen ositerajan määräämiseksi olettaen, että on käytetty optimaalista allokoointia. Jotta löydettäisiin se ositejako, joka antaisi pienimmän mahdollisen varianssin estimaatille, on tälle varianssille laskettava arvio erikseen kullekin ositerajalle. Oletetaan, että optimaalinen allokointi on suoritettu luvussa 4.3. esitetyllä menetelmällä, jolloin optimaalinen allokointi on suhteessa  $\bar{X}_h$ :hon. Optimaalisen allokoinnin kaava oli

$$(1) \quad n_h = n \frac{X_h}{\bar{X}}$$

Sijoittamalla  $n_h$ :n arvo yleiseen varianssin kaavaan saadaan:

$$\begin{aligned} (2) \quad V(\hat{Y}_{RS}) &= \sum_h \frac{N_h^2}{n_h} \cdot V'_h - \sum_h N_h V'_h \\ &= \frac{X}{n} \sum_h \frac{N_h^2}{\bar{X}_h} \cdot V'_h - \sum_h N_h V'_h \\ &= \sum_h N_h \frac{X}{n\bar{X}_h} \cdot V'_h - \sum_h N_h V'_h \end{aligned}$$

Kaavalla (1) saatu optimaalinen allokoinen perustui oletukseen, että ositteen hajonta on likimäärin suhteessa estimoitavan suureen populaation keskiarvoon  $\bar{Y}_h$ :hon. Koska yrityksen investointien arvo  $Y_{hi}$  on likimäärin suhteessa yrityksen koon mittaan  $X_{hi}$ :hin, niin hajonta ositteessa on myös likimäärin suhteessa  $\bar{X}_h$ :hon (H-H-M s. 217). Näiden oletusten todettiin pitävän varsin hyvin paikkansa luvussa 4.3. siinä käytetyn ositerajan puitteissa. Voitaneen olettaa, että myös muiden ositerajojen kohdalla ositteen hajonta on likimäärin suhteessa  $\bar{X}_h$ :hon. Tällöin on siis voimassa

$$\sqrt{V'_h} = K_1 \bar{X}_h \quad \text{eli}$$

$$(3) \quad V'_h = K \bar{X}_h^2, \quad K = K_1^2 \quad (\text{Cochran s. 135})$$

Huomattakoon, että suhde-estimaatin tapauksessa  $V'_h$  on muuttujan  $d_{hi}$  varianssi, joka määritellään  $d_{hi} = y_{hi} - R_h x_{hi}$  (Cochran s. 135).

Sijoittamalla (3) kaavaan (2) saadaan

$$\begin{aligned} (4) \quad V(\hat{Y}_{RS}) &= K \cdot \left[ \sum_h N_h \frac{X}{n\bar{X}_h} \bar{X}_h^2 - \sum_h N_h \frac{X_h^2}{N_h^2} \right] \\ &= K \cdot \left[ \sum_h N_h \frac{X \cdot \bar{X}_h}{n} - \sum_h \frac{X_h^2}{N_h} \right] \\ &= K \cdot \left[ \sum_h N_h \cdot X \cdot \frac{X_h}{N_h} - \sum_h X_h \bar{X}_h \right] \\ &= K \cdot \left[ \sum_h \frac{X \cdot X_h}{n} - \sum_h X_h \bar{X}_h \right] \end{aligned}$$

$$= K \cdot \left[ \sum_h x_h \left( \frac{X}{n} - \bar{x}_h \right) \right]$$

$$= K \cdot T$$

jossa  $T = \sum_h x_h \left( \frac{X}{n} - \bar{x}_h \right)$  ,  $h = 1, 2$

Se ositeraja, joka antaa pienimmän mahdollisen estimaatin varianssin, voidaan nyt löytää laskemalla eri ositerajoja vastaavat T:n arvot. Huomattakoon vielä, että kunkin ositerajan kohdalla kaava (4) edellyttää käytetyn optimaalista allokointia. Tilastollisesta päätoimistosta saadun puu- ja paperiteollisuutta koskevan kaikkien populaatioon kuuluvien yritysten tilastoaineiston perusteella on laskettu kaavan (4) perusteella optimaalinen ositeraja tälle sektorille. Käytetty aineisto esitetään seuraavassa taulukossa.

Taulukko.

Puu- ja paperiteollisuus

| Ositeraja<br>työntekijöitä | Työntekijöitä ositteessa 1 | Yrityksiä |
|----------------------------|----------------------------|-----------|
| 50                         | 5 111                      | 160       |
| 100                        | 8 491                      | 207       |
| 150                        | 11 544                     | 228       |
| 200                        | 12 076                     | 231       |
| 300                        | 13 517                     | 237       |
| 400                        | 15 658                     | 240       |

Koko populaatiossa oli työntekijöitä yhteensä 16 525 ja yrityksiä 245.

Optimaalinen ositeraja voidaan päätellä seuraavan taulukon perusteella:

Taulukko.

| Ositeraja  | 1000·T       | Työntekijöiden lukumäärä |         |
|------------|--------------|--------------------------|---------|
|            |              | Osite 1                  | Osite 2 |
| 50         | 2 851        | 5 111                    | 10 660  |
| 100        | 3 407        | 7 280                    | 8 491   |
| <u>150</u> | <u>2 506</u> | 10 790                   | 4 981   |
| <u>200</u> | <u>2 502</u> | 12 076                   | 3 695   |
| 300        | 2 643        | 13 517                   | 2 254   |
| 400        | 2 783        | 14 680                   | 1 090   |

Taulukon T:n arvoista voidaan päätellä, että kaavalla (4) saatu optimaalinen ositeraja olisi 150-200 työntekijää käsittävien yritysten kohdalla, joka siis poikkeaa selvästi nykyisin käytössä olevasta ositerajasta 100 työntekijän yritysten kohdalla. Tämän luvun alussa esitetty karkea ositerajan määrääminen perustui siihen, että työntekijöiden lukumäärä olisi kummasakin ositteessa sama. Taulukosta havaitaan, että ositerajaksi tulisi tällöin suunnilleen nykyisin käytössä oleva raja.

Suoritetun analyysin perusteella ei voida tehdä kovin selviä johtopäätöksiä optimaalisesta ositerajasta puu- ja paperiteollisuudessa. Kaavan (4) perusteella saatua ositerajaa voitaneen kuitenkin pitää luotettavampana kuin työntekijöiden lukumäärän perusteella saatua ositerajaa. Jotta kaavan (4) perusteella voitaisiin tarkemmin analysoida optimaalista ositerajaa, olisi analysoitava perusteellisemmin kaavan (4) johtamiseen sisältyvät olettamukset. Muiden osapopulaatioiden kohdalla ei optimaalista ositerajaa tutkittu, koska ei ollut saatavissa tarvittavaa tilastoaineistoa.

## 6. Käytännölliset johtopäätökset

Tutkimuksessa saatujen tulosten perusteella voidaan esittää seuraavat johtopäätökset tulevien kyselyjen suorittamista varten:

- 1) Selvästi tehokkain estimointimenetelmä on nykyisin käytössä oleva suhdemenetelmä. Sitä vastoin suhdemenetelmässä käytettävän apumuuttujan valintaa on tarkasteltava osapopulaatioittain. Tutkimuksen tulokset osoittivat, että tehokkain apumuuttuja ei ole sama kaikissa osapopulaatioissa. Puu- ja paperiteollisuudessa tehokkaimmaksi apumuuttujaksi osoitettiin tuotannon brutto- tai jalostusarvo ja metalliteollisuudessa työntekijöiden lukumäärä. Muun tehdasteollisuuden kohdalla tehokkain apumuuttuja voidaan selvästi määrätä vasta sitten, kun tämä sektori disaggregoidaan alaryhmiin, jolloin pääoma- ja työvoimavaltaiset alat saadaan erilleen toisistaan.
- 2) Kyselyssä käytetty allokointi poikkeaa huomattavasti optimaalisesta allokoinnista. Tulevien kyselyjen yhteydessä optimaalinen allokointi voitaneen suorittaa varsin tehokkaasti luvun 4.3. menetelmällä eli kaavalla

$$n_h = n \frac{X_h}{\bar{X}}$$

koska estimaattien varianssit eivät ole herkkiä kohtuulliselle poikkeamiselle tarkasta optimaalisesta allokoinnista.

- 3) Mikäli tavoitteeksi asetettu estimaattien varianssi (4.2.) halutaan saavuttaa, on otoksen kokoa nostettava metalli- ja muussa tehdasteollisuudessa nykyisestään, vaikka käytettäisiinkin optimaalista allokoointia.
- 4) Puu- ja paperiteollisuudessa optimaalinen ositeraja on 150-200 työntekijän kohdalla, kun sitä vastoin nykyisin käytössä oleva ositeraja on 100 työntekijän kohdalla. Optimaalista ositerajaa olisi kuitenkin tutkittava tarkemmin kaikissa osapopulaatioissa, ennen kuin nykyistä ositerajaa on syytä muuttaa.

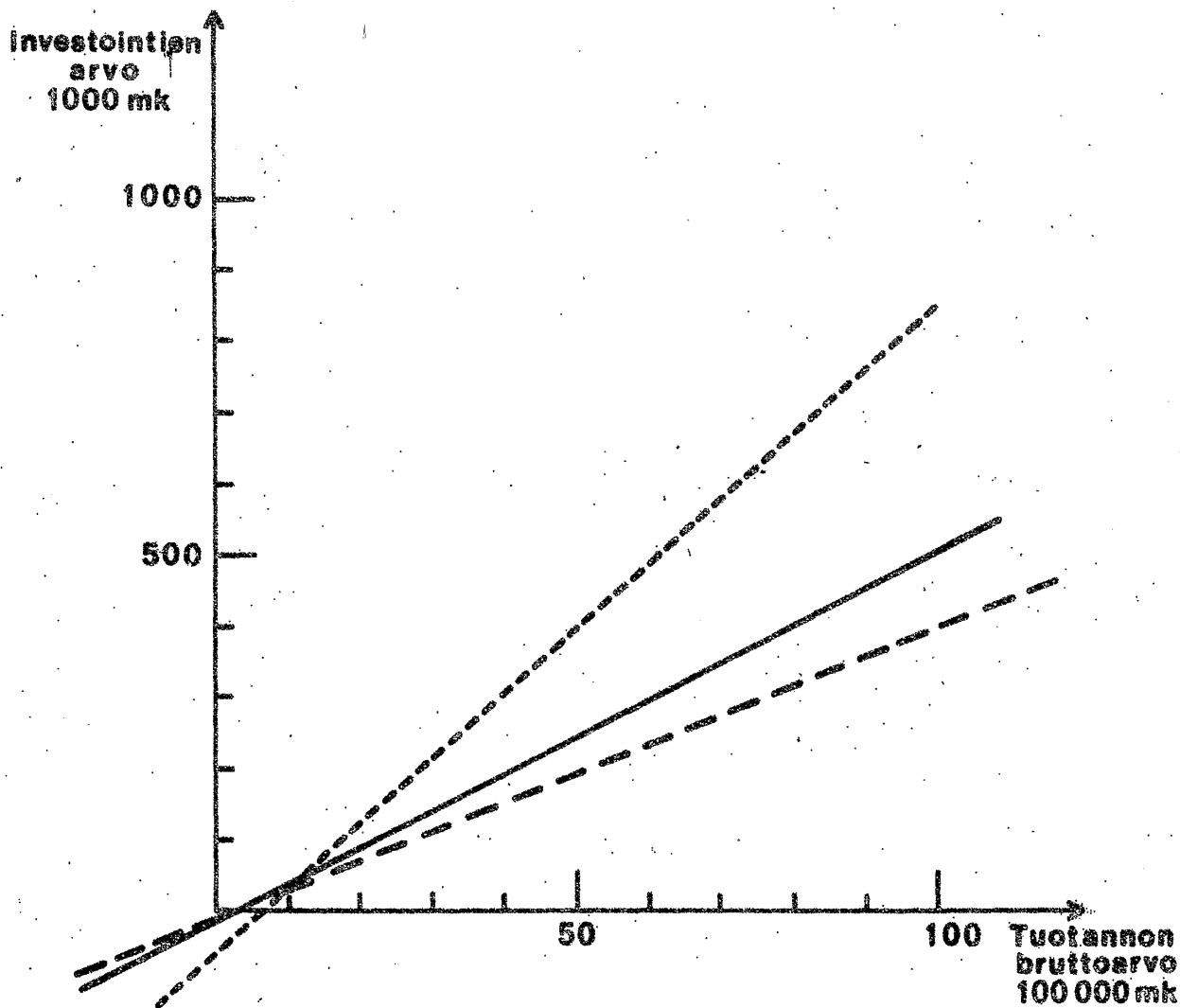


LÄHDELUETTELO

- Bowker, A.H. and Goode H.P. Sampling Inspection by Variables,  
New York 1952
- Cochran, W. Sampling Techniques, New York 1962
- Deming, W. Some Theory of Sampling,  
New York 1950
- Elfving, G. Todennäköisyyslaskenta, Helsinki 1966
- Murthy, M.N. Sampling Theory and Methods,  
Calcutta 1967
- Sukthatme, P.V. Sampling Theory of Surveys with  
Applications, New Delhi 1954
- Yates, F. Sampling Methods for Censuses and  
Surveys, London 1965

Liite A 1

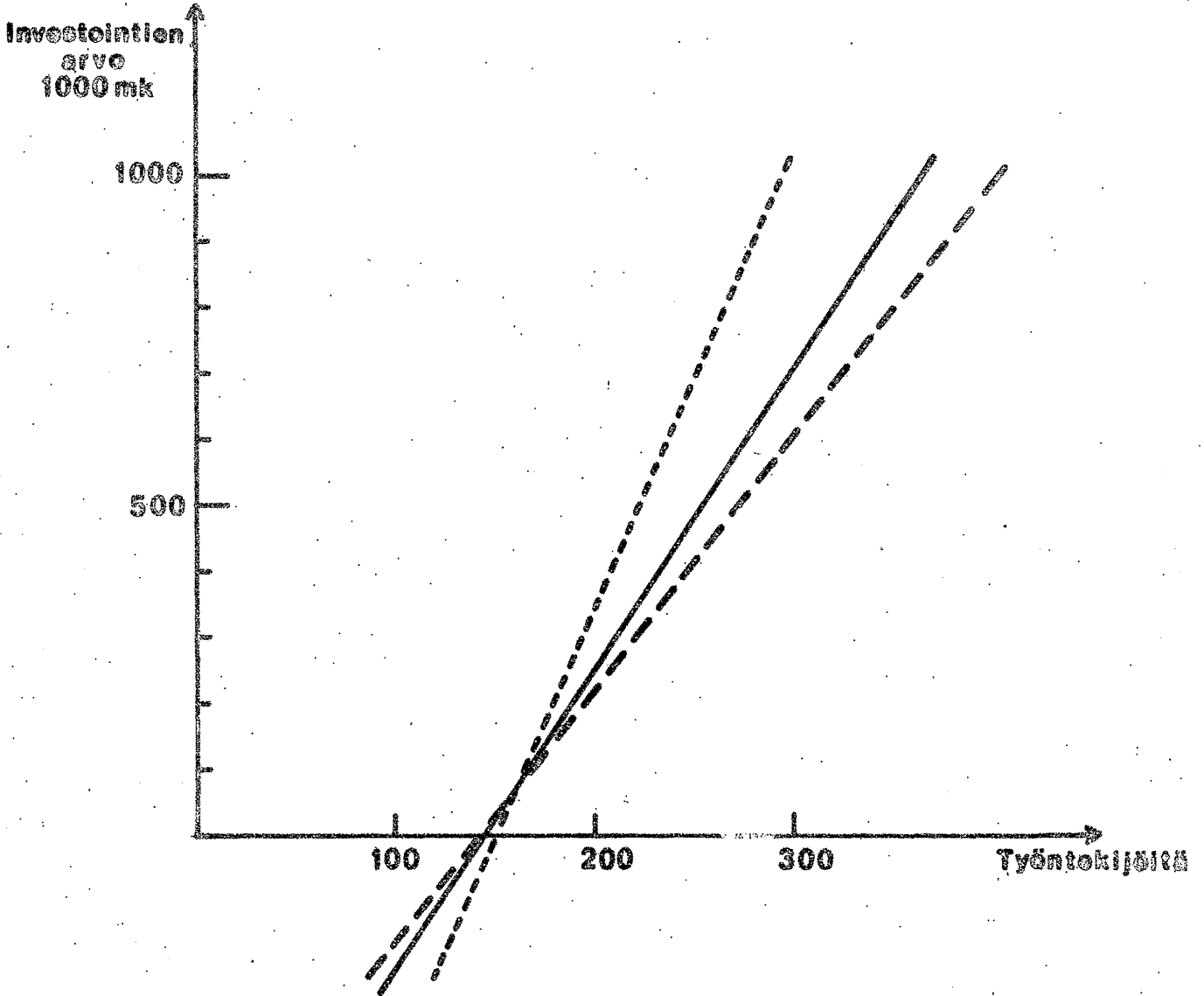
Puu- ja paperiteollisuus



|         |              |                                      |           |
|---------|--------------|--------------------------------------|-----------|
| -----   | Investoinnit | 1967: $y = -64.6 + 8.95 x$<br>(1.50) | R = 0.653 |
| - - - - | "            | 1968: $y = -12.9 + 4.16 x$<br>(0.50) | R = 0.768 |
| ————    | "            | 1969: $y = -19.1 + 5.48 x$<br>(0.95) | R = 0.645 |

Lähte B I

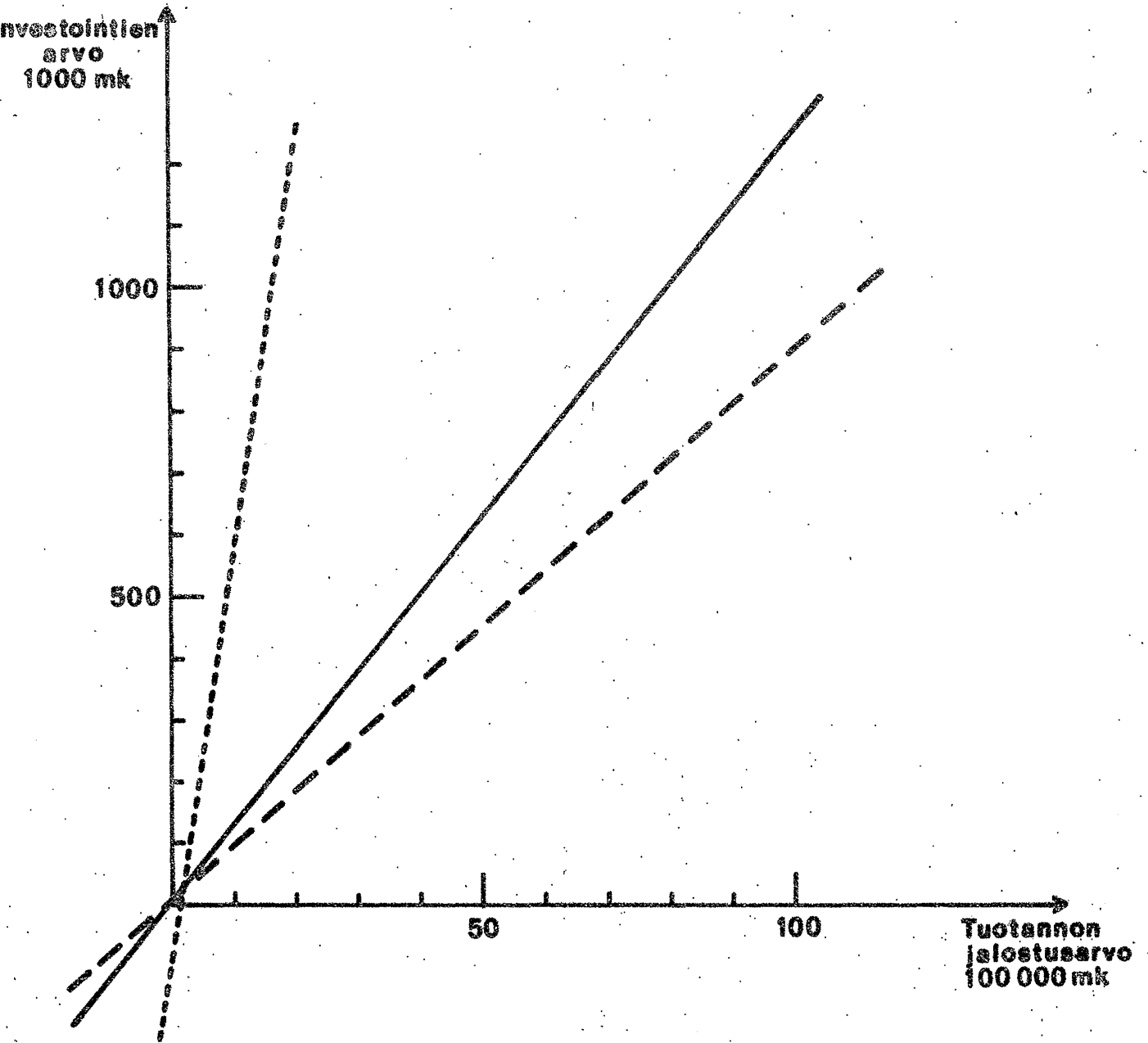
Puu- ja paperiteollisuus



|       |                    |                           |             |
|-------|--------------------|---------------------------|-------------|
| ----- | Investoinnit 1967: | $y = -1194 + 7.98 x$      | $R = 0.478$ |
|       |                    | (2.10)                    |             |
| ----- | "                  | 1968: $y = -543 + 3.74 x$ | $R = 0.570$ |
|       |                    | (0.79)                    |             |
| ----- | "                  | 1969: $y = -618 + 4.38 x$ | $R = 0.418$ |
|       |                    | (1.35)                    |             |

Kaite C 1

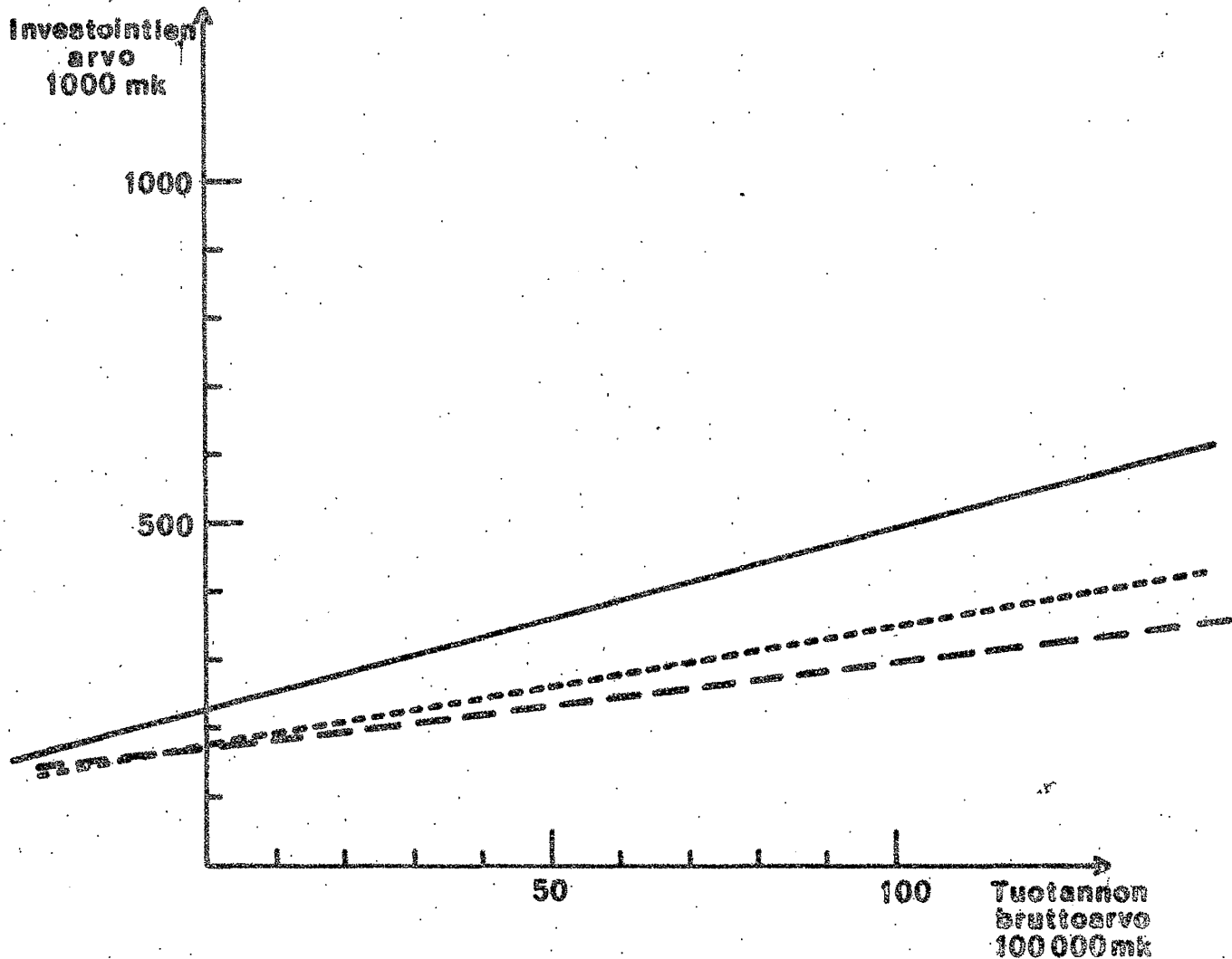
Puu- ja paperiteollisuus



|           |   |             |
|-----------|---|-------------|
| -----     | Investoinnit 1967: $y = -51.4 + 21.8 x$ | $R = 0.598$ |
|           | (4.2)                                   |             |
| - . - . - | 1968: $y = 5.1 + 9.2 x$                 | $R = 0.641$ |
|           | (1.6)                                   |             |
| —————     | 1969: $y = 0.9 + 12.6 x$                | $R = 0.552$ |
|           | (2.7)                                   |             |

Liite A 2

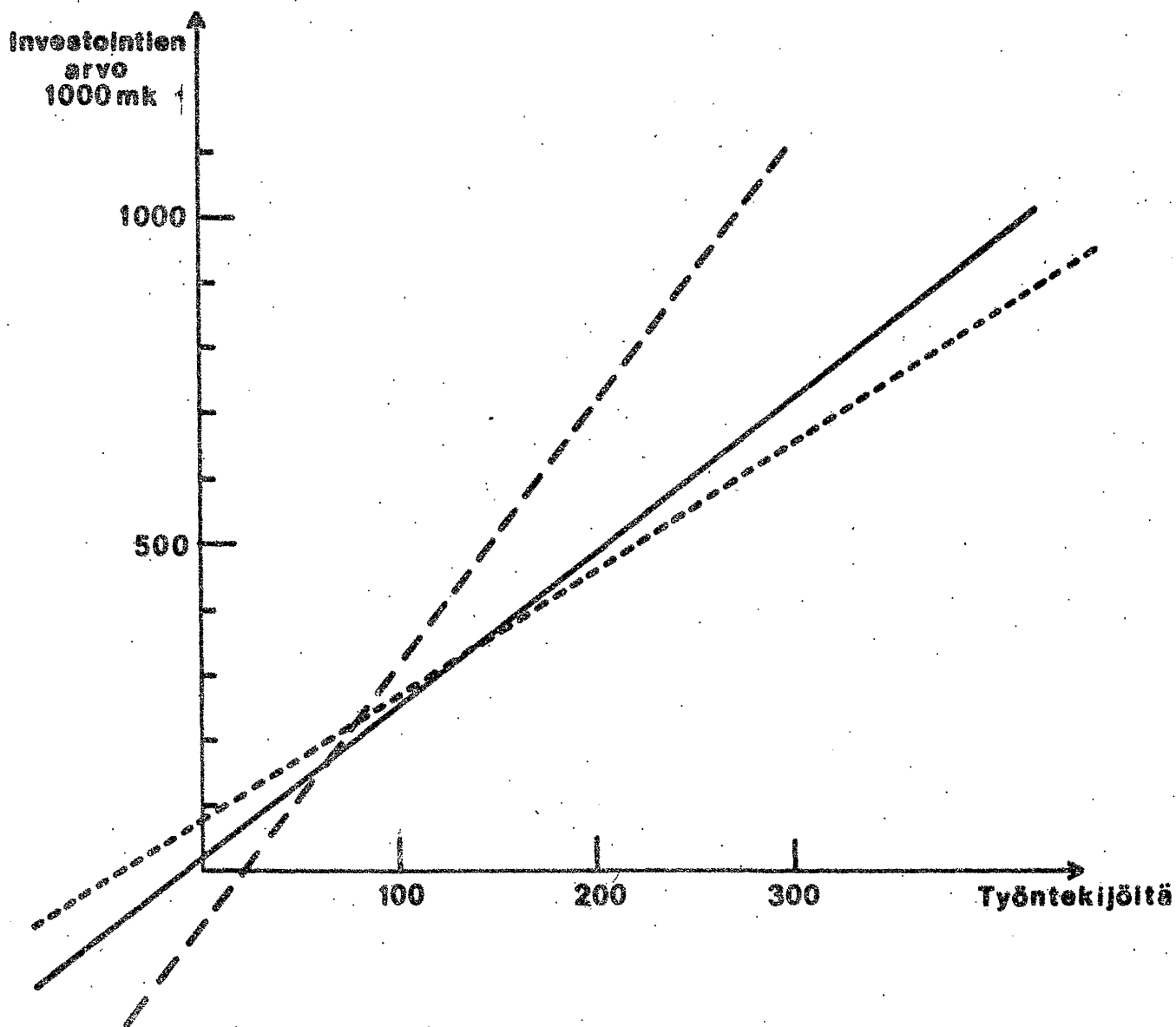
Metalliteollisuus



|       |                    |                            |             |
|-------|--------------------|----------------------------|-------------|
| ----- | Investointit 1967: | $y = 170.6 + 1.85 x$       | $R = 0.194$ |
|       |                    | (0.90)                     |             |
| ----- | "                  | 1968: $y = 168.1 + 1.32 x$ | $R = 0.135$ |
|       |                    | (0.80)                     |             |
| ----- | "                  | 1969: $y = 228.9 + 2.68 x$ | $R = 0.175$ |
|       |                    | (1.37)                     |             |

Liite B 2

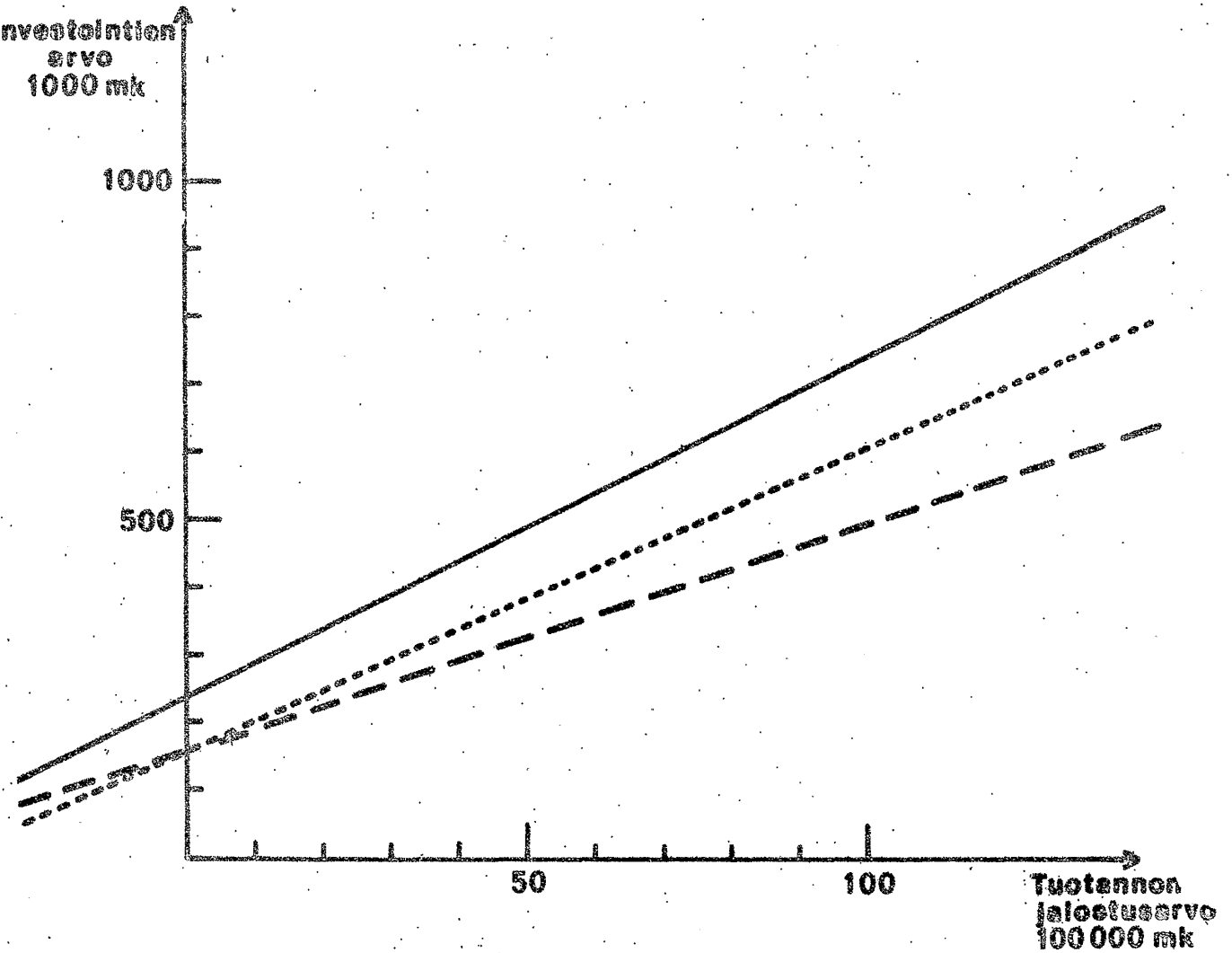
Metalliteollisuus



|       |                    |                                  |           |
|-------|--------------------|----------------------------------|-----------|
| ..... | Investoinnit 1967: | $y = 77.6 + 1.8x$<br>(0.5)       | R = 0.387 |
| ----- | "                  | 1968: $y = 97.4 + 4.1x$<br>(0.8) | R = 0.470 |
| ————— | "                  | 1969: $y = 30.1 + 2.9x$<br>(0.8) | R = 0.410 |

Liite C 2

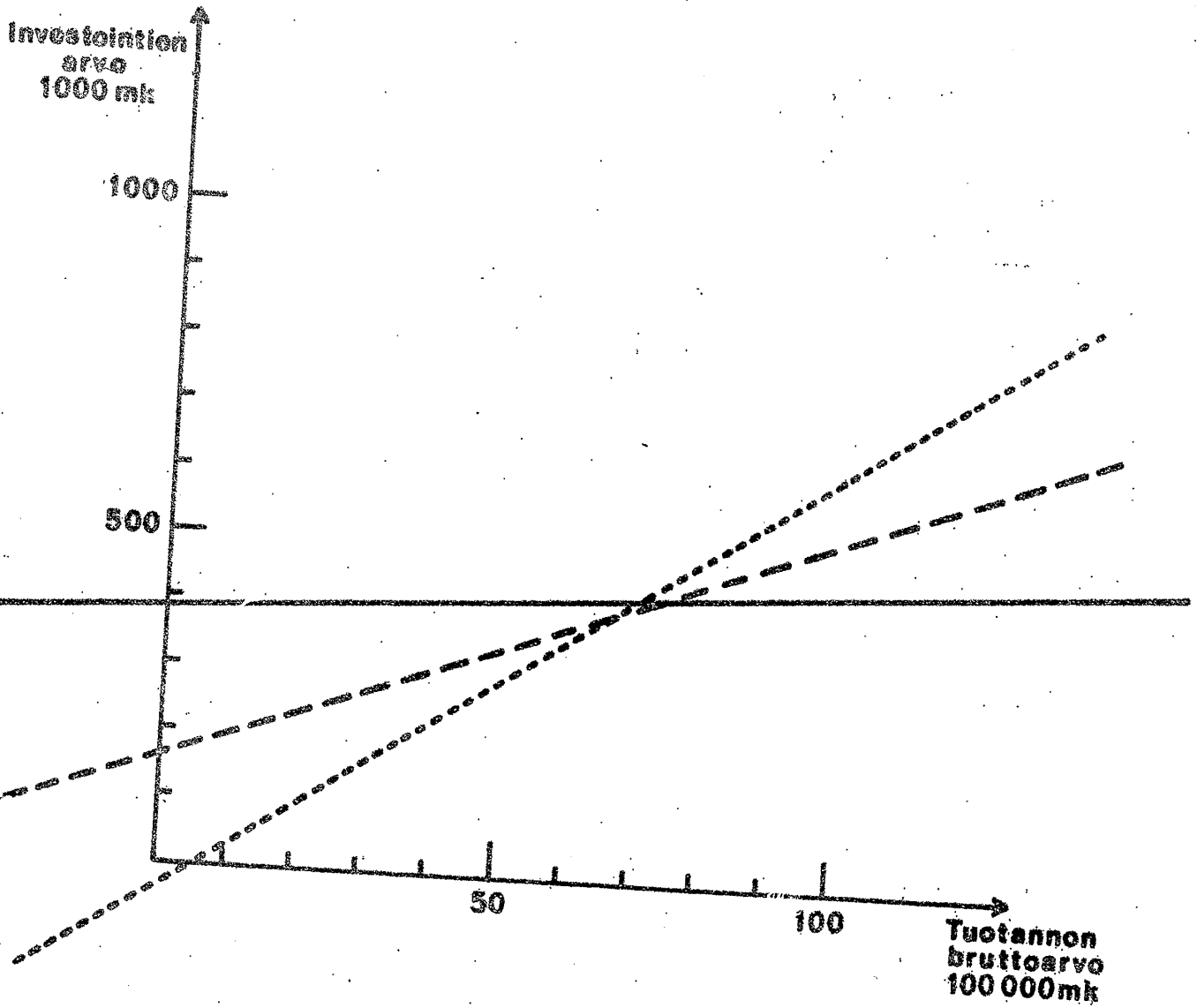
Metalliteollisuus



|       |                    |                                    |             |
|-------|--------------------|------------------------------------|-------------|
| ----- | Investoinnit 1967: | $y = 155.6 + 4.5 x$<br>(2.0)       | $R = 0.212$ |
| ----- | "                  | 1968: $y = 155.9 + 3.3 x$<br>(1.8) | $R = 0.158$ |
| ----- | "                  | 1969: $y = 233.1 + 5.1 x$<br>(3.1) | $R = 0.140$ |

Liite A 3

Maun kehittäminen

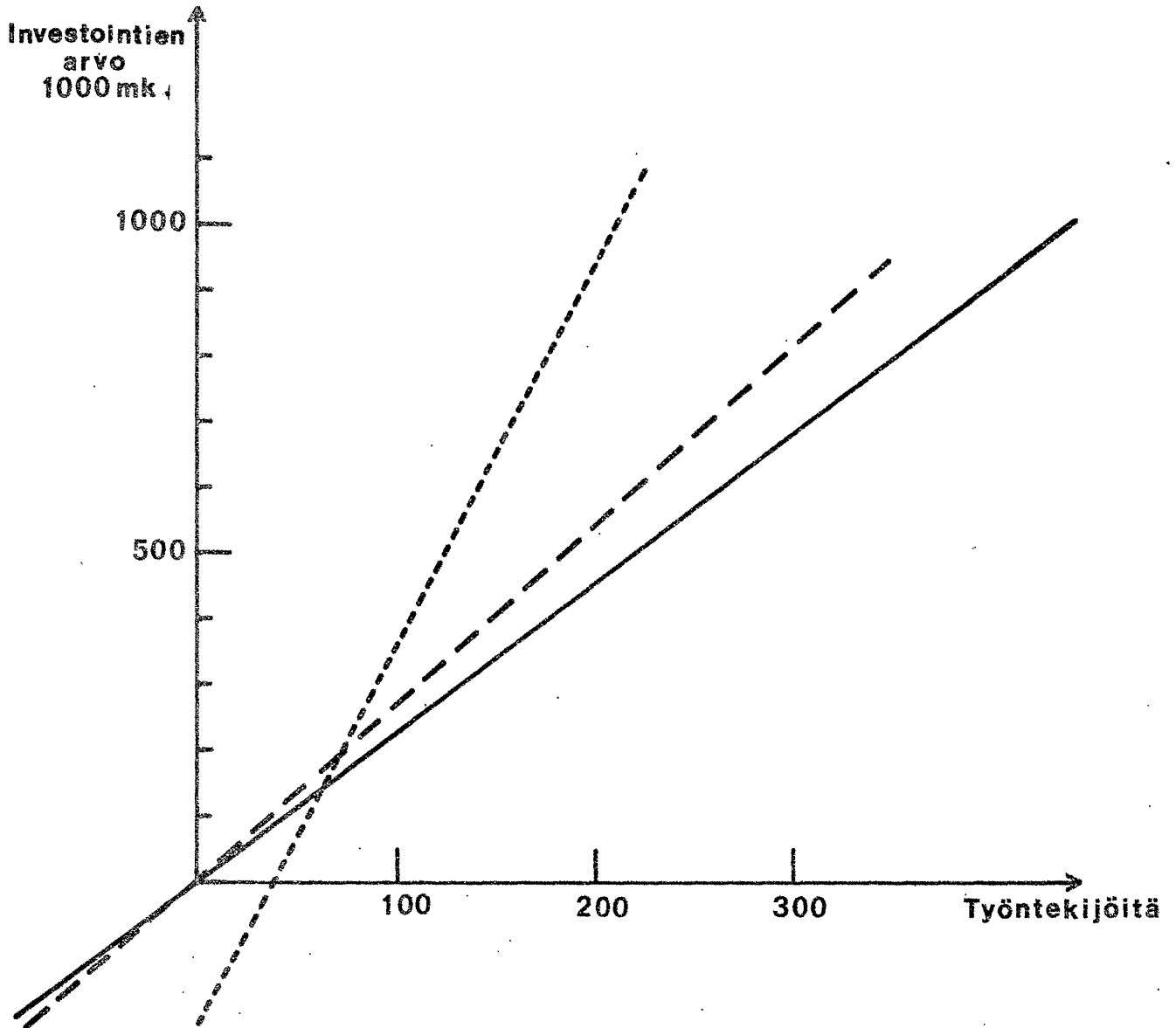


|       |                    |                     |             |
|-------|--------------------|---------------------|-------------|
| ..... | Investoinnit 1967: | $y = 41.6 + 6.6 x$  | $R = 0.786$ |
| ----- | "                  | $y = 156.4 + 3.8 x$ | $R = 0.372$ |
| ————  | "                  | $y = 386.9 + 0.6 x$ | $R = 0.136$ |



Liite B 3

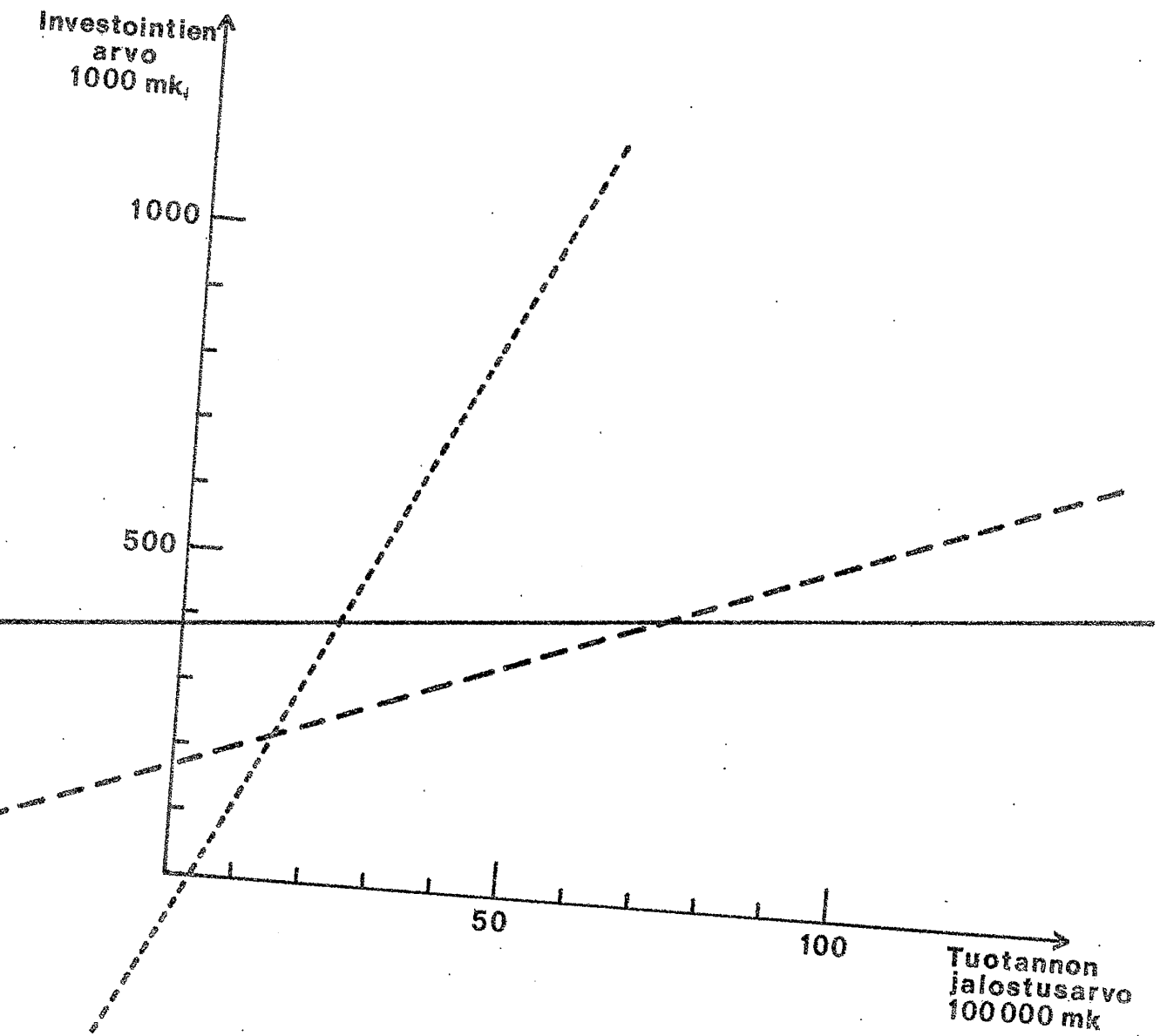
Muu tehdasteollisuus



|       |                    |                                    |           |
|-------|--------------------|------------------------------------|-----------|
| ..... | Investoinnit 1967: | $y = -220.0 + 5.8 x$<br>(0.6)      | R = 0.538 |
| ----- | "-                 | 1968: $y = 7.2 + 2.7 x$<br>(0.5)   | R = 0.376 |
| ————— | "-                 | 1969: $y = -23.3 + 2.5 x$<br>(0.4) | R = 0.435 |

Liite C 3

Muu tehdasteollisuus



|             |                    |                      |             |
|-------------|--------------------|----------------------|-------------|
| -----       | Investoinnit 1967: | $y = -90.0 + 20.3 x$ | $R = 0.745$ |
|             |                    | (1.3)                |             |
| - . - . - . | "                  | $y = 156.6 + 3.8 x$  | $R = 0.372$ |
|             |                    | (0.7)                |             |
| —————       | "                  | $y = 386.9 + 0.6 x$  | $R = 0.113$ |
|             |                    | (0.8)                |             |

Liite 1a: Estimaattien varianssien laskemisessa käytetyt kaavat

otantamenetelmä      estimointimenetelmä      varianssin kaava

|                   |                  |   |
|-------------------|------------------|---|
| simple random     | $N \bar{y}$      | $V = \frac{(N-n)}{nN} \left( \sum W_h S_h^2 - \sum \frac{W_h S_h^2}{n_h} + \sum \frac{W_h^2 S_h^2}{n_h} \right. \\ \left. + \sum W_h \bar{y}_h^2 - \left( \sum W_h \bar{y}_h \right)^2 \right) \text{ (Cochran s. 99)}$ |
| stratified random | $N \bar{y}_{st}$ | $V = \sum_h N_h (N_h - n_h) \frac{S_{yh}^2}{n_h} \text{ (Cochran s. 74)}$   |
| "                 | $\hat{Y}_{Rs}$   | $V = \sum_h \frac{N_h (N_h - n_h)}{n_h} (S_{yh}^2 + R_h^2 S_{xh}^2 - 2 R_h r_h S_{xh} S_{yh}) \\ \text{(Cochran s. 129)}$   |
| "                 | $\hat{Y}_{Rc}$   | <p>muuten sama kuin <math>\hat{Y}_{Rs}</math>:n V paitsi, että <math>R_h^2</math>:n paikalla on <math>R^2</math> ja <math>R_h</math>:n paikalla R</p> <p style="text-align: right;">(Cochran s. 131)</p>                |

$$\text{Cov}(x,y) = r_h S_{xh} S_{yh} = S_{yxh}$$

Liite 1b: Estimaattien varianssien estimaattien laskemisessa  
tarvittavia muuttujia: aineisto kevät 1968

Puu- ja paperiteollisuus

| osite | $N_h$ | $n_h$ | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-------|-------|-----------|-----------|-----------|----------|
| 1     | 205   | 40    | 51.       | 54.       | 42.       | 19.      |
| 2     | 38    | 18    | 481.      | 1160.     | 379.      | 75.      |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 612.       | 116.       | 93.        | 0.82       | 0.83       | 0.64       |
| 2     | -14576.    | 55210.     | 2069.      | 1.76       | 3.18       | 1.38       |
| koko  |            |            |            | 1.41       | 2.30       | 1.10       |

$s_{yh} = \text{est } S_{yh}$        $y = \text{investoinnit vuosina 1967, 1968, 1969}$

$x = \text{työntekijöiden lukumäärä}$

$$r'_{yh} = \frac{y}{x}$$

Metalliteollisuus

| osite | $N_h$ | $n_h$ | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-------|-------|-----------|-----------|-----------|----------|
| 1     | 461   | 90    | 200.      | 386.      | 140.      | 21.      |
| 2     | 76    | 34    | 749.      | 1243.     | 1187.     | 106.     |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 235.       | 2553.      | 588.       | 2.90       | 3.65       | 1.78       |
| 2     | 1350.      | 6471.      | 1381.      | 2.51       | 2.87       | 3.67       |
| koko  |            |            |            | 2.62       | 3.13       | 3.16       |

Muu tehdasteollisuus

| osite | $N_h$ | $n_h$ | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-------|-------|-----------|-----------|-----------|----------|
| 1     | 879   | 170   | 129.      | 165.      | 105.      | 19.      |
| 2     | 208   | 100   | 1435.     | 937.      | 737.      | 87.      |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 240.       | 505.       | 359.       | 1.87       | 2.08       | 1.51       |
| 2     | 57179.     | 14510.     | 16029.     | 4.26       | 2.95       | 2.44       |
| koko  |            |            |            | 3.73       | 2.76       | 2.24       |

$x = \text{työntekijöiden lukumäärä}$

Liite 2a: Estimaattien varianssien estimoimisessa käytetyt  
muuttujat aineisto: syksyn 1968 kysely

x = työntekijöiden lukumäärä

Puu- ja paperiteollisuus

| osite | $N_h$ | $n_h$ | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-------|-------|-----------|-----------|-----------|----------|
| 1     | 207   | 41    | 493.      | 158.      | 81.       | 23.      |
| 2     | 38    | 19    | 784.      | 327.      | 633.      | 113.     |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 3627.      | 1224.      | 292.       | 2.48       | 1.37       | 1.99       |
| 2     | 55902.     | 18128.     | 4850.      | 2.67       | 1.33       | 1.31       |

Metalliteollisuus

| osite | $N_h$ | $n_h$ | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-------|-------|-----------|-----------|-----------|----------|
| 1     | 461   | 91    | 312.      | 377.      | 731.      | 22.      |
| 2     | 74    | 36    | 692.      | 542.      | 798.      | 107.     |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 535.       | 719.       | 1227.      | 3.42       | 3.58       | 5.96       |
| 2     | 23820.     | 13872.     | 25582.     | 2.16       | 1.67       | 2.33       |

Muu tehdasteollisuus

| osite | $N_h$ | $n_h$ | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-------|-------|-----------|-----------|-----------|----------|
| 1     | 874   | 172   | 297.      | 246.      | 505.      | 19.      |
| 2     | 208   | 103   | 1261.     | 1538.     | 1746.     | 98.      |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 1351.      | 895.       | 1725.      | 3.11       | 2.72       | 3.52       |
| 2     | 55314.     | 48111.     | -37404.    | 3.11       | 3.67       | 3.74       |

Liite 2b: Estimaattien varianssien estimaattien laskemisessa  
 käytetyt muuttujien arvot: aineisto syksyn 1968  
 kysely

x = tuotannon bruttoarvo

Puu- ja paperiteollisuus

| osite | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-----------|-----------|-----------|----------|
| 1     | 49.       | 158.      | 81.       | 11.8     |
| 2     | 784.      | 327.      | 633.      | 55.6     |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 3557.      | 1218.      | 283.       | 8.65       | 4.29       | 4.23       |
| 2     | 33930.     | 14170.     | 18530.     | 6.51       | 3.60       | 5.22       |

Metalliteollisuus

| osite | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-----------|-----------|-----------|----------|
| 1     | 312.      | 377.      | 731.      | 14.6     |
| 2     | 692.      | 542.      | 798.      | 79.8     |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 7.         | -253.      | 861.       | 9.76       | 10.2       | 17.0       |
| 2     | 5450.      | 4684.      | 14424.     | 5.11       | 4.2        | 5.8        |

Muu tehdasteollisuus

| osite | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-----------|-----------|-----------|----------|
| 1     | 297.      | 246.      | 505.      | 30.0     |
| 2     | 1261.     | 1538.     | 1746.     | 142.6    |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 1402.      | 1287.      | 1157.      | 5.90       | 5.16       | 6.79       |
| 2     | 142835.    | 63953.     | -23728.    | 5.67       | 6.69       | 6.81       |

Liite 2c: Varianssien estimaattien laskemisessa tarvittavat  
muuttujien arvot: syksyn 1968 kysely

apumuuttuja x = tuotannon jalostusarvo

Puu- ja paperiteollisuus

| osite | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-----------|-----------|-----------|----------|
| 1     | 49.       | 158.      | 81.       | 4.58     |
| 2     | 784.      | 327.      | 633.      | 20.50    |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 1409.      | 423.       | 127.       | 21.07      | 10.44      | 10.31      |
| 2     | 10329.     | 3421.      | 4595.      | 16.77      | 9.28       | 13.47      |

Metalliteollisuus

| osite | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-----------|-----------|-----------|----------|
| 1     | 312.      | 377.      | 731.      | 4.7      |
| 2     | 692.      | 542.      | 798.      | 33.9     |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 69.        | -35.       | -16.       | 19.8       | 20.8       | 34.6       |
| 2     | 2273.      | 2093.      | 4964.      | 10.1       | 8.3        | 11.5       |

Muu tehdasteollisuus

| osite | $s_{67h}$ | $s_{68h}$ | $s_{69h}$ | $s_{xh}$ |
|-------|-----------|-----------|-----------|----------|
| 1     | 297.      | 246.      | 505.      | 5.34     |
| 2     | 1261.     | 1538.     | 1746.     | 42.5     |

| osite | $s_{67xh}$ | $s_{68xh}$ | $s_{69xh}$ | $r'_{67h}$ | $r'_{68h}$ | $r'_{69h}$ |
|-------|------------|------------|------------|------------|------------|------------|
| 1     | 527.       | 399.       | 443.       | 18.4       | 16.1       | 20.9       |
| 2     | 38372.     | 21518.     | 74205.     | 13.8       | 16.3       | 16.6       |

Liite 3a: Optimaalinen allokointi: tarvittavat apusuureet  
Aineisto kevään 1968 kysely

Puu ja paperiteollisuus

I ratio separate menetelmä

| estimoitava suure | $\sqrt{V'_1}$ | $\sqrt{V'_2}$ | $N_2\sqrt{V'_2}$ | $N_1\sqrt{V'_1}$ | $\sum_{h=1}^2 N_h\sqrt{V'_h}$ |
|-------------------|---------------|---------------|------------------|------------------|-------------------------------|
| investoinnit 1967 | .053          | .484          | 18.39            | 10.97            | 29.36                         |
| " 1968            | .057          | 1.026         | 38.99            | 11.80            | 50.79                         |
| " 1969            | .044          | .395          | 15.01            | 9.10             | 24.11                         |

II ratio combined menetelmä

|                   |      |       |       |       |       |
|-------------------|------|-------|-------|-------|-------|
| investoinnit 1967 | .055 | .483  | 18.35 | 11.39 | 29.74 |
| " 1968            | .059 | 1.031 | 39.18 | 12.21 | 51.39 |
| " 1969            | .046 | .386  | 14.67 | 9.52  | 24.19 |

Metalliteollisuus

I ratio separate menetelmä

|                   |      |       |       |        |        |
|-------------------|------|-------|-------|--------|--------|
| investoinnit 1967 | .210 | .795  | 58.80 | 96.80  | 155.60 |
| " 1968            | .370 | 1.260 | 93.20 | 170.60 | 263.80 |
| " 1969            | .140 | 1.240 | 91.80 | 64.50  | 156.30 |

II ratio combined menetelmä

|                   |      |       |       |        |        |
|-------------------|------|-------|-------|--------|--------|
| investoinnit 1967 | .218 | .810  | 59.90 | 100.50 | 160.40 |
| " 1968            | .381 | 1.280 | 94.70 | 175.60 | 270.30 |
| " 1969            | .150 | 1.250 | 92.50 | 69.20  | 161.70 |

Muu tehdasteollisuus

I ratio separate menetelmä

|                   |      |       |        |        |       |
|-------------------|------|-------|--------|--------|-------|
| investoinnit 1967 | .132 | 1.320 | 274.60 | 115.40 | 390.0 |
| " 1968            | .151 | .935  | 194.50 | 132.00 | 326.5 |
| " 1969            | .103 | .712  | 148.10 | 90.00  | 238.1 |

II ratio separate menetelmä

|                   |      |       |        |        |        |
|-------------------|------|-------|--------|--------|--------|
| investoinnit 1967 | .139 | 1.323 | 275.20 | 121.50 | 396.70 |
| " 1968            | .154 | .930  | 193.40 | 135.00 | 328.40 |
| " 1969            | .120 | .719  | 149.60 | 104.90 | 254.50 |



Liite 3b: Optimaalista allokoimista varten tarvittavat apusuureet  
Aineisto syksyn 1968 kysely

Puu- ja paperiteollisuus

| estimoitava suure | $\sqrt{V'_1}$ | $\sqrt{V'_2}$ | $N_2\sqrt{V'_2}$ | $N_1\sqrt{V'_1}$ | $\sum_{h=1}^2 N_h\sqrt{V'_h}$ |
|-------------------|---------------|---------------|------------------|------------------|-------------------------------|
| investoinnit 1967 | 0.476         | 0.636         | 24.17            | 19.52            | 43.69                         |
| " 1968            | 0.150         | 0.285         | 10.83            | 6.15             | 16.98                         |
| " 1969            | 0.085         | 0.640         | 24.32            | 3.48             | 27.80                         |

Metalliteollisuus

estimoitava suure

|                   |       |       |       |        |        |
|-------------------|-------|-------|-------|--------|--------|
| investoinnit 1967 | 0.315 | 0.655 | 48.47 | 145.22 | 193.37 |
| " 1968            | 0.378 | 0.530 | 39.22 | 174.26 | 213.48 |
| " 1969            | 0.738 | 0.760 | 56.24 | 340.20 | 396.44 |

Muu tehdasteollisuus

estimoitava suure

|                   |       |       |        |        |        |
|-------------------|-------|-------|--------|--------|--------|
| investoinnit 1967 | 0.288 | 1.160 | 241.30 | 251.70 | 493.00 |
| " 1968            | 0.236 | 1.468 | 305.30 | 206.30 | 511.60 |
| " 1969            | 0.535 | 1.860 | 386.90 | 467.60 | 854.50 |

Liite 4a:  $V_h$ :n arvot optimaalista allokointia varten

Aineisto kevät 1968

Apumuuttuja  $x$  = työntekijöiden lukumäärä

RS = suhdemenetelmä (separate)

RC = " (combined)

Puu- ja paperiteollisuus

| estimoitava suure | osite 1 (RS) | osite 1 (RC) | osite 2 (RS) | osite 2 (RC) |
|-------------------|--------------|--------------|--------------|--------------|
| investoinnit 1967 | .0029        | .0030        | .2339        | .2311        |
| " 1968            | .0032        | .0034        | 1.0511       | 1.0623       |
| " 1969            | .0019        | .0021        | .1542        | .1487        |

Metalliteollisuus

| estimoitava suure | osite 1 (RS) | osite 1 (RC) | osite 2 (RS) | osite 2 (RC) |
|-------------------|--------------|--------------|--------------|--------------|
| investoinnit 1967 | .0449        | .0460        | .6290        | .6450        |
| " 1968            | .1363        | .1398        | 1.5894       | 1.5930       |
| " 1969            | .1890        | .1970        | 1.5336       | 1.5380       |

Muu tehdasteollisuus

| estimoitava suure | osite 1 (RS) | osite 1 (RC) | osite 2 (RS) | osite 2 (RC) |
|-------------------|--------------|--------------|--------------|--------------|
| investoinnit 1967 | .0172        | .0193        | 1.7000       | 1.7270       |
| " 1968            | .0230        | .0260        | .8610        | .9000        |
| " 1969            | .0107        | .0135        | .5085        | .5110        |

Liite 4b:  $N_h V'_h$ :n arvot otoksen koon estimoimista varten  
 käytössä suhdemenetelmä separate

Puu- ja paperiteollisuus

| estimoitava suure | $N_1 V'_1$ | $N_2 V'_2$ | $\sum_{h=1}^2 N_h V'_h$ |
|-------------------|------------|------------|-------------------------|
| investoinnit 1967 | .594       | 8.890      | 9.480                   |
| " 1968            | .658       | 39.900     | 40.560                  |
| " 1969            | .397       | 5.860      | 6.260                   |

Metalliteollisuus

| estimoitava suure | $N_1 V'_1$ | $N_2 V'_2$ | $\sum_{h=1}^2 N_h V'_h$ |
|-------------------|------------|------------|-------------------------|
| investoinnit 1967 | 20.70      | 46.55      | 67.25                   |
| " 1968            | 62.83      | 117.62     | 180.45                  |
| " 1969            | 8.71       | 113.48     | 122.19                  |

Muu tehdasteollisuus

| estimoitava suure | $N_1 V'_1$ | $N_2 V'_2$ | $\sum_{h=1}^2 N_h V'_h$ |
|-------------------|------------|------------|-------------------------|
| investoinnit 1967 | 15.00      | 353.60     | 368.60                  |
| " 1968            | 20.10      | 179.10     | 199.20                  |
| " 1969            | 105.80     | 9.40       | 115.20                  |

Liite 5: Optimaalinen allokointi, kun  $n_h$  suhteessa  $N_h \bar{x}_h$ :hon

Puu- ja paperiteollisuus

| osite | $\bar{x}_h$ | $N_h \bar{x}_h$ |
|-------|-------------|-----------------|
| 2     | 191.6       | 7281.0          |
| 1     | 41.0        | 8487.0          |
| summa |             | 15768.0         |

Metalliteollisuus

| osite | $\bar{x}_h$ | $N_h \bar{x}_h$ |
|-------|-------------|-----------------|
| 2     | 200.8       | 14859.0         |
| 1     | 43.4        | 20007.0         |
| summa |             | 34866.0         |

Muu tehdasteollisuus

| osite | $\bar{x}_h$ | $N_h \bar{x}_h$ |
|-------|-------------|-----------------|
| 2     | 210.6       | 43804.0         |
| 1     | 43.6        | 38106.0         |
| summa |             | 81910.0         |

20. RICHARD ALAND Sijoituspankkitoiminta Yhdysvalloissa - The Investment Banking Function in the United States. 1968. 31 s.
21. TIMO HELELÄ Työnseisaukset ja teolliset suhteet Suomessa vuosina 1919 - 1939. 1969. 341 s.
22. SIRKKA HÄMÄLÄINEN Kotitalouksien säästämiseen vaikuttavista psykologisista tekijöistä ja niiden mittaamismahdollisuuksista. 1969. 177 s.
23. HEIKKI KOSKENKYLÄ An Evaluation of the Predictive Value of the Investment Survey of the Bank of Finland Institute for Economic Research. 1969. 12 s.
24. HEIKKI KOSKENKYLÄ Suomen Pankin investointikyselyn otantaan liittyviä ongelmista. 1970. 71 s.

IVA5a 1970 16446.3

Suomen

Suomen pankin taloustieteellisen  
tutkimuslaitoksen julkaisuja ; D

Koskenkylä, Heikki

Suomen Pankin investointikyselyn  
otantaan liittyvistä

1996-02-28

**SUOMEN PANKIN  
KIRJASTO**