THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Series A7

COLLECTED SCIENTIFIC PAPERS OF

LEO TÖRNQVIST

Helsinki 1981

COLLECTED SCIENTIFIC PAPERS OF

LEO TÖRNQVIST

Helsinki 1981

ISSN 0356-7435
ISBN 951-9205-74-8

FOREWORD

Professor Leo Törnqvist has acted as scientific consultant to the Research Institute of the Finnish Economy since 1974. During that time he has made a great contribution to various research projects in the Institute. The Institute has very much benefited from Leo Törnqvist´s long and successful career as a teacher and as a researcher. This collection of articles, which is published in honour of his 70th birthday, gives us an opportunity to express our gratitude to him.

This book has been edited by Dr. Matti Virén. Kari Alho, Lic.Pol.Sc., Dr. Timo Teräsvirta, Dr. Pentti Vartia and Dr. Yrjö Vartia also participated in making the selection of articles included in this volume. Tuula Ratapalo typed the unpublished papers.

Helsinki, November 3, 1981

Tauno Ranta

PREFACE

Professor Leo Törnqvist has contributed greatly to statistics and economics both as a teacher and a researcher. As regards his research work, mention should perhaps be made above all of some of his studies on the theory of index numbers and on demand theory, but Leo Törnqvist has also made significant contributions to the decision-making science, econometrics and the population science. The articles included in the present selection will even in themselves provide a picture of the width of their author's scope.

Leo Törnqvist has played a most important part in the training of researchers in the fields of statistics and economics in Finland. All of the post-war post-graduate students of statistics and economics, and especially those of the 1950s and 1960s, have almost without exception been in contact with him. Though Leo Törnqvist has served as a professor of statistics, his contributions to economics have been of very great importance. Leo Törnqvist was the initiator of econometric research in Finland, and this is why numerous doctoral dissertations in economics have been written under his guidance.

It has been characteristic of Leo Törnqvist, as a researcher, that, rather than simply evaluating and marginally extending the scientific work done elsewhere, he has always endeavoured to develop new ideas. And, as a teacher, he has tried to induce his pupils also to assume a similar, creative attitude toward research.

Leo Törnqvist's activities have to a large extent been connected with two places: the University of Helsinki Institute of Statistics and the Institute of Buseiness Economics of the National Board for Post and Telegraphy. The latter institute was in fact founded, in 1949, on Leo Törnqvist's initiative. In the course of the years, numerous students of statistics and, also, economics have had an opportunity to do advanced -level research work at that institute. It should also be mentioned that Leo Törnqvist played a very active role in the founding of a similar research unit in the National Board for the State Railways.

The papers included in the present volume have been published, with a few exceptions, in various books and periodicals. They do not necessarily do full justice to Leo Törnqvist, since a large proportion of his written work consists of various unpublished research reports and notes. In selecting articles for the present book, however, the editors have limited themselves to seeking to offer a representative picture of Leo Törnqvist's work on the basis of his publications. For practical reasons, his published books have not

been included here. A complete bibliography appended to this book is meant to complete the picture conveyed here of the author's work.

Leo Törnqvist was born in Jeppo, a Swedish-language locality in Southern Ostrobotnia, on February 14, 1911. His parents - Anders Törnqvist, a master builder, and his wife Anna Finskas - had in all eight children: one daughter and seven sons, of whom one died while still young. In 1935 Leo Törnqvist married Maj-Lis Lindberg, and three children - Maj-Britt, Nils and Anna - were born to them. Leo Törnqvist passed the matriculation examination necessary for university-level studies at Åbo Samskola in 1930, after which he started his studies at the Åbo Academy. In 1933 he passed the Candidate of Philosophy examination, majoring in mathematics and minoring in physics, chemistry and economics. During the academic year 1935-36 he studied statistics and economics under Harald Cramér and Gunnar Myrdal at Stockholm University. In 1937 he became Doctor of Philosophy, after defending his dissertation on "Kriterien für die reellen algebraischen Zahlen, aritmetische Ketten und diophantische Approximationen" at the Åbo Academy. From 1937 to 1938 he was Instructor of Statistics at the Åbo Academy and in the years 1938 to 1950 he served as a mathematician traffic inspector at the Tariffs Section of the National Board for the State Railways. In 1943 he was appointed Lecturer of Economic Statistics at the University of Helsinki. From 1945 to 1950 he served as Professor of Statistics at the same university on a part-time or full-time

ad interim basis, and in 1950 he was appointed Professor of Statistics on a permanent basis. He occupied this post up until 1974, when he retired on a pension; in the years 1949-1977 he also served as Head of the Institute for Business Economics of the National Board for Post and Telegraphy. The great significance of this institute for research in the fields of statistics and economics was already referred to above. In the years 1952-1953 Leo Törnqvist was a guest researcher of the Cowles Commission in Chicago, and in 1960-1961 he served as an expert for the United Nations technical assistance in Jakarta, Indonesia. Since 1974 he has acted as scientific consultant to the Research Institute of the Finnish Economy.

Of Professor Leo Törnqvist's numerous other activities, his membership of the Advisory Board for Statistics in the years 1954-1974 should be mentioned here. Since 1956 Leo Törnqvist has been a member of the Societas Scientiarum Fennica and, since 1951, a fellow of the Econometric Society. In 1956 he was elected a member of the International Institute for Statistics. In 1961 he was made Commander of the Order of the Lion of Finland, and in 1971 he was conferred the honorary degree of Doctor of Business Economics by the Helsinki School of Economics and Business Administration. In 1978 he received an Honorary Doctorate in Political Science from the Åbo Academy.

As editors of this book we wish to thank the personnel of the Research Institute of the Finnish Economy, especially Inkeri Happonen and Tuula Ratapalo, for the help we have been given. It has been both a pleasant task and a priviledge to gather these articles of Leo Törnqvist for this publication in honour of his 70th birthday. Finally, we wish to thank the publishers of the following articles, notes and reviews for permission to reprint them here.

Helsinki, February 14, 1981

Matti Virén
Kari Alho
Timo Teräsvirta
Pentti Vartia
Yrjö Vartia

CONTENTS

I. Index Number Theory

Prememoria angående förändring av beräkningssättet för Finlands Banks konsumtionsprisindex [julkaisematon moniste, Suomen Pankki, 1935, 1-49] (A memorandum on changing the way of calculating the Bank of Finland's consumption price index; in Swedish) [unpublished, mimeographed; Bank of Finland, 1935, 1-59] ..... 15— 76

Levnadskostnadsindexerna i Finland och Sverige, deras tillförlitlighet och jämförbarhet (The Finnish and Swedish cost-of-living indices, their reliability and comparability; in Swedish) [*Ekonomiska Samfundets Tidskrift*, 1936, No. 37, 1-35] ..... 77—112

The Bank of Finland's Consumption Price Index [*Bank of Finland, Monthly Bulletin*, 1936, No. 10, 27-34] .... 113—120

Vilket är förhållandet mellan finska markens och svenska kronans köpkraft? (yhdessä Erik Törnqvistin kanssa) (What is the relationship between the purchasing powers of the Finnish markka and the Swedish krona?; in Swedish, together with Erik Törnqvist) [*Ekonomiska Samfundets Tidskrift*, 1937, No. 39, 1-39] ..... 121—160

Säsongfluktuationerna i Finlands Banks sedel- och myntcirkulation åren 1870-1937 (Seasonal fluctuations in the Bank of Finland's notes and coins in circulation in 1870-1937; in Swedish) [*Ekonomiska Samfundets Tidskrift*, 1939, No. 44, 1-36] ..... 161—196

II. Demand Theory

Herman Wold: Efterfrågan på jordbruksprodukter och dess känslighet för pris- och inkomstförändringar (kirjaarvostelu) (a book review, in Swedish) [*Ekonomisk Tidskrift*, 1941, No. 43, 216-225.] ..... 198—208

Wold, Herman and Juréen, Lars: Demand Analysis. A Study in Ekonometrics (a book review, in Swedish) [*Ekonomiska Samfundets Tidskrift*, 1953, No. 2, 209-214] ..... 209—214

III. Population Science

En befolkningsstatistisk jämförelse mellan Finland och Sverige (A population-statistical comparative analysis of Finland and Sweden; in Swedish) [*Ekonomiska Samfundets Tidskrift*, 1948, No. 3, 139-171] ..... 215—247

The Post-War Population Development of Finland Compared with Predictions Made after the War [*World Population Conference, 1965, Vol. III, Projections, Measurement of Population Trends*, UN, Department of Economic and Social Affairs, 1967, 44-46] ..... 249—252

Models for Non-Explosive World Population Development [Helsingin yliopiston Valtiotieteellisen tiedekunnan promootio-julkaisu, 1971, 11-17] ..... 253—259

IV. Decision-Making Theory

On the Economic Theory of Lottery-Gambles [*Skandinavisk Aktuarietidskrift*, 1945, 228-246] ..... 261—279

The Interplay between Decision Makers [*Management Sciences, Models and Techniques, Vol. I, Proceedings of the 6th International Meeting of the Institute of Management Science*, 1960, 323-327] . 281—285

Beslutvetenskapen och människans strävan att styra skeendet (The decision-making science and man's endevour to steer the course of events; in Swedish) [*Societas Scientiarum Fennica*, 1970, No. 6, 3-5] ..... 286—288

V. Econometrics

A Method for Calculating Changes in Regression Coefficients and Inverse Matrices Corresponding to Changes in the Set of Available Data [*Skandinavisk Aktuarietidskrift*, 1957, 219-226] ..... 289—296

Industrieproduktionens utveckling i Finland åren 1925-1956 (The course of industrial production in Finland in 1925-1956; in Swedish) [*Ekonomiska Samfundets Tidskrift*, 1958, No. 3, 150-163] ..... 298—311

VI. Other Applications of Statistics

On the Distribution Function for a Function of n Statistic Variables and the Central Limit Theorem in the Mathematical Theory of Probability [*Skandinavisk Aktuarietidskrift*, 1946, 206-229] ..... 312—336

Some New Principles for Solving Linear Programmin Problems, [*Bulletin de L'Institut International de Statistique, 30^e Session*, 335-356] ..... 337—358

On Distribution Functions for Quantities Related to Networks [*Bulletin de L'Institut International de Statistique, 32^e Session*, 1960, 609-612] ..... 359—362

The Theory of Replicated Systematic Cluster Sampling with Random Start [*Review of International Statistical Institute*, 1963, No. 1, 11-23] ..... 363—375

Minimaxing and Optimal Programming [*Essays in Mathematical Economics* (in Honor of Oskar Morgenstern), Princeton University Press, 1967, 105—111] ..... 377—383

VII. Production of Statistical Data

An Attempt to Analyze the Problem of an Economical Production of Statistical Data [*Nordisk Tidskrift for Teknisk Økonomi*, 1948, 265-274] ..... 385—394

How to Cope with the Urgent Need for Statistical Services in Developing Countries [A paper presented in United Nations Conference of the Applications of Science and Technology for the Benefit of the less Developed Areas, 1962, unpublished, mimeographed] ..... 395—401

About Different Kinds of Research Projects [*Scientists at Work* (Festschrift in honor of Herman Wold, edited by T. Dalenius, G. Karlsson and S. Malmquist), 1970, 236-246] ..... 402—412

Bibliography ..... 413—423

Promemoria angående förändring av

beräkningssättet för Finlands Banks konsumtionsprisindex

författad av

fil. mag. Leo Törnqvist*)

*) Editors´ note: this memorandum has been retyped when preparing the manuscript, Appendix I has been omitted.

Vid en kritisk granskning av beräkningssättet för Finlands Banks konsumtionsprisindex framträder en del teoretiska svagheter. Ifall man jämför indexen med dess svenska förebild Sveriges Riksbanks konsumtionsprisindex framgår, att den svenska indexen är betydligt mera komplicerad, men även mera förfinad. (se kap. I). Jag har betraktat det som min uppgift att konstruera en index, som är möjligast enkel, men som från teoretisk synpunkt skall kunna anses lika bra eller bättre än den svenska indexen. Den lösning jag kommit till framgår av följande sammanfattning av undersökningens resultat.

§ 1. Ett vägt logaritmiskt (= geometriskt) medeltal, beräknat enligt kedjemetoden, med i princip rörliga vikter, lämpar sig bäst för beräkningen av Finlands Banks konsumtionsprisindex. Vikterna skola ange huru många promille av konsumenternas sammanlagda veckoutgift, som faller på de olika varukategorierna. I praktiken behöver dock ej smärre förskjutningar förorsaka viktförändringar.

Formel: $\log P_{t_0}^{t''} = \log P_{t_0}^{t'} + \Sigma\, c_i(\log p_i^{t''} - \log p_i^{t'}):1000$

$P_{t_0}^{t''}$ = den sökta prisindexen

$P_{t_0}^{t'}$ = föregående veckas prisindex

c_i = varukategoriens andel av totalkonsumtionens värde under veckan uttryckt i promille

$p_i^{t''}$, $p_i^{t'}$ = varuprisen vid respektive tidpunkter t'' och t'

Σ = summatecknet, varvid additionen sker över alla i indexen ingående varuslag i.

Motiveringen är dels teoretisk dels praktisk. I kap. II härledes detta beräkningssätt utgående från Fishers indexkriterier. Emedan vägningskoefficienterna äro principiellt rörliga, underlättas eventuella reformer i varuurvalet, nya varor kunna införas i indexen och vägningskoefficienterna för "gamla" varor kunna ändras, utan att några omräkningar behöver äga rum. Behovet av vägningskoefficientförändringar är dock ganska litet, emedan för de flesta varor efterfrågan är nästan normalelastisk, d.v.s. så beskaffad, att varans relativa andel i totalutgiften förblir i det närmaste oförändrad även om varupriset förändras. Ifall denna egenskap vore allmängiltig, kunde man rent matematiskt bevisa att det logaritmiska medeltalet med konstanta vägningskoefficienter är den lämpligaste indexformeln. Säsongvaror med extremt starka prisförändringar sakna likväl denna egenskap. Vid de praktiska räkningarna göres en väsentlig arbetsbesparing på grund av divisionernas bortfallande, vid oförändrade pris från vecka till vecka bortfalla även motsvarande multiplikationer med vägningskoefficienterna.

§ 2. Säsongvaror med extremt starka och höftiga prisförändringar, såsom grönsaker och färsk fisk, är det bäst att helt bortlämna från indexen. Deras vikter överflyttas lämpligen på ärter, kalja, restaurangförtäring och salt fisk.

Motivering: Ifall man önskade medtaga dem i indexen, skulle detta medföra mycket besvär med justering av vägningskoefficienter och eliminering av säsongrörelser. Denna eliminering

vore förövrigt ganska vansklig på grund av att även väderleksförhållanden och rena tillfälligheter spela en så stor roll för dessa varors vidkommande. Om dessa faktorer kunde bortelimineras skulle de återstående prisförändringarna ej mera ha något märkbart inflytande på indexen. I praktiken nås samma resultat bekvämast på i § 2 föreslaget sätt. (Se även kap. III och bilagda kurvor)

§ 3. För potatisens vidkommande föreslås, att juli-augusti prisstegringen lämnas obeaktad. Sista juni-noeringen bebehålles under denna tid.

Denna prisstegring förorsakas till största delen av övergången från gammal till ny potatis, den är brant, men av kort varaktighet. I övrigt gäller samma motivering, som för § 2.

§ 4. Skolavgifterna bortlämnas från indexen. I stället intages skolböcker i bildningsgruppen med vikten 7.

Motivering: Skolavgifterna kunna icke anses som representative för någon varugrupp. Från pristeoretisk synpunkt utgöra de icke pris på en vara, skolundervisning, utan bidrag till bekostande av densamma. Största bidraget lämnas som bekant av staten. Skolböckerna kunna däremot anses representative för hela bokgruppen, som nu helt saknas. Ifall man önskade medtaga varan skolundervisning i indexen skulle lärarlönerna vara den lämpligaste representanterna för denna grupp. Bild-

ningsgruppens index har blivit fullständigt missvisande på grund av skolavgifternas starka stegring, 2,5 gånger, sedan 1/7 1931. (En stegring, som till största delen torde vara ett iakttagelsefel beroende på för få primäruppgifter. Se även kap. III.)

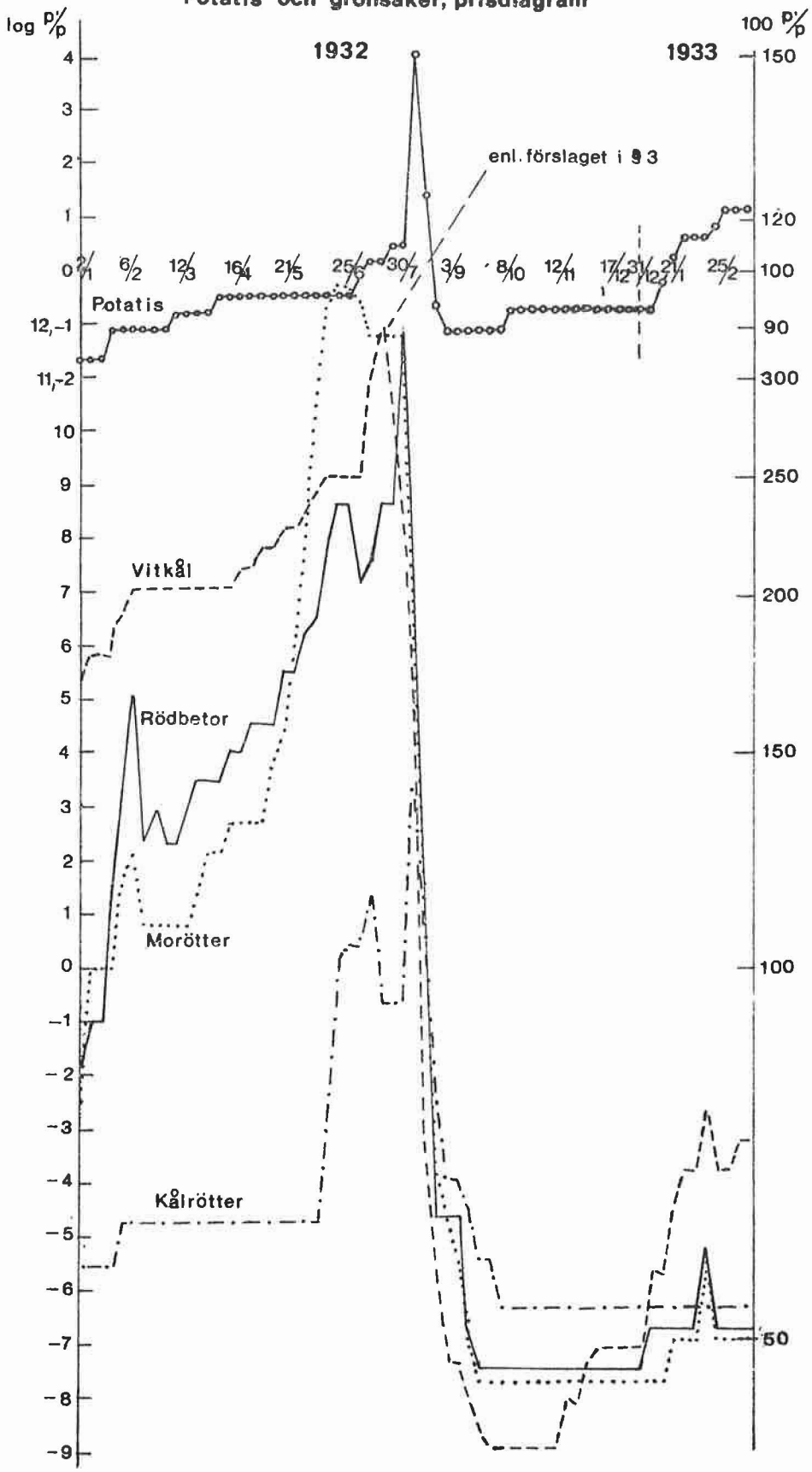
Jag har ej närmare undersökt huruvida vägningskoefficienterna i allmänhet kunna anses approximativt riktiga. Vid en jämförelse med Sveriges Riksbanks konsumtionsprisindex lägger man märke till flere markanta avvikelser, som tyda på de nu använda koefficienternas karaktär av vaga uppskattningar. En grundlig statistisk undersökning av konsumtionens verkliga sammansättning skulle behövas för vägningskoefficienternas noggrannare bestämning.

Beträffande en del prisnoteringar t.ex. hyror skulle måhända något flere primäruppgifter behövas för erhållande av en tillförlitlig bild av prisutvecklingen.

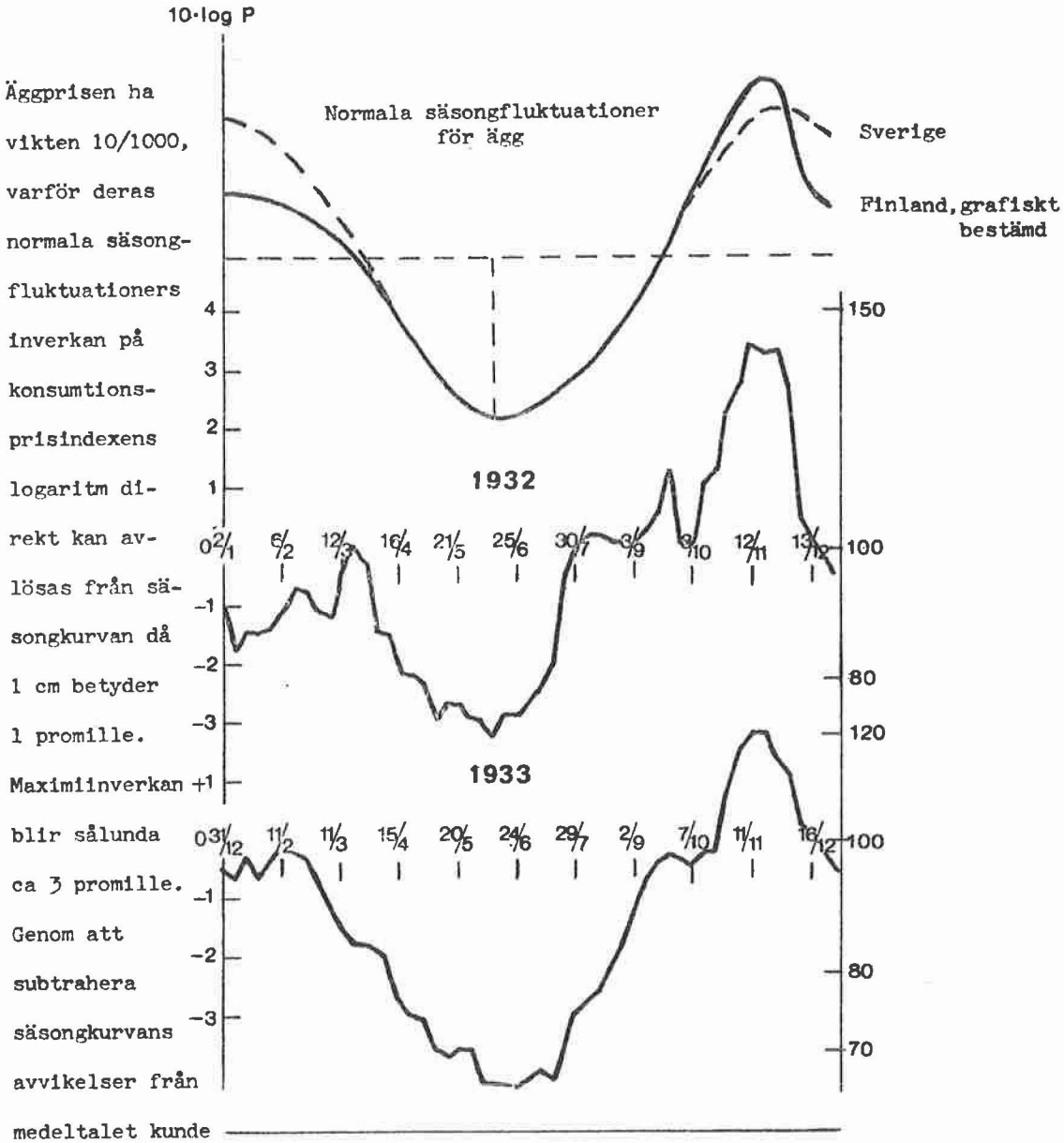
Det vore kanske skäl att medels stickprov (skenköpare) kontrollera riktigheten av primäruppgifterna. Förändring i rabatterna åt köparna måste även betraktas som prisförändringar.

Ett önskemål vore att även priser från landsorten skulle erhållas. Dessa uppgifter skulle kunna bearbetas skilt för sig och användas för kontrollering av huruvida Helsingforsprisens förändringar kunna anses representativa för hela landet om hänsyn tages till en eventuell eftersläpning för landsortens vidkommande.

Potatis och grönsaker, prisdiagram



Ägg, prisdiagram



1. BESKRIVNING OCH KRITIK AV DET NU ANVÄNDA BERÄKNINGSSÄTTET FÖR FINLANDS BANKS KONSUMTIONSPRISINDEX

Finlands Banks konsumtionsprisindex liknar mycket motsvarande index beräknad av Sveriges Riksbank. I Ekonomisk Tidskrift 1933 ingår en beskrivning av den svenska indexen författad av Erik Lindahl. Den beräknas enligt den s.k. kedjemetoden med justering av vägningskoefficienterna vid varje påvisbar förändring av desamma. Säsongrörelser och prisförändringar på grund av förändrad indirekt beskattning elimineras. Prisnoteringar hämtas från femton orter i Sverige. Varuurvalet är i stort sett detsamma som i Finlands Banks index. Den betydelsefullaste avvikelsen består i, att i den svenska indexen tjänsterna (läkararvode, husliga och offentliga tjänster) medtagits och ha en stor vikt.

Finlands Banks index beräknas som ett vägt, aritmetiskt medeltal av de enskilda varornas specialindexer. Som vägningskoefficienter användas, åtminstone i princip, respektive varugrupps andel i totalkonsumtionens värde under år 1931. I brist på en grundlig statistisk undersökning av konsumtionen ha indirekta uppskattningar över budgetanalyser och rena gissningar tillgripits. De i indexen ingående varorna äro i regel representanter för en större likartad varugrupp, varför de vikter den erhållit i indexen kunna vara något större än vad deras andel i totalkonsumtionen skulle förutsätta. Härvid har man gjort antagandet att varor tillhörande en sådan varugrupp approximativt förändras i pris på samma sätt

som den vara som representerar dem i indexen. Detta är i regel riktigt, men får denna representativa metod dock ej drivas allt för långt för att ej indexens felmarginal skall bli otillåtet stor. För en så inhomogen varugrupp som bostad (hyror) erfordras ett ganska stort antal kvaliteter och storlekar på lokaler för att hyresprisens förändringar skola bli riktigare iakttagna. En speciell hyresindex borde kanske beräknas, som skulle läggas till grund för de fortsatta räkningarna. Hyresförändringarna skulle i så fall säkerligen bli mindre diskontinuerliga (språngartade) än som nu är fallet. Samma sak torde gälla några andra varuslag.

Varje tisdag erhåller socialministeriets statistiska avdelning, som sköter om beräkningen av indexen, prisuppgifter från ett antal affärsmän i Helsingfors. Ifall flere olika uppgifter på samma vara erhålles tages ett enkelt aritmetiskt medeltal av dessa som pris på ifrågavarande vara. Dessa priser divideras därpå med ett utgångspris, som för de flesta varor är deras pris den 1/7 1931. För ett antal säsongvaror, nämligen potatis, ägg, grönsaker och färsk fisk, ha årsmedeltalet för hela år 1931 tagits till utgångspris. För mjölk, mejerismör och bondsmör har man valt ett något egendomligt utgångspris, nämligen årsmedeltalen för år 1931, minskade för mjölk med 2,1, mej. smör 2,6 och bondsmör 2,4 %.

De erhållna specialindexerna sammanvägas därpå till ett antal gruppindexer, vilka i sin tur sammanvägas till större gruppindexer och slutligen till den sökta konsumtionspris-

indexen. Genom detta förfarande underlättas i viss mån räkningarna samt en analys av prissystemets förändringar.

Förutom denna index med utgångspris år 1931 beräknas en lika byggd index, där föregående veckas pris valts till utgångspris. Den på detta sätt erhållna indexen visar ibland ganska märkliga avvikelser från det resultat man erhåller genom att dividera två på varandra följande veckoindextal. Det har till och med hänt att den ena visat ett prisfall, medan den andra visat prisstegring. Förklaringen till detta blottar den använda indexformelns teoretiska svagheter.

Dessa vecka till vecka-indextal ha ej multiplicerats till en index med fast bas. Detta förfaringssätt användes vid den s.k. kedjemetoden.

Redan på torsdagen samma vecka erhåller Finlands Bank ett sammandrag av resultaten. Indexen finnes beräknad sedan 2/1 1932.

Vad valet av utgångspris beträffar, förefaller det som om ändamålet med att välja årsmedeltalet under år 1931 till utgångspris för säsöngvarorna skulle ha varit att minska betydelsen av valet av den speciella bastidpunkten 1/7 1931.

Enär vägningskoefficienterna principiellt ange varornas relativa andel i varukonsumtionens värde under hela året 1931, är det motiverat att som utgångspris välja medeltalet under

hela året, ty härigenom finnes en viss sannolikhet för att säsongvarornas inverkan under året i medeltal skall stå i proportion till deras vikt.

Den formel,

$$P^{t}_{t_0} = \Sigma\, c_i \cdot \frac{p^{t}_{i}}{p^{t_0}_{i}} : \Sigma\, c_i ,$$

som ligger till grund för beräkningssättet, är blott en matematisk förklädnad av formeln

$$P^{t}_{t_0} = \frac{\Sigma\, q_i p^{t}_{i}}{\Sigma\, q_i p^{t_0}_{i}} , \text{ ty om man sätter } q_i = \frac{c_i}{p^{t}_{i0}}$$

erhålles den förra formeln. Av den senare formen framgår, att den ifrågavarande indexen även kunde beräknas så, att man uträknade priset för ett varuaggregat, där resp. vara inginge med kvantiteten q_i enheter. Emedan $q_i = \frac{c_i}{p^{t}_{i0}}$ framgår, att q_i är omvänt proportionellt med utgångspriset. Om priset 1/7 1931 var onormalt lågt för någon vara har denna vara erhållit en stor vikt q_i och tvärtom vid relativt höga utgångspris. Genom att välja $p^{t}_{i}0$ = årsmedeltalet motverkas detta i någon mån. Genom att såsom fallet varit med mjölk, mej.smör och bondsmör minska detta medelpris med vissa % ha deras vikter i själva verket ökats vilket knappast varit meningen.

Vid vecka till vecka-indexen, $P^{t''}_{t'} = \Sigma\, c_i \cdot \frac{p^{t''}_{i}}{p^{t'}_{i}}$, $(\Sigma\, c_i = 1)$

räknar man ej med samma varuaggregat utan med kvantiteterna $q_i' = \frac{c_i}{p_i^{t'}} = q_i \cdot \frac{p_i^{t0}}{p_i^{t'}}$, d.v.s. med kvantiteter förändrade i omvänd proportion till specialindexens ställning vid den nya tidpunkten t´. Då man dividerar två på varandra följande veckoindextal med varandra utföres räkneoperationen

$$P_{t_0}^{t''} : P_{t_0}^{t'} = \frac{\Sigma\, c_i \cdot \frac{p_i^{t''}}{p_i^{t0}}}{P_{t_0}^{t'}} = \frac{\Sigma\, c_i \cdot \frac{p_i^{t'}}{p_i^{t0}} \cdot \frac{p_i^{t''}}{p_i^{t'}}}{\Sigma\, c_i \cdot \frac{p_i^{t'}}{p_i^{t0}}} = \frac{c_i \cdot \frac{p_i^{t'}}{p_i^{t0}} \cdot \frac{p_i^{t''}}{p_i^{t'}}}{\Sigma\, c_i \cdot \frac{p_i^{t'}}{p_i^{t0}}}$$

$$\therefore P_{t_0}^{t''} : P_{t_0}^{t'} = \Sigma\, c_i' \cdot \frac{p_i^{t''}}{p_i^{t'}} \qquad c_i' = \frac{c_i \cdot \frac{p_i^{t'}}{p_i^{t0}}}{\Sigma\, c_i \cdot \frac{p_i^{t'}}{p_i^{t0}}} = c_i \cdot \frac{p_i^{t'}}{p_i^{t0} \cdot P_{t_0}^{t'}}$$

Av det erhållna uttrycket framgår, att kvoten kan betraktas såsom ett vägt medeltal av prisförändringarna från vecka till vecka, men vägningskoefficienterna ha ändrats i proportion till varornas relativa priser vid den nya jämförelsetidpunkten. Detta resultat behöver på den grund icke överensstämma med den genom direkt jämförelse beräknade vecka till veckaindexen $P_{t'}^{t''}$, vid vilken vägningskoefficienterna c_i bibehållas.

I verkligheten förändras de olika varornas relativa andel i totalkonsumtionens värde ej på detta sätt, ej ens approximativt.

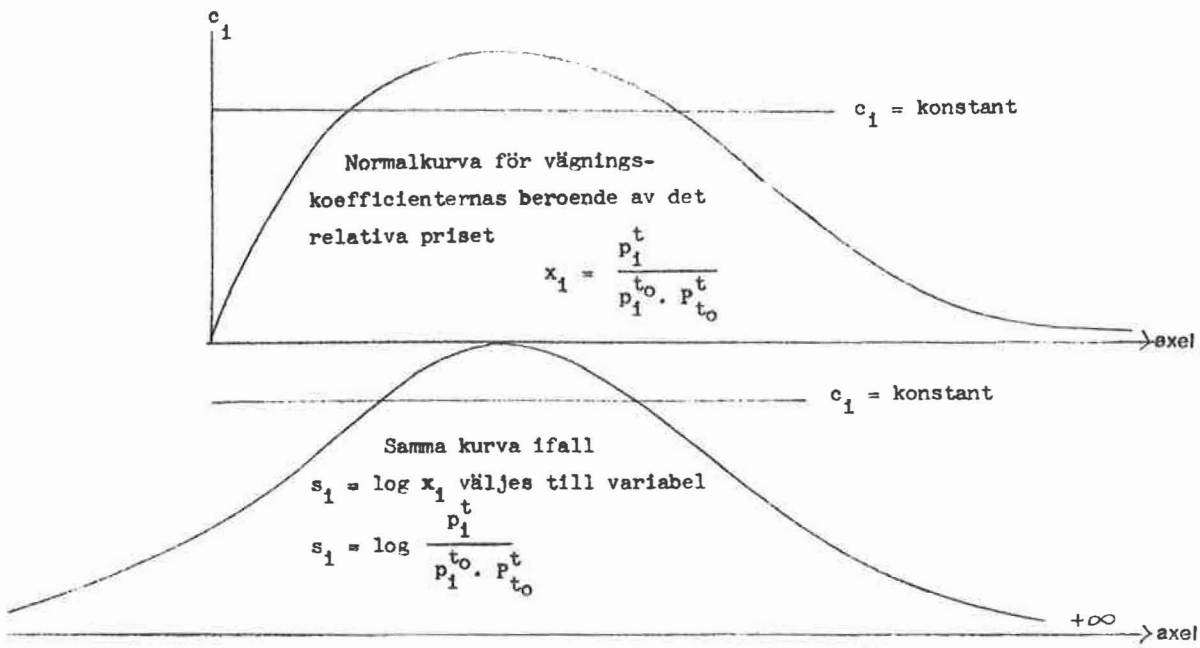
Ifall vi förutsätta, att konsumenternas medelrealinkomst förblir oförändrad, kan man vänta sig på grund av vad man vet om prisbildningslagarna, att denna andel skall variera med förändrat relativt pris $x_i = \frac{p_i^t}{p_i^{t_0} p^{t}_{t_0}}$ på ett sätt, som framgår av nedanstående kurva

Vid priset noll veta vi nämligen att konsumenternas utgifter för varan är noll, vid ett tillräckligt högt pris inträffar åter samma sak. Däremellan finnes ett maximum. Säsongvaror med extremt starka prisväxlingar såsom strömming och vitkål beskriva under ett år nästan hela denna kurva. I varje händelse förändras ej deras betydelse i konsumtionen proportionellt med priset x såsom de nuvarande beräkningssättet skulle förutsätta. Grafiskt åskådliggöres ett sådant antagande av en rät linje genom origo, vilken i regel mycket illa ansluter sig till kurvan.

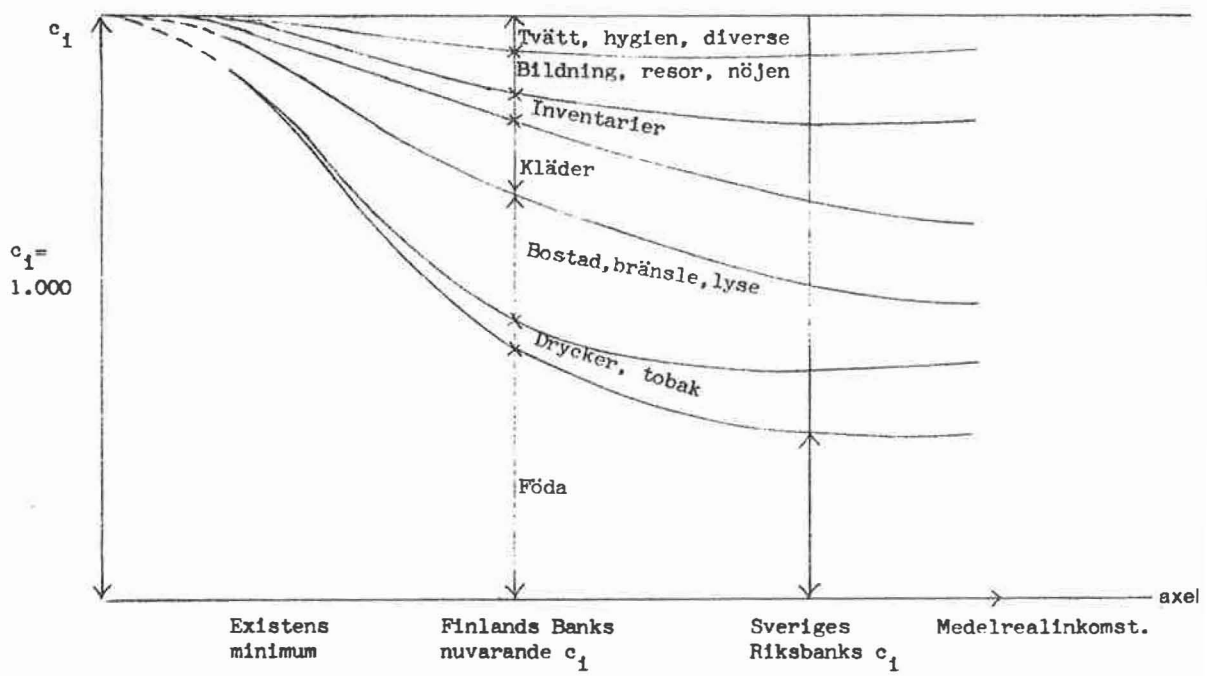
Förutom på grund av ändrat relativt pris x förändras en varas relativa andel i konsumtionens värde på grund av förändringar i medelrealinkomsten och på grund av förändringar i konsumenternas smak och levnadsförhållanden t.ex. temperatur-

växlingar, tekniska framsteg m.m. Av dessa faktorer äro förändringarna i medelrealinkomsten de betydelsefullaste. Den sekulära stegringen i medelrealinkomsten nödvändiggör förr eller senare vägningskoefficientjusteringar.

Vägningskoefficienterna



Vägningskoefficienternas beroende av medelrealinkomsten



2. MATEMATISK HÄRLEDNING AV DEN REKOMMENDERADE INDEXFORMELN

Vi betrakta en utgiftssumma $U(t) = \Sigma\, p_i^t\, q_i^t$ där p_i^t anger priset på en vara eller varukategori i och q_i^t försålda kvantiteter av denna vara till ifrågavarande pris under en kort tidsperiod i omgivningen av tidpunkten t. Vid en annan tidpunkt t_0 må motsvarande utgiftssumma var $U(t_0) = \Sigma\, p_i^{t_0}\, q_i^{t_0}$. Då p och q förändras språngvis anse vi likväl att förändringen tagit en viss ehuru godtyckligt liten tid i anspråk så att även de mellanliggande värdena antagits. De i verkligheten diskontinuerliga funktionerna p och q ersättas med i avseende å tiden kontinuerliga och deriverbara funktioner erhållna i enlighet med Lagrangé interpolationsförfarande. Indexproblemets matematiska sida består då i att finna två funktioner $P_{t_0}^t$ och $Q_{t_0}^t$ med följande egenskaper:

1) $P_{t_0}^t \cdot Q_{t_0}^t = \frac{U(t)}{U(t_0)}$

2) $P_{t_0}^t$ och $Q_{t_0}^t$ skola vara funktioner av p_i och q_i så beskaffade, att $P_{t_0}^t$ övergår i $Q_{t_0}^t$, ifall p och q byta plats. (Symmetrikriteriet)

3) $P_{t_0}^{t''} = P_{t_0}^{t'} \cdot P_{t'}^{t''}$, $Q_{t_0}^{t''} = Q_{t_0}^{t'} \cdot Q_{t'}^{t''}$. (Interkalationskriteriet)

4) Ifall $p_i^{t''} = k \cdot p_i^{t'}$ för varje i, så bör även

$$P_{t_0}^{t''} = k \cdot P_{t_0}^{t'}$$

Ifall $q_i^{t''} = k \cdot q_i^{t'}$ för varje i, så bör även

$$Q_{t_0}^{t''} = k \cdot Q_{t_0}^{t'}$$ (Proportionalitetskriteriet)

5) $P_{t_0}^{t} \neq 0$ såframt $U(t) \neq 0$, vilket i det följande förutsättes

$Q_{t_0}^{t} \neq$ " $U(t) \neq 0$, " " " "

Proportionalitetskriteriet ger som specialfall för $k = 1$ identitetskriteriet, och sammanställt med interkalationskriteriet (3) det s.k. inversionskriteriet: $P_{t_0}^{t} \cdot P_{t}^{t0} = 1$, $Q_{t_0}^{t} \cdot Q_{t_0}^{t} = 1$.

Indexproblemet formulerat på detta sätt är olösligt. Irving Fisher har gjort en vidlyftig undersökning av indexproblemet ("The making of indexnumbers"), varvid han studerat ett otal indexformler. Han uppställer ovanstående indexkriterier och påvisar, att ingen av de undersökta formlerna uppfyller dem alla på en gång. Det man kan eftersträva är, att så många som möjligt av dessa kriterier skola uppfyllas, och att felen i förhållande till de övriga skola vara möjligast små. $P_{t_0}^{t}$ kunna vi kalla prisindexen, $Q_{t_0}^{t}$ kvantitetsindexen.

En god inblick i problemets rätta natur får man genom att angripa det med den högre matematikens tillhjälp.

Vi logaritmera relationen (1) $P^t_{t_0} \cdot Q^t_{t_0} = \frac{U(t)}{U(t_0)}$ och få

$$\log P^t_{t_0} + \log Q^t_{t_0} = \log U(t) - \log U(t_0) =$$

$$\log \Sigma\, p_i^t q_i^t - \log \Sigma\, p_i^{t_0} q_i^{t_0}$$

I det vi bilda totala differentialen för vardera membrum, erhålla vi då $\log U(t_0)$ är en konstant

$$\frac{dP^t_{t_0}}{P^t_{t_0}} + \frac{dQ^t_{t_0}}{Q^t_{t_0}} = \frac{\Sigma\, q_i^t\, dp_i^t}{\Sigma\, p_i^t\, q_i^t} + \frac{\Sigma\, p_i^t\, dq_i^t}{\Sigma\, p_i^t\, q_i^t}$$

eller förlängt och täljarens termer dividerade med $\Sigma\, p_i^t\, q_i^t = U(t)$

$$\frac{dP^t_{t_0}}{P^t_{t_0}} + \frac{dQ^t_{t_0}}{Q^t_{t_0}} = \Sigma\, \frac{p_i^t q_i^t}{U(t)} \cdot \frac{dp_i^t}{p_i^t} + \Sigma\, \frac{p_i^t q_i^t}{U(t)} \cdot \frac{dq_i^t}{g_i^t}$$

På grund av symmetrikriteriet sluta vi oss till, att

$$\frac{dP^t_{t_0}}{P^t_{t_0}} = \Sigma\, \frac{p_i^t q_i^t}{U(t)} \cdot \frac{dp_i^t}{p_i^t} = \Sigma\, c_i(t) \cdot \frac{dp_i^t}{p_i^t}, \text{ då } c_i(t) = \frac{p_i^t q_i^t}{U(t)}$$

$$\frac{dQ^t_{t_0}}{Q^t_{t_0}} = \Sigma\, \frac{p_i^t q_i^t}{U(t)} \cdot \frac{dq_i^t}{q_i^t} = \Sigma\, c_i(t) \cdot \frac{dq_i^t}{q_i^t}, \text{ " " "}$$

Genom integration erhålles

$$\log P_{t_0}^{t'} = \int_{t_0}^{t'} \Sigma\, c_i(t) \cdot \frac{dp_i^t}{p_i^t} = \Sigma \int_{t_0}^{t'} c_i(t) \cdot \frac{dp_i^t}{p_i^t}$$

$$\log Q_{t_0}^{t'} = \int_{t_0}^{t'} \Sigma\, c_i(t) \cdot \frac{dq_i^t}{q_i^t} = \Sigma \int_{t_0}^{t'} c_i(t) \cdot \frac{dq_i^t}{q_i^t}$$

I analogi härmed få vi

$$\log P_{t'}^{t''} = \Sigma \int_{t'}^{t''} c_i(t) \cdot \frac{dp_i^t}{p_i^t}, \log Q_{t'}^{t''} = \Sigma \int_{t'}^{t''} c_i(t) \cdot \frac{dq_i^t}{q_i^t}$$

Enligt den bestämda integralens egenskaper är $\int_{t_0}^{t''} = \int_{t_0}^{t'} + \int_{t'}^{t''}$

och således gäller $\log P_{t_0}^{t''} = \log P_{t_0}^{t'} + \log P_{t'}^{t''} = \log P_{t_0}^{t'} \cdot P_{t'}^{t''}$.
Den erhållna indexformeln uppfyller således interkalationskriteriet under den gjorda approximationen att p och q äro kontinuerliga och deriverbara funktioner av t.

Gör man vissa speciella förutsättningar om de "ideala" vägningskoefficienterna $c_i(t)$ kunna integrationerna explicit utföras. Det enklaste antagandet är härvid, att $c_i(t) = c_i =$ en konstant för varje t. I så fall få vi omedelbart

$$\log P_{t_0}^{t'} = \Sigma\, c_i \cdot (\log p_i^{t'} - \log p_i^{t_0}) = \Sigma\, c_i \cdot \log \frac{p_i^{t'}}{p_i^{t_0}}$$

och således

$$P_{t_0}^{t'} = \Pi \left(\frac{p_i^{t'}}{p_i^{t_0}} \right)^{c_i} \qquad (\Pi = \text{produkten över alla i:n})$$

d.v.s. det vägda geometriska medeltalet. På grund av symmetriregeln är likaså

$$Q_{t_0}^{t'} = \Pi \left(\frac{q_i^{t'}}{q_i^{t_0}} \right)^{c_i}$$

Under denna förutsättning $c_i(t) = c_i$ för varje t uppfylles även proportionalitetskriteriet, ifall vi sätta $p_i^{t''} = k \cdot p_i^{t'}$ få vi nämligen

$$P_{t_0}^{t''} = \Pi \left(\frac{k \cdot p_i^t}{p_{i0}^t} \right)^{c_i} = k^{\Sigma c_i} \Pi \left(\frac{p_i^{t'}}{p_{i0}^t} \right)^{c_i} = k \cdot P_{t_0}^{t'} \cdot (\Sigma\, c_i = 1)$$

Enligt symmetriregeln gäller proportionalitetskriteriet även för kvantitetsindexen $Q_{t_0}^t$.

Att någon varas pris p_i^t vore = 0 samtidigt som $c_i(t)$ vore konstant och $\neq 0$ är en absurbidet, varför nollkriteriet saknar betydelse under den givna förutsättningen. Integralformeln däremot uppfyller nollkriteriet, ty

$$\log P_{t_0}^{t'} = \Sigma \int_{t_0}^{t'} c_i(t) \cdot \frac{dp_i^t}{p_i^t} = \Sigma \int_{t_0}^{t'} \frac{p_i^t q_i^t}{U(t)} \cdot \frac{dp_i^t}{p_i^t} =$$

$$\Sigma \int_{t_0}^{t'} \frac{q_i^t dp_i^t}{U(t)}$$

kan icke antaga värdet $-\infty$, emedan nämnaren $U(t) = \Sigma\, p_i^t q_i^t$ aldrig värdet 0 och integranden sålunda alltid är ändlig.

Om $c_i(t)$ ej är en konstant så gäller enligt integralkalkylens medelvärdssats, förutsatt att dp/pdt = D log p ej byter tecken

$$\log P_{t_0}^{t} = \Sigma \int_{t_0}^{t'} c_i(t) \cdot \frac{dp_i^t}{p_i^t} = \Sigma \bar{c}_i \cdot \int_{t_0}^{t'} \frac{dp_i^t}{p_i^t} = \Sigma \bar{c}_i \cdot \log \frac{p_i^{t'}}{p_i^{t_0}}$$

då $\bar{c}_i$ = ett medeltal av $c_i(t)$ under tidsperioden (t_0, t'). Ett gott närmevärde erhålles, om $\bar{c}_i$ väljes lika med varans andel i totalutgiften under hela perioden (t_0, t'). En ännu bättren approximation erhålles ifall man uppdelar räkningen i etapper och i enlighet med interkalationkriteriet beräknar indexen enligt kedjemetoden. Vi få

$$\log P_{t_0}^{t''} = \log P_{t_0}^{t'} + \Sigma \int_{t'}^{t''} c_i(t) \cdot \frac{dp_i^t}{p_i^t} \text{ och slutligen}$$

$$\log P_{t_0}^{t''} = \log P_{t_0}^{t'} + \Sigma c_i (\log p_i^{t''} - \log p_i^{t'}) ,$$

varvid medelvärdet c_i för $c_i(t)$ under perioden (t', t'') (= en vecka) i praktiken väljes så gott det sig göra låter = varan i:s andel i totalkonsumtionen under sagda period.

Enligt symmetriregeln fås motsvarande formel för kvantitetsindexen $Q_{t_0}^{t''}$.

Härmed ha vi härlett den rekommenderade indexformeln. Den indexformel, som nu användes för Finlands Banks konsumtionsprisindex, erhålles genom antagandet $q_i^t = k \cdot q_i$ för

varje i, d.v.s. $c_i(t) = \frac{q_i p_i^t}{\Sigma\, q_i p_i^t}$. Även i detta fall kan integrationen utföras och man får

$$p_{t_0}^{t} = \frac{\Sigma\, q_i p_i^{t'}}{\Sigma\, q_i p_i^{t_0}} = \Sigma\, \frac{q_i p_i^{t_0}}{\Sigma\, q_i p_i^{t_0}} \cdot \frac{p_i^{t'}}{p_i^{t_0}} = \Sigma\, c_i(t_0) \cdot \frac{p_i^{t'}}{p_i^{t_0}},$$

vilket kan kontrolleras genom logaritmisk derivering.

I det man då q_i^t i själva verket ej följer antagandet $q_i^t = k \cdot q_i$ väljer q_i proportionellt med ett medeltal av q_i^t för perioden (t_0, t') av ett eller annat slag, erhålles en hel klass av indexformler. I Sveriges Riksbanks konsumtionsprisindex användes principiel medeltalet $q_i = \frac{q_i^{t'} + q_i^{t_0}}{2}$, varjämte kedjemetoden användes enligt formeln

$$p_{t_0}^{t''} = p_{t_0}^{t'} \cdot p_{t'}^{t''}.$$

Härigenom erhålles en mycket god approximation till integralformeln. Det grundläggande antagandet $q_i^t = \frac{q_i^{t'} + q_i^{t_0}}{2}$ är dock en något sämre approximation av verkligheten än antagandet $c_i(t) = c_i$ för varje t, varvid c_i anger varan i:s andel i totalkonsumtionen under den ifrågavarande perioden. Se kurvan i kap. 1. c_i = konstant är en parallell rät linje till x-axeln, som i regel bättre ansluter sig till kurvan än en rät linje genom origo, som grafiskt åskådliggör det svenska antagandet.

1. Fishers "ideala" indexformel

$$p_{t_0}^{t'} = \frac{\Sigma\, q_i^{t_0} p_i^{t'}}{\Sigma\, q_i^{t_0} p_i^{t_0}} \cdot \frac{\Sigma\, q_i^{t'} p_i^{t'}}{\Sigma\, q_i^{t'} p^{t_0}}$$

erhålles, ifall man antager, att

$$c_i(t) = \frac{1}{2} \cdot \frac{q_i^{t_0} p_i^{t}}{\Sigma\, q_i^{t_0} p_i^{t}} + \frac{q_i^{t'} p_i^{t}}{\Sigma\, q_i^{t'} p_i^{t}} \quad .$$

Detta antagande är ej mera verklighetstroget än de tidigare. I praktiken saknar man vanligen kännedom om kvantiteterna $q_i^{t'}$.

Alla Fishers många indexformler kunna betraktas som specialfall eller approximationer till integralformel. Denna integrals värde är beroende av integrationsvägen fastställd av de närmare antagandena om vägningskoefficienterna $c_i(t)$. Närmast "sanningen" träffa de indexformler, vid vilka vägningskoefficienterna med minsta möjliga fel approximera de "ideala" vägningskoefficienterna $c_i(t)$:s verkliga förändring.

I kap. 3 skola vi diskutera den felmarginal, som finnes vid alla indexberäkningar. Felmarginalens storlek beror på valet av indexformel, felet i vägningskoefficienterna inbegripet varuvalets bristande representativitet, primäruppgifternas otillförlitlighet och eventuella räknefel.

3. FELMARGINALEN VID INDEXBERÄKNINGAR

Allmän teori

Vid praktiska indexberäkningar måste man använda sig av mer eller mindre goda närmevärden för de storheter, som ingå i en teoretisk indexformel. I det följande skola vi undersöka i vad mån detta påverkar slutresultatet.

$$\bar{c}_i, \bar{p}_i^t, \bar{p}_i^{t_0}, \log \bar{P}_{t_0}^{t'} = \Sigma \bar{c}_i \cdot (\log \bar{p}_i^{t'} - \log \bar{p}_i^{t_0}),$$

$$\Sigma \bar{c}_i = 1$$

må vara de "ideala", exakta storheterna och

$$c_i, p_i^{t'}, p_i^{t_0}, \log P_{t_0}^{t'} = \Sigma c_i \cdot (\log p_i^{t'} - \log p_i^{t_0}),$$

$$\Sigma c_i = 1$$

motsvarande närmevärden[1].

Skillnaden

$$\Delta = \log P_{t_0}^{t'} - \log \bar{P}_{t_0}^{t} = \Sigma c_i \cdot (\log p_i^{t'} - \log p_i^{t_0}) -$$

$$\Sigma \bar{c}_i (\log \bar{p}_i^{t'} - \log \bar{p}_i^{t_0}$$

---

1) Det vägda logaritmiska medeltalet kan anses representera samtliga indexformler, ty låter man vägningskoefficienterna vara fritt föränderliga, kan man medels denna formel erhålla varje medeltal, som kan tänkas utgående från de givna special-indexerna, varför detta val av indexformel icke minskar allmängiltigheten av vår teori.

skola vi kalla indexfelet[1].

Vi införa beteckningarna

$$\Delta c_i = c_i - \bar{c}_i$$

$$\Delta_i = (\log p_i^{t'} - \log p_i^{t0}) - (\log \bar{p}_i^{t'} - \log \bar{p}_i^{t0})$$

och få

$$\Delta = \Sigma\ (\bar{c}_i + \Delta c_i)(\log p_i^{t'} - \log p_i^{t0}) -$$

$$\Sigma\ \bar{c}_i(\log p_i^{t'} - \log p_i^{t0} - \Delta_i)$$

$$= \Sigma\ \Delta c_i(\log p_i^{t'} - \log p_i^{t0}) + \Sigma\ \bar{c}_i\Delta_i$$

Emedan

$$\Sigma\ \Delta c_i = \Sigma\ (c_i - \bar{c}_i) = \Sigma\ c_i - \Sigma\ \bar{c}_i = 1 - 1 = 0$$

är även

$$\log P_{t_0}^{t'} \cdot \Sigma\ \Delta c_i = \Sigma\ \Delta c_i \log P_{t_0}^{t'} = 0$$

och då vi subtrahera denna nolla från Δ fås

---

1) Med log avse vi naturliga logaritmen med basen e = 2,71828..., som erhålles ur de Briggska genom multiplikation med 2,302585....

$$\Delta = \Sigma\, \Delta c_i (\log p_i^{t'} - \log p_i^{t0} - \log P_{t0}^{t'}) + \Sigma\, \bar{c}_i \Delta_i$$

I det vi ytterligare införa den förkortande beteckningen

$$s_i = \log p_i^{t'} - \log p_i^{t0} - \log P_{t0}^{t'} = \log \frac{p_i^{t'}}{p_i^{t0} P_{t0}^{t'}}$$

få vi

$$\Delta = \Sigma\, \Delta c_i s_i + \Sigma\, \bar{c}_i \Delta_i = \Sigma\, \bar{c}_i \frac{\Delta c_i}{\bar{c}_i} s_i + \Sigma\, \bar{c}_i \Delta_i$$

Formeln säger, att indexfelet utgör summan av ett med de "ideala" vägningskoefficienterna vägt medeltal av produkten av det relativa felet i vägningskoefficienterna ($\frac{\Delta c_i}{\bar{c}_i}$) och specialindexernas logaritms avvikelse från totalindexens ($= s_i$) och ett lika vägt medeltal av felen i de enskilda specialindexernas logaritmer.

"Korrelationsfelet"

Den första termen kan enligt korrelationsteorien skrivas i formen

$$\Sigma\, \Delta c_i s_i = r \cdot z \cdot s$$

då

$$z = \Sigma \frac{(\Delta c_i)^2}{c_i} \text{ och } s = \Sigma\, c_i s_i^2,$$

$$-1 = r$$

$$-1 \leq r = \Sigma \frac{\Delta c_i}{z\sqrt{c_i}} \cdot \frac{s_i\sqrt{c_i}}{s} \leq 1.$$

Vissa sannolikhetsskäl tala för att felen $(\Delta c_i)^2$ skola vara närmelsevis proportionella med c_i, varför just denna uppdelning av

$$\Delta c_i s_i \text{ i } \frac{\Delta c_i}{\sqrt{c_i}} \text{ och } \sqrt{c_i} \cdot s_i \text{ har gjorts.}$$

z = ett mått på noggrannheten hos vägningskoefficienterna c_i

s = ett mått på spridningen i prissystemet.

Korrelationskoefficienten r är ungefär = 0, ifall vägninskoefficientfelen Δc_i och s_i ej stå i något sådant beroende av varandra, att deras produkter $\Delta c_i \cdot s_i$ övervägande bli positiva eller negativa.

Ifall vi antaga, att integralformeln är den "ideala" indexformeln, och medelvärdena $\bar{c}_i$ bestämts så, att likheten

$$\log \bar{P}_{t_0}^{t'} = \Sigma \int_{t_0}^{t'} c_i(t) \frac{d\bar{p}_i^t}{\bar{p}_i^t} = \Sigma \bar{c}_i(\log \bar{p}_i^{t'} - \log \bar{p}_i^{t_0})$$

gäller, och c_i i motsvarande närmeformel approximerar varugruppen i:s andel i totalkonsumtionens värde under perioden (t_0, t'), är denna korrelation med all sannolikhet mycket liten. Emedan $c_i(t)$ är en funktion av s_i så beskaffad, att

$c_i(t)$ har ett maximum för ett visst värde s_i, (jämför kap. 1, kurvan), framstår det som lika troligt, att produkten $\Delta c_i \cdot s_i$ lika ofta skall vara positiv som negativ, och deras summa sålunda ett mycket litet tal[1].

För att belysa indexfelet från korrelationsräkningens synpunkt anföra vi ett på födaindexen den 16/6 1934 tillämpat exempel. Vi antaga a) r = 0,1, ett antagande som i de flesta fall torde ange den rätta storleksordningen. För beräkningen av

$$z = n \cdot \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\Delta c_i}{\sqrt{c_i}}\right)^2 = z'\sqrt{n},$$

(n = antalet varjslag, i födaindexen = 55) antaga vi, att felen sinsemellan äro ungefär lika stora och således vart och ett ungefär lika med

$$z' = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\Delta c_i}{\sqrt{c_i}}\right)^2$$

Δc_i alltså ungefär lika med $z'\sqrt{c_i}$. I det vi antaga felet i mjölkens vikt 77/444 vara av storleksordningen 5/444 fås

$$z' = \frac{1}{444} \cdot \frac{5}{\sqrt{\frac{77}{444}}} = \frac{5}{\sqrt{77 \cdot 444}} = 0{,}027,$$

---

1) Sannolika medelspridningen i $\Sigma \Delta c_i s_i$ är ungefär = $zs\sqrt{(1+r^2)/n}$, varför $|r|$ måste vara > $\sqrt{1/n-1}$ för att vara väsentligt $\neq 0$.

vilket antagande motsvarar ett medelfel i de minsta vikterna (1/444) på 0,58/444. ($\Delta c_i = 0{,}027 \cdot \sqrt{\frac{1}{444}}$). Detta ger för

$$z = 0{,}027 \cdot \sqrt{55}$$

värdet 0,200. Spridningen s i prissystemet sedan den 1/7 1931 kan omedelbart beräknas utgående från de givna primäruppgifterna. Man får s = 0,226, en relativt stor spridning; sålunda blir korrelationsfelet $r \cdot z \cdot s = 0{,}1 \cdot 0{,}200 \cdot 0{,}226 = 0{,}0045$ (= 4,5 promille).

Ifall b) r antages = 0,5, som anger en mycket stor korrelations, fås $r \cdot z \cdot s = 0{,}0226$ (22,8 promille). Ifall åter c) grönsakerna och den färska fisken bortlämnas i enlighet med § 2, minskar spridningen s till 0,143 och korrelationsfelet under de båda antagandena för r till 0,0029 respektive 0,0143.

Bortlämnas de extremasäsongvarorna minskar även z - vilket i exemplet icke beaktats - på grund av att felen i dessa varors vägningskoefficienter $\frac{\Delta c_i}{\sqrt{c_i}}$ äro de relativt största, beroende på, att s_i för dessa varors vidkommande fluktuerar våldsamt under året.

"Iakttagelsefelet".

Vi vilja så intressera oss för den andra termen i indexfelet $\Sigma\, \bar{c}_i \Delta_i$, "iakttagelsefelet".

Iakttagelsefelen i de enskilda specialindexernas logaritmer äro dels systematiska dels osystematiska och uppkomma från följande särskilda felkällor:

De systematiska felen. 1) De erhållna prisuppgifterna kunna vara systematiskt missvisande, på grund av att kundrabatterna förändrats över hela linjen, utan att denna faktiska prisförändring framträder i de erhållna prisuppgifterna. Enligt sakkännares utlåtande föregås varje prissänkning av ökning av köprabatterna och en prisstegring av en minskning av dessa rabatter. Smärre prisförändringar ersättas helt med rabattförändringar. Denna faktor är särskilt märkbar för kläder och inventarier. Denna faktors verkan på slutresultatet består i en tendens till eftersläpning och stabilisering av totalindexen jämfört med den ideala indexen $\bar{P}^{t}_{t_0}$.

2) I stället för prisförändringar är priset stabilt, men kvaliteten hos varorna förändras. Denna faktor är av speciell betydelse för restaurangmaten. (Jfr Colin Clark: The national income). Den verkar i samma riktning som den föregående.

De systematiska felen kunna minskas genom att noggrannt fastställa de varors kvalitetsegenskaper, på vilka prisuppgifter inhämtas samt genom stickprovskontroll av primäruppgifternas tillförlitlighet.

Det systematiska felets sannolika värde i Δ_i beteckna vi med v_i och $\Sigma\, \bar{c}_i v_i$ med v. De återstående felen $\Delta_i - v_i$ äro sinsemellan utan korrelation och kallas här osystematiska fel.

De osystematiska felen

Vi beteckna $\Delta_i - v_i$ med 0_i. Dessa 0_i uppkomma på grund av följande orsaker:

1) Antalet prisnoteringar för varje enskild till gruppen i hörande vara är litet, varför de erhållna prisnoteringarna ej äro tillräckligt reprsentativa för hela marknaden.

2) Gruppen i representeras ofta blott av en enda vara eller varukvalitet, som ansetts typisk för hela gruppen.

3) Valet av tidpunkt för prisnoteringens inhämtande kan ha påverkat noteringarna. Denna sista faktor spelar en ganska stor roll för priskänsliga marknadsvaror sådana som färsk fisk, vars priser variera mycket under veckans och dagens lopp.

4) Räknefel, misskrivningar, o.d., vilka fel kunna undvikas genom kontrollräkningar.

De osystematiska felens avsaknad av inbördes korrelation minskar betydligt deras inverkan på indexfelet. Enligt sannolikhetskalkylens principer blir det sannolika osystematiska felet

$$0 = \Sigma (\bar{c}_i)^2 \cdot 0_i^2$$

Detta uttryck kan även skrivas

$$0 = \Sigma (\bar{c}_i)^2 \cdot \frac{\Sigma (\bar{c}_i)^2 \cdot 0_i^2}{\Sigma (c_i)^2} = \Sigma (\bar{c}_i)^2 \cdot 0',$$

vilken form är lämplig vid uppskattningar, som grunda sig på antagandet, att felen 0_i äro ungefär lika stora för varje i. $\sqrt{\Sigma (\bar{c}_i)^2}$ ungefär lika med $\sqrt{\Sigma (c_i)^2}$ är för Finlands Banks index = 0,2.

Exempel: 0_i må vara = 0,01, d.v.s. primäruppgifterna vore i medeltal riktiga på 1 % när. Vi få då $0 = 0{,}2 \cdot 0{,}01 = 0{,}002$ (2 promille).

Men för det fall att man inhämtat primäruppgifter på de i indexen ingående varorna i proportion till deras vikt, en metod, som vi ville rekommendera, är enligt sannolikhetskalkylen $0_i \sqrt{c_i}$ ungefär konstant. För detta antagande lämpar det sig att skriva det sannolika felet

$$0 = \Sigma (\bar{c}_i)^2 \cdot 0_i^2 = \Sigma c_i (\bar{c}_i 0_i^2) = \Sigma \bar{c}_i \cdot u_i^2$$

Exempel: $0_i = 0{,}01$ för $\bar{c}_i = 1/1000$ och få $u_i = 0_i \sqrt{\bar{c}_i} = 0{,}000316$, vilket värde för u_i enligt ovanstående antagande gäller för varje i. Således blir $0 = u_i = 0{,}0003$. Genom att förfara på detta sätt minskas det osystematiska till en obetydlighet på 0,3 promille.

Enär $0^2 = \dfrac{\Sigma (\bar{c}_i)^2 \cdot 0_i^2}{\Sigma (\bar{c}_i)^2}$ är ett vägt medeltal med kvadraten på vägningskoefficienterna $\bar{c}_i$ som vikter, domineras det osystematiska felet av felen i specialindexerna för bostaden, som har den största vikten 180/1000, och mjölken, vikt 77/1000, varför det är skäl att försöka få dessa möjligast exakta.

Bland samtliga, specialindexer torde bildningsgruppindexen vara den mest missvisande på grund av de däri ingående skolavgifternas stora vikt (7/1000) och den tvivel underkastade, starka prisstegringen på 250 % hos dessa sedan 1/7 1931. Utgångspriset 200:- gällde knappast för majoriteten skolor i landet, knappast ens i Helsingfors, utan voro avgifterna den 1/7 1931 högre. Skolavgifterna äro icke representativa för varan skolundervisning. De utgöra nämligen ej hela betalningen för densamma. Vad deras vikt (7/1000) beträffar är den säkerligen oriktig. En grov kalkyl visar, att deras andel i konsumenternas sammanlagda årsutgifter knappast kan vara mer än 3 promille. Med 45,000 elever i lärdomsskolorna á 1000:- i året fås ett belopp på 45 miljoner mk. Nationalutgiften åter är ca 15,000 milj.mk. Skolavgifternas andel bleve på så sätt inalles ca 3 promille. Folkskolorna uppbära ej alls några skolavgifter.

Det aritmetiska medeltalet

Av intresse är det måhända även att kargöra sambandet mellan det vägda aritmetiska medeltalet och det med samma vikter c_i vägda logaritmiska medeltalet.

$\log A_{t_0}^{t'} = \log \Sigma c_i \frac{p_i^{t'}}{p_i^{t_0}}$ må beteckna logaritmen för det aritmetiska medeltalet,

$\log P_{t_0}^{t'} = \Sigma c_i \log \frac{p_i^{t'}}{p_i^{t_0}}$ det logaritmiska medeltalet.

Skillnaden

$$\Delta a = \log A_{t_0}^{t'} - \log P_{t_0}^{t'} = \log \Sigma\, c_i \frac{p_i^{t'}}{p_i^{t_0}} - \Sigma\, c_i \log \frac{p_i^{t'}}{p_i^{t_0}}$$

skola vi utveckla i serie enligt potenser av s_i.

Vi ha identiskt

$$\Delta a = \log \frac{A_{t_0}^{t'}}{P_{t_0}^{t'}} = \log \Sigma\, c_i \frac{p_i^{t'}}{p_i^{t_0} P_{t_0}^{t'}} = \log \Sigma\, c_i \cdot e^{\log \frac{p_i^{t'}}{p^{t_0} P_{t_0}^{t'}}}$$

$$= \log \Sigma\, c_i \cdot e^{s_i}$$

Enligt Taylors sats är

$$e^{s_i} = 1 + s_i + \frac{s_i^2}{2} + \frac{s_i^3}{6} + \frac{s_i^4}{24} + R_i\,,\ R_i = \frac{e^{\theta s_i} \cdot s_i^5}{120}\,,\ (0 < \theta < 1)$$

sålunda blir

$$\frac{A_{t_0}^{t'}}{P_{t_0}^{t'}} = \Sigma\, c_i e^{s_i} = \Sigma\, c_i + \Sigma\, c_i s_i + \frac{1}{2}\Sigma\, c_i s_i^2 + \frac{1}{6}\Sigma\, c_i s_i^3 +$$

$$\frac{1}{24}\Sigma\, c_i s_i^4 + R$$

$$R = \Sigma\, c_i R_i = \text{litet av femte ordningen}$$

Nu är $\Sigma\, c_i = 1$ och

$$\Sigma\, c_i s_i = \Sigma\, c_i \log \frac{p_i^{t'}}{p_i^{t_0} P_{t_0}^{t'}} = \Sigma\, c_i \log \frac{p_i^{t'}}{p_i^{t_0}} - \log P_{t_0}^{t'} = 0$$

och $\Sigma\, c_i s_i^2 = s^2$ (Se sid. 20).

Uttrycket reduceras härigenom till

$$\frac{A^{t'}_{t_0}}{P^{t'}_{t_0}} = 1 + \frac{1}{2}\, s^2 + \frac{1}{6}\, \Sigma\, c_i s_i^3 + \frac{1}{24}\, \Sigma\, c_i s_i^4 + R.$$

Ifall vi ännu kunna antaga, att talen s_i äro någorlunda symmetriskt (normalt) fördelade på positiva och negativa värden är det sannolika värdet av $\Sigma\, c_i s_i^3 = 0$ och $\Sigma\, c_i s_i^4 = 3s^4$. (Jfr. J.W. Lindeberg: Todennäköisyyslasku ja sen käytäntö tilastotieteessä, sid. 160).

I det vi hänföra små avvikelser härutinnan till resttermen R, fås

$$\frac{A^{t'}_{t_0}}{P^{t'}_{t_0}} = 1 + \frac{1}{2}\, s^2 + \frac{1}{8}\, s^4 + R'. \quad \left(R' = R + \frac{1}{6}\, \Sigma\, c_i s_i^3 + \frac{\Sigma\, c_i s_i^4 - 3 \cdot s^4}{24}\right)$$

Enligt serieutvecklingen för $\log (1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3(1+x)}$ fås

$$\Delta a = \log \frac{A^{t'}_{t_0}}{P^{t'}_{t_0}} = \log \left(1 + \left(\frac{1}{2}\, s^2 + \frac{1}{8}\, s^4 + R'\right)\right) =$$

$$= \frac{1}{2}\, s^2 + \frac{1}{8}\, s^4 + R' - \frac{1}{2}\left(\frac{1}{2}\, s^2 + \frac{1}{8}\, s^4 + R'\right)^2 + \frac{(\quad)^3}{3(1 + (\quad))}\,.$$

Då räkningarna utföras taga termerna av fjärde ordningen ut varandra och vi få

$$\Delta a = \frac{1}{2}\, s^2 + R'' = R' + \text{litet av sjätte ordningen)}$$

Vid mycket stark assymetri i fördelningen domineras R'' av den däri ingående termen $\frac{1}{6}\, \Sigma\, c_i s_i^3$. I regel kan dock R'' försummas, varför man med mycket stor noggrannhet kan sätta

$$\Delta a = \frac{1}{2} s^2 .$$

Vi ha härmed erhållit satsen: Skinnaden mellan det vägda aritmetiska medeltalets och det lika vägda logaritmiska medeltalet logaritmer är alltid positiv och nästan exakt = halva spridningen $s^2 = \Sigma\, c_i s_i^2$.

Exempel: För födaindexen 16/6 1934 ha vi beräknat $\log A_{t_0}^{t'}$, $\log P_{t_0}^{t'}$,

$\frac{1}{2} s^2$ och $\frac{1}{6} \Sigma\, c_i s_i^3$.

$\log A_{t_0}^{t'} = 0{,}0306$ $\qquad \frac{1}{2} s^2 = 0{,}0259$

$\log P_{t_0}^{t} = 0{,}0081$ $\qquad \frac{1}{6} \Sigma\, c_i s_i^3 = -0{,}0045$

$\log A_{t_0}^{t'} = \log P_{t_0}^{t'} = 0{,}0225$ $\qquad \frac{1}{2} s^2$ $\qquad \frac{1}{6} \Sigma\, c_i s_i^3 = 0{,}0214$

Att termen $-\Sigma\, c_i s_i^3$ i detta exempel blivit så pass stor som 0,0045 (4,5 promille) förklaras av den färska fiskens och grönsakernas extremt stora värden s_i. Strömmingen dominerar de högre potenserna $\Sigma\, c_i s_i^n$. Den är även orsaken till, att den fjärde potensen i resttermen $\frac{\Sigma\, c_i s_i^4 - 3s^4}{24} = 0{,}0011$ skulle behöva medtagas för att Δa skulle bli riktig med fyra decimaler.

Ifall vi i enlighet med § 2 bortlämna den färska fisken och grönsakerna, minskar $\frac{1}{2} s^2$ till 0,0102 och $\log P_{t_0}^{t'}$ stiger till

0,0205 medan log $A_{t_0}^{t'}$ sjunker obetydligt till 0,0304, d.v.s. log $A_{t_0}^{t'}$ - log $P_{t_0}^{t'}$ blir = 0,099 eller ungefär = $\frac{1}{2} s^2$ = 0,0102. Assymmetrin har praktiskt taget försvunnit i det att $\frac{1}{6} \Sigma c_i s_i^3$ blir så litet som - 0,0003 (0,3 promille).

Indexformler med kumulativa fel

Det aritmetiska medeltalet är således enligt den ovanstående alltid större än det lika vägda logaritmiska medeltalet. Man kunde härav tro, att det aritmetiska medeltalet med växande avstånd från bastidpunkten skulle skilja sig mer och mer från det logaritmiska, men detta blir i regeln icke fallet, ty spridningen s, som vid bastidpunkten t_0 är lika med 0, (ifall utgångsprisen valts lika med ett medeltal för en längre tidsperiod, så att de ej samtidigt existerat vid en tidpunkt t_0, är denna spridning genast från början olika 0) växer visserligen i början hastigt med tiden, men blir snart en trögrörlig storhet, som ej visar någon tendens att växa. Carl Snyder, statistiker vid Federal Reserve Bank of New-York, har efter en vidlyftig undersökning sträckande sig över flere sekel kommit till slutsatsen, att några som helst sekulära förskjutningar i de relativa priserna, d.v.s. i storheterna s_i enligt vår terminologi, överhuvudtaget icke kunna påvisas.

Om man däremot skulle beräkna prisindexen genom multiplikation av successiva vecka till vecka-indextal

$$I_v^{v+1} = \Sigma c_i \frac{p_i^{v+1}}{p_i^v}, \quad I_0^n = \prod_0^{n-1} I_v^{v+1}$$

får den sålunda beräknade indexen I_0^n en tendens att växa ifrån det logaritmiska medeltalet med samma vägningskoefficienter P_0^n för att, ifall indexen skulle fortsättas tillräckligt länge, till slut tappa kontakten med såväl det direkt bildade logaritmiska medeltalet som det aritmetiska, ty för varje vecka är

$$\log I_v^{v+1} = \log P_v^{v+1} + \frac{1}{2} s_{v+1}^2 + R''_{v+1}$$

($\frac{1}{2} s_{v+1}^2 > 0$, så snart något pris överhuvud förändrats under veckan, R'' väsentligt mindre än $\frac{1}{2} s_{v+1}^2$). Genom addition över alla v:n från 0 till n-1 får man

$$\log I_0^n = \sum_0^{n-1} \log I_v^{v+1} = \sum_0^{n-1} \log P_v^{v+1} + \frac{1}{2} \sum_0^{n-1} s_{v+1}^2 + \sum_0^{n-1} R''_{v+1}$$

Sannolika värdet av $R_0^n = \sum_0^{n-1} R''_{v+1} = 0$.

Emedan det logaritmiska medeltalet med konstanta vikter uppfyller interkalationskriteriet (se kap. II) är

$$\sum_0^{n-1} \log P_v^{v-1} = \log P_0^n = \Sigma c_i \cdot \log \frac{p_i^n}{p_i^0}$$

Vi beteckna ännu

$$s^2(n) = \frac{1}{n} \cdot \sum_0^{n-1} s_{v+1}^2 ,$$

vilken storhet är i det närmaste en konstant, och få

$$\log I_0^n = \log P_0^n + \frac{1}{2} n \cdot s^2(n) + R_0^n$$

eller

$$\log I_0^n - \log P_0^n = \frac{1}{2} n \cdot s^2(n) + R_0^n .$$

Skillnaden växer således i det närmaste proportionellt med n och blir den till slut större än varje godtyckligt stort tal.

Ifall man således skulle försöka beräkna en prisindex I_0^n enligt kedjeförfarandet utgående från Finlands Banks nuvarande vecka till vecka-indextal skulle denna index ha en tendens att kumulativt växa över alla gränser, trots att ett direkt bildat aritmetiskt eller logaritmiskt medeltal ej alls skulle ha förändrat sig. Kedjeförfarandet tillämpat på aritmetiska medeltal av typen $\Sigma\, c_i \cdot \frac{p_i^t}{p_i^{t_0}}$ leder sålunda till absurda resultat. Ifall man såsom fallet är med Sveriges Riksbanks konsumtionsprisindex justerar vägningskoefficienterna, kan man med god tur motverka denna tendens, vilket skulle förutsätta, att termen $\frac{1}{2} s^2$ skulle kompenseras av termen $r \cdot z \cdot s$ i uttrycket för indexfelet genom att korrelationen r ständigt vore väsentligt negativ. Men med hänsyn tagen till att dessa justeringar i praktiken äro mycket sällsynta, kommer denna kumulativa faktor dock att få ett stort inflytande. Beräkningssättet för Sveriges Riksbanks konsumtionsprisindex besitter sålunda ett icke obetydligt teoretiskt fel. Vid de tidtals företagna kontrollräkningarna med fast bastidpunkt försöker

man få bort sådana kumulativa fel, utan att dock vara klart medveten om, att i själva beräkningssättet insmugit sig en kumulativt verkande felkälla.

För att belysa det kumulativa felets stora praktiska betydelse ha vi beräknat I_0^n utgående från Finlands Banks vecka till veckaindextal för föda-gruppen. Såsom första index I_0^1 har härvid valts $A_{1/7\ 1931}^{2/1\ 1932} = 112{,}1$, men har denna första tidsperiod bortlämnats vid beräkningen av halva medelspridningen vecka till vecka. Räkningarna ha utförts i naturliga logaritmer. (Se bilaga III.)

t	n	$\log I_0^n$	$\log A_0^n$	$\log P_n^n$	$\log \frac{N_0^n}{A_0^n}$	$\log \frac{I_0^n}{P_0^n}$	$\log \frac{A_0^n}{P_0^n}$
1/7-31	0	0,000	0,000	0,000	0,000	0,000	0,080
2/1-32	1	0,114	0,114	0,107	0,000	0,007	0,007
31/12-32	53	0,100	0,046	0,035	0,054	0,065	0,011
30/12-33	105	0,163	0,042	0,032	0,121	0,131	0,010
29/12-34	157	0,287	0,099	0,086	0,188	0,201	0,013
(16/6-34	129	0,159	0,031	0,008	0,128	0,151	0,023)

Halva medelspridningen vecka till vecka

$$\frac{1}{2}\, s(n)^2 = \frac{1}{n}\left(\log \frac{I_0^{v+n}}{P_0^{v+n}} - \log \frac{I_0^v}{P_0^v}\right)$$

var för år 1932

$$\frac{0{,}065 - 0{,}007}{52} = \frac{0{,}058}{52} = 0{,}00112.$$

År 1933 var motsvarande storhet

$$\frac{0,131 - 0,065}{52} = \frac{0,066}{52} = 0,00127$$

och år 1934

$$\frac{0,201 - 0,131}{52} = \frac{0,070}{52} = 0,00135.$$

Under hela treårsperioden var halva medelspridningen vecka till vecka 0,00124.

Finlands Banks vecka till vecka index (födagruppen) visar således i medeltal c:a 1,24 promille för högt.

Halva spridningen i förhållande till utgångsprisen (1/7-31) erhålles ur tabellen för $\Delta a = \log \frac{A_0^n}{P_0^n} \cdot \Delta a$ visar ej någon tendens att växa och är, frånsett 16/6-34, av storleksordningen 0,012 eller c:a tio gånger större än halva vecka till vecka spridningen. Den 2/1 var $\Delta a = 0,007$. Detta låga värde förklaras delvis av det ringa avståndet från bastidpunkten. Den 16/6 1934 var spridningen ovanligt stor, vilket torde vara typiskt för sommarmånaderna, i det att den färska fisken, grönsakerna och potatisen då ha extremt stora s_i^2 värden. Vecka till vecka spridningan var något mindre under första halvåret 1934 (= 0,0008) än under det senare halvåret, tydande på att prissystemet under det senare halvåret varit i livligare rörelse.

Ifall kedjeförfarandet tillämpas på logaritmiska medeltal och samma vägningskoefficienter användas hela tiden kommer man till precis samma resultat som genom direkt jämförelse med bastidpunkten. Men även om man förändrar vägningskoefficienterna i enlighet med § 1 finnes det ej något skäl att vänta sig kumulativa fel. Att här djupare ingå på denna fråga skulle dock föra för långt. Ett giltigt skäl att ändra vägningskoefficienterna, då de ej mera kunna anses som goda approximationer för respektiva varugrupps andel i totalkonsumtionens värde, utgör det faktum, att intresset för en riktig jämförelse av konsumtionsprisnivån mellan två nära nutiden belägna tidpunkter i regel är större än för en riktig jämförelse mellan nutiden och en långt avlägsen bastidpunkt. En index, i vilken logaritmiska medeltal och principiellt rörliga vikter användas, saknar bastidpunkt i vanlig bemärkelse och möjliggör samma precision vid beräknandet av prisnivåförändringen mellan tidpunkter, vilkas inbördes tidsavstånd äro lika stora, oberoende av avståndet från indexens begynnelsetidpunkt. Indexen blir med andra ord aldrig "omodern".

4. DETALJBESKRIVNING AV DET NYA BERÄKNINGSSÄTTET

Antalet primäruppgifter.

Såsom i kap. III framhållits domineras iakttagelsefelet av felen i de varupris, som ha de största vikterna, varför det är skäl att iakttaga dessa noggrannare än andra varupris. Ifall varugruppen är mycket heterogen såsom fallet är med bostadsgruppen, bör man uppdela den i flere undergrupper och skilja mellan olika varukvaliteter. Ett tiotal bostadstyper borde medtagas i indexen. Sedan man på detta sätt fått mindre och homogenare grupper fördelas dessa gruppers vikter ytterligare på de enskilda primäruppgifterna i enlighet med uppgiftsgivarens relativa andel i försäljningen av ifrågavarande vara. En grov skala vore härvid tillräcklig, mindre vikter än 1 torde knappast behövas för att iakttagelsefelet skall bli så litet att det kan försummas. Antalet prisuppgifter för en vara bleve sålunda i regel mindre än varans vikt. För varor, vilkas prisbildning domineras av ett fåtal storfirmor, skulle det enligt denna metod erfordras blott ett fåtal uppgifter, medan varor med mindre enhetlig prisbildning skulle fordra ett större antal uppgifter. För mjölk och mej.smör, vilkas prisbildning domineras av Valio, kunna tre eller fyra prisuppgifter vara tillräckliga medan bondsmöret, som har vikten 25, måhända skulle behöva att tiotal uppgifter för att ej iakttagelsefelet för denna varas vidkommande bleve relativt stort.

Beräkning av medelpriset.

Ifall olika prisuppgifter för samma vara erhållits beräknas medelpriset på följande sätt: Prisen multipliceras med de vikter, som erhållits, då varans vikt fördelats på de enskilda uppgifterna enligt uppgiftsgivarens relativa andel i försäljningen av varan. Produkternas summa, som kan erhållas direkt, ifall man ej tömmer räknemaskinen mellan de enskilda multiplikationerna, divideras därpå med dessa vikters summa. Den erhållna kvoten avkortas, om den första siffran är 1 eller 2 till ett fyrsiffrigt tal, i annat fall till ett tresiffrigt tal. Det så erhållna medelpriset antecknas i arbetstabellens första kolumn för ifrågavarande vecka.

Logaritmuppslagningen.

I arbetstabellens andra kolumn antecknas prisernas naturliga logaritmer med tre decimaler. Decimalkomman utsättas ej, siffrorna ange sålunda direkt tusendedelar (promille). De erhållas från en logaritmtabell uppställd så som av bilaga I framgår. Denna logaritmtabell, som av mig utarbetats utgående från en fyrställig tabell (J. Hoüel, Tables Numériques), har jag försökt göra speciellt lämpad för indexberäkningar. En treställig naturlig logaritmtabell är noggrann på en promille när, vilket med hänsyn till indexfelet är fullt tillräckligt. Naturliga logaritmer äro bekvämare än vanliga (Briggska) logaritmer för indexberäkningar, emedan man måste använda fyrställiga Briggska logaritmer för att erhålla en noggrannhet på en promille när.

De naturliga logaritmerna äro speciellt "naturliga" för indexberäkningar, på grund av att

$$\log \frac{p'}{p} = \frac{p' - p}{\bar{p}},$$

varvid $\bar{p}$ är ett medelvärde till p' och p, närmare bestämt är

$$\frac{p' + p}{2} > \bar{p} > \sqrt{p' \cdot p},$$

d.v.s. $\bar{p}$ ligger mellan det aritmetiska och det geometriska medeltalet till p' och p, och som exempel visa betydligt närmare det geometriska.

Exempel: $p' = 1{,}21$, $p = 1$. Vi ha då

$$\log \frac{1{,}21}{1} = \frac{0{,}21}{p} = 0{,}19062, \quad \therefore \bar{p} = \frac{0{,}21}{0{,}19062} = 1{,}1017,$$

$$\frac{0{,}21}{\frac{1 + 1{,}21}{2}} = \frac{0{,}21}{1{,}105} = 0{,}19005, \quad \frac{0{,}21}{\sqrt{1 \cdot 1{,}21}} = \frac{0{,}21}{1{,}1} = 0{,}19091,$$

således är $\frac{p' + p}{2} = 1{,}105 > \bar{p} = 1{,}1017 > \sqrt{p' \cdot p} = 1{,}100$ och $\bar{p}$ betydligt närmare det geometriska medeltalet. $\bar{p}$ kan även med mycket stor precision sättas

$$= \sqrt[3]{p' \cdot p \cdot \left(\frac{p' + p}{2}\right)}$$

Den naturliga logaritmen för en index $\frac{p'}{p}$ är ett bättre mått på relativa prisförändringen än angivandet av huru många procent p' är större eller mindre än p. Procenttalens asymmetri,

varmed vi förstå att en prisstegring kan anges av ett numeriskt godtyckligt stort tal medan ett prisfall endast kan bli hundra procent, föranleder en lätt att underskatta prisfallen jämfört med prisstegringarna, och gör procenttal mindre väl lämpade för angivande av relativa prisförändringar. Den naturliga logaritmen däremot är symmetrisk i avseende å talet 1, i det att log p´/p = - log p/p´. Ifall således ett pris först stiger med 100 % och därpå faller med 50 %, d.v.s. tillbaka till utgångspriset, så anges såväl prisstegringen som prisfallet vid räkning med naturliga logaritmer av numeriskt samma tal. Vi anse det ej vara något skäl att ange indextalen i form av procenttal, indextalens naturliga logaritmer äro åtminstone för vetenskapsmannen betydligt intressantare. Med kännedom om log p´/p, som vi beteckna med d, kan han lätt beräkna p´/p enligt serien $p'/p = 1 + \frac{d}{1} + \frac{d^2}{2} + (\frac{d^3}{2 \cdot 3} + \ldots)$ även utan tillgång till naturlig logaritmtabell. Ifall d är ett litet tal räcker det med de två första termerna. Ex. log 1,105 = 0,100... (se tabell I,2) $1{,}105 = 1 + 0{,}100 + \frac{1}{2}(0{,}100)^2$.

Ifall indextalen anges i naturliga logaritmer underlättas övergången till ny bastidpunkt, i det att log p´´/p - log p´/p = log p´´/p´. Man kan således använda det enkla räknesättet subtraktion för att få indexen förändrad till den bastidpunkt, som för tillfället intresserar en. Vid vetenskpaliga undersökningar över prissystemets förändringar och deras samband med andra ekonomiska företeelser äro indextalens naturliga logaritmer det bekvämaste utgångsmaterialet, ty de ekonomiska sammanhangen äro av logaritmisk natur, d.v.s. det är de

ekonomiska storheternas relativa förändringar m.a.o. logaritmer, och icke storheterna själva uppmätta i sinsemellan ojämförbara, mer eller mindre godtyckligt valda enheter, som ingå i funktionssammanhangen.

Räkneregler.

För användningen av vår specialkonstruerade logaritmtabell gäller följande regler:
Huvudregel. Ifall den första siffran i priset - man fäster sig ej alls vid decimalkomman - är större än eller lika med 7 uppsökes logaritmen på tabell I,1, i vilken tabell logaritmerna äro negativa. De två första siffrorna uppsökas i den första kolumnen, över vilken bokstaven N finnes antecknad. Därpå går man åt höger till dess man kommer till den kolumn, över vilken den tredje siffran i priset finnes antecknad. Det funna talet utgör den sökta logaritmen och antecknas i arbetstabellens andra kolumn. Ex. log 8:67 = -143, log -:85 = -163, log 70:38 Obs! avrundas till log 704 = -351. Ifall den första siffran är 1 och den andra mindre än fyra, uppsökes logaritmerna på tabell I, I denna tabell finner man de tre första siffrorna under N, den fjärde siffran anger åter i vilken av kolumnerna man skall söka logaritmen. Om de två första siffrorna ligga i intervallen 14 - 55 uppsökes logaritmen på tabell I,3. De två första siffrorna uppsökas under N den tredje anger åter kolumnen. Ifall man mot vanligheten skulle ha ett fyrsiffrigt tal, vilket enligt regeln för beräkning av medelpriset förutsätter, att den första siffran är mindre än 3,

kan man interpolera, d.v.s. till logaritmen för det tresiffriga tal, som erhålles då den fjärde siffran icke beaktas, addera det tal, som fås, då skillnaden mellan denna logaritm och det följande tresiffriga talets logaritm multipliceras med den fjärde siffran i talet, produkten divideras med tio och avkortas till en siffra. Vi ha konstaterat, att man interpolerar snabbare på dett sätt en genom att använda en interpolationstabell. Interpolationerna äro ytterst sällsynta och kunna helt undvikas, om man alltid avkortar pris, som icke uppsökas på tabell I,2 till tresiffriga tal. Logaritmerna för tal, vilkas två första siffror ligga i intervallen 55 - 70 uppsökas på tabell I,4.

Huvudregeln gäller utan undantag blott för utgångsprisen (1/7 - 31). Vid de fortsatta räkningarna måste man även beakta följande specialregler:

Sp.r. 1). Om priset på en vara sjunkit under utgångspriset, så att den första siffran blivit mindre än 7, uppsökes dess logaritm på tabell I, till dess priset sjunker så mycket, att den första siffran blir mindre än 6, vilken händelse torde vara synnerligen sällsynt. Ifall detta dock inträffar, uppsökes logaritmen på tabell I,4 eller I,3, men från den funna logaritmen subtraheras talet 1303, så att resultatet blir begativt. Ifall priset sjunker så otroligt kraftigt, att det tal, som bildas av de två första siffrorna blir mindre än 35, subtraheras 2303 från denna logaritm.

Ex. utg.pris 14:33, log 5:53 = 710 - 1303 = -593. Sp.r.2). Ifall ett pris, vars två första siffror vid utgångstidpunkten

bildade ett tal, som var mindre än 35, eller som tidigare uppsökts på tabell I,1, stigit i pris så att de två första siffrorna bilda ett tal större än 35, adderas 1000 till den logaritm, som man finner i tabellen, vilket innebär, att den i denna del av tabellen underförstådda första siffran 1 i detta fall utsättes. I motsatt fall, om priset sjunkit, så att det kritiska stället 35 passerats, subtraheras 1000 från den på vanligt sätt uppslagna logaritmen.

Ex. 1) Utg.pris 3:08 stigit till 3:83, log 3:83 = 343 + 1000 =
2) Utg.pris 36:50 sjunkit " 30:50, log 30:50 = 1115 - 1000 = 115

Denna specialregel skulle bortfalla, om man räknade med en logaritmtabell, där de efter det kritiska stället följande logaritmerna skulle vara 1000 större, men motvägas denna förenkling av, att i så fall en stor del av logaritmerna skulle bestå av fyrsiffriga tal, vilka sulle öka skrivvarbetet.

Sp. r. 3). Ifall priset stigit över utgångspriset, så att den första siffran blivit större än 7, uppsökes dess logaritm på tabell I,4. Blir den första siffran 9, i vilket fall den andra siffran anger kolumnen, får man medels interpolation det tresiffriga talets logaritm. Om priset stiger ytterligare, så att den första siffran ånyo blir mindre än 7 adderas talet 1303 till den enligt huvudregel uppslagna logaritmen. Sp.r. 2) gäller samtidigt och oberoende av sp. r. 3). Vid tiofaldig prisstegring stiger prisets logaritmer med 2303.

Ex. Utg. pris 55:00 stiger till 150:00, log 150:00 = 405 + 1303 = 1708

Utg. pris 25:00 stiger till 150:00, log 150:00 = 405 + 1000 + 1303 = 2708

I alla övriga fall gäller huvudregeln.

Dessa regler förefalla kanske invecklade, men möjligheterna för feluppslagning äro dock praktiskt taget obefintliga blott man ihågkommer minnesregeln:

Skillnaden mellan logaritmerna för två på varandra följande veckopris är nästan aldrig ett fyrsiffrigt tal. Fyrsiffrig blir den endast, om förändringen är mycket våldsam, d.v.s. vid ca en tredubbling av prisen eller ett prisfall till ca en tredjedel av föregående veckas pris. Sedan de extrema säsongvarorna bortlämnats från indexen inträffar detta ej alls, om icke en verkligt våldsam inflation eller deflation förekommer. Skillnaden är positiv, om priset stigit sedan föregående vecka och negativ om det sjunkit. Denna skillnad beräknas som det följande steget i räkningen och skrives i arbetstabellens tredje kolumn.

Vecka till vecka-differensen.

Ifall logaritmen för den senaste veckans pris har samma tecken som för den föregående veckans subtraherar man det mindre talet från det större. Om logaritmerna ha motsatt tecken adderas deras talvärden. Tecknet, som blir riktigt om man utför subtraktionen enligt algebrans teckenregler, erhålles lättast enligt ovanstående minnesregel. Emedan skillnaderna

i regel äro tvåsiffriga tal, utföres subtraktionen snabbast utan användning av räknemaskin. Då ingen prisförändring sedan föregående vecka ägt rum, blir skillnaden noll, varför det är praktiskt att bortlämna alla anteckningar för ifrågavarande vara. Jämför bilaga II, där kolumnen för vecka till vecka-differensen betecknats med d.

Uträkningen av indexens vecka till vecka-förändring.

De erhållna vecka till vecka-differenserna multipliceras med respektive varas vägningskoefficienter, varvid multiplikationsmaskinen ej tömmes mellan de enskilda multiplikationerna, är differensen negativ vevas bakåt, är den positiv framåt. Den i räkneverket befintliga summan, vilken, ifall den är negativ, fås till sitt talvärde genom att bestämma den addend, som gör summan lika med noll, antecknas för varje varugrupp, vars specialindex är av intresse. För lokalisering av eventuella räknefel kan ytterligare anteckningar över summan göras under räkningens gång. Maskinen tömmes ej förrän logaritmen för konsumtionsprisindexens vecka till vecka-förändring framträder tusen gånger förstorad i räkneverket. Logaritmerna för de enskilda gruppindexernas vecka till vecka-förändringar erhållas genom att dividera tillväxten för respektive varugrupp i den antecknade, löpande summan med gruppens vikt. Såväl tillväxten, som självfallet även kan vara negativ, som den erhållna kvoten antecknas i arbetstabellen. De erhållna kvoterna anger i regel direkt indexernas förändring i promille (se tabell I,1 el. I,2).

Den löpande indexen

Nu återstår blott att till föregående veckas indexlogaritm (algebraiskt) addera vecka till vecka-indexlogaritmen för att erhålla logaritmen för Finlands Banks konsumtionspris-index. Vid denna addition medtagas samtliga decimaler. På samma sätt förfares med specialindexerna. De tre sista decimalerna bortlämnas självfallet vid eventuell publicering, men böra medtagas vid räkningen för att kontrollen skall bli exakt. För specialindexernas vidkommande adderas de på varandra följande vecka till vecka-tillväxterna i bråkformen för undvikande av att avkortningarna vid divisionen med gruppens vikt skola åstadkomma att kontrollen misstämmer på de sista decimalerna.

Om man önskar angiva konsumtionsprisindexen såsom hittills i procentform, vilket vi icke rekommendera, erhålles denna procent genom att från tabell I,1 eller I,2 uppsöka det tal, som har indexlogaritmen till logaritm.

Kontrollen.

Genom multiplikation av prislogaritmerna (andra kol.) med vägningskoefficienterna på samma sätt som vid beräkningen av vecka till vecka-indexen och genom att från den erhållna slut-summan subtrahera en lika beräknad summa för en godtyckligt vald vecka sedan den senaste vägningskoefficientförändringen och jämföra den så erhållna skillnaden med tillväxten i den löpande indexens logaritm, erhålles en mycket god kontroll.

Siffrorna skola nämligen, ifall räkningarna utförts riktigt, exakt överensstämma. Denna metod att beräkna indexens förändring med överhoppande av mellanliggande veckor är den bekvämaste och säkraste metoden för uträkning av indexen, ifall man blott är intresserad av enstaka, från varandra långt avlägsna tidpunkter. Subtraktionerna bortfalla nämligen vid detta beräkningssätt.

Bilaga II.

Närmare detaljer över beräkningssättet framgår av bilaga II. Första kolumnen, betecknad med p, upptar utgångsprisen 1/7 1931 för födagruppen. Andra kolumnen, betecknad log p, innehåller prisens naturliga logaritmer bestämda enligt huvudregeln. 203141 är den för kontrollen enligt ovanstående beräknade summan av vägningskoefficienterna gånger prislogaritmerna. De närmast efterföljande kolumnerna hänföra sig till 2/1 1932. Vid bestämmande av log 3:83 = 1343 har specialregeln 2) kommit till användning. Summan 248341 har beräknats för kontrollens skull. Den tredje kolumnen för denna vecka, betecknad med e (ero), anger logaritmdifferenserna i förhållande till utgångsprisen. De motsvara för denna första vecka till vecka-differensen. 45200 utgör den algebraiska summan av vägningskoefficienterna c och talen e. Denna summa dividerad med 444 ger födaindexens logaritmtillväxt sedan 1/7-31 lika med log P = 102. Denna logaritms Numerus uppslagen från tabell I,2 är födaindextalet för 2/1-32, lika med 110,7. Den fjärde kolumnen c anger vägningskoefficienterna.

Den färska fiskens vikter ha i enlighet med § 2 överflyttats på salt fisk och fördelats så, att den salta strömmingens, sillens och anjovisens vikter fördubblats, d.v.s. ökat från 3 till 6. Grönsakernas vikter ha överflyttats på ärter (1 till 3), kalja (4 till 5), middag (9 till 10) och kaffe (9 till 10).

Den följande kolumnen e^2 ha vi använt för beräkning av s^2 enligt formeln (jfr J.W. Lindeberg, Todennäköisyyslasku ja sen käytäntö tilastotieteessä, sid. 14, formel 6)

$$s^2 = \Sigma\, c_i \cdot e_i^2 - (\log P)^2 .$$

Härvid ihågkommes, att logaritmerna ange tusendelar. I kolumnen e^2 ha vi avkortat kvadraterna på talen i kolumnen e, vilka skulle ange milliontedelar till tusendedelar, varför dessa kvadrater äro små tal. Kvadreringen utföres snabbast medels en tabell över kvadraterna på talen 1 - 300, avkortade så att de tre sista siffrorna bortfalla.

För den 2/1 1932 blir spridningen

$$s^2 = \frac{11540}{444} \cdot \frac{1}{1000} - \left(\frac{45200}{444}\right)^2 \cdot \frac{1}{1000000} = 0{,}016$$

I de tre följande kolumnerna, som hänföra sig till 9/1-32 ha inga anteckningar gjorts för oförändrade varupris. Kolumnen d anger vecka till vecka-differensen.

Kolumnen e anger såsom tidigare skillnaden mellan denna veckas prislogaritmer och utgångsprislogaritmerna. log P = $\frac{43341}{444}$ = 98 har erhållits så, att till den föregående veckans log P = $\frac{45200}{444}$ adderats (algebraiskt) vecka till veckaindexens logaritm = $\frac{-1859}{444}$ = -. Till samma resultat kommer man genom att såsom föregående vecka multiplicera talen e med vikterna c. Indexens procentsiffror fås genom Numerus uppslagning log 110,3 = 98, log 99,6 = -4.

Kontrollen, som det är onödigt att utföra mer än ett par gånger i året, framgår av exemplet tillämpat på priserna 2/1-32.

De två återstående kolumnerna s och s^2 uträknas vid en eventuell grundlig analys av prissystemets förändringar sedan utgångstidpunkten. Talen s utgöra skillnaden mellan e-kolumnens tal och log P = 98. Den följande kolumnen innehåller kvadraterna på dessa tal. Spridningen inom födaggruppen $s^2 = \Sigma\, c_i s_i^2$ är lika med 0,016, som väl överensstämmer med föregående veckas spridning, vilken beräknades enligt en annan metod. En sådan grundlig analys vore det måhända skäl att utföra en gång i vartalet.

Vägningskoefficientförändringarna.

Då det konstaterats, att de använda vikterna ej mera kunnat anses som goda närmevärden för ifrågavarande varugrupps andel i totalkonsumtionens värde, justeras vägningskoeffi-

cienterna. Detta göres helt enkelt så, att man efter justeringen räknar med de nya vikterna vid uträkningen av vecka till vecka-differensen. Om man vill få reda på viktjusteringarnas inverkan på indexen, kan förändringen i log P beräknas enligt formeln för indexfelet

$$\Delta = \Delta c_i e_i = \Sigma \Delta c_i s_i \text{ , då } \Delta_i = 0$$

Man konstaterar härvid lätt, att viktförändringarna skola vara ganska stora för att märkbart påverka indexen. Ifall antalet viktjusteringar sedan indexens början blir ett hundratal, kan man för upptäckande av eventuella, av slumpen åstadkomna korrelationsfel, beräkna en index direkt utgående från utgångsprislogaritmerna och den senaste veckans prislogaritmer, vilken väges med medeltalet av alla under den förflutna tiden använda vikter. Denna index bör aj avvika från den förra indexen med mer än spridningen s^2. Ifall större avvikelse upptäckes justeras indexen så att detta villkor uppfylles.

Bilaga II.

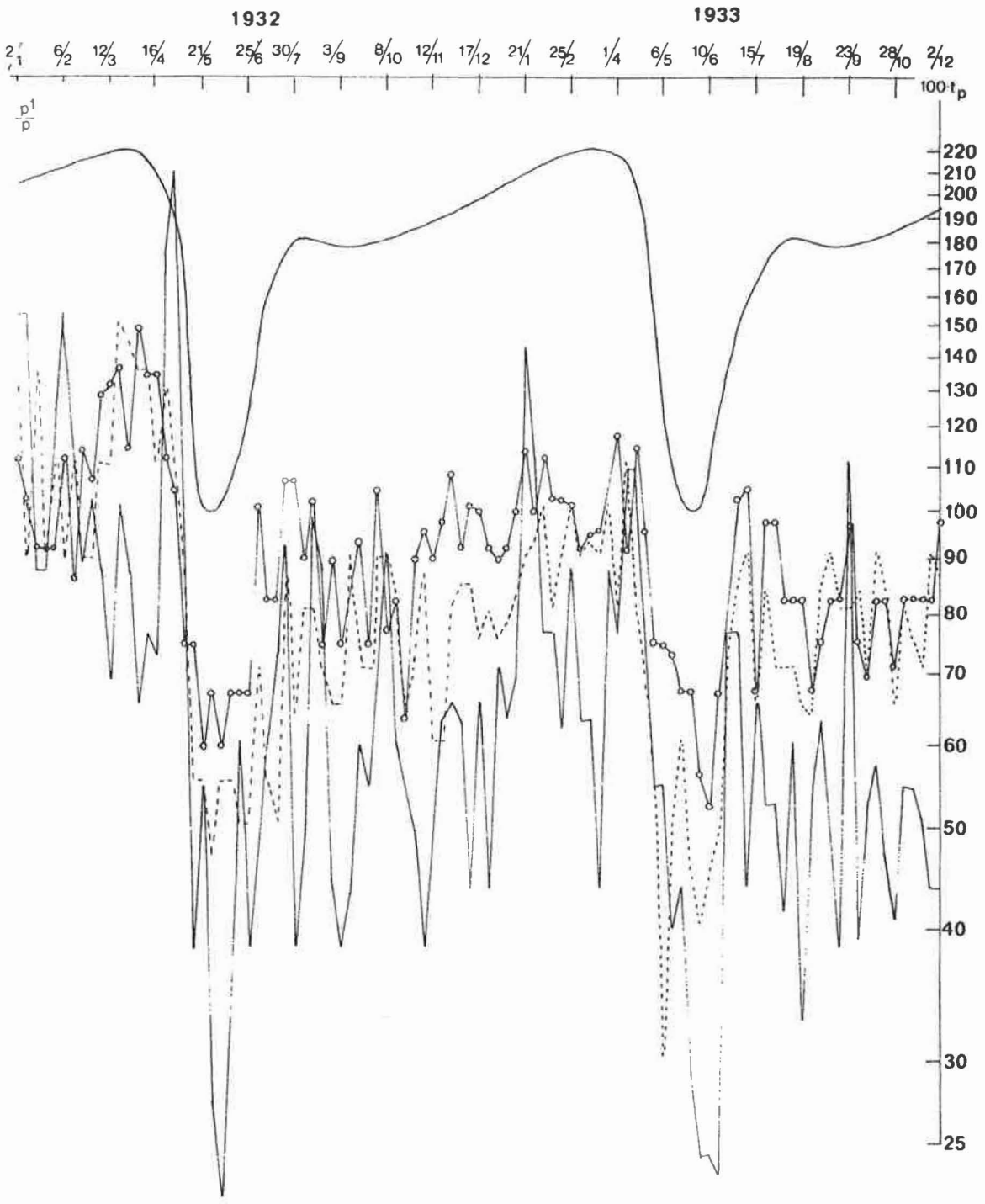
9/1

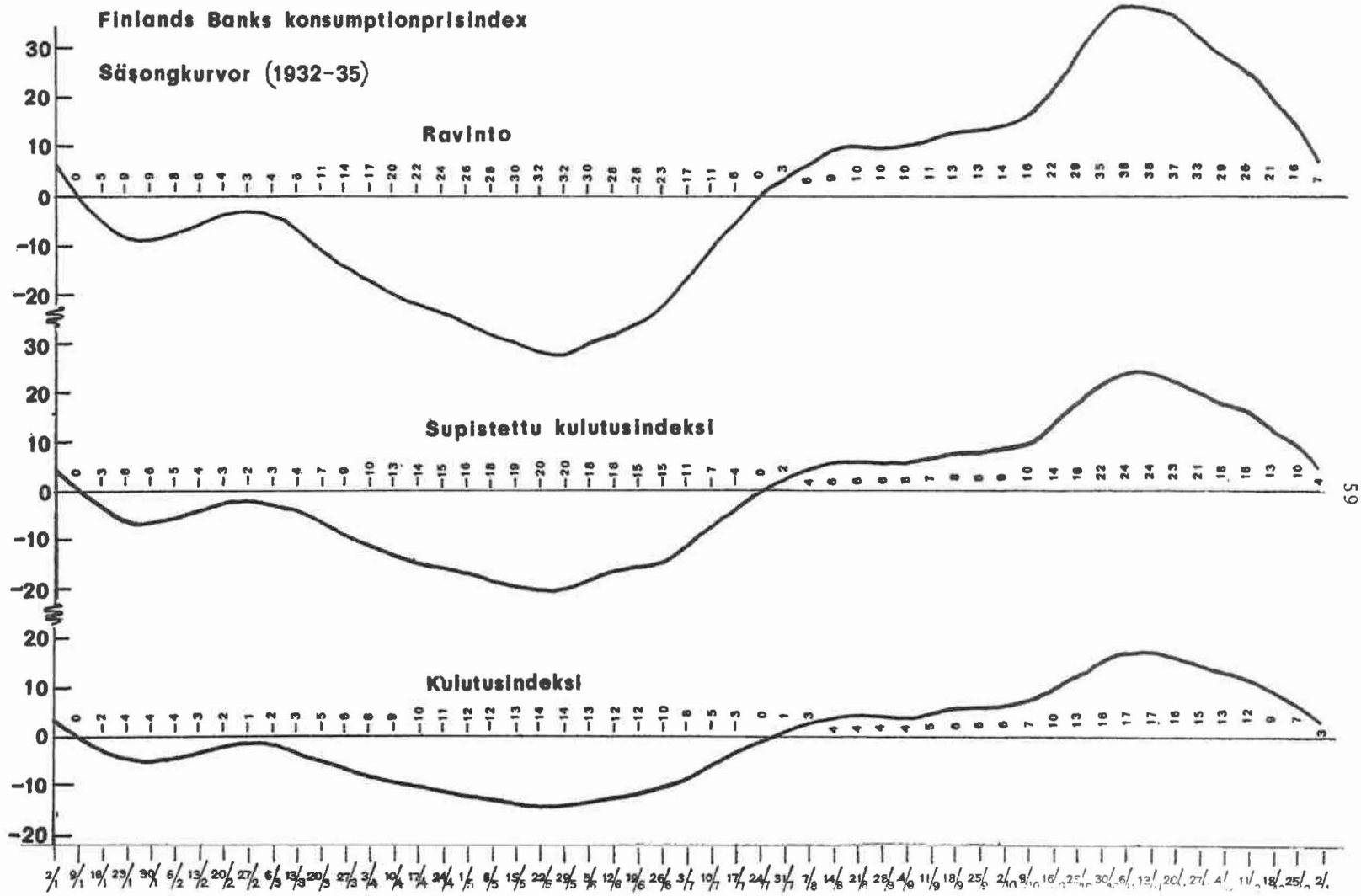
Tavaralaji	p	e	c	e^2	p	log p	d	e	s	s^2
Maito		215	77	46				215	117	13
Meijerivoi		200	26	40	27:-	883	-37			4
Maatiaisvoi		139	25	19	23:-	833	-43	96	-2	0
Kerma		134	6	18				134	36	1
Juusto		-104	6	11	17:33	550	55	-59	-157	25
Margariini		177	8	31				177	79	6
Paistinrasva		68	4	5	15:67	449	22	90	-8	0
Vehnäjauhot		216	17	47	6:17	820	3	219	121	15
Ruisjauhot		140	8	20				140	42	2
Ruislestyt		189	8	36				189	91	8
Kaurasuurimot		211	4	45	4:63	533	16	227	129	17
Ohra "		63	1	4				63	-35	1
Riisi "		143	4	20				143	45	2
Manna "		254	4	65				254	156	24
Ruisleipä, näkki		81	6	7	6:43	861	-12	69	-29	1
" pehmeä		110	14	12				110	12	0
Hiivaleipä		67	10	4				67	-31	1
Ranskanleipä		92	10	8				92	-6	0
Kahvi		47	10	2				47	-51	3
Naudanliha, tuore		-114	18	13	5:73	746	-39	-153	-251	63
Vasikan " "		-100	4	10	12:-	182	6	-94	-192	37
Lampaan " "		-47	4	2	10:50	49	16	-31	-129	17
" " palvattu		-80	2	6				-80	-178	32
Makkara		-58	11	3				-58	-156	24
Sianliha, tuore		-48	8	2	10:67	65	57	9	-89	8
" suolattu		-51	4	3	10:-	000	51	0	-98	10
" savustettu		-49	4	2				-48	-147	31
Silakka, suolattu		218	6	46				218	120	14
Silli		25	6	1	8:50	-163	-9	16	-82	7
Anjovis		0	6	0				0	-98	10
Munat		-109	10	12	17:33	550	-65	-174	-276	76
Kahvi		244	26	60				244	146	21
Palasokeri		122	13	15	9:25	-78	11	133	35	1
Kristallisokeri		224	13	50				237	139	19
Perunat		-166	17	28				-166	-264	70
Luumut		254	2	65				254	156	24
Sekahedelmät		109	2	10	16:-	470	25	134	36	1
Omenat		14	3	0	15:40	432	82	96	-2	0
Appelsiinit		-93	3	8	7:33	-311	-75	-168	-266	71
Banaanit		99	2	10	16:80	519	37	136	38	1
Herneet		95	3	9				95	-3	0
Kalja		-43	5	2				-43	-141	20
Pilsneri		0	1	0				0	-98	10
Suklaa		142	1	20				142	44	2
Karamellit		152	1	23	43:75	476	29	181	83	7
Suola		0	1	0	-:90	-105	58	58	-40	2
Päivällinen		0	10	0				0	-98	10
Kahvi		0	10	0				0	-98	10
		$\frac{45200}{444}$ = 102 (log P)		$\frac{11540}{444}$	110,3		$\frac{-1859}{444}$ =-4 (99,6)	$\frac{43341}{444}$ =98		0,016

248341 - 203141 = 45200

Bilaga III. Uträkningar för tabellen på sid. 28.

Tavaralaji	Suhde-luku	31/12-32 log	30/12-33 log	29/12-34 log	16/6-34 log
Maito	77	8	151	215	8
Meijerivoi	26	46	-42	126	-42
Maatiaisvoi	25	29	-95	139	-150
Kerma	6	134	195	195	134
Juusto	6	-50	-16	-68	-50
Margariini	8	74	-67	-45	-67
Paistinrasva	4	-23	-178	209	-23
Vehnäjauhot	17	109	36	91	61
Ruisjauhot	8	153	90	100	100
Ruislestyt	8	177	148	121	148
Kaurasuurimot	4	66	0	-36	-36
Ohra "	1	-17	-75	-77	-61
Riisi "	4	176	193	159	159
Manna "	4	186	177	145	186
Ruisleipä	6	77	77	64	41
" pehmeä	14	138	138	153	138
Hiivaleipä	10	95	95	100	95
Ranskanleipä	10	148	148	156	142
Kahvileipä	10	73	73	54	54
Naudanliha	18	-297	-143	115	180
Vasikanliha	4	-148	-111	-25	-61
Lampaan "	4	-282	-185	-185	88
" palvattu	2	-80	-69	-49	-145
Makkara	11	185	37	216	216
Sianliha	8	140	231	140	111
" suolattu	4	95	210	33	210
" savustettu	4	134	201	229	240
Silakka	3	-337	277	682	-1514
" suolattu	3	176	-26	127	78
Hauki	3	-106	-106	11	-395
Ahven	3	-277	-499	-277	-1039
Silli	3	-189	-143	-143	-143
Anjovis	3	-14	0	0	0
Munat	10	-38	-52	-62	-462
Kahvi	26	241	98	129	129
Palasokeri	13	170	157	94	135
Kristallisokeri	13	206	148	133	162
Perunat	17	-70	-300	-246	-316
Luumut	2	147	140	131	131
Sekahedelmät	2	69	69	55	42
Omenat	3	0	-74	42	-138
Appelsiinit	3	-214	0	-158	-288
Banaanit	2	-65	-83	-47	-29
Herneet	1	131	190	282	268
Porkkanat	1	-770	-770	-770	697
Punajuuret	1	-747	-53	-747	456
Lantut	1	-633	-609	-656	-19
Valkokaali	2	-704	-566	-1101	1149
Kalja	4	-117	-150	-150	-150
Pilsneri	1	0	-172	-172	-172
Suklaa	1	64	64	30	30
Karamellit	1	128	92	92	92
Suola	1	57	-61	-61	0
Päivällinen	9	0	-19	-19	-19
Kahvi	9	29	54	54	54
		log P = 0,035(46)	0,031(89)	0,086(26)	0,008(1)
		log A = 0,046	0,042	0,099	0,031
		Δ a = 0,011	0,010	0,013	0,023


1932
1933
2/1 6/2 12/3 16/4 21/5 25/6 30/7 3/9 8/10 12/11 17/12 21/1 25/2 1/4 6/5 10/6 15/7 19/8 23/9 28/10 2/12
100·t_p
p¹/p
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
25


Finlands Banks konsumptionprisindex
Säsongkurvor (1932-35)
Ravinto
Supistettu kulutusindeksi
Kulutusindeksi

Särtryck ur Ekonomiska Samfundets Tidskrift. Ny serie. H. 37.

Levnadskostnadsindexerna i Finland och Sverige, deras tillförlitlighet och jämförbarhet

Av Leo Törnqvist.

Om man icke enbart vill nöja sig med ett kvalitativt omdöme om en prisindex tillförlitlighet och användbarhet för olika ändamål, utan vill ha ett kvantitativt uttryck för det osäkerhetsmoment, som finnes hos indexen, och storleksordningen hos de fel, som uppkomma, då indexen utnyttjas för ändamål, för vilka annorlunda konstruerade indexer vore att föredraga, står man inför en av de mest komplicerade frågorna inom indexproblemet. Denna för indextalens praktiska användning viktiga fråga har blivit nästan helt försummad i den litteratur angående indexproblemet, som kommit till min kännedom. [1]) Då vi i det följande skola

[1]) Ansatser till en teori för osäkerhetsmarginalen hos indexberäkningar kan man finna t.ex. i

A. Bowley: Elements of Statistics. London 1926. The influence on the precision of index numbers of correlation between the prices of commodities. Journal of the royal statistical society, Mars 1926.

L. V. Bortkiewicz: Die Kaufkraft des Geldes. Nordic Statistical Journal 1932.

J. W. Lindeberg: Todennäköissyyslasku ja sen käytäntö tilastotieteessä. Helsingfors 1927.

I. Fisher: The making of index numbers, A study of their varieties, tests and reliability. Cambridge, Mass. 1922.

G. Haberler: Der Sinn der Indexzahlen. Tübingen 1927.

K. G. Hagstroem: Prisernas spridning. Skandinaviska kreditaktiebolagets kvartalsskrift, 1923.

H. Eneborg: Statistisk undersökning rörande partiprisrörelsen i Sverige 1920—23. Stockholm 1924.

undersöka levnadskostnadsindexerna i Finland och Sverige från denna synpunkt, måste vi först utveckla en teori för osäkerhetsmarginalen vid indexberäkningar.

Osäkerhetsmomentet vid indexberäkningarna härflyter ur många olika felkällor: indexformelvalet, viktssystemets olämplighet för det betraktade ändamålet, inbegripet den representativa karaktären hos de i indexen ingående varorna, och slutligen iakttagelsefelen vid prisinsamlingen. Dessa fels möjlighet att påverka slutresultatet beror på prisspridningen inom det betraktade prissystemet, det är den bristande proportionaliteten i de olika varuprisens förändringar.

En omständighet, som påverkar felmarginalen, utgör valet av indexformel. De aritmetiska indexformlerna kunna skrivas i formen

$$ {}_aP_0^1 = \sum c_i \frac{p_i^{(1)}}{p_i^{(0)}} ; \sum c_i = 1 . $$

En indexformel, som uppträder i samband med den aritmetiska, är det harmoniska medeltalet

$$ {}_hP_0^1 = 1 : \sum c_i \frac{p_i^{(0)}}{p_i^{(1)}} . $$

Man får nämligen genom att låta de genom (0) och (1) angivna tidpunkterna eller ekonomikretsarna byta plats

$$ {}_aP_1^0 = \sum c_i \frac{p_i^{(0)}}{p_i^{(1)}} = 1 : {}_hP_0^1 , $$

$$ {}_aP_0^1 \cdot {}_hP_1^0 = {}_hP_0^1 \cdot {}_aP_1^0 = 1 . $$

Det harmoniska medeltalet är alltid mindre än det aritmetiska, om icke samtliga talen $p_i^{(1)} : p_i^{(0)}$ äro $= {}_aP_0^1 = {}_hP_0^1$. Det harmoniska medeltalet kan nämligen skrivas som ett aritmetiskt medeltal

$$ {}_hP_0^1 = \sum c_i : \sum c_i \frac{p_i^{(0)}}{p_i^{(1)}} = \sum c_i \frac{p_i^{(0)}}{p_i^{(1)}} \cdot \frac{p_i^{(1)}}{p_i^{(0)}} : \sum c_i \frac{p_i^{(0)}}{p_i^{(1)}} $$

$$ = \sum \acute{c}_i \frac{p_i^{(1)}}{p_i^{(0)}}, $$

och då vikterna $\acute{c}_i = c_i \frac{p_i^{(0)}}{p_i^{(1)}} : \sum c_i \frac{p_i^{(0)}}{p_i^{(1)}}$ ge mindre vikt åt de relativt dyra prisen $p_i^{(1)}$, minskas härigenom medeltalet. Vi ha således ${}_aP_0^1 \, {}_aP_1^0 \geqq 1$, ${}_hP_0^1 \cdot {}_hP_1^0 \leqq 1$, varvid likhetstecknet endast gäller, då prisen $p_i^{(1)}$ och $p_i^{(0)}$ äro proportionella.

Det geometriska medeltalet, som definieras genom formeln

$$ \log {}_gP_0^1 = \sum c_i \log \frac{p_i^{(1)}}{p_i^{(0)}}, $$

har den vackra egenskapen

$$ {}_gP_0^1 \cdot {}_gP_1^0 = 1 . $$

Sambandet mellan dessa medelvärden framgår genom ett studium av de s.k. momentmedeltalen

$$ ({}_aP_0^1)^\alpha = \sum c_i \left(\frac{p_i^{(1)}}{p_i^{(0)}} \right)^\alpha = \sum c_i \, e^{\alpha \log (p_i^{(1)} / p_i^{(0)})} . $$

Vi ha tydligen ${}_aP_0^1 = {}_1P_0^1$, ${}_hP_0^1 = {}_{-1}P_0^1$. Det geometriska medeltalet erhålles som gränsvärde, då α går mot noll. Dessa momentmedeltal utgöra kontinuerligt växande funktioner av α.

I det vi dividera vartdera membrum med $({}_gP_0^1)^\alpha$ och utveckla högra membrum i serie enligt potenser av α fås

$$ ({}_aP_0^1 : {}_gP_0^1)^\alpha = \sum c_i \, e^{\alpha \log (p_i^{(1)} / p_i^{(0)} \, {}_gP_0^1)} = \sum c_i \, e^{\alpha \cdot s_i} $$

$$ = 1 + \frac{\alpha^2}{2!} \sum c_i s_i^2 + \frac{\alpha^3}{3!} \sum c_i s_i^3 + \cdots + \frac{\alpha^\nu}{\nu!} \sum c_i s_i^\nu + \cdots . $$

Genom att logaritmera vartdera membrum fås härav

$$(1) \qquad \log {}_{\alpha}P_0^1 - \log {}_{g}P_0^1 = \tfrac{1}{2}\alpha \sum c_i s_i^2 + \frac{\alpha^2}{6} \sum c_i s_i^3 + \cdots + k_\nu \frac{\alpha^{\nu-1}}{\nu!} + \cdots .$$

Då $\alpha = 0$, blir högra membrum 0, varför ${}_{0}P_0^1 = {}_{g}P_0^1$. Talen k_ν kallas i statistiken semiinvarianterna till den av talen $s_i = \log (p_i^{(1)}/p_i^{(0)}\ {}_{g}P_0^1)$ representerade fördelningen. Talen c_i uppfattas härvid som frekvenstal. Om fördelningen är normal, vilket approximativt torde vara fallet, då det rör sig om prisnivåjämförelser i tiden, äro semiinvarianterna alla $= 0$, förutom den första

$$s^2 = \sum c_i s_i^2 ,$$

som utgör ett mått på spridningen. Med kännedom om det aritmetiska, harmoniska och geometriska medeltalet kan man beräkna närmevärden för s och $\sum c_i s_i^3$. Vi få nämligen

$$(2) \qquad \log {}_{a}P_0^1 - \log {}_{h}P_0^1 = s^2 + \cdots$$

$$(3) \qquad \tfrac{1}{2}(\log {}_{a}P_0^1 + \log {}_{h}P_0^1) - \log {}_{g}P_0^1 = \frac{1}{6} \sum c_i s_i^3 + \cdots$$

Denna senaste formel innehåller satsen: Då talen s_i är normalt fördelade, är ${}_{g}P_0^1 = \sqrt{{}_{a}P_0^1 \cdot {}_{h}P_0^1}$ (Jfr. Fishers ideala indexformel).

Av (1) framgår att det geometriska medeltalet är det enda av momentmedeltalen, som kan användas för bildande av kedjeindextal, ty om $\alpha \neq 0$, och ett stort antal indextal ${}_{\alpha}P_\nu^{\nu+1}$ multipliceras med varandra, så divergerar produkten mot $+\infty$ för $\alpha > 0$ och mot 0 för $\alpha < 0$, även under antagandet att det geometriska medeltalet $P_\nu^{\nu+1}$ skulle förbli konstant $= 1$.

Det har knappast någonsin tagits i betraktande att använda momentmedeltal med $|\alpha| > 1$ såsom indexformler.

Dessa formler påverkas nämligen starkt av de extrema s_i värdena. Man kan nämligen även skriva momentmedeltalens formel i följande form

$$ {}_{\alpha}P_0^1 = \sum c_i \left(\frac{p^{(1)}}{p_i^{(0)}\, {}_{\alpha}P_0^1} \right)^{\alpha-1} \cdot \frac{p_i^{(1)}}{p_i^{(0)}}, $$

varav framgår, att vid växande α de relativt dyraste varorna få allt större vikt; då α går mot $+\infty$, närmar sig ${}_{\alpha}P_0^1$ det största prisförhållandet $p_i^{(1)} : p_i^{(0)}$, och då α avtar mot $-\infty$, går ${}_{\alpha}P_0^1$ mot det minsta talet $p_i^{(1)} : p_i^{(0)}$. Det geometriska medeltalets centrala roll bland momentmedeltalen berättigar en att betrakta det som ett presumtivvärde framom det aritmetiska och harmoniska medeltalet. Det aritmetiska medeltalet kan lämpligen betraktas som en övre gräns och det harmoniska som en undre gräns för de indextal, som åtminstone skenbart använda samma vikter c_i. På grund av det samband, som existerar mellan dessa tre medelvärden och spridningen och snedheten i fördelningen hos talen s_i, får man med kännedom om dessa medelvärden fram de mest karaktäristiska dragen hos denna fördelning. Som mått på den osäkerhet, som uppstår på grund av indexformelvalet, använda vi

$$ \log {}_{\alpha}P_0^1 - \log {}_{h}P_0^1 = s^2 + \cdots = \Delta \text{ (formel)}. $$

Då vi nu gå att undersöka det osäkerhetsmoment, som uppstår på grund av att vikterna c_i ej utgöra de för det ändamål indexen skall användas teoretiskt riktigaste vikterna, kunna vi nöja oss med att enbart studera det geometriska medeltalet.

De för det betraktade ändamålet riktigaste vikterna må vara $\bar{c}_i$. Felen i vikterna c_i beteckna vi

$\Delta c_i = \bar{c}_i - c_i$, och ha, då viktsumman fortfarande skall vara $= 1$,

$$ \Sigma \Delta c_i = 0, \quad \Sigma \Delta c_i \log {}_{g}P_0^1 = 0. $$

Förändringen i resultatet vid övergången till det nya viktsystemet är

$$\Delta(c) = \sum (c_i + \Delta c_i) \log \frac{p_i^{(1)}}{p_i^{(0)}} - \sum c_i \log \frac{p_i^{(1)}}{p_i^{(0)}} = \sum \Delta c_i \log \frac{p_i^{(1)}}{p_i^{(0)}}$$

$$= \sum \Delta c_i \log \frac{p_i^{(1)}}{p_i^{(0)}} - \sum \Delta c_i \log {}_gP_0^1 = \sum \Delta c_i \log \left(\frac{p_i^{(1)}}{p_i^{(0)}\, {}_gP_0^1} \right)$$

$$= \sum \Delta c_i \, s_i .$$

Denna strängt riktiga formel

$$\Delta(c) = \sum \Delta c_i \, s_i = \sum c_i \frac{\Delta c_i}{c_i} \cdot s_i$$

har man dock ej stor glädje av, om talen Δc_i icke äro kända, eller ifall talen s_i för de varuslag, som tidigare haft vikten $c_i = 0$, icke äro kända. Summauttrycket kan tänkas uppdelat i ett antal delsummor, i det att närbesläktade varuslag, som kännetecknas av att motsvarande s_i-tal blott relativt litet avvika från varandra, sammanföras till grupper. De varor, som ej alls medtagits i indexen, kan tänkas samlade kring de i indexen medtagna varorna, vilkas vikter ökas med vikterna på de varor, vilkas s_i-tal komma ett i indexen förekommande s_i-tal närmast. Genom denna viktöverföringsprocess påverkas $\Delta(c)$ blott obetydligt. Denna tankegång ligger till grund, då man låter stora varugrupper representeras av ett fåtal varuslag, som dock ges en vikt, som skulle tillkomma hela varugruppen. Talen Δc_i må numera beteckna skillnaden mellan de vikter, som enligt denna viktöverföringsmetod borde tillkomma varuslaget och de faktiskt använda vikterna c_i.

Uttrycket för $\Delta(c)$ kan skrivas i formen

$$\Delta(c) = \sum c_i \frac{\Delta c_i}{c_i} \cdot s_i = r z s,$$

varvid vi betecknat

$$z^2 = \sum c_i \left(\frac{\Delta c_i}{c_i}\right)^2, \quad s^2 = \sum c_i s_i^2, \quad r = \Delta(c) : z s.$$

Enligt korrelationsteorin veta vi, att r alltid till sitt talvärde är mindre än eller lika med 1.

Om felen $\frac{\Delta c_i}{c_i}$ och talen s_i äro inbördes oberoende variabler, är det sannolika värdet av $\Delta(c) = 0$, och det sannolika värdet av detta tals kvadrat

$$= \Sigma (c_i)^2 \cdot z^2 \cdot s^2.$$

Om de icke äro oberoende, utan det sannolika värdet av korrelationskoefficienten r är $\bar{r} \neq 0$, så tillkommer en term av storleksordningen $\bar{r}^2 z^2 s^2$. Korrelationskoefficienten $\bar{r}$ är sannolikt rätt liten, då det ej finnes något skäl, varför felen i vikterna skulle vara i någon större mån beroende av talen s_i; stigande relativa pris kunna nämligen i medeltal lika ofta tänkas medföra minskad som ökad relativ andel i totalutgiften. Det synes därför, som om man skulle få ett mått på den osäkerhetsmarginal, som beror av att vikterna c_i icke äro de bästa tänkbara, genom att antaga, att $\Delta(c)$ till sitt talvärde är av storleksordningen

$$\left(\Sigma c_i^2\right)^{\frac{1}{2}} \cdot z s.$$

Enär det dock kan stöta på svårigheter att på ett förnuftigt sätt överflytta de varors vikter, som ej förekomma i indexen, på de däri förekommande, kan det vara att föredraga att uppdela summan $\Sigma \Delta c_i s_i$ i två termer

$$\Sigma \Delta c_i s_i = \Delta c_0 s_0 + \Sigma' \Delta c_i s_i.$$

Δc_0 betecknar härvid summan av de varors vikter, som icke blivit representerade i indexen. s_0 betecknar de icke medtagna varornas prisindexlogaritms avvikelse från de medtagna varornas prisindexlogaritm. I brist på närmare kunskaper om s_0 måste man försöka med en grov uppskattning t.ex.

$$s_0^2 \approx (1 + \Sigma c_i^2) \cdot s^2.$$

Faktorn $1 + \Sigma c_i^2$ ha vi tillfogat för att antyda att spridningsmåttet s sannolikt ger ett för litet värde på det sannolika värdet av s_0^2. Då talen c_i äro stora, är s ett otillförlitligare mått på spridningen inom prissystemet än då vikterna c_i alla äro små. Om man är rädd för att undervärdera $s_0 : s$ inflytande kan man ännu tillfoga någon faktor beroende av den uppskattande personens större eller mindre pessimism. De icke representerade varornas bidrag till storleksordningen hos talet $\Delta(c)^2$ uppskatta vi sålunda till

$$(1 + \Sigma c_i^2)\Delta c_0^2 s^2 .$$

Termen $\Sigma'\Delta c_i s_i$ kan man uppskatta på samma sätt som i det fall då Δc_0 antogs vara $= 0$. Enär $\Sigma'\Delta c_i = -\Delta c_0 \neq 0$ får man dock en bättre uppskattning av denna term genom att tillfoga en nollterm

$$\Delta c_0 \Sigma c_i s_i = 0 .$$

Vi få på detta sätt

$$\Sigma' \Delta c_i s_i = \Sigma' \Delta c_i s_i - \Sigma' c_i \Delta c_0 s_i = \Sigma' (\Delta c_i - c_i \Delta c_0) s_i$$
$$= \Sigma' \Delta' c_i s_i = r z' s.$$

Härvid ha vi betecknat $\Delta' c_i = \Delta c_i - c_i \Delta c_0$,

$$z'^2 = \sum c_i \left(\frac{\Delta' c_i}{c_i}\right)^2 = \sum c_i \left(\frac{\Delta c_i}{c_i}\right)^2 - \Delta c_0^2, \quad s^2 = \sum c_i s_i^2.$$

Korrelationskoefficienten r, som till sitt talvärde alltid är mindre än 1, kan härvid uppskattas till storleksordningen

$$r^2 \approx \Sigma c_i^2 .$$

Genom att sammansätta de två termernas bidrag till $\Delta(c) : s$ storleksordning, under antagande av att de äro inbördes oberoende, fås

$$(\Delta(c))^2 \approx \left(\sum c_i^2\right) z'^2 s^2 + \left(1 + \sum c_i^2\right) \Delta c_0^2 s^2 =$$
$$\left(\sum c_i^2\right) \cdot \sum c_i \left(\frac{\Delta c_i}{c_i}\right)^2 s^2 + \Delta c_0^2 s^2 .$$

Ifall vi antaga, att talen

$$\log p_i^{(1)}/p_i^{(0)}$$

äro behäftade med felen σ_i, och antaga att den systematiska delen i dessa fel är liten, bidrager denna felkälla till slutresultatets osäkerhet med ett tillskott av storleksordningen

$$\Sigma c_i^2 \sigma_i^2 = \sigma^2 \cdot \Sigma c_i^2 = (\Delta(\text{prisen}))^2,$$

d.v.s. Δ (prisen) $= (\Sigma c_i^2)^{\frac{1}{2}} \sigma$.

Ifall vi sammanställa dessa från olika felkällor härrörande osäkerhetsmoment i enlighet med schemat för av varandra oberoende spridningstal, fås som mått på indextalets osäkerhetsmoment

$$\Delta(\log P) \approx \sqrt{(\Delta(\text{formel}))^2 + (\Delta(\text{vikterna}))^2 + (\Delta(\text{prisen}))^2}$$
$$\approx \sqrt{s^4 + (\Sigma c_i^2) z'^2 s^2 + (1 + \Sigma c_i^2) \Delta c_0^2 s^2 + (\Sigma c_i^2) \sigma^2}.$$

Detta osäkerhetsmoment tolkas härvid som en matematisk förväntan beträffande storleksordningen av det fel, som göres, då man i stället för det för ett givet ändamål teoretiskt bästa indextalet använder ett surrogat. Om detta surrogat utgöres av ett geometriskt medeltal, synes det rimligt att minska det på grund av valet av indexformel härrörande osäkerhetsmomentet med hälften och sätta Δ (formel) $= \frac{1}{2} s^2$.

De räkneoperationer, som bruka företagas med indextal, utgöra multiplikationer och divisioner. Genom att tänka sig dessa räkneoperationer utförda medelst logaritmer inses, att resultatets logaritm blir osäker av storleksordningen $\Delta(\log P)$. Utföras flere operationer med indextal, sammansättes slutresultatets osäkerhetsmoment så, att kvadraten på detta utgör summan av kvadraterna för de enskilda indextalens osäkerhetsmått. Genom att använda detta sätt att

sammansätta osäkerhetsmomenten, som bygger på antagandet om att de betraktade indextalens fel äro av varandra oberoende statistiska variabler, kommer man dock i det ofta förekommande fallet, att ett indextal ${}_{\alpha}P_1^2$ ($\alpha = -1, 0, 1$) med de teoretiskt riktiga vikterna $\bar{c}_i$ ersättes med kvoten mellan indextalen ${}_{\alpha}P_0^2 : {}_{\alpha}P_0^1$, vilka vägts med vikterna c_i, att i regel överskatta slutresultatets osäkerhet.

Om vi betrakta kvoten mellan två geometriska indextal ${}_{g}P_0^2 : {}_{g}P_0^1$, vilka vardera vägts med vikterna c_i, så överensstämmer denna kvot exakt med det lika vägda geometriska medeltalet ${}_{g}P_1^2$, i det att

$$\log ({}_{g}P_0^2 : {}_{g}P_0^1) = \log {}_{g}P_0^2 - \log {}_{g}P_0^1 = \Sigma c_i (\log p_i^{(2)} - \log p_i^{(0)})$$
$$- \Sigma c_i (\log p_i^{(1)} - \log p_i^{(0)}) = \Sigma c_i (\log p_i^{(2)} - \log p_i^{(1)}) = \log {}_{g}P_1^2 .$$

Kvoten ${}_{g}P_0^2 : {}_{g}P_0^1$ är således endast skenbart beroende av prisen $p_i^{(0)}$. Vid bedömandet av osäkerhetsmomentet har man i detta fall att lägga talen $\log (p_i^{(2)}/p_i^{(1)}\,{}_{g}P_1^2)$ till grund för kalkylen, vilket rätt väsentligt förbättrar uppskattningen. Om det åter är fråga om osäkerhetsmomentet för kvoten mellan två aritmetiska medeltal

$${}_{a}P_0^2 : {}_{a}P_0^1 = \sum c_i \frac{p_i^{(2)}}{p_i^{(0)}} : {}_{a}P_0^1 = \sum c_i \frac{p_i^{(1)}}{p_i^{(0)}\,{}_{a}P_0^1} \cdot \frac{p_i^{(2)}}{p_i^{(1)}} = \sum c_i' \frac{p_i^{(2)}}{p_i^{(1)}},$$

får man ett bättre mått på osäkerhetsmarginalen genom att uppfatta kvoten som ett aritmetiskt medeltal med vikterna

$$c_i' = c_i \frac{p_i^{(1)}}{p_i^{(0)}\,{}_{a}P_0^1} .$$

Efter dessa teoretiska betraktelser skola vi försöka oss på en granskning av levnadskostnadsindexerna i Finland och Sverige.

I Social Tidskrift 1923 nr 10 ingår en beskrivning på Finlands officiella levnadskostnadsindex' byggnad. Vi anföra de för vårt ändamål betydelsefulla fakta.

Finlands officiella levnadskostnadsindex beräknas på grundval av de minuthandelspris, som Statistiska Centralbyråns Socialstatistiska avdelning [1]) insamlar genom sina ombud i 21 städer och orter med sammanträngd befolkning.

Grunden för indexberäkningarnas viktsystem bildas av förbrukningen inom en arbetarfamilj tillhörande den mellersta välståndsklassen med en årsutgift år 1908—09 på 1600—2,000 mark. De använda vikterna grunda sig på den officiella redogörelse rörande konsumtionsförhållandena i Finland, som yrkesinspektören Vera Hjelt enligt uppdrag av Industristyrelsen offentliggjorde under titeln: »Undersökning av yrkesarbetares levnadsvillkor i Finland 1908—09.» Vid beräkningen tages ej förbrukningens förändring med i betraktande, indexen anger huru stora levnadskostnaderna skulle bliva jämförda med kostnaderna under förra hälften av år 1914 under antagande av att förbrukningen vore densamma som åren 1908—09. Först beräknas utgående från den officiella minutprisstatistiken gruppindexer, som därpå sammanvägas sålunda, att de multipliceras med respektive varugrupps procentuella andel i förbrukningen år 1908—09, varpå produkterna adderas och divideras med vikternas summa. De använda vikterna framgå av följande tablå kolumn 1.

	1908—09			1928
	1	*c*	*c'*	
Föda	55,0	61,9	55,6	44,7
Beklädnad och skodon	11,6	13,1	12,0	11,9
Bostadshyra	11,8	13,3	16,0	14,0
Bränsle och belysning	4,1	4,6	4,9	3,9
Tobak	1,0	1,1	1,3	2,0
Tidningar	0,9	1,0	1,0	1,2
Skatter	4,5	5,0	9,2	4,6
Övriga utgifter	11,1			17,7
	100,0	100,0	100,0	100,0

[1]) Numera Byrån för social forskning vid Socialministeriet.

Vi ha tillfogat trenne kolumner av vilka den första (c) anger de vikter, som de olika varugrupperna kommit att erhålla, emedan gruppen övriga utgifter saknar representant i indexen. Den följande kolumnen (c') anger de vikter, varmed de olika gruppernas prisrörelser påverkat levnadskostnadsindexen i medeltal under åren 1926—34, då hänsyn tages till gruppernas indextals medeltal under dessa år. ($c_i' = c_i p_i^{(m)} / p_i^{(0)} P_0^{(m)}$, då $p_i^{(m)} / p_i^{(0)}$ = medeltalet av specialindexen under dessa år och $P_0^{(m)}$ är medeltalet av levnadskostnadsindexen under samma tid. Jfr. s. 73). Den sista kolumnen anger utgiftsfördelningen för arbetare och befattningshavare i medeltal enligt 1928 års levnadskostnadsundersökning [1]). 1928 års budget hänför sig till en något högre levnadsstandard än 1908—09 års budget.

Födagruppen representeras av följande varuaggregat:

	Kvantitet	Medelpris 1926—34	% vikt
Mjölk, oskummad	87,6 l	1: 69	23,8
Smör, mejeri-	1,8 kg	28: 85	8,4
» , bond-	1,8 kg	25: 62	7,4
Ägg	0,5 tjog	21: 07	1,7
Potatis	45,0 l	0: 70	5,0
Ärter	0,7 kg	5: 53	0,6
Rågmjöl	10,0 »	2: 96	4,8
Rågbröd, mjukt, billigaste slag	28,9 »	3: 71	17,2
Havregryn	1,5 »	4: 41	1,1
Nötkött, färskt, stek	3,0 »	11: 30	5,5
Fläsk, färskt, skinka	1,5 »	15: 79	3,8
» , salt, sidfläsk	1,5 »	16: 00	3,9
Strömming, färsk	0,83 »	4: 05	0,5
» , salt	0,83 »	4: 12	0,6
Sill, salt	0,83 »	7: 28	1,0
Kaffe, obränt (Santos & Rio)	1,5 »	31: 74	7,7
Socker, bit-	4,9 »	9: 00	7,1
			100,0.

[1]) Sociala specialundersökn. 14. Levnadskostnadsundersökningen av år 1928. Helsingfors 1936.

Man fäster sig vid rågbrödets stora vikt. Vetemjölets utelämnande och brödgruppens ensidiga representation torde vara de största bristerna i denna budget.

Äggpriset säsongrensas sålunda, att det aktuella priset multipliceras med en faktor, som anger förhållandet mellan årsmedelpriset år 1913 och medelpriset under motsvarande månad år 1913.

Med avseende å beklädnaden har man nöjt sig med att följa prisen på en manskostym och på manskängor av boxläder. Det stöter på svårigheter att från landsbygden införskaffa prisuppgifter på varor av närmare angivet slag.

Hyresuppgifterna hänföra sig till två rum och kök. Hyresräkning företages varje vår.

Gruppen bränsle och lyse representeras av en enda vara: björkved, lång, hemkörd.

Tobakspriserna hänföra sig till fyra enklare cigarettsorter.

Gruppen tidningar och litteratur representeras av medeltalet av prenumerationspriserna per månad för några i arbetarkretsar vanliga tidningar.

Skatterna beräknas med hänsyn såväl till de statliga, som kommunala och kyrkliga skatterna. Skatten beräknas härvid på en inkomst, som erhålles så, att man multiplicerar det år 1914 till grund för beräkningen lagda beloppet med levnadskostnadsindexen för det ifrågavarande skatteåret och dividerar med 100.

Såsom av denna beskrivning framgår är levnadskostnadsindexen i Finland byggd på ett ganska litet prismaterial. Då viktsystemet är något föråldrat är det naturligt, att man ej kan hysa någon högre grad av förtroende för den precision, varmed den mäter levnadskostnadernas förändringar. Byrån för social forskning kommer att i en snar framtid modernisera indexen.

I det följande komma vi att intressera oss för den precision, varmed den kan anses mäta levnadskostnadernas förändringar under åren 1926—34.

De olika gruppernas prisindexer ha under dessa år utvecklat sig på följande sätt (Statistisk Årsbok för Finland 1935):

	Föda	Beklädn.	Hyra	Bränsle	Tobak	Tidningar	Skatter	Levnadsk.
1926	1107,8	1041,9	1305,7	1270,6	1297,7	1079,4	2058,1	1182,9
27	1115,1	1035,7	1378,9	1404,9	1299,7	1127,0	2135,3	1206,9
28	1150,2	1047,2	1422,0	1433,9	1297,0	1158,7	2038,7	1232,7
29	1123,5	1054,8	1457,1	1453,7	1299,4	1174,6	2085,4	1225,3
30	971,2	1044,1	1470,9	1393,2	1301,3	1174,6	2096,6	1129,2
31	869,0	1000,8	1404,4	1050,3	1379,0	1174,6	2135,7	1038,6
32	897,3	978,8	1305,4	878,3	1439,0	1174,6	1975,6	1024,7
33	894,0	966,6	1182,1	886,2	1334,2	1174,6	1912,7	1000,7
34	875,8	957,7	1103,2	901,8	1332,9	1174,6	1996,5	982,7
medeltal	1000,4	1014,2	1336,6	1185,9	1331,2	1157,0	2048,3	1113,7.

För en närmare analys av dessa siffror övergå vi till att betrakta dessa tals naturliga logaritmer. De av oss angivna logaritmtalen kunna tolkas som promillesiffror angivande huru många promille respektive indextal överskjuter (+) eller underskjuter (—) talet 1000, då överskottet eller underskottet jämförts med ett medeltal till 1000 och indextalet.

$\log p \approx \frac{p - 1000}{\sqrt{1000 \cdot p}} \cdot 1000$ promille, när decimalkomman hos talet p blivit placerad så att $10{,}000 > p > 100$.

Tabell II. Levnadskostnadsindexen och gruppindexerna uttryckta i naturliga logaritmer.

	Föda	Beklädn.	Hyra	Bränsle	Tobak	Tidningar	Skatter	Levnadsk.
1926	103	41	267	240	261	76	723	168
27	109	35	321	340	262	120	759	188
28	140	46	352	360	260	148	712	209
29	116	54	376	374	262	161	735	203
30	— 29	43	385	331	263	161	740	121
31	—140	1	339	49	321	161	759	38
32	—109	—21	266	—130	364	161	681	24
33	—112	—34	167	—121	288	161	648	1
34	—132	—43	98	—103	287	161	691	— 17
medeltal	—6	14	286	150	285	146	716	104
log(aritm.)	0	14	290	171	286	146	718	108.

I stället för att bibehålla 1914 som basår betrakta vi den index, som har det aritmetiska medeltalet åren 1926—34 till bas = 1000. Indexformeln

$$P_m^{(t)} = P_0^{(t)} : P_0^{(m)} = \sum c_i \frac{p_i^{(m)}}{P_0^{(m)}\, p_i^{(0)}} \cdot \frac{p_i^{(t)}}{p_i^{(m)}} = \sum c_i' \cdot p_i^{(t)} / p_i^{(m)}$$

lägges till grund för den fortsatta analysen.

I följande tabell ha vi sammanställt de naturliga logaritmerna av talen $p_i^{(t)}/p_i^{(m)}$ och $P_o^{(t)} : P_o^{(m)}$ samt de nya vikterna c_i'. Talen erhållas ur föregående tabell genom att från talen i respektive kolumn subtrahera logaritmerna för de aritmetiska medeltalen under åren 1926—34, vilka antecknats i den sista raden i tabell II.

Tabell III.

	Föda	Beklädn.	Hyra	Bränsle	Tobak	Tidningar	Skatter	Levnadsk.	
Vikt	55,6	12,0	16,0	4,9	1,3	1,0	9,2	*a*	*g*
År									
1926	103	27	—23	69	—25	—70	5	60	59,6
27	109	21	31	169	—24	—26	41	80	79,5
28	140	32	62	189	—26	2	— 6	101	100,0
29	116	40	86	203	—24	15	17	95	94,4
30	— 29	29	95	160	—23	15	22	13	12,2
31	—140	—13	49	—122	35	15	41	— 70	— 73,1
32	—109	—35	— 24	—301	78	15	—37	— 84	— 85,6
33	—112	—48	—123	—292	2	15	—70	—107	—108,2
34	—132	—57	—192	—274	1	15	—27	—125	—126,7.

Den sista med (g) betecknade kolumnen innehåller de med vikterna c_i' vägda medeltalen till gruppindexernas logaritmer, d.v.s. logaritmen för det vägda geometriska medeltalet till specialindexerna $p_i^{(t)}/p_i^{(m)}$. Enär avkortningsfelen hos logaritmerna spela in, ger dock skillnaden mellan kolumn (*a*) och (*g*) icke något tillförlitligt uttryck för $\frac{1}{2}\, s^2$.

Dock kan man se att $\frac{1}{2} s^2$ är av storleksordningen några promille. Grupperna beklädnad, tobak, tidningar och skatter ha under dessa år varit ganska stabila. Bränslegruppens kraftiga prisförändringar beror till en del på att den endast representeras av en enda vara. Skulle även de utelämnade varorna gas och lyse medtagits, skulle denna grupp visat större stabilitet. Födagruppen dominerar levnadskostnadsindexen. För att lättare kunna bedöma i vad mån gruppindexernas förändringar skett proportionellt med levnadskostnadsindexens, beräkna vi talen log $(p_i^{(t)}/p_i^{(m)}P_m^{(t)})$ genom att från dessa kolumner subtrahera den näst sista kolumnens siffror. Dessa tal ge en bild av de olika gruppernas relativa prisrörelser.

Tabell IV. $s_i = \log (p_i^{(t)}/p_i^{(m)} P_m^{(t)})$

	Föda	Beklädn.	Hyra	Bränsle	Tobak	Tidningar	Skatter	s^2	s
Vikt	55,6	12,0	16,0	4,9	1,3	1,0	9,2		
1926	43	—33	— 83	9	— 85	—130	— 55	2,12	46
27	29	—59	— 49	89	—104	—106	— 39	1,95	45
28	39	—69	— 39	88	—127	— 99	—107	3,40	58
29	21	—55	— 9	108	—119	— 80	— 78	2,00	45
30	—42	16	82	147	— 36	2	9	3,17	56
31	—70	57	119	— 52	105	85	111	6,86	83
32	—25	49	60	—217	162	99	121	5,28	73
33	— 5	59	— 16	—185	109	122	47	2,65	51
34	— 7	68	— 67	—149	126	140	98	3,67	61
							medeltal	3,44	58.

I diagram I ha vi grafiskt åskådliggjort dessa siffror.

Grupperna beklädnad, tobak, tidningar och skatter ha stigit i pris relativt till levnadskostnadsindexen, övriga gruppers prisrörelser ha skett mera oregelbundet, hyra och bränsle uppvisa ett maximum åren 1931 respektive 1930. Födagruppens rörelser ha skett i nära överensstämmelse med levnadskostnadsindexen.

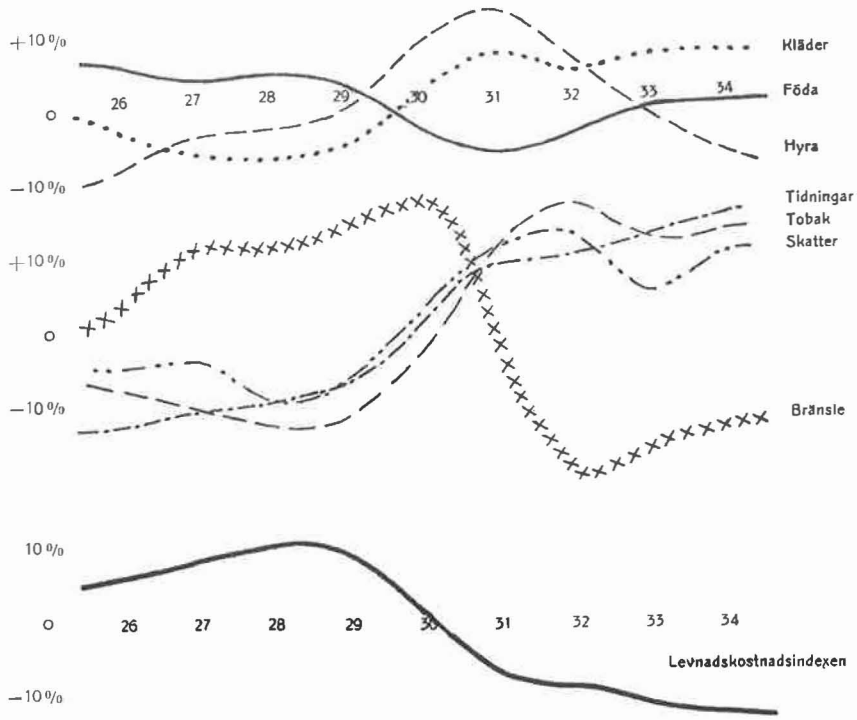


Diagram I. Gruppindexernas utveckling relativt till levnadskostnadsindexen.

Spridningsmåttet s var störst år 1931, då det uppgick till 83 $^0/_{00}$. Spridningen synes växa, då prissystemets rörelsehastighet uppåt eller nedåt ökar. Indexformelfelet s^2 är för detta år c:a 7 $^0/_{00}$. Om även gruppindexernas bidrag till detta skulle tagas i betraktande, skulle indexformelfelet ytterligare stiga med några $^0/_{00}$. I medeltal under dessa år kan indexformelfelet uppskattas till c:a 3,5 $^0/_{00}$.

Vi skola så undersöka storleksordningen av det osäkerhetsmoment, som härrör av viktsystemets olämplighet för olika ändamål. Härvid är måhända en jämförelse mellan det använda viktsystemet och det viktsystem, som skulle erhållas, om resultatet av 1928-års levnadskostnadsunder-

sökning beträffande arbetare och befattningshavare skulle läggas till grund för viktbestämningen av intresse. En jämförelse med den av K. Socialstyrelsen i Sverige [1]) använda gamla och nya budgeten är av intresse för bedömandet av jämförbarheten mellan Finlands och Sveriges levnadskostnadsindexer. Viktsystemen ha vi sammanställt i följande tablå:

	Finland		Sverige	
	c'	1928	a	b
Föda	55,6	44,7	42,8	36,1
Beklädnad	12,0	11,9	11,85	14,1
Bostadshyra	16,0	14,0	15,0	14,6
Bränsle och belysning ...	4,9	3,9	4,1	4,1
Tobak	1,3	2,0	0,0	0,9
Tidningar	1,0	1,2	0,0	0,0
Skatter	9,2	4,6	7,95	6,1
Övriga utgifter		17,7	18,3	24,1
	100,0	100,0	100,0	100,0

Kolumnen a innehåller de enligt den gamla budgeten (juli 1914) begagnade procentvikterna, kolumnen b de efter omläggningen av K. Socialstyrelsens levnadskostnadsindex 1931 begagnade procentvikterna. För att kunna bedöma storleken av den variation i indexen, som skulle åstadkommas genom en övergång från c' till de tre andra viktsystemen, behöva vi framförallt kännedom om det med z' betecknade talet. Gruppen övriga utgifter behandlas som en fristående felkälla.

I följande tabell ha vi sammanställt talen Δc_i för de tre budgeterna, samt de för uppskattningen av termen $\Sigma \Delta c_i s_i$ erforderliga talen

$$(\sum c'^2_i)^{\frac{1}{2}}, \; z' = \sqrt{\sum \frac{(\Delta c_i)^2}{c'_i} - \Delta c_0^2} \text{ och } \Delta c_0^2 .$$

[1]) Sociala Meddelanden 1932 nr 4.

	c'	1928	a	b
Föda	55,6	—10,9	—13,8	—17,5
Beklädnad	12,0	— 0,1	—0,15	2,1
Bostad	16,0	— 2,0	— 1,0	— 1,4
Bränsle och lyse	4,9	— 1,0	— 0,8	— 0,8
Tobak	1,3	0,7	— 1,3	— 0,4
Tidningar	1,0	0,2	— 1,0	— 1,0
Skatter	9,2	— 4,6	— 1,25	— 3,1
Övriga		17,7	18,3	24,1
	100	0,0	0,0	0,0.
		$\sum \frac{(\Delta c_i)^2}{c_i'} = 0{,}053$	0,061	0,084
		$\Delta c_0^2 = 0{,}031$	0,036	0,058
$\sum c_i'^2 = 0{,}36$		$z'^2 = 0{,}022$	0,025	0,026
$(\sum c_i'^2)^{\frac{1}{2}} = 0{,}6$		$z' = 0{,}15$	0,16	0,16.

I det vi bortse från det på grund av utelämnandet av gruppen övriga utgifter härrörande osäkerhetsmomentet, fås, då vi för spridningen s använda medeltalet under åren 1926—34, följande värden på

$$(\sum c_i'^2)^{\frac{1}{2}} \cdot z' \cdot s$$

(1928) $0{,}6 \cdot 0{,}15 \cdot 58 = 5{,}2$ promille
(a & b) $0{,}6 \cdot 0{,}16 \cdot 58 = 5{,}6$ »

Gruppen övriga utgifters bidrag till osäkerhetsmomentet uppskatta vi till $\Delta c_0 \cdot s\sqrt{1 + \sum c_i'^2}$ och få för de betraktade budgeterna ett fel av storleksordningen 12, 13, 16 promille. Då de båda feltermerna adderas under antagande av inbördes oberoende fås ett »fel» av storleksordningen

$$\sqrt{(\sum c_i'^2)\, z'^2 s^2 + \Delta c_0^2 s_i^2 (1 + \sum c_i'^2)} = 13,\ 14,\ 17 \text{ promille.}$$

Genom att direkt uträkna talen $\sum' \Delta c_i s_i$ med användande av de i tabell IV angivna talen s_i får man de i nedan-

stående tabell sammanställda förändringarna i levnadskostnadsindexens naturliga logaritm. De tre sista kolonnerna innehålla tal, som komma till användning vid uppskattningen av gruppen övriga utgifters inverkan på resultatet ($\Delta c_0 \cdot s_0$).

Tabell V.

			$\Sigma' \Delta c_i s_i$		$-\Delta c_0 \log P_m^{(t)}$		
	$\log P_m^{(t)}$	(1928)	*a*	*b*	(1928)	*a*	*b*
1926	60	—1,4	—2,0	—3,8	—10,6	—11,0	—14,5
27	80	—2,1	—1,2	—3,7	—14,2	—14,6	—19,3
28	101	—0,5	—2,6	—3,6	—17,9	—18,5	—24,3
29	95	—0,5	—0,3	—1,9	—16,8	—17,4	—22,9
30	13	0,8	4,1	5,2	— 2,3	— 2,4	— 2,8
31	— 70	1,5	5,2	6,5	12,4	12,8	15,0
32	— 84	—0,2	0,9	0,9	14,9	15,4	18,0
33	—107	1,5	—1,0	0,7	18,9	19,6	22,9
34	—125	0,2	—1,5	—0,2	22,1	22,8	26,9.

Den största förändringen i resultatet skulle erhållas vid övergång till K. Socialstyrelsens nya budget b. Vid jämförelse mellan år 1926 och 1931 skulle en enligt budget *b* vägd index visat ett 6,5 + 3,8 = 10,3 promille lägre prisfall än det, som levnadskostnadsindexen (log $P_m^{(t)}$) utvisade. Övergången till 1928 års budget skulle ha påverkat resultatet med blott ett par promille. Härvid ha vi dock icke ännu tagit den förändring i betraktande, som medtagandet av gruppen övriga utgifter i indexen skulle ha åstadkommit. Ett sätt att uppskatta denna terms inflytande på indexen fås genom att antaga, att denna varugrupps priser förhållit sig konstanta under den betraktade perioden. Detta antagande torde träffa rätt nära det riktiga, ty de viktigaste varorna inom gruppen övriga utgifter bestå av medicin, hygien, alkoholdrycker, resor, nöjen, inventarier och personliga tjänster. Av dessa varor höra de flesta till den mycket trögrör-

liga prisgruppen. Enär det dock är bekant, att gruppen inventarier och personliga tjänster följt med i prisfallet, kommer man genom detta antagande i viss mån att överskatta gruppen övriga utgifters inflytande på levnadskostnadsindexen. Antagandet, att gruppen övriga utgifters prisindex förblivit konstant under den betraktade perioden, innebär att $s_0 = -\log P_m^{(l)}$, varför indexen komme att förändras med $-\Delta c_0 \log P_m^{(l)}$. Dessa tal ha antecknats för de tre olika budgeterna i de tre sista kolumnerna av tabell V. I följande tabell ha vi sammanställt de indexlogaritmer, som man skulle erhålla, ifall de tre andra betraktade budgeterna skulle ha lagts till grund för beräkningen och gruppen övriga utgifters index skulle ha förblivit konstant.

Tabell VI.

	log P	(1928)	a	b
1926	60	48	47	42
27	80	64	64	57
28	101	83	80	73
29	95	78	77	70
30	13	11	15	15
31	— 70	— 55	— 52	—49
32	— 84	— 69	— 68	—65
33	—107	— 87	— 88	—83
34	—125	—103	—104	—98.

Medtagandet av gruppen övriga utgifter i levnadskostnadsindexen skulle sålunda ha dämpat prisfallet, ehuru knappast så mycket, som ovanstående serier skulle ge vid handen. Den redan tidigare givna uppskattningen av storleksordningen av de förändringar i indexen, som skulle uppkomma genom att företaga någorlunda rimliga vägningskoefficientförändringar, (13, 14, 17 promille) synes träffa rätt nära det riktiga. Detta »fel» domineras för levnadskostnadsindexen helt av den utelämnade varugruppen övriga utgifter. I för-

hållande till levnadskostnadsindexens förändringar kan man vänta sig, att medtagandet av denna varugrupp i indexen skulle minska svängningarna med närmare 20 %. Om svängningsperioden är lång, anpassa sig även de trögrörliga varorna efter de förändrade prisförhållandena.

Vi gå så att betrakta specialindexernas tillförlitlighet.

Analys av födagruppen.

Prisutvecklingen för de i födaindexen ingående varorna framgår av tabell VII. Siffrorna äro frånsett medeltalen, som uträknats av oss, hämtade från Social Tidskrift nr 1, 1936. Siffrorna ange penni. De hänföra sig för åren 1926—31 till 40 orter och åren 1932—34 till 30 orter, varför de ej helt sammanfalla med de årsmedeltal, som skulle erhållas genom att lägga det prismaterial, som begagnas för levnadskostnadsindexberäkningen (21 orter), till grund.

På grundval av det i tabell VII givna prismaterialet ha vi beräknat en födaindex, som i följande tabell jämförts med den officiella födaindexen.

	Index enl. VII	Officiella	Skillnad
1926	1102,4	1107,4	— 5,0
1927	1109,8	1114,7	— 4,9
1928	1139,3	1149,8	—10,6
1929	1122,8	1123,1	— 0,3
1930	980,8	970,8	10,0
1931	869,0	868,6	0,4
1932	900,0	896,9	3,1
1933	895,8	893,6	2,2
1934	879,5	875,4	4,1
Medeltal	1000,0	1000,0	0.

Tabell VII.

Varuslag	vikt		1926	1927	1928	1929	1930	1931	1932	1933	1934	M:tal
Mjölk, oskummad ...	87,6	l	187	193	198	193	169	144	146	144	146	169
Smör, mejeri-	1,8	kg	3412	3421	3533	3444	2784	2378	2321	2328	2349	2885
» , bond-	1,8	»	3054	3028	3150	3030	2478	2073	2046	2092	2116	2562
Ägg	0,5	tjog	2519	2470	2667	2568	2157	1822	1704	1537	1522	2107
Potatis	9,0	5 l	385	343	473	493	333	280	279	304	247	349
Ärter	0,7	kg	527	538	589	630	569	516	530	537	537	553
Rågmjöl	10,0	»	300	320	336	317	268	260	299	289	272	296
Rågbröd, mjukt	28,9	»	379	388	397	394	365	340	365	361	349	371
Havregryn	1,5	»	443	451	487	469	444	418	445	423	393	441
Nötkött, färskt.....	3,0	»	1184	1232	1336	1352	1261	1005	926	890	986	1130
Fläsk, färskt	1,5	»	1941	1814	1815	1826	1737	1291	1216	1290	1280	1579
» , salt	1,5	»	1946	1797	1795	1825	1754	1377	1278	1318	1306	1600
Strömming, färsk ...	0,83	»	439	418	474	496	485	350	340	322	333	405
» , salt ...	0,83	»	436	403	430	468	459	407	403	351	348	412
Sill, salt	0,83	»	730	732	736	744	743	737	716	717	697	728
Kaffe, obränt	1,5	»	3605	3539	3243	3283	2732	2712	3278	3175	2996	3174
Socker, bit-	4,9	»	949	1010	858	797	772	856	955	970	937	900.

Ifall det i tabell VII givna prismaterialet skulle ha helt sammanfallit med de årsmedeltal, som skulle ha erhållits med stöd av de vid levnadskostnadsindexberäkningen begagnade prisen, skulle vi ha erhållit samma födaindextal som enligt de officiella beräkningarna. Differenserna ge en uppfattning av storleksordningen av det osäkerhetsmoment, som uppkommer hos indextalet på grund av medeltalsbildningen över de i indexen ingående orterna. Genom att variera ortsvalet och beräkningssättet av medelprisen för hela landet skulle indextalet kunna förskjutas med ibland ända till 10 promille. Då antalet orter ökas, synes det som om prisfallet under den betraktade perioden skulle komma att dämpas. Att döma av den tabell över den relativa födoämnesdyrhetens på olika orter utveckling under den betraktade perioden, som finnes publicerad i Social Tidskrift, 1936, nr 1, sid. 50, har dock prisutvecklingen på de i indexen ingående 21 orterna skett förvånansvärt parallellt.

Förutom medeltalsbildningen över de olika i indexen ingående orterna utföres en medeltalsbildning i tiden för att erhålla årsmedeltalen. För bedömandet av det härur härflytande osäkerhetsmomentet, borde man ha kunskap om prisens variation under årens lopp. Härvid är det främst de säsongmässiga prisrörelserna som intressera. Den officiella levnadskostnadsindexen har en rätt kraftigt utpräglad *säsongrörelse*. Genom att begagna den visserligen något grova, men såsom det synes för detta ändamål tillräckligt goda, metoden att bilda medeltalen av respektive månadsindextal under åren 1921—35 och dividera med årsmedeltalet, få vi nämligen följande *säsongindex för levnadskostnadsindexen*

jan.	febr.	mars	april	maj	juni	juli	aug.	sept.	okt.	nov.	dec.
99,9	99,1	99,1	98,2	97,6	98,2	99,7	101,9	101,6	101,9	101,9	100,7.

Årsmedeltalet av dessa siffror är = 100.

Säsongrörelsen hos födaindexen har en nästan dubbelt större amplitud. Enligt en undersökning som vi utfört beträffande Finlands Banks konsumtionsprisindex framgår, att

födaindexen ensam är ansvarig för denna säsongrörelse. Enär födaindexens säsong blivit dämpad i förhållandet (55,6: 100) i levnadskostnadsindexen, är denna sålunda i det närmaste dubbelt kraftigare hos födaindexen.

»Säsongrensningen» av äggpriset har sålunda ej haft avsedd verkan. Det förefaller oss, som om man skulle träffa den rätta storleksordningen av det tillskott till födaindexens osäkerhetsmarginal, som uppkommer på grund av medeltalsbildningen i rummet och tiden, genom att uppskatta detta till c:a 10 promille.

För att kunna bedöma övriga felkällors inflytande behöva vi kännedom om i vad mån de olika varuslagens priser utvecklat sig proportionellt med födaindexen. Detta framgår tydligast genom att beräkna talen $\check{s}_i = \log (p_i^{(t)} / p_i^{(m)} P_m^{(t)})$, där $p_i^{(t)}$ betecknar varans medelpris under året t, $p_i^{(m)}$ medelpriset under hela den betraktade perioden och $P_m^{(t)}$ den officiella födaindexen justerad så att medeltalet under hela perioden = 1000. Talen framgå av tabell VIII.

Medelvärdet av s^2 under denna period är 6,8 promille. Vi torde därför kunna uppskatta indexformelfelet till c:a 7 promille.

Medelvärdet av spridningen s är 80 promille. Spridningen inom födagruppen var störst åren 1932—33.

För att kunna bedöma i vad mån viktsystemet »föråldrats», jämföra vi detsamma med en budget, som vi sammanställt med stöd av 1928-års levnadskostnadsundersökning. Siffrorna hänföra sig till arbetare och befattningshavare och ha tagits från den på s. 51 i Sociala Specialundersökningar 14, Helsingfors 1936, befintliga tabellen 16. För att underlätta jämförelsen ha härvid en del varuslag sammanförts till större grupper. Den andra kolumnen innehåller de vikter, som dessa gruppers representanter i indexen erhållit.

Tabell VIII. Naturliga logaritmer. $\hat{s}_i = \log (p_i^{(t)}/p_i^{(m)}\acute{P}_m^{(t)})$

Varuslag	vikt	1926	1927	1928	1929	1930	1931	1932	1933	1934
Mjölk, oskummad	23,8	— 2	24	18	17	29	— 21	— 38	— 49	— 15
Smör, mejeri-	8,4	63	62	73	61	— 6	— 53	—108	—102	— 73
» , bond-	7,4	72	58	66	52	— 5	— 72	—116	— 91	— 60
Ägg	1,7	75	48	95	81	52	— 6	—104	—204	—194
Potatis	5,0	— 5	—126	164	229	— 18	— 80	—115	— 26	—214
Ärter	0,6	—151	—136	— 77	15	58	71	67	83	103
Rågmjöl	4,8	— 89	— 31	— 13	— 47	— 70	11	119	88	48
Rågbröd, mjukt	17,2	— 82	— 64	— 72	— 56	13	53	93	85	71
Havregryn	1,1	— 99	— 87	— 41	— 55	36	86	118	70	17
Nötkött, färskt	5,5	— 56	— 22	28	64	139	23	— 90	—127	— 4
Fläsk, färskt	3,8	103	29	— 2	29	124	— 62	—152	— 90	— 78
» , salt	3,9	93	7	— 25	15	121	— 10	—116	— 82	— 71
Strömming, färsk	0,5	— 23	— 78	17	86	209	— 16	— 66	—188	— 64
» , salt	0,6	— 46	—131	— 97	11	137	128	87	— 48	— 37
Sill, salt	1,0	—100	—103	—129	— 94	50	153	93	97	89
Kaffe, obränt	7,6	24	0	—118	— 82	—121	— 17	141	113	74
Socker, bit	7,1	— 50	6	—188	—238	—125	90	168	187	174
log (officiella födaindexen)		103	109	140	116	— 29	—140	—109	—112	—132
log (aritm. medeltal $_a\acute{P}_m^{(t)}$)		97	104	130	116	— 19	—140	—105	—110	—128
log ($_g\acute{P}_m^{(t)}$ = geometriskt)		95	103	126	112	— 22	—141	—111	—115	—132
$(\Sigma c_i^2)^{\frac{1}{2}} = 0{,}3$	$s^2 =$	3,8	2,7	7,2	8,9	5,4	2,5	11,1	11,4	7,8
	$s =$	62	52	85	96	74	50	105	106	88.

Tabell IX.

	(1928)	c	Δc
Mjölk, grädde och ost	20,6	23,8	—3,2
Smör och margarin	15,2	15,8	—0,6
Ägg	2,1	1,7	0,4
Potatis	3,7	5,0	—1,3
Ärter och grönsaker	1,2	0,6	0,6
Mjöl	8,9	4,8	4,1
Bröd	9,1	17,2	—8,1
Gryn	2,6	1,1	1,5
Kött, charkuteriv., fett	14,7	13,2	1,5
Fisk	3,0	2,1	0,9
Kaffe	6,2	7,6	—1,4
Socker	6,5	7,1	—0,6
Övrigt (frukt, svagdricka m.m.)	6,2		6,2
	100,0	100,0	0,0.

För z' erhåller man värdet 0,3. Talet domineras av avvikelserna beträffande mjöl och bröd. Enär dock summan av dessa gruppers vikter 17,2 + 4,8 = 23,0 blott relativt litet avviker från motsvarande summa enligt 1928 års budget 9,1 + 8,9 = 20,0, och dessa båda varugruppers prisutveckling enligt tabell VIII att döma varit rätt likartad, synes detta värde på z ge en något för oförmånlig bild av det begagnade viktsystemet. Utelämnandet av vetemjölet, som enligt 1928 års undersökning haft större vikt än rågmjölet, förstärker dock svagheterna beträffande bröd och mjölgruppen. Vetemjölets prisutveckling har på grund av sitt starkare beroende av växelkurserna och utlandsförhållandena varit i viss mån rätt avvikande från rågmjölets och rågbrödets prisutveckling. Av följande serier över rågmjölets, vetemjölets och rågbrödets $\acute{s}_i$ tal framgår i vad mån de utvecklat sig parallellt.

	1926	1927	1928	1929	1930	1931	1932	1933	1934
Rågmjöl	—89	—31	— 13	— 47	—70	11	119	88	48
Vetemjöl	34	8	—161	—191	—42	42	136	98	109
Rågbr., mjukt	—82	—64	— 72	— 56	13	53	93	85	71.

En överflyttning av 7 % från rågbrödets vikt till vetemjölets skulle ha förändrat födaindexen med för de olika åren följande promille

år	1926	1927	1928	1929	1930	1931	1932	1933	1934
Δ	8	5	—6	—9	—4	—1	3	1	3.

Övriga vid övergången till en modernare budget uppkomna förskjutningar kunna såväl förstärka som motverka mjöl och brödgruppens inflytande.

Genom att för z' använda värdet 0,3, för spridningen s medelvärdet 80 och för Δc_0 värdet 6,2 % fås som uppskattning för storleksordningen av det på grund av viktförskjutningar uppkommande »felet»

$$\sqrt{(\Sigma c_i^2)\, z'^2 s^2 + (1 + \Sigma c_i^2)\, \Delta c_0^2 s^2} \approx 9 \text{ promille.}$$

I det vi sammanställa indexformelfelet, viktförskjutningsfelet och det på grund av medeltalsbildningen i tiden och rummet uppkommande osäkerhetsmomentet enligt schemat för av varandra oberoende variabler fås som resultat

$$\sqrt{7^2 + 9^2 + 10^2} \approx 15 \text{ promille.}$$

Födaindexens osäkerhetsmoment torde sålunda kunna uppskattas till c:a 15 promille.

En jämförelse av födaindexens budget med den i Sverige av K. Socialstyrelsen använda budgeten försvåras av att denna upptager mycket flere varuslag. Den efter 1931 års revision av indexen använda livsmedelsbudgeten innehåller följande varukvantiteter.[1])

[1]) Sociala Meddelanden, 1932, nr 4.

Budget för socialstyrelsens levnadskostnadsindex (1931)

Livsmedel	Sverige 1931	Finland 1928
Mjölk, oskummad	1000 l	(1007,8)
» separerad	60 »	(64,9)
Smör	36 kg	(47,0)
Ost, helfet	12 »	{ 4,5
» halvfet	5 »	}
Margarin	28 »	(17,3)
Ägg	30 tjog	(11,0)
Potatis, gammal	560 l	(531,6)
Ärter, gula	6 kg	(5,0)
Bönor, bruna	2 »	
Mjöl, vete-	176 »	(103,8)
» råg-	20 »	{ 125,9
» rågsikt	53 »	}
Gryn, havre-	22 »	(23,4)
» ris-	8 »	(20,0)
Rågbröd, hårt	34 »	(20,4)
» sötlimpa	53 »	{ 73,3
» grovbröd	34 »	}
Vetebröd, skorpor	6 »	{ 39,3
» mjukt matbröd	11 »	}
» mjukt kaffebröd	27 »	
Nötkött	29 kg	(38,1)
Kalvkött	22 »	(5,7)
Fårkött	5 »	{ 15,5
Malet nötkött	13 »	}
Fläsk, färskt	37 »	(28,1)
» salt	11 »	(5,2)
Fisk		(55,5)
Abborre och gädda	7 »	
Sill, färsk	9 »	
Strömming, färsk	15 »	(15,5)
Torsk	15 »	
Sill, salt	18 »	(9,5)
Kabeljo	4 »	

Kaffe, rostat	20 kg	(19,6)
Socker, bit-	109 »	(101,5)
Äpplen	20 »	
Apelsiner	7 ds.	
Bananer	7 kg	
Katrinplommon	5 »	
Vitkål	25 »	
Pilsner	75 fl.	
Svagdricka	45 l.	

Övriga varugrupper äro likaledes betydligt fylligare representerade än i den finländska levnadskostnadsindexen.

Det torde knappast löna sig mödan att förvandla dessa kvantitetsvikter till procentvikter. Budgeten liknar rätt mycket den budget som enligt 1928 års levnadskostnadsundersökning erhölls för arbetare och befattningshavare i Finland. Dessa siffror äro hämtade från den på sid. 66 i Sociala Specialundersökningar 14 befintliga tabellen. I siffran för mjölk, oskummad ingår ej grädde (20,3 l), osten har ej närmare specificerats, talet 73,3 hänför sig till mjukt rågbröd, 39,3 till annat bröd, 15,5 till annat kött. Siffran 19,6 innefattar både rostat (5,1 kg) och orostat kaffe (11,1 kg).

En övergång till den svenska budgeten torde ge ett något högre värde på det av viktsystemets variation härrörande osäkerhetsmomentet än som vi erhöllo medelst jämförelse med 1928 års budget. Ett värde på z om låt oss säga 0,5 torde dock vara tillräckligt för att så kraftiga förändringar i budgetens sammansättning som mellan den svenska budgeten och levnadskostnadsindexens födabudget skola vara möjliga. Detta värde på z skulle ge ett osäkerhetsmoment på c:a 13 promille för »viktsystemfelet», födaindexens totala osäkerhetsmoment skulle stiga till c:a 17 promille.

Vad övriga gruppindexer beträffar kan man på grund av det fåtal varuslag, som ingå i dem, ej utföra någon liknande detaljanalys. Vi måste därför göra uppskattningen mera på en höft. Då det synes troligt, att de gruppindexer, som förhållit sig relativt stabila under perioden, ha ett mindre osä-

kerhetsmoment än de rörligare gruppindexerna, förefaller det rimligt att antaga, att osäkerhetsmomenten hos de olika gruppindexerna utgöra ungefär samma procent av indexens medelavvikelse från periodmedeltalet som hos födaindexen, d.v.s. c:a 15 %. Vi få enligt detta antagande (se tabell III) att gruppindexen för beklädnad skulle vara osäker på c:a 5, hyra c:a 15, bränsle c:a 30 samt tobak, tidningar och skatter c:a 5 promille. Med tillhjälp av dessa uppskattningar får man för storleksordningen av det osäkerhetsmoment hos levnadskostnadsindexen, som uppkommer på grund av gruppindexernas otillförlitlighet under antagande av att de olika gruppindexernas »fel» äro av varandra oberoende

$$\sqrt{\Sigma c_i^2 \sigma_i^2} = \tfrac{1}{100}\sqrt{(55{,}6)^2 \cdot 15^2 + (12{,}0)^2 \cdot 5^2 + (16{,}0)^2 \cdot 15^2} + (4{,}9)^2 \cdot 30^2 + ((1{,}3)^2 + (1{,}0)^2 + (9{,}2)^2) \cdot 5^2 \approx 9 \text{ promille.}$$

Resultatet domineras fullständigt av födaindexens osäkerhetsmoment. Även om man varierar uppskattningarna för de övriga grupperna med 5—10 promille, påverkas resultatet ej härav med mer än 1 promille. Genom att för födaindexens osäkerhetsmoment använda uppskattningen 17 promille (jämförelse med svenska förhållanden) växer det på grund av gruppindexernas otillförlitlighet uppkommande osäkerhetsmomentet hos levnadskostnadsindexen till 10 promille.

I det vi sammanställa de olika bidragen till levnadskostnadsindexens osäkerhetsmoment enligt additionssatsen för oberoende variablers spridningsmoment fås

$$\Delta \text{ (levn. kostn. index)}^2 = \Delta \text{ (formel)}^2 + \Delta \text{ (gruppvikterna)}^2 + \Delta \text{ (gruppindexerna)}^2 \approx (3{,}5)^2 + 17^2 + 10^2 \approx 401 \approx 20^2$$

d.v.s. Δ (levn.kostn.index) $\approx$ 20 promille.

Vi ha således kommit till resultatet:

Genom att variera beräkningssättet för levnadskostnadsindexen i Finland inom rimliga gränser innefattande det av K. Socialstyrelsen i Sverige använda skulle man ha erhållit indexserier för åren 1926—34 som, förutsatt att samtliga serier jus-

terats så att periodmedeltalet = 1000, med stor sannolikhet skulle till största delen ha förblivit inom ett område, som icke mer än 20 promille avviker från den officiella levnadskostnadsindexens serie. Genom övergång till ett viktsystem som bättre överensstämmer med de nutida konsumtionsförhållandena i Finland och Sverige skulle prisfallet 1928—34 sannolikt ha blivit dämpat med ett par procent (jfr. tabell VI).

För åren 1932—35 finnes det en möjlighet att jämföra levnadskostnadsindexen med Finlands Banks konsumtionsprisindex. Denna index, som har en byggnad, som mycket påminner om K. Socialstyrelsens levnadskostnadsindex och Sveriges riksbanks konsumtionsprisindex, har under dessa år utvecklat sig på följande sätt (årsmedeltal av 52 veckors indextal) [1])

	1/7—31	1932	1933	1934	1935
F. B:s index	(100,0)	103,3	100,2	98,1	100,3
Levn. kostn.[2])	(102,1?)	102,5	100,1	98,3	99,7.

Konsumtionsprisindexen anger sålunda en prisstegring för tiden 1/7 1931 till 1932. Till en del kan den bero på säsongrörelse hos konsumtionsprisindexen (c:a 5 promille) till en del på att de pris, som i indexen valts till utgångspris och som daterats till den 1 juli 31, icke äro fullt jämförbara med prisuppgifter, som inhämtats för de senare veckorna. Till en del utgöra utgångsprisen årsmedeltal för år 1931, varför de ej äro några pris som verkligen existerat. Huvudsakligen beror dock prisstegringen på växelkursernas stegring i samband med övergivandet av guldmyntfoten hösten 1931. I konsumtionsprisindexen ingår ett icke ringa antal importvaror (vetemjöl, kaffe, socker, frukter, medicin, bensin, automobiler m.m.). Levnadskostnadsindexen reagerade ej märkbart för valutakursernas stegring. Ännu en detalj har

[1]) Bank of Finland Monthley Bulletin, 1936, nr 10.

[2]) Decimalkomman flyttad ett steg, 102,1 är medeltalet av juni och juli månads indextal.

man att ihågkomma vid en jämförelse mellan levnadskostnadsindexen och Finlands Banks konsumtionsprisindex. Vid beräkningen av konsumtionsprisindexen inhämtas prisuppgifter enbart från Helsingfors. Enligt den officiella beräkningen av det relativa dyrhetstalet för föda på olika orter i Finland har dyrhetstalet för Helsingfors (Social Tidskrift, 1936, nr 1, s. 50, tabell II) utvecklat sig på följande sätt

1931	1932	1933	1934	1935
974,8	988,3	1003,1	1008,7	1026,7.

Om vi justera konsumtionsprisindexen med hänsyn härtill (födagruppen har i konsumtionsprisindexen vikten 44,4 procent) få vi följande justerade serie

1/7—31	1932	1933	1934	1935
100,0	102,7	98,9	96,6	98,0.

Denna serie överensstämmer något sämre med levnadskostnadsindexen än den ojusterade, vilket tyder på att övriga varugruppers prisindexer i Helsingfors jämfört med övriga orter rört sig i motsatt riktning. Enär dyrhetstalet för föda beräknas på samma prismaterial som levnadskostnadsindexen är dess tillförlitlighet ej särdeles stor.

I diagram II ha vi sammanställt levnadskostnadsindexen, den i tabell VI kolumn b (K. Socialstyrelsens nya budget) beräknade hypotetiska indexserien och Finlands Banks konsumtionsprisindex. Vi ha ännu tillfogat en kurva över partiprisindexens utveckling. Denna avviker åren 1931—35 rätt väsentligt från levnadskostnadsindexen. Orsaken härtill står att söka i partiprisindexens från levnadskostnadsindexen starkt avvikande byggnad. Importvarorna ha i partiprisindexen stor vikt, prisstegringen 1931—32 beror av valutakursernas höjning i samband med övergivandet av guldmyntfoten. Hyresgruppens kraftiga nedgång 1932—34 förklarar den avvikande utvecklingen hos levnadskostnadsindexen jämfört med partiprisindexen sedan år 1932. Kur-

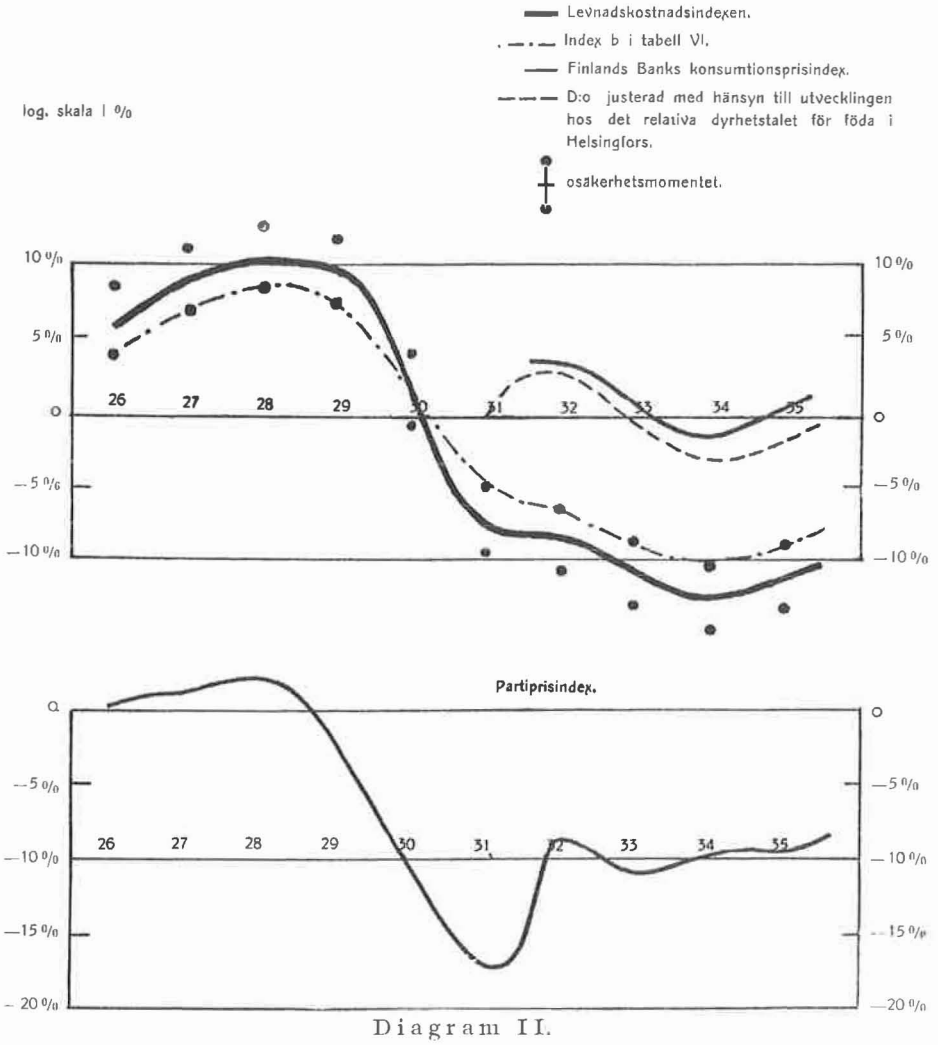


Diagram II.

vorna ha ritats med jämn krökning enbart med stöd av årsmedeltalen.

Beträffande K. Socialstyrelsens levnadskostnadsindex bry vi oss ej om att utföra en liknande grundlig analys av dess tillförlitlighet. En blick på de olika gruppindexernas ut-

veckling visar, att spridningsförhållandena äro ganska lika under den betraktade perioden i Sverige och Finland. Indexens modernare byggnad och komplettare varuförteckning gör, att de för olika ändamål önskvärda viktförändringarna äro mindre än beträffande levnadskostnadsindexen i Finland. Det synes därför som om dess osäkerhetsmoment vore något mindre än levnadskostnadsindexens i Finland. En uppskattning på 10—15 promille torde träffa den rätta storleksordningen. Kommerskollegii partiprisindex visar under denna period ett betydligt kraftigare prisfall än levnadskostnadsindexen. Orsaken står att söka däri att hyreskostnaderna företett en relativt till övriga grupper starkt stigande utveckling. Den stabila utvecklingen hos gruppen övriga utgifter samt den stigande skatteposten ha även bidragit till att minska det prisfall som levnadskostnadsindexen visat.

Finlands officiella levnadskostnadsindex och K. Socialstyrelsens levnadskostnadsindex i Sverige torde sålunda med reservation för ett par procents fel kunna användas för beräknandet av reallönernas utveckling under perioden 1926—34. Om man vill jämföra de med användande av dessa indexer framräknade reallönernas utveckling i Sverige och Finland med varandra, adderas osäkerhetsmomenten enligt antagande om inbördes oberoende, varigenom man får en osäkerhetsmarginal på c:a 3—4 %. Om det vore fråga om att få tag i ett relationstal mellan reallöner i Sverige och Finland, måste man ha tillgång till ett tal, som anger förhållandet mellan kronans och finska marks köpkraft gentemot konsumtionsvaror. Vid ett försök att praktiskt bestämma detta tal visar det sig, att detta indextals osäkerhetsmoment är en hel storleksordning större än levnadskostnadsindexernas osäkerhetsmoment. Man får vara nöjd, om man kan bestämma förhållandet mellan kronans och finska marks köpkraft gentemot konsumtionsvaror med en felmarginal på 10 %, beroende på den stora spridningen i förhållandena mellan prisen på ekvivalenta varor i Sverige och Finland.

THE BANK OF FINLAND'S CONSUMPTION PRICE INDEX

BY
LEO TÖRNQVIST, M. A.

INTRODUCTION.

After the abandonment of the gold standard in the autumn of 1931 the necessity was felt of a more subtle measure for the fluctuations of the internal price level and one that could be calculated with less delay than the official cost of living index and the wholesale price index. The Bank of Finland decided upon the calculation of a consumption price index intended to illustrate the movement of prices for all goods consumed in the country. For practical reasons, however, particulars of prices were only obtained for the City of Helsinki (Helsingfors), but these may be taken, broadly speaking, as representative of the movement of prices as a whole. The weighting system was in this case to indicate the share of the respective class of goods in the total value of consumption. An index constructed in such a manner has been advocated, i. a., by J. M. Keynes [1]) as an index of considerable importance in present day monetary policy. The consumption price index of the Swedish State Bank [2]), which closely resembles the Bank of Finland's consumption price index as regards the selection of classes of goods, is an instance of an index of this kind.

The calculation of the Bank of Finland's consumption price index was begun on January 2nd, 1932. It was originally calculated as a weighted arithmetical average of the special indices of the individual commodities and classes of goods. The index formula employed was the wellknown one:—

$$P\,{}^{t}_{t_o} = \Sigma\, c\,\frac{p_t}{p_{t_o}} : \Sigma\, c$$

The basic prices p_{t_o} refer to July 1st, 1931, with the exception of some seasonal commodities for which the average annual price in 1931 was taken as a base. p_t signifies the last week's prices. In determining the weights c some preliminary results of the investigation into the cost of living made in 1928 were made use of and were supplemented by estimates. In addition to this index with July 1st, 1931, as a base, a similarly weighted index was calculated with the previous week's prices as a base. The latter index at times recorded very appreciable differences from the result obtained by dividing the former index number by the corresponding index number of the previous week. The one might show a rise in price, while the other recorded a fall. This circumstance made an investigation into the method of calculation desirable. The Bank of Finland entrusted the author of this article with the task of improving this method and in the autumn of 1935 a new scheme was adopted by the Bank. It was also decided to recalculate the series of indices for the past.

[1]) J. M. Keynes, A Treatise on Money, Vol. I, p. 57, London, 1930.
[2]) Erik Lindahl, Sveriges Riksbanks konsumtionsprisindex, Ekonomisk tidskrift, 1933.

THE PRESENT CONSTRUCTION OF THE INDEX.

The Bank of Finland's present consumption price index is calculated as a weighted logarithmic (= geometric) average, vulgar logarithms being used. The weights are variable in principle and the chain method is employed in the calculation. As soon as it is established that the weighting system employed no longer constitutes a good approximation for the per mille rates of the consumers' total expenditure that fall to the share of the respective class of goods, an adjustment of the weights is undertaken and the adjusted weighting system is applied in calculating the changes from week to week.

Mathematically the calculations are made according to the following formula:—

$$\log P^{t}_{t_o} = \log P^{t'}_{t_o} + \Sigma\, c\, (\log p_t - \log p_{t'}) : \Sigma\, c$$

Thus, to the logarithm for the index number for the preceding week the weighted average of the difference between the logarithm for last week's price and the logarithm for the previous week's price, is added. As most of the commodity prices do not alter at all from one week to another, the work of calculation is very slight.

The derivation of this index formula was made by starting from Fisher's various tests of index formulæ [1]) by means of adopting the view of a continuous transition from one constellation of prices to another. In this way we obtain as a solution of the theoretical aspect of the index problem a curve integral [2]) which the index formula given above most closely approximates for practical purposes. The value of this curve integral depends on the manner in which we imagine the proportionate distribution of the expenditure on different classes of commodities to change in the transition from one constellation of prices to another. If we assume that this distribution remains unaltered from one week to another, the integral gives the logarithmic average as a result for the week-to-week change in the index.

The arithmetical average cannot be employed for calculations by the chain method, as it would result in the index calculated in this manner growing beyond all bounds owing to the fact that the arithmetical average has an upward bias. In order to investigate this phenomenon the product of the arithmetical week-to-week index numbers for the food index was reckoned out and it was discovered that an index number calculated by the chain method would in three years (156 weeks) exceed the directly calculated arithmetical index number by about 20 per cent. The arithmetical week-to-week index numbers were thus on an average about 1 1/4 per cent too high.

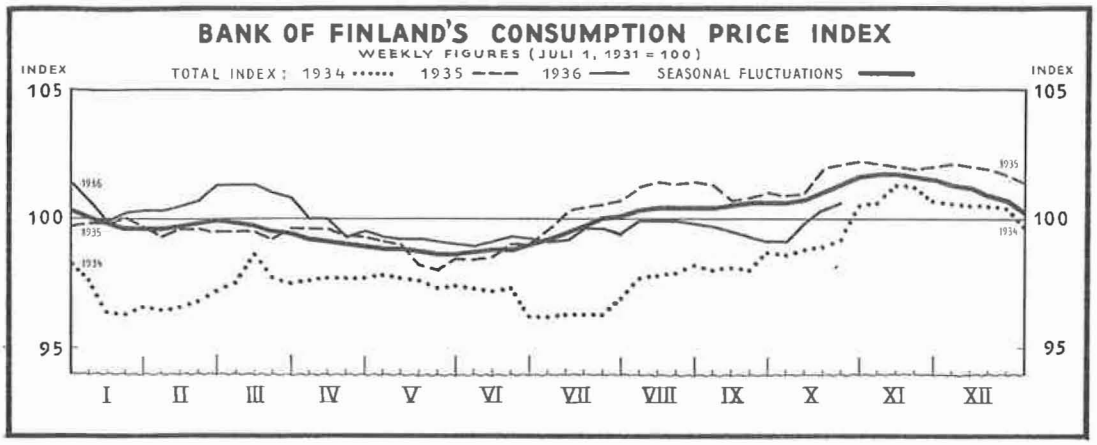
In addition to this change in the actual method of calculation some small alterations were made in the weighting system and the choice of commodities. For instance, some commodities of an exceptionally seasonal nature, such as fresh fish and vegetables, were omitted. The importance of these commodities in the household budget varies very greatly in the course of the year in connection with the great and fairly irregular fluctuations in price, so that their inclusion in the index is not justified without fairly extensive calculations. In addition it was decided in regard to potatoes, the price of which records a powerful temporary rise in July and August owing to the change from old to new potatoes, that their price should remain stable for that period.

The possibility of altering the weighting system was only utilised once on the occasion, when the class of alcoholic beverages was included in the index for the first time on July 16th, 1932. In case no changes in the weights are made, precisely the same result is

[1]) Irving Fisher, The Making of Index Numbers. Cambr., Mass., 1922.

[2]) This solution by means of an integral first occurs in F. Divisia: L'Indice monétaire et la théorie de la monnaie, Revue d'Economie politique, 1927. The curve integral may be defined by the formula:—

$$\log P^{t}_{t_o} = \Sigma \int_{t_o}^{t} c\,(t)\, d\,(\log p_t)\,;\ \Sigma\, c\,(t) = 1$$



obtained by the chain method as by a direct comparison with the basic date.

The Bank of Finland's consumption price index is calculated once a week, the calculations being made by the Research Office of the Ministry for Social Affairs. The details of prices that are obtained solely for Helsinki, refer to 155 kinds of commodities. A list of the principal commodities and main classes of goods as well as their respective weights will be found in appendix III.

In connection with the rearrangement a new division into classes was introduced, the commodities most susceptible to changes in the business cycle being combined in one class entitled *"sensitive consumption price index"*. The remaining class of rarely changing prices in which the index for housing predominates, represents 30 per cent of the total expenditure.

THE COURSE OF THE CONSUMPTION PRICE INDEX.

The course of the total consumption price index from January 2nd, 1932, to October 17, 1936, is shown in the diagram above. The monthly averages of the index are compiled in appendix I.

The index indicated a rise in price of 5.7 per cent from the basic date July 1st, 1931, to January 2nd, 1932. This rise was due partly (1/2 to 1 per cent) to the index being seasonally lower on July 1st than the average for the year. Principally, however, it was a consequence of the rise in foreign exchange rates in connection with the suspension of the gold standard in the autumn of 1931. The imported goods included in the index (wheat flour, coffee, sugar, fruit, petrol, motor tyres, medicines, etc.) rose in price at that time. During 1932—1933 the index recorded a fall in prices that was mainly due to the course followed by the level of rents. In the autumn of 1934 there was a rise in price in the index of about 2 per cent. It occurred exclusively in the food class in which the price of milk and butter rose chiefly. During 1935 there was still a rising tendency which, however, grew weaker in the course of 1936.

As it is interesting to compare the consumption price index with the official cost of living index, the relative course of both indices is given in broad lines in the following table showing their annual averages:—

Year	Bank of Finland's consumption price index	Official cost of living index
1931, July 1	100.0	100.0 [1])
1932	103.3	100.4
1933	100.2	98.1
1934	98.1	96.3
1935	100.3	97.7

[1]) Average for the index numbers for June and July.

The correspondence in the course of development is very good in view of the fact that the cost of living index only refers to about thirty commodities, while the Bank of Finland's consumption price index, as already stated, includes 155 commodities. Besides, the prices in the cost of living index refer to 21 places, whereas the consumption price index is based solely on prices in the capital. The cost of living index only includes coffee and sugar among imported goods, so that the rise in foreign exchange rates after the suspension of the gold standard scarcely affected this index.

The diagram on the opposite page illustrates the monthly averages of the total index and the indices of sensitive prices, food prices and rarely changing prices.

The course of the sensitive price index is very similar to that of the total consumption price index, as is quite natural in view of its great preponderance, 70 per cent, in the total index. However, it avoids some of the sudden changes in the consumption price index and is throughout higher. The annual averages for this index are:—

1932	1933	1934	1935
104.7	104.0	103.3	107.0

They indicate a fall in prices up to 1934 and subsequently a rise of a little more than 3 1/2 per cent.

The main class in the sensitive consumption price index consists of food and, except for some great seasonal changes, the course of the indices for sensitive prices in general and for food prices are very similar. The weight of the food index, 44.4 per cent, is over half of the weight of the sensitive index. The annual averages of the food index are:—

1932	1933	1934	1935
105.4	105.3	105.0	111.3

These figures likewise indicate a slight fall in price up to 1934 and subsequently a rise of 6 per cent.

The sudden alterations of the total index are due to the index for commodity prices that are rarely subject to change. This index includes the sub-classes Housing, electricity and gas, Hygiene, Alcoholic beverages, Travel, postage, etc., Education, and Recreation. The index for these prices followed a course that differed very considerably from the consumption price index. Their curve has the form of steps. The annual averages for 1932—1935 are as follows:—

1932	1933	1934	1935
100.3	91.9	87.0	86.3

The averages indicate a constant fall in prices. The fall is mainly due to the housing class that represents more than half the weight of this group. The level of rents which is only determined once a year is calculated as a weighted average of rents in Helsinki for two rooms and a kitchen with and without central heating. The housing index records the following changes:—

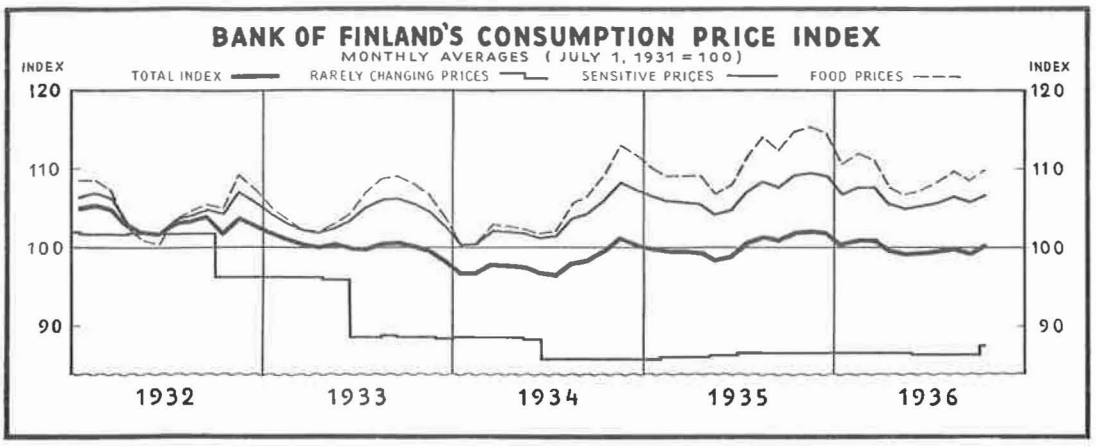
1932 Oct. 1	1933 June 17	1934 June 23	1935 June 29	1936 Oct. 3
91.0	79.7	75.7	77.7	79.1

The annual averages for all commodity classes in the consumption price index will be found in appendix II. The class of medicine has the highest index number. This great rise in price from the basic date is due to the rise in the rates of foreign exchange.

SEASONAL FLUCTUATIONS.

The consumption price index records an obvious seasonal variation. In order to study the character of this phenomenon we endeavoured to determine a normal seasonal index by using a modified Harvard method [1]). Link relatives were replaced in our modified method by week-to-week logarithm differences for 4-week moving averages. Instead of the median recommended by Person we employed an average weighted by binomial coefficients. The method was applied to the sensitive consumption price index as well as to the food index and it appeared that the food index dominated the seasonal variations almost entirely. The seasonal

[1]) Warren M. Person, Indices of General Business Conditions, Cambr., Mass., 1919.



index for the total consumption price index was then obtained by multiplying the seasonal index of the sensitive index by 0.7 and described the curve shown in the diagram on page 29. The consumption price index is thus in May normally slightly over 1 per cent below the annual average, but in November it is fully 1 1/2 per cent above the average. By varying the method of calculating the seasonal index we came to the conclusion that the extent of the seasonal amplitude, based on the available material that only refers to 4 1/2 years, can be determined with a precision of about 3 per mille.

DEVIATIONS IN THE DEVELOPMENT OF PRICES.

With a view to measuring the reliability with which the consumption price index illustrates the general development of prices of articles of consumption, it is interesting to know, to what extent the different commodity prices developed in proportion to the total consumption price index. In order to be able to measure the deviations from this course of development we calculated the standard deviations of the logarithms of the special indices for the various commodities. Mathematically the calculations were made according to the following formula:—

$$s^2 = \Sigma\, c \left(log\ p_t - log\ p_{t_o} - log\ P^t_{t_o} \right)^2 : \Sigma\, c$$

The results obtained, which in reality represent natural logarithms, but can be regarded as per mille numbers with good approximation, are compiled in the following table:—

STANDARD DEVIATION IN THE MAIN CLASSES

Natural logarithms (= per mille numbers)

	1932 July 1	1933 Jan. 7	1934 Jan. 6	1935 Jan. 5	Average
Food index	125	110	120	115	118
Sensitive consumption index	125	116	120	125	122
Total consumption price index	120	123	150	177	143

The increase in the standard deviation in the case of the consumption price index from 1933 to 1934 is due to the great fall in the housing index in relation to the consumption price index. This fall can be explained as a belated adjustment of the level of rents to the general fall in prices in 1928—1932. On the whole prices for the different commodities seem to have developed in a fairly good proportion to the general trend.

In order to examine, to what extent a seasonal movement can be observed in the dispersion, we calculated the standard deviation in the food class quarterly.

QUARTERLY DISPERSION IN THE FOOD CLASS.

	1932			1934	
Jan. 2	.....	125	Jan. 6	.....	120
Mch. 26	.....	120	Mch. 31	.....	129
June 25	.....	131	June 30	.....	157
Oct. 1	.....	124	Sept. 29	.....	143
	1933			1935	
Jan. 7	.....	110	Jan. 5	.....	115
April 1	.....	135	Mch. 30	.....	102
July 1	.....	120	July 6	.....	112
Sept. 30	.....	155	Oct. 5	.....	123
Dec. 30	.....	128	Dec. 28	.....	109

The figures indicate that the deviations in the development of food prices are rather larger during the autumn months than at other times of the year. In examining the dispersion it should be remembered that the selection of the basic prices also influences the results. If, instead of the prices on the basic date of July 1st, 1931, the average for the whole period under review had been chosen as a starting point, the dispersion would have been reduced on the average by about 30 per cent. The development of the dispersion would then also have changed to some extent.

CONCLUSION.

The logarithmic method of calculation provides the possibility, thanks to the weighting system being changeable, of an elastic adjustment to the alterations in the composition of consumption that occur in the course of time. The particulars of prices and the weighting system employed at present in calculating the Bank of Finland's consumption price index are, of course, not the best conceivable. By examining, to what extent the result is affected, if the weighting system is varied within reasonable limits, we came to the conclusion that the margin of uncertainty, which grows with the deviations in the development of prices, is $^1/_2$ to 1 per cent in extent. The consumption price index therefore gives a fairly reliable idea of the development of prices for articles of consumption in Helsinki. The possibility of error in estimating the development of prices in the whole country that arises, if the course of this index alone is examined, is counterbalanced partly by the fact that it seems probable that Helsinki prices are likely to react more rapidly to forces that change the level of prices. By this means attention can be drawn at an earlier stage to such tendencies in the course of prices for articles of consumption as may possibly call for attention from the point of view of monetary policy.

APPENDIX I

BANK OF FINLAND'S CONSUMPTION PRICE INDEX

MONTHLY AVERAGES. JULY 1, 1931=100

Year and month	*Total consumption price index*		Index of sensitive prices	Index of rarely changing prices	Food	Housing, electricity, & gas
	Seasonally adjusted	Unadjusted for seasonal variations				
1932						
January	*105.2*	*105.0*	106.4	101.8	108.6	100.0
February	*105.7*	*105.3*	106.9	101.7	108.6	100.0
March	*105.2*	*104.9*	106.3	101.7	107.3	100.0
April	*103.7*	*102.8*	103.3	101.8	102.6	100.0
May	*103.2*	*101.9*	101.9	101.8	101.0	100.0
June	*102.9*	*101.7*	101.6	101.8	100.5	100.0
July	*103.4*	*103.0*	103.5	101.8	103.6	100.0
August	*103.0*	*103.4*	104.1	101.8	104.6	100.0
September	*103.5*	*104.0*	104.8	101.8	105.5	100.0
October	*100.9*	*101.8*	104.3	96.2	105.1	91.7
November	*102.0*	*103.7*	107.0	96.2	109.3	91.7
December	*101.8*	*102.8*	105.8	96.2	107.5	91.7
1933						
January	*102.0*	*101.8*	104.3	96.2	105.1	91.8
February	*101.4*	*101.0*	103.1	96.2	103.7	91.8
March	*100.7*	*100.4*	102.2	96.2	102.4	91.8
April	*101.0*	*100.1*	101.9	96.1	102.0	91.8
May	*101.8*	*100.5*	102.5	95.9	103.0	91.8
June	*101.1*	*99.9*	103.4	92.2	104.3	86.5
July	*100.2*	*99.8*	105.1	88.5	107.1	81.2
August	*100.1*	*100.5*	106.1	88.6	108.7	81.2
September	*100.1*	*100.6*	106.2	88.6	109.1	81.2
October	*99.3*	*100.2*	105.6	88.6	108.2	81.2
November	*97.9*	*99.6*	104.6	88.6	106.7	81.2
December	*97.2*	*98.2*	102.8	88.4	103.9	81.2
1934						
January	*96.9*	*96.7*	100.4	88.5	100.3	81.2
February	*97.2*	*96.8*	100.5	88.5	100.5	81.2
March	*98.1*	*97.8*	102.1	88.5	102.9	81.2
April	*98.6*	*97.7*	102.0	88.5	102.7	81.2
May	*98.8*	*97.5*	101.8	88.4	102.4	81.2
June	*98.0*	*96.8*	101.2	87.3	101.8	79.7
July	*96.9*	*96.5*	101.4	85.7	102.1	77.4
August	*97.5*	*97.9*	103.7	85.7	105.7	77.4
September	*97.8*	*98.3*	104.3	85.7	106.6	77.4
October	*98.5*	*99.4*	105.9	85.7	109.2	77.4
November	*99.3*	*101.0*	108.2	85.7	113.0	77.4
December	*99.3*	*100.3*	107.4	85.7	111.6	77.4
1935						
January	*100.0*	*99.8*	106.6	85.7	110.2	77.4
February	*99.9*	*99.5*	105.9	86.0	109.1	77.4
March	*99.8*	*99.5*	105.8	86.0	109.1	77.4
April	*100.2*	*99.3*	105.6	86.0	109.2	77.4
May	*99.7*	*98.4*	104.2	86.3	106.9	77.4
June	*100.1*	*98.9*	104.8	86.4	107.9	77.5
July	*100.9*	*100.5*	107.1	86.5	111.7	77.7
August	*100.9*	*101.3*	108.4	86.5	113.9	77.7
September	*100.4*	*100.9*	107.7	86.5	112.4	77.7
October	*100.9*	*101.8*	109.1	86.5	114.7	77.7
November	*100.3*	*102.0*	109.5	86.5	115.3	77.7
December	*100.8*	*101.8*	109.1	86.5	114.6	77.7
1936						
January	*100.5*	*100.3*	106.8	86.5	110.6	77.7
February	*101.2*	*100.8*	107.6	86.5	111.9	77.7
March	*101.1*	*100.8*	107.6	86.5	111.1	77.7
April	*100.4*	*99.5*	105.5	86.5	107.6	77.7
May	*100.4*	*99.1*	104.9	86.5	106.7	77.7
June	*100.4*	*99.2*	105.3	86.4	107.3	77.7
July	*99.9*	*99.5*	105.7	86.4	108.3	77.7
August	*99.4*	*99.8*	106.5	86.4	109.7	77.7
September	*98.8*	*99.3*	105.8	86.4	108.5	77.7

5

APPENDIX II

BANK OF FINLAND'S CONSUMPTION PRICE INDEX

YEARLY AVERAGES. JULY 1, 1931=100

Class of commodities	1932	1933	1934	1935
Food	105.4	105.3	105.0	111.3
Tobacco	106.9	98.0	98.0	98.0
Clothing	104.9	104.8	104.4	103.0
Furniture	98.9	95.8	95.8	95.4
Means of transport	114.9	114.9	104.6	103.9
Writing materials	116.8	118.4	121.3	122.0
Laundry and cleaning	96.7	92.8	90.3	89.4
Firewood and oil fuel	95.9	94.4	91.9	95.2
Index of sensitive prices	***104.7***	***104.0***	***103.3***	***107.0***
Housing, electricity and gas	97.8	85.9	79.2	77.6
Housing	97.6	84.7	77.5	75.9
Electricity and gas	100.0	102.5	102.5	102.5
Hygiene	105.9	103.8	103.6	104.3
Medicine	137.6	133.0	132.4	132.5
Alcoholic beverages	101.5[1])	104.8	104.4	120.2
Travel, postage etc.	105.8	106.0	106.0	106.0
Education	107.5	104.1	107.5	107.5
Recreation	101.8	99.0	90.3	89.6
Index of rarely changing prices	***100.3***	***91.9***	***87.0***	***86.3***
TOTAL CONSUMPTION PRICE INDEX	***103.3***	***100.2***	***98.1***	***100.3***

[1]) Average for the period $^{16}/_{7}$—$^{31}/_{12}$.

APPENDIX III

LIST OF COMMODITIES IN THE BANK OF FINLAND'S CONSUMPTION PRICE INDEX

155 KINDS

Commodity	Weight	Commodity	Weight	Commodity	Weight
Food	*444*	Anchovies	6	Petrol, rectified	1
Milk	77	Eggs	10	Petrol, crude	3
Dairy butter	26	Coffee	26	*Writing materials*	*6*
Peasant butter	25	Lump sugar	13	*Laundry and cleaning*	*20*
Cream	6	Granulated sugar	13	*Firewood and oil fuel*	*26*
Cheese	6	Potatoes	17	Firewood	22
Margarine	8	Plums	2	Oil fuel	4
Lard	4	Mixed fruit	2	***Index of sensitive prices***	***700***
Wheat flour	17	Apples	3	*Housing, electricity and gas*	*194*
Rye flour	8	Oranges	3	Housing	180
Rye groats	8	Bananas	2	Electricity	7
Oatmeal	4	Peas	3	Gas	7
Barley meal	1	Malt beverages	5	*Hygiene*	*20*
Rice meal	4	Beer	1	Hospital fees	8
Semolina meal	4	Chocolate	1	Medicine	4
Rye bread	6	Caramels	1	Other expenses	8
Rye bread, soft	14	Salt	1	*Alcoholic beverages*	*10*
Yeast bread	10	Dinner	10	*Travel, postage etc.*	*41*
French loaves	10	Coffee	10	Train tickets	21
Buns	10	*Tobacco*	*20*	Tram fares	7
Beef	18	*Clothing*	*130*	Postage, telegrams	7
Veal	4	Wearing apparel	63	Motor bus fares	3
Mutton	4	Underclothing	35	Motor car fares	3
Mutton, cured	2	Footwear	24	*Education*	*21*
Sausages	11	Rubber wear	8	*Recreation*	*14*
Pork	8	*Furniture*	*38*	***Index of rarely changing prices***	***300***
Pork, salted	4	*Means of transport*	*16*	***TOTAL CONSUMPTION PRICE INDEX***	***1000***
Pork, smoked	4	Motor cars	4		
Baltic herrings, salt	6	Motor tyres	4		
Herrings	6	Bicycles	4		

Vilket är förhållandet mellan finska markens och svenska kronans köpkraft?

Av

Leo och Erik Törnqvist.

En erfarenhet, som de vilka vistats såväl i Sverige som Finland gjort är, att växelkursen (1 kr = 11,71 Fmk) icke ger något riktigt uttryck för förhållandet mellan kronans och markens köpkraft. Det är dyrt att leva i Sverige, säger man i Finland, medan svensken reser till Finland för att leva billigt. Dessa kvalitativa omdömens riktighet betvivlas knappast, men då det gäller att fastställa vilken växelkurs, som skulle göra det lika dyrt att leva i Finland som i Sverige, ställes man inför ett rätt invecklat problem. De svårigheter, som möta vid ett försök att utgående från resultat av konsumtionsundersökningar och tillbudsstående prisstatistik beräkna denna »köpkraftsparitetsväxelkurs», äro dels av principiell dels av praktisk art.

Ett karaktärsdrag hos indexproblemen, till vilken problemkrets uppgiften hör, är att de tal, som man önskar bestämma, icke låta sig fastställas med godtycklig noggrannhet. Det ligger i sakens natur, att de sökta indextalen äro mer eller mindre obestämda inom ett visst variationsområde. Beroende på vilken av flere sinsemellan konkurrerande synpunkter och bestämningsmetoder man väljer, förskjutes resultatet i den ena eller andra riktningen; vanligen dock ej särdeles mycket. Det synes förnuftigast att ej uppfatta indextalen såsom tal i den rena matematikens mening, vilka

teoretiskt kunna bestämmas med godtyckligt många decimaler, utan såsom empiriska tal med en viss utsträckning på talskalan. Varje försök att reducera variationsområdet till en punkt är dömt att misslyckas, i det att det begrepp, som man försöker mäta, icke får rum innanför alla detaljbestämmelser, som äro nödvändiga för att indextalet skall bli entydigt bestämt. Sådana begrepp som allmän prisnivå, valutans köpkraft, lönenivå, produktionsvolym, nationalinkomst m.m., vilka man svårligen kan undvara i en ekonomisk analys, utgöra storheter, som tåla att mätas i en grov skala, men som ifall de »fixeras med förstoringsglas» bli mer eller mindre meningslösa. Genom att tillordna indextalen en intervall inom vilken de ej äro determinerade, får man för de sökta storheterna kvantitativa uttryck, som på bästa möjliga sätt motsvara de till sin natur diffusa begreppen. För att rätt kunna värdesätta indextal är det av intresse att känna storleksordningen av denna osäkerhetsmarginal.

Begreppet penningenhetens köpkraft kan uppfattas från åtminstone tvenne i vårt fall intressanta synpunkter, konsumentens och producentens.

Från konsumentens synvinkel framstår penningens köpkraft såsom en i tid och rum varierande egenskap hos penningenheten att köpa olika mängder direkt konsumerbara varor och tjänster. Frågan i vilken mån hänsyn skall tagas till de olika nyttigheternas pris vid det kvantitativa fastställandet av denna egenskap bildar det så kallade vägningsproblemet. Innan man dock överhuvud kan ge sig i kast med detta problem, måste man av det obestämda konsumentbegreppet skapa sig en statistiskt gripbar storhet. Man måste avgöra huru den konsumentgrupp skall vara beskaffad, inom vilken de olika nyttigheternas ekonomiska betydelse skall studeras. Man måste m.a.o. avgränsa en ekonomikrets som motsvarighet till konsumentbegreppet vid fastställandet av variationerna i penningenhetens köpkraft.

Ur producentsynpunkt åter är det intressant att avgöra vilka kvantiteter reala produktionskostnadselement som i på ett eller annat sätt avgränsade ekonomikretsar vid olika

tidpunkter kunna köpas för en enhet av penningen. Besvarandet av den uppställda frågan möter i detta fall i huvudsak samma svårigheter som vid anläggandet av konsumentsynpunkten. De praktiska svårigheterna i att erhålla jämförbara prisuppgifter äro mycket stora.

Förutom dessa tvenne grundinställningar till problemet att fastställa vad som skall förstås med begreppet köpkraft kan naturligtvis en mängd andra ståndpunkter vara av intresse. Intressant kunde det t.ex. vara att undersöka huru köpkraftspariteten ställer sig för ekonomikretsar, som avgränsats i avseende å levnadsstandardens höjd, olika sociala grupper såsom olika slag av arbetare, tjänstemän o.s.v., eller för skilda producentkretsar såsom hemmamarknadsproducenter, industrier underkastade utländsk konkurrens, exportindustrier o.s.v. För de olika synpunkterna erhållas i någon mån olika indextal. Genom att enligt någon mer eller mindre motiverbar princip sammanväga dessa resultat kan man erhålla nya indextal, som kontinuerligt variera från det största till det minsta av de primärt erhållna resultaten. Den intervall på talskalan, som sålunda erhålles och vars alla punkter genom variation av problemuppläggningen kunna komma ifråga som numeriska lösningar, är att betrakta som det kvantitativa uttrycket för det allmänna, diffusa köpkraftsparitetsbegreppet. Samtliga delintervaller av denna osäkerhetsintervall äro dock icke att betrakta såsom inbördes lika goda uttryck för det allmänna begreppet, utan böra de förses med en begreppsmassa varierande mellan 0 och 1 beroende på delintervallens läge och bredd. Massan noll tillordnas en punkt och massan 1 hela osäkerhetsintervallen. Ju större massa en delintervall har desto större del av det diffusa begreppet motsvarar intervallen. Det synes naturligt att vänta sig att begreppsmassans förhållande till intervallbredden, d.v.s. begreppstätheten, skall vara större i osäkerhetsintervallens mitt än vid dess gränser. Förutom den av begreppens obestämdhet härrörande osäkerhetsintervallen tillkomma vid praktiska försök att bestämma ett indextal osäkerhetsmoment, som bero på de metodtekniska

och praktiska svårigheterna vid lösandet av indexproblemen. Dessa osäkerhetsmoment ha en något annan karaktär än den principiella osäkerhetsintervallen, i det att man genom förbättrade beräkningsmetoder och genom tillgång till fullständigare och bättre primärmaterial kan minska osäkerheten hos resultatet. Den principiella osäkerhetsintervallen kan däremot icke minskas på annat sätt än genom att uppspjälka det allmänna begrepp, som indextalet skall motsvara, i flere, snävare definierade begrepp. Detta kan vara motiverat vid sådana tillfällen, då man på grund av storleken hos den principiella osäkerhetsintervallen icke kan draga de önskade slutsatserna. Det kan sålunda t.ex. visa sig nödvändigt att i växelkursteorin operera med ett snävare begrepp än det allmänna köpkraftsparitetsbegreppet för att man statistiskt skall kunna verifiera teorins satser.

De metodtekniska lösningarna av indexproblemet.

Vi antaga att vi löst den principiella synpunktsfrågan och lyckats avgränsa en statistiskt gripbar ekonomikrets i Finland och en motsvarande krets i Sverige, t.ex. arbetare och befattningshavare i de båda länderna. De för problemet att fastställa köpkraftsparitetsväxelkursen relevanta ekonomiska storheterna äro, ifall vi hålla oss till konsumentsynpunkten: 1) de sammanlagda utgifterna för konsumtionen under den betraktade perioden, 2) utgiftsfördelningen på olika varukategorier, 3) antalet hushåll inom kretsen och deras sammansättning av konsumerande individer, 4) konsumerade kvantiteter av olika varuslag, 5) varuprisen på olika orter och vid olika tider. Med utgångspunkt från dessa storheter erhållas genom medeltalsbildningar de för det s.k. normalhushållet gällande, konsumtionen belysande talen. Genom att företaga levnadskostnadsundersökningar söker man bestämma närmevärden för dessa tal. Härvid utväljes enligt någon princip ett antal familjer, vilkas affärshändelser under en bokföringsperiod av i regel ett år studeras.

De metodtekniska lösningarna till indexproblemet kunna med begagnande av matematiska symboler sammanfattas i s.k. indexformler. I det följande förse vi de till den finländska ekonomikretsen hörande storheternas matematiska symboler med det övre tecknet (1) de till den svenska kretsen hörande med tecknet (2). Pris betecknas med bokstaven p, kvantiteter med q. Den bråkdel av totalutgiften, som kommer på de olika varukategorierna, betecknas med bokstaven c. För angivande av olika varuslag förses dessa symboler med det undre tecknet i, som betecknar varans nummer i en varuförteckning. Summation över samtliga konsumerade varuslag betecknas genom att framför det allmänna uttrycket för de termer, som skola adderas, sätta summatecknet Σ. Symbolen $p_i^{(1)}$ betecknar sålunda ett pris för varan i hänförande sig till den finländska ekonomikretsen, närmare bestämt det medelpris, som normalhushållet anses ha erlagt och som erhålles genom att dividera penningvärdet av de olika hushållens sammanlagda konsumtion med antalet inom kretsen under den betraktade tidsperioden konsumerade enheter av varan i. Då denna totalkvantitet divideras med antalet hushåll inom kretsen, erhålles den kvantitet $q_i^{(1)}$, som normalhushållet anses ha konsumerat under den ifrågavarande perioden. De motsvarande till den svenska kretsen hörande med övre tecknet (2) försedda storheterna behöva icke hänföra sig till samma tidsperiod som för den finländska kretsen.

Vi tänka oss att ett för den finländska ekonomikretsen representativt normalhushåll överflyttar till Sverige och införlivas med den svenska kretsen. Normalhushållets inkomster antagas fortfarande bli utbetalade i finska mark till samma belopp per tidsenhet som före överflyttningen. I Sverige växlas den finländska valutan i kronor. Ekonomikretsarna betraktas som platser där varuslaget i har priset $p_i^{(1)}$ mark respektive $p_i^{(2)}$ kronor. En viss anpassning efter prisförhållandena i Sverige antages äga rum. Problemet består i att avgöra vilken växelkursen borde vara för att normalhushållets levnadsstandard varken skall försämras eller

förbättras genom överflyttningen till den nya ekonomikretsen.

Ifall vi beteckna den sökta växelkursen med X_f och de varukvantiteter, som det finländska normalhushållet konsumerar efter överflyttningen till Sverige och som representera samma levnadsstandard som de i den finländska kretsen konsumerade kvantiteterna $q_i^{(1)}$, med symbolen $q_i^{(1,2)}$ ha vi under ovannämnda antaganden:

$$\Sigma p_i^{(1)} q_i^{(1)} = X_f \Sigma p_i^{(2)} q_i^{(1,2)}, \text{ varav följer } X_f = \frac{\Sigma p_i^{(1)} q_i^{(1)}}{\Sigma p_i^{(2)} q_i^{(1,2)}}.$$

Emedan man saknar kännedom om kvantiteterna $q_i^{(1,2)}$ måste man för att man skall kunna erhålla en praktiskt användbar indexformel försöka eliminera dessa storheter. Ifall man gör det rimliga antagandet att övergången från det i Finland konsumerade varuaggregatet till det nya aggregatet (huvudsakligen) förorsakats av att detta senare ställt sig billigare än det tidigare konsumerade, gäller relationen:

$$\Sigma p_i^{(2)} q_i^{(1,2)} \leq \Sigma p_i^{(2)} q_i^{(1)}.$$

Genom att i uttrycket för X_f ersätta summan $\Sigma p_i^{(2)} q_i^{(1,2)}$ med den enligt antagandet större summan $\Sigma p_i^{(2)} q_i^{(1)}$ får man en undre gräns för den sökta växelkursen:

$$X_f \geq \frac{\Sigma p_i^{(1)} q_i^{(1)}}{\Sigma p_i^{(2)} q_i^{(1)}}.$$

En övre gräns för X_f fås genom att antaga, att orsaken, varför normalhushållet icke redan i Finland konsumerade de varukvantiteter, som efter överflyttningen till Sverige förbrukades, var, att det i Finland konsumerade aggregatet ställde sig billigare, d.v.s. att

$$\Sigma p_i^{(1)} q_i^{(1)} \leq \Sigma p_i^{(1)} q_i^{(1,2)}.$$

Om man i uttrycket för X_f ersätter $\Sigma p_i^{(1)} q_i^{(1)}$ med $\Sigma p_i^{(1)} q_i^{(1,2)}$ får man sålunda med stor sannolikhet ett tal som är större

än X_f. Den sökta växelkursen torde alltså ligga mellan gränserna.

$$\frac{\Sigma p_i^{(1)} q_i^{(1,2)}}{\Sigma p_i^{(2)} q_i^{(1,2)}} \geqq X_f \geqq \frac{\Sigma p_i^{(1)} q_i^{(1)}}{\Sigma p_i^{(2)} q_i^{(1)}}$$

Vi tänka oss på samma sätt det svenska normalhushållet överflyttat till Finland och låta X_s beteckna det antal mark, som hushållet borde erhålla för en krona, och $q_i^{(2,1)}$ de efter flyttningen konsumerade varukvantiteterna, vilka motsvara samma levnadsstandard som de i Sverige konsumerade kvantiteterna $q_i^{(2)}$. Enligt detta betraktelsesätt fås sannolikhetsrelationerna:

$$\frac{\Sigma p_i^{(1)} q_i^{(2)}}{\Sigma p_i^{(2)} q_i^{(2)}} \geqq X_s \geqq \frac{\Sigma p_i^{(1)} q_i^{(2,1)}}{\Sigma p_i^{(2)} q_i^{(2,1)}}.$$

Växelkurserna X_s och X_f behöva naturligtvis icke vara identiska. Ifall vi till de tidigare förutsättningarna foga antagandet, att det svenska normalhushållet och det motsvarande finländska icke väsentligt skilja sig från varandra i avseende å levnadsstandardens höjd och de till desamma hörande konsumerande individernas behov, synes det rimligt att vänta sig att skillnaden mellan dessa växelkurser är relativt liten. Under denna sista förutsättning torde kvantiteterna $q_i^{(1,2)}$ vara approximativt proportionella med $q_i^{(2)}$ och $q_i^{(2,1)}$ med $q_i^{(1)}$. Vore de exakt proportionella skulle den övre gränsen för X_f sammanfalla med den statistiskt lättare gripbara övre gränsen för X_s, medan den undre gränsen för X_s vore lika med den undre gränsen för X_f, ty proportionalitetsfaktorn kan bortförkortas. Vid ungefärlig proportionalitet gäller samma sak approximativt. Om X betecknar ett medelvärde till de båda växelkurserna X_f och X_s förefaller det sålunda mycket sannolikt att detta tal ligger mellan gränserna:

$$\frac{\Sigma p_i^{(1)} q_i^{(2)}}{\Sigma p_i^{(2)} q_i^{(2)}} \geqq X \geqq \frac{\Sigma p_i^{(1)} q_i^{(1)}}{\Sigma p_i^{(2)} q_i^{(1)}}$$

Dessa gränser, som gå under namn av *Paasche-Laspeyre*'s indexformler, kunna via förlängningar och divisioner även skrivas i formen:

$$\Sigma c_i^{(2)} \frac{p_i^{(1)}}{p_i^{(2)}} \geqq X \geqq 1 : \Sigma c_i^{(1)} \frac{p_i^{(2)}}{p_i^{(1)}}.$$

Vägningskoefficienterna

$$c_i^{(\alpha)} = \frac{p_i^{(\alpha)} q_i^{(\alpha)}}{\Sigma p_i^{(\alpha)} q_i^{(\alpha)}}; \quad \Sigma c_i^{(\alpha)} = 1, \quad \alpha = 1 \text{ eller } 2$$

ange resp. varuslags värdeandel i normalhushållets totalutgift.

De ovan betraktade gränserna ha i hög grad intresserat indexteoretikerna. *G. Haberler* ägnar en stor del av sin bok »Der Sinn der Indexzahlen» åt de ovan behandlade relationerna *L. V. Bortkiewicz* [1]), *A. A. Konüs* [2]), *J. M. Keynes* [3]), m.fl. ha begrundat och försökt utreda betingelserna för deras giltighet. Tvivelaktigt är, huruvida det ovan använda betraktelsesättet skall kunna tillämpas på abstrakta ekonomikretsar, vilket fall är det från statistisk synpunkt intressanta. Då hänsyn tages till, att talen $p_i^{(\alpha)}$, $q_i^{(\alpha)}$, $c_i^{(\alpha)}$ erhållas medelst stickprovmetoder och att dessa tal till sin natur äro synnerligen invecklade medeltal, är det ej att förvåna om den sökta växelkursen X icke skulle ligga mellan de gränser, som erhållas enligt denna metod. Därtill kommer att varukvantiteterna q_i resp. vikterna c_i sällan hänföra sig till de år, för vilka prisen betraktas. Under sådana omständigheter blir innebörden av de skenbara gränser för det metodtekniska felet vid beräkningen av »köpkraftsparitetsväxelkursen», som fås genom *Paasche*'s och *Laspeyre*'s formler mycket dunkel. Avståndet mellan gränserna kan betraktas som en storhet av slumpkaraktär, vilket i bästa fall kan

[1]) *L. V. Bortkiewicz*, Nordic Statistical Journal, 1923, 1924, 1932.

[2]) *A. A. Konüs*, Economic Bulletin, Conjuncture Institute of Moscow, nr 9, 10, 1924.

[3]) *J. M. Keynes*, A Treatise on Money, London 1930, Vol. I, Book V.

ge en viss föreställning om osäkerhetsmarginalen vid beräkningen.

Genom att bilda det geometriska medeltalet av de ovan betraktade gränserna erhålla vi *Fisher*'s ideala indexformel [1])

$$X \approx \sqrt{\frac{\Sigma p_i^{(1)} q_i^{(1)}}{\Sigma p_i^{(2)} q_i^{(1)}} \times \frac{\Sigma p_i^{(1)} q_i^{(2)}}{\Sigma p_i^{(2)} q_i^{(2)}}} = \sqrt{\frac{\sum c_i^{(2)} \frac{p_i^{(1)}}{p_i^{(2)}}}{\sum c_i^{(1)} \frac{p_i^{(2)}}{p_i^{(1)}}}}.$$

Denna formel rekommenderas av *H. Staehle* för internationella jämförelser av kostnaderna för föda [2]).

Ett annat sätt att effektivt angripa indexproblemet har angivits av *Fr. Divisia* [3]). Det bygger på tanken om en kontinuerlig övergång från den prissituation, som förefanns i den första ekonomikretsen, till den i den andra kretsen rådande situationen. Man får genom att slutföra uppslaget en indexformel, som kan framställas i form av en kurvintegral:

$$\log X = \Sigma \int_2^1 c_i^{(\alpha)} \frac{d\, p_i^{(\alpha)}}{p_i^{(\alpha)}} = \Sigma \int_2^1 c_i^{(\alpha)}\, d\,(\log p_i^{(\alpha)}).$$

Parametern α varierar från 2 till 1. För $\alpha = 2$ fås den för den svenska ekonomikretsen karakteristiska priskonstellationen och för $\alpha = 1$ den motsvarande finländska. För övriga värden på α beteckna talen $p_i^{(\alpha)}$ mellan gränserna $p_i^{(1)}$ och $p_i^{(2)}$ liggande, tänkta pris. $c_i^{(\alpha)}$ ange den del av totalutgiften, som vid det av talen $p_i^{(\alpha)}$ angivna prisläget faller på varan i. Beroende på vilket sätt övergången mellan de båda prissituationerna tänkes ske erhålles ett något olika resultat för $\log X$.

[1]) *Irving Fisher*, The Making of Indexnumbers, A study of their Varietes, Tests and Reliability, Cambr., Mass., 1922.

[2]) *Hans Staehle*, International comparisons of food costs, 1934, Econometrica, 1934. Jfr. även Review of Economic Studies, juni, 1935.

[3]) *Francois Divisia*, L'Indice monétaire et la théorie de la monnaie, Revue d'Economie politique, 1927.

Vid normalelastisk efterfrågan, d.v.s. $c_i^{(\alpha)} = c_i =$ konstant, ger denna indexformel det vägda geometriska medeltalet:

$$\log X = \Sigma c_i \log \frac{p_i^{(1)}}{p_i^{(2)}}.$$

Genom att låta c_i representera ett medelvärde av $c_i^{(\alpha)}$ kommer man med stor sannolikhet att till resultat erhålla ett indextal tillhörande en delintervall med stor begreppsmassa i förhållande till intervallens bredd. (Jfr. sid. 3. I detta sammanhang betrakta vi det snävare köpkraftsparitetsväxelkursbegrepp, vars obestämdhet enbart beror på omöjligheten att entydigt lösa vägningsproblemet.) Medeltalet till det med vikterna $c_i^{(1)}$ och det med $c_i^{(2)}$ vägda geometriska medeltalet synes även ge ett gott värde för X. Om de båda ekonomikretsarna skilja sig från varandra blott ifråga om tidpunkterna, då prisen betraktas, men geografiskt sammanfalla, är det naturligt att låta prisen $p_i^{(\alpha)}$ och vikterna $c_i^{(\alpha)}$ beskriva det historiska förloppet hos priserna och vikterna vid övergången till det nya prisläget. Talet X kommer i detta fall att beteckna en prisnivåindex. Detta betraktelsesätt leder osökt till kedjeindextal av geometrisk typ med rörliga vikter. Finlands Banks konsumtionsprisindex har denna byggnad. [1])

En ur teoretisk synpunkt intressant metod att angripa problemet erhålles genom att studera en levnadsstandardfunktion

$$L = L(q_1, q_2, q_3, \ldots),$$

vilken funktion uppfattas som entydigt bestämd av de per tidsenhet av normalhushållet konsumerade kvantiteterna q_i. Så snart prisen p_i äro givna, finnes det en entydigt bestämd minsta penningkostnad, för vilken levnadsstandarden L kan uppnås. Genom att sätta denna av L och priskonstella-

[1]) Då förf. skrev den P. M., som åstadkom att Finlands Banks konsumtionsprisindex fick denna byggnad, kände han ännu ej till *Divisia*'s arbete. En kort beskrivning av indexen ingår i Bank of Finland Monthly Bulletin, nr 10, 1936.

tionen beroende minsta kostnad i relation till samma kostnad vid en annan priskonstellation, men samma levnadsstandard L, kan man på ett förnuftigt sätt definiera ett indextal angivande huru många penningenheter i den ena situationen som motsvara en enhet i en annan situation för ett normalhushåll med en levnadsstandard angiven av talet L. Medelst detta betraktelsesätt inser man, att olika hög levnadsstandard sannolikt även medför olika indextal. Dessa tal kunna dock väl tänkas förbli inom en relativt liten intervall trots kraftiga variationer i levnadsstandarden L. Genom att tillfoga vissa plausibla antaganden om L-funktionens byggnad kan man genom en matematisk behandling av det minimeringsproblem, som fastställandet av ovannämnda minsta kostnad, innebär, visa, att ifall man försummar en del som relativt små betraktade termer, de på detta sätt definierade indextalen låta sig approximativt framställas medelst indexformler av samma karaktär som de tidigare angivna. [1]) Entydigheten vid lösningen av minimeringsproblemet förutsätter dock att levnadsstandardfunktionens byggnad förblir densamma även vid de eventuella förändringar i miljö, ekonomikretsens sammansättning och de konsumerande individernas behov, som äga rum mellan de två tillfällen, vilkas motsvarande minimikostnader skola jämföras med varandra.

En intressant metod för lösandet av indexproblemet bygger på det faktum att vägningskoefficienterna c_i med växande levnadsstandard visa en karakteristisk vridning, i det att existensutgifternas andel minskar och de kulturbetonade utgifternas ökar. Denna iakttagelse kallas Engels lag. Matematiskt kan denna lag uttryckas genom formeln

$$\frac{c_i^{(1)}}{c_i^{(2)}} = \frac{C_i(L^{(1)})}{C_i(L^{(2)})} + \varepsilon ,$$

där resttermen ε är en relativt till den första termen liten kvantitet beroende bl.a. av talet $\log (p_i^{(1)}/X p_i^{(2)})$.

[1]) Ragnar Frisch, Annual Survey of General Economic Theory: The Problem of Index Numbers, Econometrica, jan. 1936.

Genom att sammanföra likartade varor till en större grupp, t.ex. födoämnena till en födagrupp, och från samma synpunkt studera gruppens vägningskoefficient minskar resttermens relativa betydelse. Genom att för de båda kretsarna jämföra en serie av utgifter per normalhushåll och motsvarande viktfördelning kan man erhålla ett närmevärde för förhållandet mellan kostnaderna för upprätthållande av samma levnadsstandard i de båda kretsarna. Matematiskt uttryckt: Ifall

$$c_i^{(1)} = c_i^{(2)} \text{ är } 1 = \frac{C_i(L^{(1)})}{C_i(L^{(2)})} + \varepsilon,$$

varav erhålles

$$L^{(1)} \approx L^{(2)}.$$

Om den till vikten $c_i^{(1)} = c_i^{(2)}$ hörande motsvarande kostnaden i det ena landet är $U^{(1)}$ och i det andra landet $U^{(2)}$, är följaktligen

$$U^{(1)} \approx X\, U^{(2)}, \text{ d.v.s. } X \approx \frac{U^{(1)}}{U^{(2)}}.$$

Engels lag kan sålunda utnyttjas för bestämning av köpkraftsparitetsväxelkursen. Man har dock att märka, att vid lika höga inkomster kroppsarbetarnas existensutgifter äro större än de intellektuella arbetarnas, varför ekonomikretsarnas sammansättning, miljö och individernas behov böra vara möjligast likartade för att metoden skall kunna ge ett gott resultat.

På grund av att de för denna metod behövliga utgiftssiffrorna med tillhörande viktfördelning i praktiken blott kunna fastställas till ett ringa fåtal, uppstår ett rätt invecklat interpolationsproblem vid uppgiftens numeriska lösande.

Den senast behandlade metoden är speciellt intressant ur synpunkten, att den tillåter jämförelse av levnadskostnaderna mellan olika länder även i det fall, att man saknar

tillgång till som jämförbara (ekvivalenta) betraktade varors priser.

Efter denna korta översikt över de matematiskt statistiska metoder, som stå till buds, skola vi behandla de praktiska svårigheter, som möta vid försöket att fastställa köpkraftsparitetsväxelkursen mellan Finland och Sverige.

De speciella praktiska svårigheterna vid utnyttjandet av det tillbudsstående materialet.

Valet av tidpunkt eller tidsperiod, som skall läggas till grund för beräkningen, inverkar på resultatet. Det förefaller vara enklast att välja årsmedelprisen till utgångspunkt och samma år för vartdera landet. Nödvändigt är det dock icke att ha tillgång till priser hänförande sig till samma år. Man kan nämligen med tillhjälp av levnadskostnadsindexerna i Finland och Sverige justera de primärt erhållna resultaten till en gemensam jämförelseperiod. Slutresultatet kan därefter med tillhjälp av dessa indexer fortskrivas till önskad tidpunkt. Det osäkerhetsmoment, som härflyter ur dessa och liknande räkneoperationer, bildar föremål för en särskild undersökning publicerad i denna tidskrift H. 37 [1]),

Då frågan gäller att numeriskt bestämma en storhet sådan som köpkraftsparitetsväxelkursen mellan Finland och Sverige betraktad t.ex. ur konsumentsynpunkt, är man, ifall man icke utnyttjar Engels lag, främst i behov av jämförbara prisuppgifter både från Finland och Sverige på varor, som kunna betraktas som ekvivalenta, d.v.s. varor, som kunna utbytas mot varandra, utan att det kan bestämt avgöras vilkendera, som är värdefullare. Svårigheten att finna sådana ekvivalenta varor och de fel, som misslyckanden häri åstadkomma, ge upphov till den detalj i indexproblemet, som kallas kvalitetsproblemet. Vid jämförelse mellan varor i

[1]) Leo Törnqvist, Levnadskostnadsindexerna i Finland och Sverige, deras tillförlitlighet och jämförbarhet, Ekon. Samf. Tidskr., Ny s., H. 37, 1936.

Finland och Sverige finnes en risk för att de svenska varorna skola i medeltal vara av högre kvalitet än motsvarande finska varor, trots att de ha samma beteckningar. Detta fel skulle medföra en undervärdering av kronans relativa köpkraft, ehuru knappast med någon mera betydande procent.

En annan svårighet består i prisens lokalisering. På olika orter i Finland gälla olika pris på samma vara och detsamma gäller för Sverige. Det synes vara en utväg att jämföra medelpriser för städer och bosättningscentra i Finland med motsvarande priser i Sverige. Byråns för social forskning vid Socialministeriet medelpriser för livsförnödenheter på 36 orter i Finland [1]) och K. Socialstyrelsens medelpriser för 53 orter i Sverige bilda härvid det naturligaste utgångsmaterialet. Vid jämförelse av varuförteckningarna finner man ca 40 gemensamma varuslag. Någon garanti för att varukvaliteterna överensstämma så att de kunna betraktas som ekvivalenta finnes dock icke.

Förutom ett någotsånär användbart prismaterial är man i behov av en rätt ingående kännedom om konsumtionsförhållandena i Finland och Sverige för att kunna jämföra levnadskostnaderna i de båda länderna med varandra. Tack vare levnadskostnadsundersökningen av år 1928 i Finland och liknande undersökningar i Sverige av omkring år 1933 fås en rätt god bild av konsumtionens sammansättning i de båda länderna. Vid det i det följande gjorda försöket att utnyttja dessa undersökningar för bestämmande av vikter för de olika medtagna prisen ha vi dock icke, dels på grund av brist på uppgifter dels på grund av att vi i någon mån försökt fördela de icke medtagna varornas vikter på de varor, som medtagits, kunnat undgå ett visst godtycke.

Vid våra räkningar ha vi begränsat oss till konsumentsynpunkten och därvid betraktat den ekonomikrets, som i levnadskostnadsundersökningen av år 1928 betecknats »arbetare och befattningshavare» i Finland, samt den motsvarande kretsen »arbetare och lägre tjänstemän» i Sverige.

[1]) Åren 1921—31, 40 orter, 1932—34, 30 orter, 1935—, 36 orter.

Själva det begagnade materialets tillförlitlighet skola vi diskutera i samband med frågan om osäkerhetsmarginalen hos resultaten.

Beräkningar av förhållandet mellan finska markens och svenska kronans köpkraft gentemot födoämnen.

I tabell I ha vi sammanställt material för beräkning av förhållandet mellan markens och kronans köpkraft gentemot födoämnen, varvid till jämförelseår valts år 1933 för Finland och 1931 för Sverige. Såsom vi redan tidigare framhållit är det icke nödvändigt att prisuppgifterna hänföra sig till samma år för vartdera landet. Valet är icke på något sätt förmånligt för precisionen i jämförelsen. De valda åren tillhöra olika skeden i den senaste prisfallsperioden. Genom att sedermera jämföra resultaten enligt denna generella metod med motsvarande siffror erhållna medelst direkt prisjämförelse hänförande sig till samma år för vartdera landet kunna vi erhålla en viss inblick i det osäkerhetsmoment, som beror på valet av jämförelseår.

De i tabell I ingående varuslagen ha erhållits genom en jämförelse av de i Social Tidskrift, Helsingfors, och Sociala Meddelanden, Stockholm, förekommande varuförteckningarna. Från dessa källor ha även årsmedelprisen tagits.

För talen i kolumn $q^{(1)}$ ha vi utnyttjat tab. 21, sid. 66*, och för talen $c^{(1)}$ tab. 16, sid. 51* i Sociala specialundersökningar 14, Levnadskostnadsundersökningen år 1928, A. Bokföringsfamiljerna i städer och övriga bosättningscentra, Helsingfors 1936. Talen $q^{(2)}$ ha tagits från budgeten för K. Socialstyrelsens levnadskostnadsindex (1931), sid. 321, Sociala Meddelanden, nr 4, 1932. Talen $c^{(2)}$ ha erhållits från Sociala Meddelanden, 1934, sid. 797, tab. E. Sådana siffror, som vi mer eller mindre godtyckligt uppskattat, ha vi satt inom parentes. Godtycket hos de av oss uppskattade siffrorna inskränker sig blott till fördelningen av vikterna på representanterna för undergrupperna mjölk och

Tab. I. *Markens och kronans köpkraft gentemot födoämnen* (1933 : 1931).

Varuslag		Finland 1928 $q^{(1)}$	Sverige 1931 $q^{(2)}$	Finland 1928 $c^{(1)}$	Sverige -32/33 $c^{(2)}$	Penni 1933 $p^{(1)}$	Öre 1931 $p^{(2)}$	$\frac{p^{(1)}}{p^{(2)}}$	‰ $\log \frac{p^{(1)}}{10\,p^{(2)}}$
Mjölk, oskummad	l	(1154)	1000	(19,4)	(16,0)	144	21	6,86	—377
» separerad	l	65	60	(0,3)	(0,4)	46	8	5,75	—554
Smör, mejeri-	kg	(24)	(18)	(6,5)	(4,0)	2328	255	9,13	— 91
» lands-	»	(23)	(18)	(6,0)	(3,4)	2092	240	8,72	—137
Margarin	»	21	28	3,2	4,2	1343	129	10,41	40
Ost, helfet	»	5	(16)	0,9	2,5	1997	190	10,51	50
Ägg, färska	tj	11	30	2,1	4,4	1537	193	7,96	—228
Potatis, gammal	5 l	106	112	3,7	2,5	304	44	6,91	—370
Ärter, gula	kg	5	6	(1,2)	(1,5)	537	32	16,78	518
Vetemjöl, prima	»	104	176	(5,0)	(4,5)	558	39	14,31	358
Rågmjöl, oskrätt	»	126	73	(3,9	(1,2)	289	24	12,04	186
Havregryn, ångprep.	»	23	22	(1,1)	(0,6)	423	36	11,75	161
Risgryn	»	20	8	(1,5)	(0,3)	635	46	13,81	322
Rågbröd, knäcke-	»	8	34	(0,6)	(1,1)	665	88	7,56	—280
» grovt, spis-	»	13	34	(0,7)	(0,9)	477	64	7,45	—294
» mjukt, billigaste	»	73	53	(3,1)	(3,0)	361	46	7,85	—242
Vetebröd, franskt bröd	»	(15)	11	(1,6)	(1,0)	886	75	11,81	166
» kaffebröd	»	(24)	27	(3,1)	(3,2)	1104	(103)	(10,72)	70
Nötkött, färskt, stek-	»	(20)	(22)	(2,8)	(2,6)	890	181	4,92	—710
» » sopp-	»	(18)	(20)	(1,9)	(3,0)	655	150	4,37	—828
Kalvkött, gödkalv	»	(3)	(11)	(0,5)	(1,2)	1039	225	4,62	—773
» spädkalv	»	(3)	(11)	(0,3)	(0,7)	546	138	3,96	—927
Fårkött, färskt, stek	»	(10)	5	(1,8)	(0,6)	1132	218	5,19	—656
Fläsk, färskt, sid-	»	28	37	(5,8)	(4,8)	1290	147	8,78	—130
» salt	»	5	11	(1,0)	(1,2)	1318	153	8,61	—150
Fisk, färsk, gädda	»	(10)	(4)	(0,9)	(0,7)	878	182	4,82	—730
» » abborre	»	(12)	(3)	(0,8)	(0,5)	612	121	5,06	—682
» » strömming	»	16	15	(0,5)	(0,6)	322	68	4,74	—747
Sill, salt	»	10	18	(0,8)	(1,5)	717	56	12,80	247
Kaffe, obränt	»	20	(20)	6,2	5,7	3175	(210)	(15,12)	413
Bitsocker	»	102	109	6,5	4,6	970	42	23,09	837
Övriga				6,3	17,6				
				100,0	100,0				

grädde, smör, rotfrukter och grönsaker (ärter), mjöl, gryn, bröd, kött och fläsk inklusive charkuterivaror samt fisk. Margarin representerar även fett. Vid fördelningen ha vi utnyttjat bl.a. Finlands Banks konsumtionsprisindex och Sveriges Riksbanks konsumtionsprisindex' vikter samt tab. II, sid. 14, i tabellbilagan till Sociala specialundersökningar 14. Summorna av vikterna c_i för representanterna för de

nämnda undergrupperna överensstämma med resultaten enligt de officiella undersökningarna. Uppgifterna i den finländska levnadskostnadsundersökningen hänföra sig till 581 arbetar- och 242 befattningshavarhushåll, sammanlagt 823 bokföringsfamiljer. Den motsvarande svenska levnadskostnadsundersökningen av år 1932/33 omfattade 349 industriarbetar- och 261 lägre tjänstemannahushåll, sammanlagt 610 bokföringsfamiljer.

Vad beträffar den av oss uppskattade siffran för ostens kvantitetsvikt i Sverige (16) bör nämnas, att vi erhållit densamma genom att addera kvantitetsvikten för ost, helfet, 12 kg, och ost halvfet, 5 kg, samt på grund av den halvfeta ostens lägre pris minskat summan med 1 kg. Kvantitetsvikterna för nötkött i kolumnen $q^{(1)}$ ha erhållits så att vi adderat vikten för nötkött, 29 kg, och för malet nötkött, 13 kg, i Socialstyrelsens budget av år 1931 samt fördelat summan 42 kg på de båda slagen av nötkött. Vikten för kaffe, obränt (20) hänför sig i Socialstyrelsens budget till bränt kaffe. Det av oss uppskattade priset för kaffebröd (103) hänför sig till januari 1932, emedan några prisuppgifter för år 1931 för kaffebröd icke publiceras i Sociala Meddelanden. Prissiffran (210) för obränt kaffe har erhållits genom att multiplicera medelpriset för januari 1932 för bränt kaffe, santosblandning, 289 öre, med talet 0,73, vilket utgör förhållandet mellan priserna för obränt och bränt kaffe, vilka pris i sin tur erhållits genom att dividera det finländska normalhushållets utgifter för obränt och bränt kaffe år 1928 med motsvarande konsumerade kvantiteter. Den angivna siffran kan ej anses särdeles tillförlitlig.

I kolumnen $\frac{p^{(1)}}{p^{(2)}}$ angivas förhållandena mellan prisen i Finland och Sverige. De relativt högsta talen hänföra sig till varor, vilkas pris påverkats av valutakurserna och olikheterna i tullbeskattningen i de båda länderna. Kött- och fiskprisen synas vara relativt billiga i Finland. Den sista kolumnen innehåller prisförhållandenas naturliga logaritmer

uttryckta i promille. Den konstanta addenden $^e\log 10 = 2303$ promille har härvid försummats.

Genom att på materialet i tabell I tillämpa de i det föregående betraktade indexformlerna ha vi erhållit ett flertal olika indextal. Deras spridning kring medeltalet ger en god inblick i indexformelvalets inverkan på resultatet av jämförelsen mellan den finska markens och den svenska kronans köpkraft gentemot födoämnen åren 1933 och 1931.

Tab. II. *Resultaten beträffande förhållandet mellan födoämneskostnaderna i Finland år 1933 och Sverige år 1931 uträknade enligt olika indexformler.*

Medeltal	Indexformel	Resultat
Harmoniskt	$X_1 = 1 : \Sigma c_i^{(1)} \frac{p_i^{(2)}}{p_i^{(1)}}$	8,61
Geometriskt	$\log X_2 = \Sigma c_i^{(2)} \log \frac{p_i^{(1)}}{p_i^{(2)}}$	9,07
Aritmetiskt	$X_3 = \frac{\Sigma p_i^{(1)} q_i^{(1)}}{\Sigma p_i^{(2)} q_i^{(1)}}$	9,08
Fishers ideala (q)	$\log X_4 = \frac{1}{2} (\log X_3 + \log X_7)$	9,14
Geometriskt »ideala»	$\log X_5 = \frac{1}{2} (\log X_2 + \log X_8)$	9,19
Fishers ideala (c)	$\log X_6 = \frac{1}{2} (\log X_1 + \log X_9)$	9,20
Aritmetiskt	$X_7 = \frac{\Sigma p_i^{(1)} q_i^{(2)}}{\Sigma p_i^{(2)} q_i^{(2)}}$	9,21
Geometriskt	$\log X_8 = \Sigma c_i^{(1)} \log \frac{p_i^{(1)}}{p_i^{(2)}}$	9,32
Aritmetiskt	$X_9 = \Sigma c_i^{(2)} \frac{p_i^{(1)}}{p_i^{(2)}}$	9,84
Aritmetiskt av samtliga	$X_{10} = \Sigma X_\mu : 9$	9,18

Att döma av tabellen synas resultaten koncentrera sig kring talet 9,2. De avvika frånsett de två yttersta med en-

dast ca 1 % från detta värde. Enligt *Divisia*'s betraktelsesätt skulle man vara böjd för att anse att det sökta indextalet skulle tillhöra intervallen mellan de geometriska medeltalen X_2 och X_8. Ökad säkerhet för att undgå förhastade slutsatser vinnes genom att på grund av omöjligheten att på ett rationellt sätt välja indexformel tillordna resultatet intervallen $(9{,}2 \pm 0{,}6)$.

Med hänsyn tagen till att kvantitetsvikterna icke på ett förnuftigt sätt låta sig adderas till gruppvikter samt att de svårligen låta sig överflyttas på de representanter för grupperna som medtagas, synas indexformler, i vilka kvantitetstal utgöra vikter, vara mindre väl ägnade för erhållande av tillförlitliga resultat. En fullt tillräcklig belysning av spridningen i resultaten till följd av indexformelvalet erhålles genom att uträkna de indextal, som i tabell II betecknats med X_1, X_2, X_5, X_8 och X_9. I fortsättningen komma vi att nöja oss med dessa fem indextal. I omgivningen av talet X_5 synes begreppstätheten vara större än invid de övriga här betraktade medeltalen.

Utgående från de ovannämnda fem indextalen skola vi med tillhjälp av livsmedelsindexen i K. Socialstyrelsens levnadskostnadsindex beräkna köpkraftspariteten gentemot födoämnen år 1933. Från år 1931 till 1933 sjönk nämnda index från 100 till 94. Livsmedelskostnaderna i Sverige voro sålunda ca 6 % lägre 1933 än 1931, varför vi för att erhålla den sökta pariteten ha att multiplicera de tidigare erhållna siffrorna med kvoten $100 : 94 = 1{,}06(4)$. Resultaten ha antecknats i den andra sifferkolumnen av tabell III. De för år 1935 gällande resultaten ha i sin tur erhållits genom multiplikation av dessa tal med faktorn 0,96(3), som anger den relativa prisutvecklingen för födan i Finland och Sverige mellan åren 1933 och 1935. Faktorn har beräknats med stöd av de officiella levnadskostnadsindexerna för föda enligt formeln:

$$\frac{\text{Föda (Finland, 1935)}}{\text{Föda (Finland, 1933)}} : \frac{\text{Livsmedel (Sverige, 1935)}}{\text{Livsmedel (Sverige, 1933)}}$$

$$= \frac{911}{894} : \frac{128}{121} = 0{,}96(3)\,.$$

Tab. III. *Pariteter för föda, indirekt beräknade, åren 1933 och 1935.*

Medeltal	Vägning	X_μ	1933/31	1933	1935
Harmoniskt	$c^{(1)}$	X_1	8,61	9,16	8,82
Geometriskt	$c^{(2)}$	X_2	9,07	9,65	9,29
Geometriskt »ideala»	$\frac{1}{2}(c^{(1)}+c^{(2)})$	X_5	9,19	9,78	9,42
Geometriskt	$c^{(1)}$	X_8	9,32	9,92	9,55
Aritmetiskt	$c^{(2)}$	X_9	9,84	10,47	10,08

Tab. IV. *De relativa födoämneskostnaderna i Finland och Sverige åren 1933 och 1935.*

Varuslag		Finland 1928 $c^{(1)}$	Sverige 1933 $c^{(2)}$	Penni 1933 $p^{(1)}$	Öre 1933 $p^{(2)}$	$\frac{p^{(1)}}{p^{(2)}}$	Penni 1935 $p^{(1)}$	Öre 1935 $p^{(2)}$	$\frac{p^{(1)}}{p^{(2)}}$
Mjölk, oskummad	l	(19,4)	(16,0)	144	20	7,20	157	21	7,48
» separerad	»	(0,3)	(0,4)	46	7	6,57	49	7	7,00
Smör, mejeri-	kg	(6,5)	(4,0)	2328	244	9,54	2564	268	9,57
» lands-	»	(6,0)	(3,4)	2092	224	9,34	2319	253	9,13
Margarin	»	3,2	4,2	1343	127	10,57	1483	135	10,99
Ost, helfet	»	0,9	2,5	1997	147	13,59	1998	177	11,29
Ägg, färska	tj	2,1	4,4	1537	161	9,55	1610	184	8,75
Potatis, gammal	5 l	3,7	2,5	304	36	8,44	284	43	6,60
Ärter, gula	kg	(1,2)	(1,5)	537	33	16,27	523	48	10,90
Vetemjöl, prima	»	(5,0)	(4,5)	558	36	15,50	548	38	14,42
Rågmjöl, oskrätt	»	(3,9)	(1,2)	289	24	12,04	267	24	11,13
Havregryn, ångprep.	»	(1,1)	(0,6)	423	32	13,22	388	35	11,09
Risgryn	»	(1,5)	(0,3)	635	42	15,12	623	40	15,58
Rågbröd, knäcke-	»	(0,6)	(1,1)	665	85	7,82	665	82	8,11
» grovt, spis	»	(0,7)	(0,9)	477	61	7,82	467	59	7,92
» mjukt, billigaste	»	(3,1)	(3,0)	361	46	7,85	341	45	7,58
Vetebröd, franskt bröd	»	(1,6)	(1,0)	886	72	12,31	849	72	11,79
» kaffebröd	»	(3,1)	(3,2)	1104	100	11,04	1053	97	10,86
Nötkött, färskt, stek-	»	(2,8)	(2,6)	890	148	6,01	1088	179	6,08
» » sopp-	»	(1,9)	(3,0)	655	114	5,75	805	143	5,63
Kalvkött, gödkalv	»	(0,5)	(1,2)	1039	188	5,53	1143	217	5,27
» spädkalv	»	(0,3)	(0,7)	546	114	4,79	628	129	4,87
Fårkött, färskt, stek-	»	(1,8)	(0,6)	1132	187	6,05	1229	205	6,00
Fläsk, färskt, sid-	»	(5,8)	(4,8)	1290	142	9,08	1704	177	9,63
» salt	»	(1,0)	(1,2)	1318	148	8,96	2245	182	12,34
Fisk, färsk, gädda	»	(0,9)	(0,7)	878	157	5,59	898	153	5,85
» » abborre	»	(0,8)	(0,5)	612	109	5,61	636	110	5,78
» » strömming	»	(0,5)	(0,6)	322	59	5,46	350	57	6,14
Sill, salt	»	(0,8)	(1,5)	717	40	17,93	700	62	11,29
Kaffe, obränt	»	6,2	5,7	3175	(228)	(13,93)	2871	(212)	(13,54)
Bitsocker	»	6,5	4,6	970	47	20,64	911	46	19,78
Övriga	»	6,3	17,6						
		100,0	100,0						

Då det är av intresse att undersöka i vad mån de erhållna resultaten överensstämma med de indextal, som fås genom en direkt prisjämförelse för år 1933 resp. 1935, skola vi med tillhjälp av det i tabell IV sammanställda materialet beräkna förhållandena mellan födoämneskostnaderna i Finland och Sverige.

Viktsystemen $c^{(1)}$ och $c^{(2)}$ äro desamma som i tabell I. Årsmedelprisen äro hämtade ur Social Tidskrift resp. Sociala Meddelanden. Kaffepriset har för Sveriges vidkommande uppskattats genom multiplikation av priset för bränt kaffe, Santosblandning, med den tidigare nämnda faktorn 0,73. Beträffande de relativa prisen på de olika födoämnena i Finland och Sverige gör man samma iakttagelse som tidigare på grundval av tabell I, nämligen, att varor, som i större utsträckning gå i internationell handel ha relativt höga prisförhållanden $p^{(1)} : p^{(2)}$. De särskilt höga förhållandena för socker och kaffe förklaras av de relativt högre importtullarna på dessa varor i Finland.

I tabell V ha vi sammanställt de på grundval av det i tabell IV givna materialet beräknade fem olika medeltalen jämte de tidigare, på indirekt väg erhållna, motsvarande talen.

Tab. V. *Pariteter för föda åren 1933 och 1935.*

Medeltal	Vägning	X_μ	1933		1935	
			Indirekt	Direkt	Indirekt	Direkt
Harmoniskt	$c^{(1)}$	X_1	9,16	9,19	8,82	9,09
Geometriskt	$c^{(2)}$	X_2	9,65	9,62	9,29	9,37
Geometriskt »ideala»	$\frac{1}{2}(c^{(1)} + c^{(2)})$	X_5	9,78	9,68	9,42	9,46
Geometriskt	$c^{(1)}$	X_8	9,92	9,75	9,55	9,56
Aritmetiskt	$c^{(2)}$	X_9	10,47	10,31	10,08	10,10

Tabell V synes visa, att den indirekta metoden ger resultat, som blott obetydligt avvika från de direkt beräknade. Spridningen i de senare resultaten är något mindre än i de indirekt erhållna, vilket är rätt naturligt med hänsyn till det

tidigare nämnda sakförhållandet beträffande prisen 1933 i Finland och 1931 i Sverige.

För att kontrollera i vad mån de vid den indirekta beräkningen använda multiplikatorerna 1,06(4) och 0,96(3) äro tillförlitliga jämföra vi dem med motsvarande siffror erhållna genom att betrakta det »ideala» geometriska medeltalets utveckling. Förhållandet mellan det för år 1933 gällande talet 9,68 och det för 1933/31 erhållna värdet 9,19 är 1,053, d.v.s. ca 1 % mindre än talet 1,06(4). Detta tyder på att den svenska livsmedelsindexen kan utnyttjas för fortskrivningsändamål, om man tillordnar de erhållna resultaten en osäkerhetsmarginal på ca 1 %. Detta överensstämmer väl med de erfarenheter vi gjort vid den tidigare nämnda undersökningen beträffande tillförlitligheten och jämförbarheten hos levnadskostnadsindexerna i Finland och Sverige. [1]) Förhållandet mellan de direkt erhållna paritetsiffrorna för åren 1935 och 1933 är 9,46 : 9,68 = 0,977, d.v.s. det överstiger den av oss använda multiplikatorn 0,96(3) med ca 1,5 %. Även denna osäkerhetsprocent är av en storleksordning, som med hänsyn tagen till att den finländska födaindexen har ett osäkerhetsmoment på ca 1 $^1/_2$ %, överensstämmer med våra förväntningar.

Mellan de ovan betraktade medeltalen och spridningen i prisförhållandenas naturliga logaritmer existerar ett intimt samband, i det att logaritmdifferensen mellan det aritmetiska och det lika vägda geometriska medeltalet med god approximation är lika med halva kvadraten på dessa logaritmers medelkvadratavvikelse från logaritmen för det geometriska medeltalet. Logaritmdifferensen mellan det geometriska och det lika vägda harmoniska medeltalet är approximativt lika med samma differens mellan det aritmetiska och geometriska. Summan av dessa differenser ger ett gott uttryck för kvadraten på spridningen. [2])

---

[1]) a.a., Ekon. Samf. Tidskr., Ny s., H. 37.

[2]) Jfr. a. a., Ekon. Samf. Tidskr. Ny s. H. 37, formlerna (1), (2) och (3).

På grundval av de i tabell V givna medeltalen ha vi beräknat de i tabell VI redovisade spridningstalen.

Tab. VI. *Spridningstal för födagruppen i naturliga, 3-ställiga logaritmer.*

Jämförelseår	1933/31	1933	1935
$\log X_9 - \log X_2 \approx \frac{1}{2} s^2$	.082	.070	.075
$\log X_8 - \log X_1 \approx \frac{1}{2} s^2$	.080	.059	.050
Summa $\approx s^2$	.162	.129	.125
Spridning $s \approx$	.402	.359	.353

Spridningstalen s ange, att prisförhållandena i tabell I i medeltal avvika med omkring 40 % från indextalet $(9{,}2 \pm 0{,}6)$ samt att motsvarande tal $p^{(1)} : p^{(2)}$ för åren 1933 och 1935 i tabell IV ha en medel(kvadrat)avvikelse av ca 35 %. [1])

Med kännedom om dessa spridningstal kan man uppskatta den inverkan på resultatet, som de i vår jämförelse icke medtagna varorna skulle utövat om de hade kunnat medtagas. Det synes nämligen rimligt att antaga, att de obekanta prisförhållandena inom gruppen »övriga» i medeltal ej avvika med väsentligt mer än spridningen från de erhållna paritetstalen. Ifall vi för gruppen »övriga» använda medelvikten $\frac{1}{2}(c^{(1)} + c^{(2)}) = \frac{1}{2}(6{,}3 + 17{,}6) = 12{,}0$ och antaga, att logaritmen för gruppens paritetsindex t.ex. överstiger logaritmen för det ideala geometriska medeltalet år 1935 $(= {}^{e}\log 9{,}46)$ med $s = 0{,}35$ skulle resultatet förskjutas med $12 \cdot 0{,}35 = 4{,}2\%$ De få uppgifter, som stått oss till buds beträffande de varor, som ingå i gruppen »övriga», nämligen frukt och bär, rotfrukter, konserver, svagdricka, m.m., tyda på att denna grupp faktiskt skulle utöva en något höjande inverkan på slutre-

[1]) De naturliga logaritmerna ha egenskapen: $\log (X \pm s) = \pm \frac{s}{X'}$, där $X \pm \frac{1}{2} s \geqq X' \geqq \sqrt{X (X \pm s)}$. Med stöd av denna egenskap betrakta vi 3-ställiga naturliga logaritmer som approximativt angivande promilletal och omvänt.

sultatet. Förskjutningen torde dock vara mindre än de ovan uppskattade fyra procenten.

En omständighet, som även bör beaktas vid bedömningen av osäkerheten i våra beräkningar, är, att de använda vikterna hänföra sig till en bestämd levnadsstandard, som icke är densamma för vartdera viktsystemet. Genom att begagna den metod, som även kom till användning vid undersökningen över levnadskostnadsindexernas i Finland och Sverige tillförlitlighet och jämförbarhet, nämligen beräknandet av det med z betecknade talet, vilket utgör ett mått på tvenne viktsystems divergens från varandra, ha vi kommit till uppfattningen att om vi i stället för det begagnade viktsystemet $c^{(1)}$ skulle använt de vikter, som höra till den lägsta välståndsgruppen (under 5,000: — mks årsinkomst per konsumtionsenhet) skulle resultatet sannolikt förskjutits med ($rzs = \pm 0{,}2 \cdot 0{,}14 \cdot 0{,}35 =$) ca $\pm$ 1 %. Ifall vi åter använt det viktsystem, som enligt levnadskostnadsundersökningen år 1928 tillhör den högsta välståndsgruppen, arbetare och befattningshavare i Finland med över 10,000: — marks årsinkomst per konsumtionsenhet, skulle resultatet troligen förskjutits med inemot $\pm$ 2 %. Om vi sålunda till osäkerhetsintervallen foga en addend utgörande ett par procent av medeltalet, synes det troligt, att de förskjutningar i födapariteten, som uppkomma genom att lägga olika hög levnadsstandard till grund för beräkningen, till största delen skola hållas innanför intervallen. I samband med konjunkturväxlingarna ske förändringar i den betraktade ekonomikretsens levnadsstandard. Dessutom är att märka att de vid levnadskostnadsundersökningarna studerade bokföringsfamiljerna endast omfatta en liten del av samtliga hushåll, varför det är skäl att taga hänsyn till det osäkerhetsmoment i jämförelsen, som står i samband med levnadsstandardförändringar.

Förrän vi slutligt kunna fastställa storleksordningen hos osäkerhetsmomentet i resultaten beträffande födapariteten, måste vi bilda oss en uppfattning om tillförlitligheten i de årsmedelpris, som lagts till grund för jämförelsen. Det före-

faller att vara mycket svårt att få något begrepp om huru stora »iakttagelsefelen» äro. Det är likväl möjligt att för en del varor jämföra de officiella siffrorna med pris erhållna ur de tidigare nämnda hushållsbudgetundersökningarna i Finland och Sverige. 1928 års levnadskostnadsundersökning innehåller för Finlands vidkommande uppgifter om såväl normalhushållets utgifter för de olika varuslagen som de konsumerade kvantiteterna. Genom att dividera normalhushållets årsutgifter för en vara med motsvarande konsumerade kvantitet erhålles till resultat ett årsmedelpris, som vägts i enlighet med den konsumerade kvantiteten och sålunda bättre överensstämmer med de pris, som betraktas vid den teoretiska härledningen av indexformlerna. Dessa årsmedeltal kunna sedan jämföras med de i Social Tidskrift ingående motsvarande priserna. För Sveriges vidkommande kan man beträffande en del varor göra en liknande jämförelse med användande av resultat erhållna vid K. Socialstyrelsens levnadskostnadsundersökningar och de i Sociala Meddelanden publicerade årsmedelprisen. I tabell VII ha vi sammanställt det material vi samlat för bedömandet av prisförhållandenas $p^{(1)} : p^{(2)}$ tillförlitlighet.

Tab. VII. *Material för bedömning av prisförhållandenas tillförlitlighet.*

Varuslag	Finland 1928		Sverige 1932/33		I	II	I—II
	Social Tidskr.	Budget	Soc. Medd.	Budget	$\frac{p^{(1)}}{p^{(2)}}$	$\frac{\bar{p}^{(1)}}{\bar{p}^{(2)}}$	Avvikelse i procent
	$p^{(1)}$	$\bar{p}^{(1)}$	$p^{(2)}$	$\bar{p}^{(2)}$			
Mjölk, oskummad	198	200	20	20	9,9	10,0	— 1
Smör, mejeri-	3533	3204	240	225	14,7	14,2	2
» lands-	3150		222		14,2		
Margarin	1941	1844	126	120	15,4	15,4	0
Ost	2477	2289	161	144	15,4	15,9	— 3
Ägg, färska, tjog	2667	2309	166	134	16,1	17,2	— 7
Potatis, gammal, 5 l	473	408	40	40	11,8	10,2	15
Socker (bit-)	858	765	47	42	18,3	18,2	1
Vägt medelfel i prisförhållandena $p^{(1)} : p^{(2)}$							ca 3 %

Talen $p^{(2)}$ utgöra medeltal av årsmedelprisen för åren 1932 och 1933. Trots det synnerligen bristfälliga materialet för bedömningen se vi, av tabell VII, att storleksordningen hos de fel, som vidlåda prisförhållandena $p^{(1)} : p^{(2)}$ är ganska liten. Med hänsyn tagen till att felen icke synas vara systematiskt negativa eller positiva, är den inverkan de ha på medelvärdena relativt ringa, sannolikt mindre än 1 %. De ganska stora diskrepanserna mellan de officiella årsmedelprisen och de enligt budgetundersökningarna erhållna kunna till största delen förklaras av att budgetsiffrorna utgöra medelpris för flere olika kvaliteter och speciellt beträffande säsongvarorna hushållens inköp koncentrera sig till den tid av året, då varorna äro billigast. Enär samma sak gäller för vartdera landet elimineras dessa diskrepanser till största delen vid bildandet av prisförhållandena $p^{(1)} : p^{(2)}$ och $\bar{p}^{(1)} : \bar{p}^{(2)}$.

Resultaten beträffande förhållandet mellan finska markens och svenska kronans köpkraft gentemot födoämnen.

I det föregående ha vi beräknat förhållandet mellan markens och kronans köpkraft gentemot födoämnen åren 1933 och 1935. De använda viktsystemen, som vi i rätt stor utsträckning varit tvungna att komplettera med uppskattningar och indirekta beräkningar, hänföra sig till ekonomikretsarna arbetare och befattningshavare i Finland resp. arbetare och lägre tjänstemän i Sverige. Enär dessa ekonomikretsar representera en stor del av befolkningarna i de båda länderna, ge de erhållna resultaten tillika en rätt god bild av huru förhållandet ställer sig mellan köpkraften för ekonomikretsar, som omfatta samtliga invånare i Finland och Sverige.

Genom att utföra såväl en direkt prisjämförelse för år 1933 som en indirekt jämförelse, där priserna hänförde sig till år 1933 i Finland och 1931 i Sverige, kunde vi konstatera, att den indirekta metoden gav resultat, som relativt väl över-

ensstämde med de direkt beräknade (jfr. tab V). För år 1933 gav det direkt beräknade, s.k. ideala geometriska medeltalet värdet 9,7. Enligt samma metod uträknades födapariteten för år 1935 och erhölls det motsvarande talet 9,5. Kring dessa medeltal grupperade sig de övriga, enligt olika indexformler erhållna resultaten tämligen symmetriskt.

I fortsättningen gjordes vissa uppskattningar beträffande storleksordningen hos de olika osäkerhetsmoment, som böra beaktas vid bedömandet av resultatens tillförlitlighet. De olika använda indexformlerna gåvo för åren 1933 och 1935 resultat, som med mindre än 7 % avveko från de ideala geometriska medeltalen. Men hänsyn till indexformelvalet skulle det sålunda vara motiverat att tillordna resultatet av jämförelsen en osäkerhet av ca $\pm$ 6 %.

De i jämförelsen icke representerade varornas sannolika inverkan på resultaten uppskattades med tillhjälp av de i tabell VI angivna spridningstalen till ca $\pm$ 4 %, motsvarande en förskjutning av indextalen på omkring ($\pm$ 0,4).

Genom att undersöka i vad mån förändringar i levnadsstandarden och därmed följande viktförskjutningar skulle påverka jämförelsen kommo vi till uppfattningen att detta osäkerhetsmoment utgör ca $\pm$ 2 %, d.v.s. kunde åstadkomma en förskjutning av medeltalen med ungefär ($\pm$ 0,2).

Slutligen försökte vi med tillhjälp av ett antal stickprov, (tabell VII) uppskatta inflytandet av de fel, som helt naturligt alltid måste vidlåda årsmedelpris byggande på ett begränsat material. De i tabell VII spårade divergenserna äro även påverkade av kvalitetsolikheterna i de medtagna varorna samt av sättet att bilda årsmedelpris genom medeltalsbildningar i såväl tid som rum. Att döma av dessa stickprov synes denna felkälla vara den relativt minst betydelsefulla. En resultatförskjutning på ca $\pm$ 1 % torde dock kunna härflyta ur prisnoteringarnas bristande tillförlitlighet och jämförbarhet.

Frågan, huru man skall sammanfatta de olika felkällornas inverkan till ett enda osäkerhetsmoment, är även det ett

invecklat problem. Genom enkel addition skulle vi erhålla en sammanlagd osäkerhet av

$$\pm(6+4+2+1) = \pm 13\,\%_0 \text{ av medeltalet.}$$

Detta sätt att addera beaktar likväl icke att det är högst osannolikt, att de olika osäkerhetsfaktorerna alla skulle verka i samma riktning. Ifall man betraktar dem såsom av varandra helt oberoende böra de enligt sannolikhetsteorin adderas så, att det totala osäkerhetsmomentets kvadrat blir lika med summan av de enskilda momentens kvadrater. På detta sätt erhölle vi, att medeltalens relativa osäkerhetsmoment vore lika med

$$\pm\sqrt{6^2+4^2+2^2+1^2} = \pm 7{,}4\,\%_0.$$

Då det dock finnes en risk för att osäkerhetsfaktorerna icke äro helt oberoende av varandra kan det vara skäl att uppskatta osäkerhetsmomentet till en något större procent än 7,4, låt oss säga 10 %. För åren 1933 och 1935 få vi sålunda följande resultat av vår jämförelse mellan finska markens och svenska kronans köpkraft gentemot födoämnen: Det synes mycket sannolikt att största delen av det studerade begreppets kvantitativa motsvarighet ligger inom intervallerna (9,7 ± 1,0) för år 1933 och (9,5 ± 0,9) för år 1935.

Med tillhjälp av de officiella födaindexerna i Finland och Sverige fortskriva vi dessa resultat till åren 1926—36. Resultatet (9,7 ± 1,0) tillbakaskrives härvid till år 1926 genom multiplikation med de faktorer, som ange den relativa prisutvecklingen för föda i Finland och Sverige, när år 1933 tages till basår, d.v.s. detta års multiplikator sättes lika med 1,00. Faktorerna, som torde ha en osäkerhet på omkring (± 0,02), har beräknats enligt den redan tidigare utnyttjade formeln:

$$\text{Multiplikator} = \frac{\text{Föda (Finland)}}{\text{Föda (Finland, 1933)}} : \frac{\text{Livsmedel (Sverige)}}{\text{Livsmedel (Sverige 1933)}}$$

Med hänsyn till dessa tals osäkerhetsmoment ha vi ökat de indirekt beräknade paritetstalens osäkerhetsmarginal till 11 %. Talet för år 1934 har beräknats medelst interpolation mellan 1933 års och 1935 års resultat. Siffran för år 1936 har erhållits genom framskrivning av resultatet för år 1935. Multiplikatorerna för åren 1934, 1935 och 1936 överstiga med 0,02 de tal, som skulle erhållas enligt den ovan angivna formeln.

I tabell VIII ha vi jämfört de sålunda erhållna paritetstalen för föda med årsmedeltalen för Finlands Banks avista försäljningskurs för svenska kronan.

Tab. VIII. *Förhållandet mellan markens och kronans köpkraft gentemot födoämnen åren 1926—36.*

År	Finland Föda-index	Sverige Livsm.-index	Multiplikator	Föda-paritet	Avista kronkurs	Differens	% avvikelse från föda-paritet
1926	1108	158	0,95	9,2 ± 1,0	10,64	1,4 ± 1,0	15 ± 13
27	1115	149	1,01	9,8 ± 1,1	10,66	0,9 ± 1,1	9 ± 12
28	1150	153	1,02	9,9 ± 1,1	10,65	0,8 ± 1,1	8 ± 12
29	1124	150	1,01	9,8 ± 1,1	10,66	0,9 ± 1,1	9 ± 12
30	971	140	0,94	9,1 ± 1,0	10,67	1,6 ± 1,0	18 ± 13
31	869	129	0,91	8,8 ± 1,0	10,81	2,0 ± 1,0	23 ± 14
32	897	125	0,97	9,4 ± 1,0	11,95	2,6 ± 1,0	28 ± 14
33	894	121	1,00	9,7 ± 1,0	11,84	2,1 ± 1,0	22 ± 12
34	876	122	(0,99)	9,6 ± 1,1	11,71	2,1 ± 1,1	22 ± 13
35	911	128	(0,98)	9,5 ± 0,9	11,71	2,2 ± 0,9	23 ± 12
36	902	133	(0,94)	9,1 ± 1,0	11,71	2,6 ± 1,0	29 ± 14

Födaindexen för Finland, vars bas är förra halvåret 1914, ingår som gruppindex i den av Byrån för social forskning beräknade levnadskostnadsindexen. Livsmedelsindexen för Sverige beräknas av K. Socialstyrelsen kvartalsvis. Årsmedeltalen ha uträknats av oss. Juli 1914 utgör bastidpunkt. År 1931 omlades indexberäkningen, bl.a. utökades varuförteckningen över livsmedel med frukter och grönsaker samt pilsner, svagdricka och tobak, vilka varor ej ingå i den finländska födaindexen. Från och med 1932 publiceras den nya indexen med år 1931 som bas. Tillika användes densamma

för fortskrivning av den äldre serien med 1914 som basår.[1])

Av tabell VIII framgår, att födoämneskostnaderna i Sverige omräknade i finska mark med ca (9 ± 12) % överstego motsvarande kostnader i Finland åren 1927—29. År 1936 var motsvarande procent (29 ± 14). Enligt våra beräkningar kan man sålunda icke med någon större säkerhet påstå, att livsmedlen i Sverige före år 1930 voro dyrare än i Finland. Däremot synes det vara ställt utom allt tvivel, att det med hänsyn tagen endast till livsmedelskostnaderna för närvarande är väsentligt dyrare att leva i Sverige än i Finland.

Uppskattning av förhållandet mellan levnadskostnaderna i Finland och Sverige.

För de till hushållsbudgetens övriga huvudgrupper hörande varorna är det svårare att få jämförbara prisuppgifter för Sverige och Finland. I tabell IX ha vi sammanställt det material för år 1935, som stått oss till buds beträffande varugrupperna tobak, bostad, bränsle och lyse, beklädnad, skodon och hygien.

I tabellen ha de av oss uppskattade eller beräknade siffrorna satts inom parentes ( ). För de inom klammer [ ] satta talen ha vi att tacka överaktuarien vid Byrån för social forskning, Helsingfors, fil. mag. Maija Tudeer. Hennes siffror för Finland bygga på prisuppgifter från Helsingfors, vilka med undantag av kol- och koksprisen insamlats för beräkningen av Finlands Banks konsumtionsprisindex.

De icke inom parentes satta vikterna $c^{(1)}$ äro hämtade från tab. 12, sid 35*, Sociala specialundersökningar 14 och $c^{(2)}$ från tab. D i Sociala Meddelanden, 1934, sid. 796. De icke uppskattade prisen inom gruppen bränsle och lyse äro hämtade ur Soc. Tidskr. resp. Soc. Medd. Hyrespriset för Finland utgör ett vägt medeltal av månadshyrorna för två rum och kök såväl med som utan centralvärme. De förra

[1]) En beskrivning på indexen ingår i Soc. Medd., 1932, nr, 4, sid. 318 ff.

Tab. IX. *Markens och kronans köpkraft gentemot konsumtionsvaror 1935.*

Varuslag	Enhet	Finland 1928 $c^{(1)}$	Sverige 1932/33 $c^{(2)}$	Penni 1935 $p^{(1)}$	Öre 1935 $p^{(2)}$	$\frac{p^{(1)}}{p^{(2)}}$	$^{0}/_{00}$ $^{e}\log\frac{p^{(1)}}{10\,p^{(2)}}$
Föda		44,7	35,3				
Tobak		2,0	(1,2)				
Twenty Gold/ Stambul	20 st.	2,0	(1,2)	(500)	(135)	3,70	—995
Bostad		14,0	13,3				
Hyra, H:fors/St:holm	mån.	14,0	15,3	[77956]	(10322)	7,55	—281
Bränsle och lyse		3,9	3,9				
Elektricitet	kwt	(0,6)	(0,8)	325	35	9,29	— 74
Gas	m³	(0,1)	(0,2)	160	17	9,41	— 61
Petroleum	l	(0,1)	(0,1)	164	20	8,20	—198
Björkved	m³	(1,5)	(1,0)	5195	945	5,50	—598
Barrved	m³	(1,3)	(1,3)	4125	730	5,65	—571
Stenkol, antracit	hl	(0,0)	(0,1)	[5500]	613	8,97	—109
» hushålls-	»	(0,1)	(0,1)	[3000]	291	10,31	31
Koks, 20–40 mm.	100 kg	(0,1)	(0,2)	[6000]	452	13,27	283
» 40–80 »	»	(0,1)	(0,1)	[6000]	436	13,76	319
Beklädnad		8,3	9,8				
Herrkostym, konfektion	st	(4,2)	(4,6)	[77193]	[8686]	8,89	—118
Herrskjorta	»	(1,6)	(1,9)	[7477]	[542]	13,79	321
Strumpor, herr-, ylle	par	(1,2)	(1,5)	[2600]	[194]	13,40	293
» dam-, konsts.	»	(1,3)	(1,8)	[3473]	[355]	9,78	— 22
Skodon		3,6	3,0				
Herrskod., box	par	(1,6)	(1,3)	[14796]	[1534]	9,66	— 35
Galoscher	»	(0,8)	(0,5)	[4500]	[366]	12,29	206
Skoreparation	gg	(1,2)	(1,2)	[3377]	[583]	5,79	—547
Hygien		(2,0)	(2,0)				
Simbad, H:fors/St:holm	gg	(1,0)	(1,0)	(900)	(100)	9,00	—105
Barberare, H:fors/St:holm	»	(1,0)	(1,0)	(900)	(175)	5,14	—666
Övriga		(21,5)	(31,5)				
		100,0	100,0				

representeras av 225 och de senare av 123 bostäder. Motsvarande hyrespris för Stockholm ha vi beräknat med stöd av årsmedelhyror för två rum och kök med och utan centralvärme, vilka anges i en opublicerad undersökning verkställd av dr Sten Wahlund vid Stockholms stads statistiska kontor. Prisuppgifterna för tobak och hygien utgöra resultat av privatstatistiska iakttagelser gjorda vid vistelser i Helsingfors och Stockholm.

Resultaten av räkningarna ha vi sammanställt i tabell X. Medeltalen för föda äro desamma som ingå i tabell V (direkt beräknade). Då endast ett prisförhållande för gruppen tobak medtagits vid jämförelsen äro alla fem medeltalen för denna grupp lika. Detsamma gäller för gruppen bostad.

Tab. X. *Förhållandet mellan levnadskostnaderna i Finland och Sverige år 1935.*

Medeltal	Vägning	Föda	Tobak	Bostad	Bränsle och lyse	Beklädnad	Skodon	Hygien	Samtliga
Harmoniskt	$c^{(1)}$	9,09	3,70	7,55	6,34	10,23	8,20	6,54	8,28
Geometriskt	$c^{(2)}$	9,37	3,70	7,55	7,07	10,49	8,19	6,80	8,61
Geometriskt ideala	$\frac{1}{2}(c^{(1)}+c^{(2)})$	9,46	3,70	7,55	6,81	10,45	8 39	6,80	8,68
Geometriskt	$c^{(1)}$	9,56	3,70	7,55	6,56	10,41	8,59	6,80	8,75
Aritmetiskt	$c^{(2)}$	10,10	3,70	7,55	7,42	10,69	8,55	7,07	9,31
Gruppvikt	$c^{(1)}$	44,7	2,0	14,0	3,9	8,3	3,6	(2,0)	78,5
	$c^{(2)}$	35,3	(1,2)	13,3	3,9	9,8	3,0	(2,0)	68,5

På grundval av variationerna i resultatet vid användande av de olika medeltalsformlerna skulle vi för förhållandet mellan levnadskostnaderna i Finland och Sverige år 1935 erhålla intervallen $(8{,}7 \pm 0{,}6)$. Den metodtekniska osäkerhetsmarginalen vore sålunda av storleksordningen $\pm$ 7 %.

För bedömning av det totala osäkerhetsmomentet i våra beräkningar begagna vi samma metod som vid den tidigare uppskattningen beträffande födagruppen. Vi skola först bedöma det inflytande de i jämförelsen icke representerade

varorna kunna förväntas utöva. Gruppen »övriga» omfattar utgifterna för inventarier, bildning, sjukvård, personliga tjänster, skatter, förenings- och försäkringsavgifter, m.m. För en stor del av dessa utgiftskategorier är det omöjligt att angiva ekvivalenta varuenheter för de båda länderna. I gruppen ingår en mängd sådana varor, vilka allmänt erkännas vara väsentligt dyrare i Sverige än i Finland, främst personliga tjänster, nöjen och resor. Gruppen torde därför snarast utöva ett sänkande inflytande på slutresultatet. För uppskattningen av dess inverkan på jämförelsens osäkerhet antaga vi, att specialindexen för »övriga» ligger inom intervallen $(8{,}7 \pm s)$, där s utgör ett mått på gruppindextalens spridning från det ideala geometriska medeltalet. På grundval av de i tabell X angivna, med $^1/_2\,(c^{(1)} + c^{(2)})$ vägda medeltalen erhålles $s = 1{,}5$. Om vi t.ex. antaga, att indextalet för gruppen »övriga» vore $8{,}7 - 1{,}5 = 7{,}2$ och att detta tal vid sammanvägningen med de övriga gruppindexerna skall erhålla medelvikten $^1/_2\,(c^{(1)} + c^{(2)}) = {}^1/_2\,(21{,}5 + 31{,}5) = 26{,}5$ skulle detta sänka slutresultatet 8,7 med 0,4(3), d.v.s. med ungefär 5 %. Gruppen »övriga» torde sålunda vid vår jämförelse kunna åstadkomma en resultatförskjutning på ca $\pm$ 5 %.

För uppskattning av en levnadsstandardförskjutnings inverkan på resultaten skola vi beräkna de indextal man erhåller genom att begagna viktsystem hänförande sig till olika välståndsgrupper i Finland enligt 1928 års levnadskostnadsundersökning (tab. 12, sid. 35*, Sociala specialundersökningar 14). Beräkningen utföres sålunda, att de olika gruppernas ideala geometriska medeltal sammanvägas logaritmiskt med begagnande av de i tabell XI angivna viktsystemen. Vikten för hygien utgör härvid summan av utgiftsprocenterna för bad, barberare och tvätt, renhållning m.m. De tre sista raderna i tabellen innehålla resultaten enligt de olika viktsystemen. X' bygger härvid på hypotesen att indextalet för gruppen »övriga» är 7,2. Den mellersta raden anger de medeltal X, som erhållas då gruppen »övriga» icke beaktas vid jämförelsen. Den sista radens siffror ha slut-

ligen erhållits genom antagandet att »övriga»:s indextal vore 10,2.

Tab. XI. *Inflytandet av förändringar i levnadsstandardens höjd på förhållandet mellan levnadskostnaderna i Finland och Sverige år 1935.*

Utgiftsgrupp	Gruppindex		Arbetare och befattningshavare: Utgift per konsumtionsenhet Fmk					Tjänstemän
							Samtliga	
	X_5	$\log X_5$	under 5000	5000–7500	7500–10000	över 10000		Samtliga
Föda	9,46	— 56	54,4	46,5	40,0	32,1	44,7	26,7
Tobak	3,70	—995	2,2	2,0	2,0	1,7	2,0	0,9
Bostad	7,55	—281	13,0	14,3	14,2	13,4	14,0	14,9
Bränsle och lyse	6,81	—384	3,9	4,0	3,6	3,3	3,9	3,5
Beklädnad	10,45	46	7,1	8,4	8,7	9,3	8,3	9,3
Skodon	8,39	—176	3,9	3,7	3,5	2,8	3,6	2,5
Hygien	6,80	—386	1,9	2,0	2,1	2,1	2,0	1,9
Övriga	7,2 / 10,2	—329 / 20	13,6	19,1	25,9	35,3	21,5	40,3
Viktsumma			100,0	100,0	100,0	100,0	100,0	100,0
Resultat: X'			8,52	8,40	8,26	8,11	8,35	8,02
Resultat: X			8,71	8,67	8,65	8,65	8,70	8,65
Resultat: X''			8,94	8,98	9,04	9,17	9,00	9,24

Resultaten i tabell XI äro rätt märkliga. Ifall gruppen »övriga» icke medtages är stabiliteten mycket god. Alla viktsystemen skulle avkortat till två siffror ge resultatet 8,7. Med stigande levnadsstandard ökar osäkerheten i jämförelsen på grund av de icke medtagna varugruppernas (kulturutgifternas) växande andel i totalutgiften. Detta osäkerhetsmoment ha vi redan tidigare beaktat i samband med diskussionen om gruppen »övriga»:s inverkan på resultatet, i det vi genom att använda medelvikten 26,5 för »övriga», vilken motsvarar en rätt hög levnadsstandard, erhöllo osäkerhetsmarginalen $\pm$ 5 %. Det synes därför icke vara nödvändigt att öka den totala osäkerheten med något moment härflytande ur variationer i levnadsstandarden.

Återstår att uppskatta tillförlitligheten hos de erhållna gruppindextalen och det bidrag till osäkerheten i slutresultatet, som härrör ur deras osäkerhetsmoment.

Födagruppen har vi i det föregående rätt ingående behandlat. Vi erhöllo ett totalt osäkerhetsmoment av storleksordningen ± 10 %. Den del av detta moment (± 6 %), som härrör ur de metodtekniska svårigheterna har redan beaktats vid sammanvägningen av grupperna till en totalindex. Det återstående osäkerhetsmomentet utgör ca

$$\pm \sqrt{4^2 + 2^2 + 1^2} \approx \pm 5\ \%.$$

För de övriga grupperna är det svårare att beräkna storleksordningen hos indextalens osäkerhet. De följande uppskattningarna äro närmast att uppfatta som ett försök att kvantitativt uttrycka den grad av misstro vi hysa mot de av oss angivna siffrorna.

Prisförhållandet för tobak anse vi ligga mellan 4,5 och 3,0. Motsvarande osäkerhetsmarginal utgör ca ± 20 %.

Prisförhållandet 7,6 för gruppen bostad torde med ett par procents noggrannhet ange förhållandet mellan hyresnivån i Helsingfors och Stockholm. Enär dock hyrorna variera kraftigt på olika orter i Finland och Sverige är det tämligen stor risk för att denna siffra rätt mycket avviker från det resultat man skulle erhålla, ifall man ägde fullständig kännedom om hyresnivåernas relativa ställning. Vi förmoda dock att en sådan jämförelse skulle ge resultat tillhörande intervallen (7,6 ± 1,0). Bostadsgruppens osäkerhetsmoment uppskatta vi sålunda till ca ± 13 %.

Gruppen bränsle och lyse är rätt väl representerad i vår jämförelse. Vi uppskatta därför dess osäkerhetsmarginal till ca 5 %.

Prisjämförelsen för beklädnadsgruppen måste anses som rätt otillförlitlig. Kvalitetsolikheter hos varorna torde i denna grupp vara större än i de övriga. Dessutom ingå blott ett fåtal varor i jämförelsen. Vi uppskatta gruppens osäkerhetsmarginal till ± 15 %.

Siffrorna för skodon betrakta vi som något tillförlitligare än för den förra gruppen och tillordna desamma därför en osäkerhetsmarginal på ± 10 %.

Till hygiengruppen ha vi vid viktbestämningen hänfört förutom bad och barberare även tvätt och renhållning m.m. Det bristfälliga materialet tillåter oss att tillordna siffrorna en osäkerhetsmarginal på ± 25 %.

Vi antaga, att de enskilda gruppindextalens osäkerhetsmoment äro oberoende av varandra, ett antagande, som i detta fall torde vara berättigat. För sammanvägningen av osäkerhetsmomenten kunna vi under denna förutsättning begagna den tidigare utnyttjade satsen för addition av oberoende variablers spridningsmoment. På detta sätt erhålles ett sammanlagt osäkerhetsmoment av storleksordningen

$$\pm \frac{1}{73{,}5} \sqrt{(5 \cdot 40{,}0)^2 + (20 \cdot 1{,}6)^2 + (13 \cdot 13{,}7)^2 + (5 \cdot 3{,}9)^2 + (15 \cdot 9{,}0)^2 + (10 \cdot 3{,}3)^2 + (25 \cdot 2{,}0)^2} \approx \pm 4{,}5\ \%.$$

I det föregående ha vi uppskattat gruppen »övriga»:s sannolika inflytande till ± 5 % samt den metodtekniska osäkerhetsmarginalen till ± 7 %. Addera vi alla dessa moment enligt regeln för addition av oberoende variablers moment erhålla vi ett sammanlagt osäkerhetsmoment i slutresultatet på ca

$$\pm \sqrt{4{,}5^2 + 5^2 + 7^2} \approx \pm 10\ \%.$$

Det torde dock kunna ifrågasättas, huruvida det kvadratiska additionsförfarandet är tillämpligt i detta fall. Genom enkel addition erhålles ett resultat på ca ± 16 %. Enligt vår uppfattning träffar den förra uppskattningen dock närmare den riktiga storleksordningen.

Den största osäkerhetsfaktorn i vår jämförelse utgör den metodtekniska. Ifall man kunde enas om, att enbart indexformler av den geometriska typen och tvåbasiska formler av samma typ som *Fischer*'s ideala indexformel, kunna komma ifråga som metodtekniska lösningar på indexproble-

met, skulle detta osäkerhetsmoment minska väsentligt. I vårt fall skulle detsamma sålunda utgöra endast ca ± 1 % av medeltalet.

Med tillgång till ett fullständigare prismaterial skulle det osäkerhetsmoment, som beror på att de medtagna varorna representera endast en del av totalkonsumtionen, kunna minskas betydligt. Den osäkerhetsfaktor, som härrör ur variationer i levnadsstandardens höjd, och därmed följande viktförskjutningar, kan icke reduceras på annat sätt än genom ett noggrannt avgränsande av de bägge enonomikretsarna i avseende å den levnadsstandard som betraktas. I vårt fall bestämmes denna osäkerhetsfaktor huvudsakligast av skillnaden mellan gruppen »övriga»:s indextal och medeltalet för de i jämförelsen representerade varornas prisförhållanden. (Jfr. tab. XI.)

Det osäkerhetsmoment, som härflyter ur svårigheten att finna ekvivalenta varor, torde vara av mindre betydelse med undantag för grupperna beklädnad och bostad. De medeltalsbildningar av olika orters pris, som företages vid uträknandet av årsmedelprisen för ett helt land synas även ge tillförlitliga resultat. Bestämningen av hyresnivån, vilken varierar kraftigt på olika orter, stöter dock på speciella svårigheter. [1])

I enlighet med den ovan förda diskussionen beträffande osäkerheten i resultatet av vår jämförelse mellan levnadskostnaderna i Finland och Sverige år 1935 synes det vara förnuftigt att tillordna slutresultatet 8,7 en osäkerhetsmarginal på ca ± 10 %. Den kvantitativa motsvarigheten till det studerade begreppet utgöres sålunda enligt våra beräkningar av intervallen (8,7 ± 0,9).

Med tillhjälp av de officiella levnadskostnadsindexerna i Finland och Sverige fortskriva vi resultatet till åren 1926—36. I tabell XII ha vi jämfört de sålunda erhållna siffrorna med årsmedeltalen för Finlands Banks avista försäljningskurs för svenska kronan.

[1]) R. Guye: A further contribution to the international comparison of rents, International labour review, 1936.

Tab. XII. *Förhållandet mellan levnadskostnaderna i Finland och Sverige åren 1926—36.*

År	Levnadskostnadsindex		Multiplikator	Paritetskurs X	Avista kronkurs	Differens	% avvikelse från paritet	% avvikelse från kronkurs
	Finland	Sverige						
1926	1183	172	1,08	9,4 ± 1,0	10,64	1,2 ± 1,0	13 ± 12	12 ± 10
27	1207	171	1,10	9,6 ± 1,1	10,66	1,1 ± 1,1	11 ± 12	10 ± 10
28	1233	172	1,12	9,7 ± 1,1	10,65	1,0 ± 1,1	10 ± 12	9 ± 10
29	1225	170	1,13	9,8 ± 1,1	10,66	0,9 ± 1,1	10 ± 12	8 ± 10
30	1129	165	1,07	9,3 ± 1,0	10,67	1,4 ± 1,0	15 ± 13	13 ± 10
31	1039	159	1,02	8,9 ± 1,0	10,81	1,9 ± 1,0	21 ± 13	18 ± 9
32	1025	156	1,03	9,0 ± 1,0	11,95	3,0 ± 1,0	33 ± 15	25 ± 8
33	1001	153	1,02	8,9 ± 1,0	11,84	2,9 ± 1,0	33 ± 15	25 ± 8
34	983	154	1,00	8,7 ± 1,0	11,71	3,0 ± 1,0	35 ± 15	26 ± 8
35	997	156	1,00	8,7 ± 0,9	11,71	3,0 ± 0,9	35 ± 14	26 ± 7
36	998	158	0,99	8,6 ± 0,9	11,71	3,1 ± 0,9	36 ± 15	27 ± 8

Multiplikatorerna ha uträknats enligt samma metod som vid fortskrivningen av paritetssiffrorna för födagruppen. De torde ha en osäkerhetsmarginal på ± 0,03. Med hänsyn härtill har osäkerheten i de fortskrivna siffrorna ökats med ± 1 %.

Av tabell XII framgår, att man ej med full säkerhet kan påstå, att finska marken betraktad ur konsumentsynpunkt var undervärderad i förhållande till svenska kronan före år 1930, ehuru det förefaller sannolikt. Däremot torde man med bestämdhet kunna säga, att för närvarande förefinnes en betydande undervärdering. År 1936 voro sålunda enligt våra beräkningar levnadskostnaderna i Finland (27 ± 8) % lägre än motsvarande kostnader i Sverige. Sett ur motsatt synvinkel voro dessa kostnader (36 ± 15) % dyrare i Sverige än i Finland.

Inom den osäkerhetsintervall vi erhållit för våra slutliga siffror över förhållandet mellan levnadskostnaderna i Finland och Sverige torde med mycket stor sannolikhet de resultat falla, som man skulle erhålla, ifall man hade tillgång till ett i alla avseenden tillfredsställande prismaterial och full-

ständig kännedom om konsumtionens sammansättning och utgifternas fördelning.

För belysande av de erhållna resultaten ha vi även försökt utnyttja Engels lag för beräknande av förhållandet mellan levnadskostnaderna. Genom att undersöka vilka utgifter per konsumtionsenhet i Sverige och Finland, som motsvara varandra i det avseende, att utgiftsprocenterna för föda äro lika stora i de båda länderna, erhöllo vi ett resultat, som fortskrivet till år 1935 skulle givit en paritetskurs på 1 kr = ca 10 Fmk. Denna kurs överstiger med något mer än 1 Fmk det enligt vår åsikt betydligt tillförlitligare resultatet (8,7 $\pm$ 0,9). Det relativt höga värdet förklaras av att födapariteten med ca 10 % överstiger motsvarande paritet för levnadskostnaderna. (Jfr. tab. X). På grund av de relativt billiga födoämnena i Sverige är det nämligen naturligt, att vid lika hög levnadsstandard i de båda länderna utgiftsprocenten för föda är mindre i Sverige än i Finland. Ifall vi antaga, att denna faktor skulle åstadkomma en minskning av utgiftsprocenten för föda i Sverige från t.ex. 38 till 35, skulle Engels lag ge en paritetskurs på 1 kr = ca 9 Fmk. Resultatet enligt denna lag synes sålunda väl förenligt med det tidigare erhållna. Intressant vore att, då den utlovade utförligare redogörelsen för de senaste levnadskostnadsundersökningarna i Sverige blir publicerad, noggrannare undersöka i vad mån man kan utnyttja viktförskjutningsfenomenet för beräkning av förhållandet mellan levnadskostnaderna i Finland och Sverige.

Säsongfluktuationerna i Finlands Banks sedel- och myntcirkulation åren 1870—1937.

Av fil. dr LEO TÖRNQVIST.

Säsongfluktuationerna i det ekonomiska livet ha icke tilldragit sig någon större uppmärksamhet i den inhemska litteraturen. Ett glädjande undantag bildar en undersökning av dr F. Burjam »Om säsongfluktuationerna inom bankverksamheten i Finland», som publicerats i Ekonomiska Samfundets Tidskrift, ny serie, H. 23, 1932. Den metod han begagnat vid undersökningen är den enklast möjliga, nämligen beräkning av aritmetiska medeltal av de olika månadernas siffror för perioderna 1874—1883, 1904—1913 och 1924—1929. Någon trendjustering gör han icke. Hans uppsats innehåller en hel del värdefulla synpunkter på de omständigheter, som kunna tänkas förklara de iakttagna förskjutningarna i säsongfluktuationerna under den betraktade perioden.

I det följande skola vi redogöra för en undersökning [1]) av säsongfluktuationerna i sedel- och myntcirkulationen i Finland. Den utgör en del av en större undersökning av den

[1]) Vid de rätt vidlyftiga räkningarna har jag assisterats av min hustru och min broder Gösta. Ett antal normalsäsongindexar ha beräknats som övningsuppgifter i en av författaren ledd kurs i statistik vid Åbo Akademis statsvetenskapliga fakultet. Till mina medhjälpare och till Finlands Bank, som bekostat undersökningen och ställt material till förfogande, ber jag att få framföra mitt varma tack. Ett särskilt tack är jag skyldig min hustru, som hjälpt mig med renskrivningen och ritandet av diagrammen.

utelöpande sedelmängdens variationer och det samband, som existerar mellan sedelomlopp och andra ekonomiska tidsserier.

Materialet för undersökningen.

Utgångsmaterialet för säsongindexberäkningarna har utgjorts av de i bihanget, tab. IX,1 och IX,2, angivna talen över Finlands Banks utelöpande sedlar och bronsmynt vid utgången av de olika månaderna åren 1870—1937. Bronsmynten ha medtagits från och med december 1929, då de i större mängd begynte utsläppas i rörelsen ersättande 5, 10 och 20 mks sedlarna. Siffrorna ha hämtats ur Finlands Banks bokföring. Även sedlar av äldre modell, vilka sannolikt upphört att cirkulera, ha medtagits. Några korrektioner för de ur cirkulationen försvunna sedlarna och mynten ha sålunda icke gjorts, men detta torde blott i ringa mån påverka säsongindexarna. De absoluta talen äro dock på grund av denna omständighet något för stora. Endast de emitterade sedlarna ha medtagits, icke hela den begagnade sedelutgivningsrätten. Postremissväxlarna ha t. ex. icke medräknats. Under 1870-talet emitterade även Föreningsbanken i Finland sedlar. Dessa nådde ett maximum i nov. 1882 på 1.42 milj. mk (3 % av Finlands Banks sedlar, som då utgjorde 45.2 milj. mk). Utgångssiffrorna få således icke uppfattas såsom angivande hela beloppet betalningsmedel, som cirkulerat i landet. Blott i den mån de av Finlands Bank utgivna sedlarna och bronsmynten kunna antagas ha utvecklat sig proportionellt med kvantiteten cirkulerande betalningsmedel i vidsträcktare bemärkelse, ge de en riktig bild av de relativa förändringarna i betalningsmedlens kvantitet. Finlands Banks vid anfordran betalbara förbindelser kunde man med fog uppfatta såsom utelöpande betalningsmedel, enär de torde göra samma tjänst som en lika stor summa sedlar. Möjligen ha de en något mindre omloppshastighet.

Det hade varit önskvärt, att undersökningen skulle kunnat

baseras på summan av sedlar och hårt mynt. Någon väsentlig skillnad synes nämligen icke förefinnas mellan det hårda myntet och sedlarna i avseende å deras ekonomiska verkningar; de torde, utan att detta märkbart skulle förändra betalningsströmmarnas storlek, kunna ersätta varandra. Utbytandet av 5, 10 och 20 mks sedlarna mot bronsmynt, vilket ägde rum åren 1929—1937, minskar väsentligt jämförbarheten mellan siffrorna för den utelöpande sedelmängden före och under denna period. Med hänsyn härtill ha vi ansett det vara skäl att addera bronsmynten till de utelöpande sedlarna. I brist på statistiska uppgifter ha vi icke kunnat beräkna summan av de utelöpande sedlarna och mynten vid utgången av varje månad under hela den betraktade perioden. För att möjliggöra ett bedömande av, i vad mån ett medtagande av mynten skulle förändra den bild av konjunktur- och säsongrörelserna, som enbart de utelöpande sedlarna giva, ha vi i tabell 1 sammanställt en del siffror, som belysa myntens och sedlarnas relativa betydelse vid utgången av åren 1872, 1874, . . . , 1928—1937. De siffror vi medtagit hänföra sig i regel till år, för vilka årsskiftet ligger nära ett maximi- eller minimivärde för sedelcirkulationen. De årtal, för vilka detta gäller, ha tryckts med fet stil. Siffrorna för åren 1872—1898 ha hämtats ur Finlands Banks årsberättelser. I dessa angivas siffror för de utelöpande silver- och kopparmynten blott för perioden 1872—1890. Siffrorna för perioden 1901—1913 äro hämtade ur Statistisk Årsbok för Finland. Siffrorna för åren 1921—1936 ha tagits ur Finlands Banks årsböcker, med undantag av siffrorna för 5 och 10 pennis kopparmynten, och talen för år 1937, för vilka jag har att tacka Finlands Banks statistiska avdelning, som välvilligt ställt till mitt förfogande siffror över de utelöpande myntens och sedlarnas sammansättning vid utgången av varje kvartal åren 1926—1934 och vid utgången av varje månad åren 1935—1937. Dessa siffror skola vi senare analysera. Vi omnämna dock redan nu ett av resultaten, nämligen att mynten visa en betydligt mindre och från sedlarnas säsongrörelser rätt mycket avvikande säsongmässig variation.

Tabell 1. *Finlands Banks utelöpande sedlar och mynt den* $^{31}/_{12}$ *åren 1872—1937 i milj. mark.*

År	I	II	III	IV	V	VI	VII	VIII	IX	X	År
	Sedlar	Guld	Silver	Koppar	=I+II	=III+IV+V	Mynt	III+IV	VII : I	VIII : VI	
1872	49,83	0	6,22	0,48	49,83	56,53	6,70	6,70	0,135	0,119	1872
1874	73,45	0	7,14	0,58	73,45	81,17	7,72	7,72	0,105	0,095	1874
1879	36,99	2,72	5,97	0,66	39,71	46,34	9,35	6,63	0,253	0,143	1879
1880	48,61	3,02	6,52	0,69	51,63	58,84	10,23	7,21	0,210	0,123	1880
1883	47,15	4,89	6,69	0,76	52,04	59,49	12,34	7,45	0,262	0,125	1883
1886	39,97	4,55	6,47	0,80	44,52	51,79	11,82	7,27	0,296	0,140	1886
1889	55,31	4,44	7,21	0,90	59,75	67,86	12,55	8,11	0,227	0,120	1889
1893	44,44	7,38	.	.	51,82	.	.	.	.	.	1893
1898	76,82	7,48	.	.	84,30	.	.	.	.	.	1898
1901	62,36	8,79	8,13	1,44	71,15	80,72	18,36	9,57	0,294	0,119	1901
1904	72,50	15,29	9,00	1,57	87,79	98,36	25,86	10,57	0,357	0,107	1904
1907	95,03	14,85	11,06	1,86	109,88	122,80	27,77	12,92	0,292	0,105	1907
1908	85,99	14,87	10,87	1,92	100,86	113,65	27,66	12,79	0,322	0,113	1908
1911	115,26	19,48	11,44	2,13	134,74	148,31	33,05	13,57	0,287	0,091	1911
1913	113,00	32,98	12,06	2,31	145,98	160,35	47,35	14,37	0,419	0,090	1913
1921	1356,11	Guld och silvermynt cirkulera ej		Koppar 5 och 10 p	Sedlar och bronsmynt	Sedlar och mynt	Myntcirkulation	Nickel och kopparmynt	Mynt / Sedlar	Småmynt / Summa	1921
1922	1420,92										1922
1924	1249,95	Brons	Nickel	IV	V=I+II	III+IV+V	II+III+IV	III+IV	VII : I	VIII : VI	1924
1928	1513,21	0	39,56	6,18	1513,21	1558,95	45,74	45,74	0,030	0,029	1928
1929	1360,63	4,74	40,48	6,33	1365,37	1412,18	51,55	46,81	0,038	0,033	1929
1930	1279,35	5,78	41,34	6,43	1285,13	1332,90	53,55	47,77	0,042	0,036	1930
1931	1292,91	10,02	40,02	6,46	1302,93	1349,41	56,50	46,48	0,044	0,034	1931
1932	1085,26	42,01	39,62	6,43	1127,27	1173,32	88,06	46,05	0,082	0,039	1932
1933	1183,64	53,60	41,25	6,52	1237,24	1285,01	101,37	47,77	0,086	0,037	1933
1934	1277,38	63,89	44,27	6,77	1343,27	1392,31	114,93	51,04	0,090	0,037	1934
1935	1380,68	69,12	47,56	7,00	1449,80	1504,36	123,68	54,56	0,090	0,036	1935
1936	1630,12	77,52	51,66	7,30	1707,64	1766,60	136,48	58,96	0,084	0,033	1936
1937	2051,84	99,69	59,05	7,73	2151,53	2218,31	168,47	66,78	0,082	0,030	1937

Såsom av de relativa talen i de två sista kolumnerna av tabell 1 framgår, var det hårda myntets betydelse väsentligt större under 1880-talet än efter inflationen 1915—1920. Man observerar vidare, att kvoten mellan värdet av de utelöpande mynten och sedlarna var större under lågkonjunktur- än under högkonjunkturår. Guldmyntens och kopparmyntens variationer stå icke i något märkbart samband med de konjunkturmässiga fluktuationerna. Silvermynten och seder-

mera nickelmynten visa däremot tydliga konjunktursvängningar. Summan av sedlarna och mynten har en relativt mindre konjunktursvängningsamplitud än enbart sedlarna.

Emedan de säsongmässiga variationerna i de utelöpande mynten att döma av vår undersökning för perioden 1926—1937 äro relativt små, torde man kunna antaga, att summan av sedlarna och mynten har en något mindre amplitud i säsongvariationerna än enbart sedlarna. Före år 1913 skulle myntets dämpande inflytande ha varit väsentligt större än efter år 1920. De relativt små belopp nickel- och kopparmynt, som efter inflationen utgivits, skulle icke förändra säsongindextalen för de utelöpande sedlarna och bronsmynten med mer än 1—2 promille. Om man sålunda måste antaga, att ett medtagande av mynten skulle minska säsongfluktuationerna, så har man å andra sidan att räkna med, att en korrigering av utgångssiffrorna med hänsyn till de ur rörelsen försvunna sedlarna (borttappade, m.m.) skulle påverkat säsongindextalen i motsatt riktning.

Det är kanske värt att påpeka, att ifall man till de utelöpande sedlarna tillfogar Finlands Banks vid anfordran betalbara förbindelser, så blir bilden av säsongrörelserna väsentligt förändrad. Till en icke ringa del motverkas nämligen en ökning i de utelöpande sedlarna av en minskning i övriga vid anfordran betalbara förbindelser och tvärtom.

Under den betraktade perioden har de kontantfria betalningarnas relativa betydelse ökat. De stora betalningarna ske numera huvudsakligen medelst checker. Denna omständighet torde vara en väsentlig orsak till de kraftiga förskjutningar i säsongfluktuationerna, som ägt rum under den studerade perioden.

Om beräkning av säsongindextal.

Problemet att finna en i alla avseenden tillfredsställande metod för beräkningen av goda säsongindextal är ganska

komplicerat [1]). Uppgiften försvåras av, att det gäller att få ett kvantitativt mått på en storhet, som man icke kan fullt entydigt definiera. De säsongmässiga variationerna utgöra en komponent hos en tidsserie, som dessutom innehåller flere andra komponenter, långtidstrend, konjunkturmässiga variationer, tillfälliga oregelbundenheter m.m. Om säsongfluktuationerna regelbundet återkomma år för år utan större förändringar, är det jämförelsevis lätt att beräkna goda indextal. Men om säsongrörelserna, såsom i regel torde vara fallet, äro föränderliga från ett år till ett annat, blir det svårare att söka »renodla» säsongrörelserna ur den givna tidsserien. För att kunna dämpa ner de oregelbundna korttidsfluktuationernas inverkan på säsongindextalen, vore man frestad att lägga en möjligast lång period till grund för beräkningen. Men då säsongrörelserna variera med tiden, kan man genom en medeltalsbildning över ett flertal år lätt väsentligt förvanska också de rent säsongmässiga rörelserna. Därtill kommer olägenheten, att den betraktade periodens mittpunkt förskjutes allt längre tillbaka i tiden, räknat från periodens slut, ju längre tidsintervall, som lägges till grund för beräkningen. Detta minskar den erhållna säsongindexens användbarhet för elimineringen av säsongfluktuationerna vid periodens början och slut. Ju kortare period man betraktar, desto instabilare bli säsongindextalen i avseende å förskjutningar av mittpunkten. Betraktar man blott ett bestämt års säsongfluktuationer, så är det ofta möjligt att jämförelsevis väl eliminera långtidstrenden genom att sätta de olika månadernas siffror i relation till ett löpande årsmedeltal eller en grafiskt eller analytiskt bestämd trend. Men däremot ställer det sig svårt att eliminera de språngartade förändringarna i konjunkturkomponenten och de oregelbundna korttidsfluktuationerna ur de på detta sätt erhållna säsongindextalen. Genom att

[1]) En god översikt över de metoder, som kommit till användning vid beräkningar av säsongindextal, ges i *Horst Mendershausen, Annual Survey of Statistical Technique: Methods of Computing and Eliminating Changing Seasonal Fluctuations*, Econometrica, July, 1937.

för en viss månad beräkna dylika »ettåriga» säsongindextal för en hel serie på varandra följande år erhåller man en ny tidsserie, som man kan försöka dekomponera i långtidstrend, »konjunkturrörelser» och oregelbundna korttidsfluktuationer. Med en så kallad normal säsongindex avses i regel en index, som beräknats på grundval av siffror för en så lång period, att indexen kan betraktas som ett närmevärde för långtidstrenden i de »ettåriga» säsongindextalens tidsserie, avläst vid periodens mittpunkt. Detta uppnås i allmänhet först för periodlängder på omkring 15 år, i bästa fall vid 7—10 år. I attributet normal inbegriper man i regel en förhoppning om, att indexen skall ha ett prognostiskt värde, i det att man tror, att den anger ett slags matematisk förväntan för de »ettåriga» säsongindexarna i den närmaste framtiden. Tager man fasta på det prognostiska elementet hos en normalsäsongindex, är det mest konsekvent att ställa en prognos för säsongindextalens trendutveckling och begagna den sålunda erhållna säsongindextrenden vid eliminerингen av de »normala» säsongrörelserna ur den betraktade tidsserien. Vill man gå ett steg längre och ställa en prognos, som även strävar till att taga hänsyn till de »konjunkturmässiga» variationerna i de ettåriga säsongindextalen, ökas riskerna för att misslyckas betydligt. Till frågan, huruvida det faktiskt finnes något samband mellan de konjunkturmässiga variationerna i sedelcirkulationen och de svängningar med c:a 5—10 års periodlängd, som kunna observeras hos säsongindextalen, skola vi senare återkomma.

Vid våra beräkningar av säsongindextal för sedelcirkulationens säsongfluktuationer ha vi begagnat oss av flere olika metoder. Först beräknade vi säsongindexar för årtiondena 1870—1879, 1880—1889, . . . , 1920—1929, 1928—1937 enligt en modifierad Harvardmetod. Då det likväl är praxis i litteraturen, att man begagnar indelningssättet 1871—1880, 1881—1890 o.s.v., beslöto vi att beräkna också de till dessa årtionden hörande säsongindextalen i enlighet med samma metod. Härigenom fingo vi tillika en möjlighet att bedöma de erhållna säsongindextalens stabilitet i

avseende å förskjutningar av den betraktade periodens mittpunkt med ett år. De närmare detaljerna vid beräkningen och de numeriska resultaten framgå av tabellbilagorna.

Den begagnade modifikationen av Warren M. Persons' berömda kedjekvotmetod [1]) har av författaren ursprungligen utarbetats för Affärsekonomiska forskningsinstitutets (A. E. F.) räkning. Den skiljer sig från Persons' metod däri, att de s.k. kedjekvoterna (d.v.s. kvoterna mellan talen för februari och januari, mars och februari o.s.v.) ersatts med differenserna mellan på varandra följande månaders utgångstals naturliga logaritmer (promilletal), och att det av Persons rekommenderade medianvärdet för bestämningen av den normala månad-till-månad-kedjekvoten ersatts av en generellare medeltalsbildningsmetod, nämligen ett medelst binomialkoefficienter vägt medeltal av månad-till-månadlogaritmdifferenserna. Vikterna fördelas härvid så, att de två största och minsta differenserna erhålla vikten noll, de övriga i storleksföljd (1, 5, 10, 10, 5, 1). Trenden elimineras genom att justera normaldifferenserna så, att deras summa blir = 0.

Metodens styrka ligger i, att den väl eliminerar exponentiellt växande eller avtagande trender (i logaritmisk skala räta linjer) samt stora sannolikt icke säsongmässiga språng i tidsserien. Däremot kan den icke eliminera trendlinjer, som i logaritmisk skala äro jämnt krökta, annat än i den mån de kunna approximeras av en i logaritmisk skala bruten rät linje med årsskiftena såsom hörnpunkter. Denna begränsning kan i vissa fall ge upphov till ett litet metodtekniskt fel. Professor Oskar Andersson har kritiserat Persons' kedjekvotmetod med hänsyn till, att den inregistrerar säsongmässiga variationer även hos en i aritmetisk skala lineärt växande tidsserie. [2]). För att eliminera det metod-

[1]) *W. M. Persons, Indices of Business Conditions,* Review of Economic Statistics, January, 1919.

[2]) *Oskar Andersson, The Logic of the Decomposition of Statistical Series,* Journal of the Royal Statistical Society, Part III, 1927 Page 548—569.

tekniska felet kunde man begagna sig av en dubbel trendeliminering, i det man före medeltalsberäkningarna skulle eliminera trenden i månad-till-månad-logaritmdifferenserna genom att från dessa subtrahera medeltillväxten per månad hos logaritmen för sedelomloppet under det år (eller bekvämare 2-årsperiod), som har de ifrågavarande månaderna såsom mittpunkt, varpå man fortsätter enligt det ursprungliga schemat. I regel står dock ej den obetydliga förbättringen av slutresultatet i en rimlig proportion till det ökade räknearbetet. Vid relativt starka månatliga variationer i de icke säsongmässiga komponenterna är den förbättring, som kan uppnås på detta sätt, av större betydelse.

En svaghet hos den modifierade Harvardmetod, som vi i det föregående i korthet beskrivit, är att den icke är tillräckligt bekväm för en beräkning av säsongindexar för löpande perioder. För detta ändamål ha vi föredragit att begagna en annan metod. Vid denna eliminerar man först trendkomponenten hos månad-till-månad-logaritmdifferenserna genom att från dessa subtrahera medeltillväxten under en två års period med de ifrågavarande månaderna såsom mitt. Därpå söker man eliminera starka språngvisa förändringar i tidsserien genom att ersätta den för det ifrågavarande året erhållna trendrensade månad-till-månad-logaritmdifferensen med den i storleksordningen mellersta av differenserna för de tre på varandra följande år, av vilka det betraktade året är det mellersta. För att ytterligare dämpa ner de oregelbundna korttidsfluktuationerna beräknas fem års löpande medeltal av de sålunda erhållna talen. Räkningarna slutföras enligt samma schema som vid den förstnämnda metoden. I fråga om närmare detaljer och siffror hänvisas till bihanget.

De erhållna säsongindexserierna ha vi åskådliggjort grafiskt. Med tillhjälp av Ragnar Frisch's inflexionspunktsmetod [1])

[1]) *R. Frisch, Changing Harmonics and other General Types of Components in Empirical Series,* Skandinavisk Aktuarietidskrift, H. 4, 1928, och *A Method of Decomposing an Empirical Serie into its Cyclical and Progressive Components,* American Statistical Journal, March Supplement 1930.

ha vi därpå tecknat långtidstrenderna för säsongindextalen. De grafiskt bestämda trenderna ha vi avläst för vart femte år (se tabell II,3 och III,3 i bihanget). Trendserierna ha vi justerat så, att årssumman av de olika månadernas säsongindextals logaritmer blivit lika med noll. Vid justeringen iakttogs tillika att månad-till-månaddifferensernas trender skola kunna beräknas ur trendserierna för säsongindextalens logaritmer och omvänt.

Vid beräkningen av säsongindextalen för kopparmynten ha vi med hänsyn till de förhållandevis små och långsamma rörelserna hos dessa ansett det tillåtligt att begagna en mindre arbetskrävande metod bestående i en beräkning av enkla aritmetiska medeltal av de olika månadernas (kvartalens) siffror, vilka därpå justerats med hänsyn till trenden sålunda, att vi subtraherat från de olika månadernas medeltal en korrektionsterm, erhållen genom att multiplicera medeltillväxten per år i serien med månadens avstånd från årets mitt, uttryckt i bråkdelar av ett år (negativt tecken för månaderna jan.—juni, positivt för månaderna juli—dec.) samt slutligen förvandlat de sålunda erhållna korrigerade månadsmedeltalen till indextal genom division med deras medeltal.

Finlands Banks statistiska avdelning har för perioden 1925—1934 beräknat en säsongindex enligt metoden att 1) beräkna löpande tolv månaders medeltal av sedel- och bronsmyntcirkulationen, 2) dividera de olika månadernas utgångssiffror med det av tolv månaders medeltalen, för vilket den betraktade månaden utgör den sjunde i ordningen, samt därpå 3) beräkna enkla aritmetiska medeltal av de olika månadernas sålunda erhållna »ettåriga» säsongindextal, vilka medeltal slutligen 4) justeras så, att deras summa blir = 1200,0. Metoden eliminerar väl jämnt krökta trendlinjer, men gör det möjligt för en språngvis skeende icke säsongmässig variation att få ett stort inflytande på normalsäsongindextalen. En jämförelse mellan Finlands Banks index och den av mig för samma period beräknade har verkställts i tabell I,3.

Resultaten av säsongindexberäkningarna.

Säsongfluktuationerna i Finlands Banks utelöpande sedlar har under perioden 1870—1937 icke uppvisat någon större stabilitet. De i tabellbilagorna I—V sammanställda säsongindextalen ha grafiskt åskådliggjorts i diagrammen 1—5. Vi ha ritat kurvorna i diagram 1 och 5 med jämn

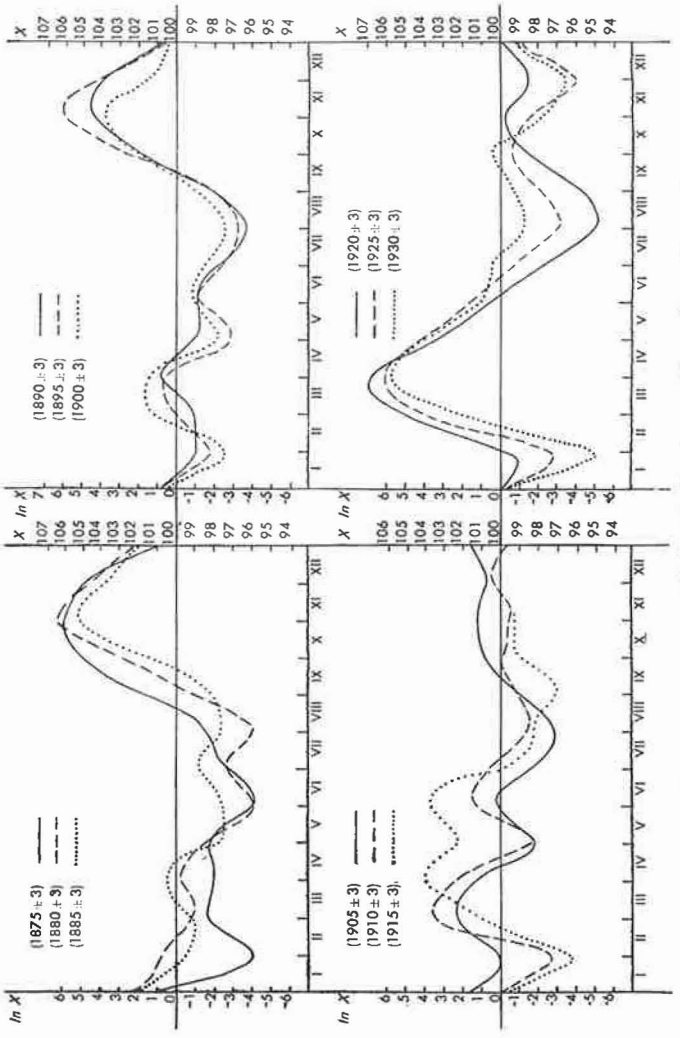


Diagram 1. Säsongfluktuationerna i Finlands Banks utelöpande sedlar och bronsmynt. Säsongindextalen *ln X* återfinnas i tabell II, 1 och II, 2.

krökning mellan punkterna för respektive månads ultimo säsongindextal. I själva verket förefinnes i sedelcirkulationen en korttidsfluktuation med månadsskiftena såsom maxima. Dessa stå naturligtvis i samband med löneutbetalningarna och härvid främst utbetalningarna till statens tjänstemän.

I början av perioden visade sedelcirkulationen i slutet av oktober månad ett kraftigt säsongmässigt maximum, som småningom minskades för att försvinna omkring år 1910. Av höstmaximet återstår blott en mindre utvidgning av cirkulationen i augusti och september månad. Höstmaximets minskning torde i huvudsak förklaras av checkrörelsens ökade betydelse. Checkclearingen har en kraftig stegring i oktober månad. Åren (1875 ± 3) var sedelcirkulationen minst i maj månad. I juni förekom en svag stegring, vilken redan omkring år 1880 utvecklats till ett tydligt maximum. F. Burjam, som även observerade junimaximet i sin A-kurva för perioden 1875—1884, skriver om detta: »Med seglationstidens början vidtar och vidtog på 1870-talet i, relativt taget, ännu högre grad än nuförtiden, en livlig import och exportverksamhet. Betalningsrörelsen vidgades av denna anledning och den utelöpande betalningsmedelsmängden tilltog i motsvarande grad. Men samtidigt ökades betalningsmedlen också därför att icke blott importörerna, utan också deras avnämare, minuthandlarna, skyndade sig komplettera sina lager. Därtill var verksamheten inom jordbruket på grund av vår- och såningsarbetena ytterst intensiv. Dessa företeelser finna sitt uttryck i A-kurvans juni-stegring»[1]). Skulle denna förklaring vara riktig, förefaller det egendomligt, att junimaximet blott kan iakttagas under den relativt korta perioden (1879 ± 3) till (1886 ± 3). Av en granskning av de trendrensade månad-till-månad-differenserna framgår, att junimaximet förekom närmare bestämt under åren 1878—1889, dessförinnan steg sedelomloppet säsongmässigt från juni till oktober.

[1]) *F. Burjam, Säsongfluktuationerna i bankverksamheten.* Ekon. samf. Tidskr., Ny serie, H. XXIII, s. 77.

Därefter faller det från 31 maj till 31 juli utan någon junistegring. Förklaringen till denna utveckling är, att junimaximet under 1890-talet försköts till maj månad. Seglationsperioden började tidigare på året? Åren 1917—1918 försvinner majmaximet, statistiskt sett, något oväntat. Om man bortser från perioden 1870—1874, har säsongen under hela perioden varit fallande från 31 mars till 30 april. Aprilmaximet var kraftigast utvecklat omkring år 1910, därefter har mars—april differensen småningom minskats. Aprilminimet försvinner åren 1917—1918 i samband med majmaximets försvinnande. Februari—mars maximet, vilket åren 1870—1876 befann sig vid månadsskiftet februari—mars, förflyttade sig därefter till den 31 mars, där det förblev till omkring år 1900. Under perioden 1900—1917 var det åter beläget i slutet av februari därefter i mars. Maximet var långsamt stigande före världskriget, för att omkring år 1915 snabbt växa till ett kraftigt vårmaximum, som utfyllt aprilminimet. Numera utvisar säsongindexen en starkt växande sedelcirkulation till den 31 mars och därpå en återströmning av sedlarna till den 31 juli.

Marsmaximet torde i huvudsak kunna förklaras av den ökade efterfrågan på betalningsmedel, som skogsarbetena i norra Finland förorsaka. Sedlarna uttagas främst via statens upp- och avskrivningsräkning.

En detalj i säsongfluktuationernas historia kan det ännu vara värt att anföra. De första antydningarna till det numera väl utpräglade årsskiftesmaximet kunna observeras under perioden 1900—1910. Under krigsåren försvinner detsamma, men återkommer därpå betydligt kraftigare utvecklat. Detta maximum kan förklaras, förutom av julmarknadens anspråk, av bankernas ökning av sina kassor vid årsskiftet.

Diagrammen över säsongrörelserna ge en ganska intressant belysning av de kraftiga strukturförändringar, som ägt rum i Finlands näringsliv under åren 1870—1937. Förändringarna i betalningsvanorna, speciellt partihandelns övergång till att begagna checker vid affärstransaktionerna

och industrins kraftiga uppsving finna i dem ett åskådligt uttryck.

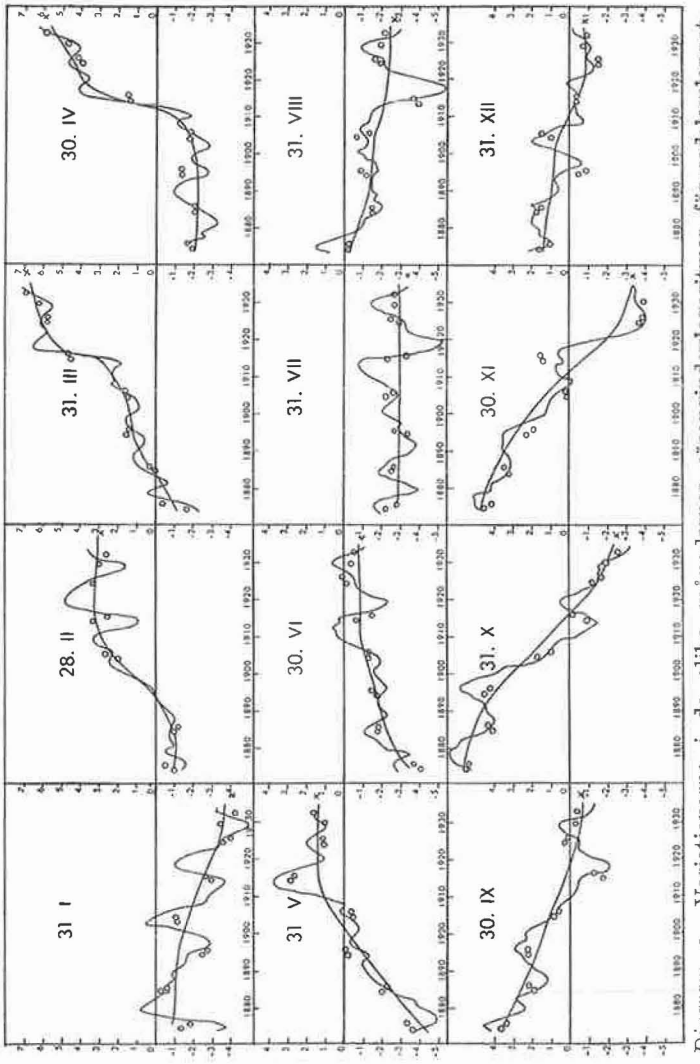


Diagram 2. Variationerna i de olika månadernas säsongindexlogaritmer för sedelomloppet. De små cirklarna angiva säsongindextal beräknade för tioårs perioder enligt en modifierad Harvard metod (Tabellerna I, 1 och IV, 1—8). Prognosen för perioden 1933—34 har betecknats med ett kryss. De krokiga kurvorna åskådliggöra de i tabellerna II, 1 och II, 2 sammanställda löpande sjuårs säsongindexlogaritmerna. De jämnt krökt kurvorna utgöra grafiskt bestämda trendlinjer, vilka avlästa för vart femte år återfinnas i tabell II, 3.

Ett närmare studium av de olika månadernas trendutveckling, sådan den framgår av de i diagram 2 inritade trendlinjerna, vilka avlästa för vart femte år angivas i

tabell II,3, ger av de olika månadernas säsongindextals utveckling följande bild:

Trenden för det till ultimo januari hörande säsongindextalet har hela tiden varit under årsmedeltalet. Efter att till omkring år 1900 hava varit nästan konstant har den därefter visat en fallande tendens, som dock synes hava avstannat omkring år 1930.

Februari månads säsongindextals trend har varit växande till omkring år 1920. Därefter har den varit nästan konstant med en måhända svagt nedåtgående tendens.

Trenden för mars månads säsongindex har varit stigande under hela perioden. Före år 1890 var stigningen relativt snabb, därefter avmattades den för att — efter en nästan språngartad tillväxt år 1917 — efter år 1920 visa en svagt stigande tendens.

April månads säsongindextal voro ända till omkring år 1910 c:a 2 % under årsmedeltalet för att under krigsåren snabbt stiga. Den uppåtgående tendensen har fortsatt i något långsammare tempo efter år 1920.

Maj månads säsongindextal ha under perioden varit stigande. Efter kriget synes stigningstendensen ha upphört.

Trenden för juni månads säsongindextal har varit svagt stigande. Stegringen synes ha avstannat omkring år 1910.

Juli månads säsongindextals trend har under perioden ifråga varit i det närmaste horisontal (c:a 3 % under årsmedeltalet).

Säsongindextalen för augusti månad ha varit svagt fallande.

September månads säsongindextals trend har varit något snabbare fallande än augusti månads.

Trenden för oktober månads säsongindextal har under perioden 1870—1900 varit svagt, åren 1900—1915 brant nedåtgående. Den fallande tendensen synes ha något dämpats efter år 1920.

Trenden för november månads säsongindextal är mycket lik oktober månads trend. Omkring år 1920 minskas november månads säsongindex snabbt. En stabilisering inträder omkring år 1925.

December månads säsongindextals trend har varit svagt fallande under den betraktade perioden.

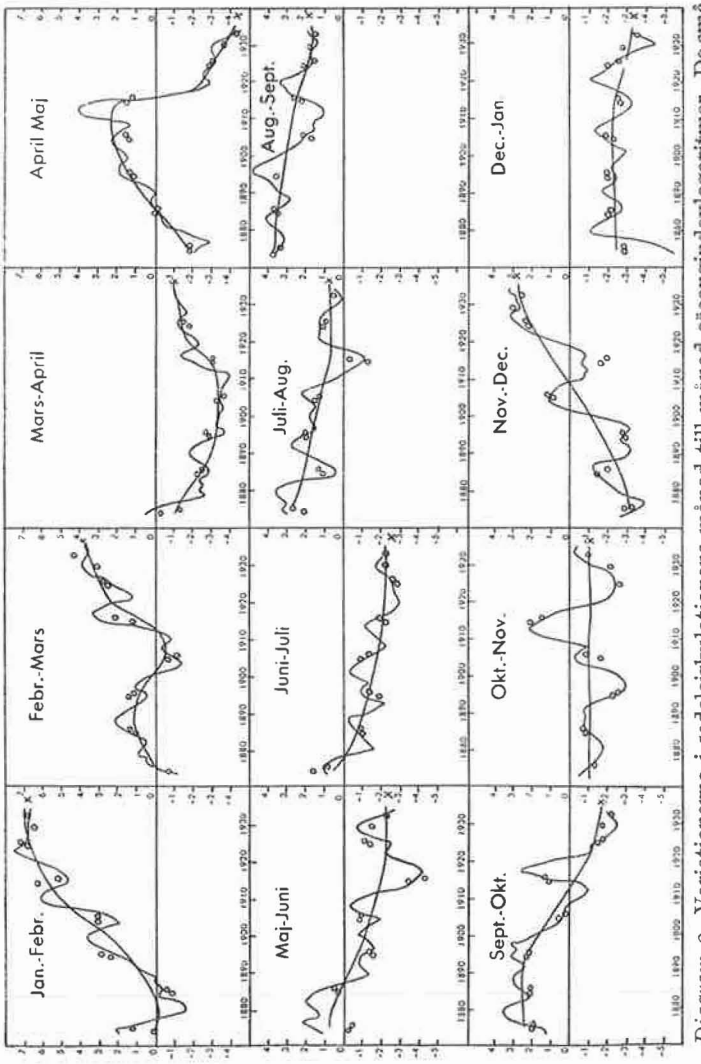


Diagram 3. Variationerna i sedelcirkulationens månad till månad säsongindexlogaritmer. De små cirklarna angiva de i tabell I, 2, de krokiga linjerna de i tabell III, 1 och III, 2 samt de jämnt böjda trendlinjerna de i tabell III, 3 sammanställda talen. Kryssen åskådliggöra prognosen för perioden 1933—43.

I fråga om långtidsutvecklingen hos månad-till-månad-logaritmdifferenserna för säsongindextalen kan man göra följande iakttagelser:

Januari—februari differensen har under perioden varit stigande. Trenden förefaller att numera ha stabiliserat sig på en nivå av c:a 7 %.

Februari—mars differensen var till omkring år 1885 växande. Därpå följde en nedgångsperiod med ett minimum omkring år 1905. Sedan dess har trenden varit stigande. Den är nu c:a 4 %, d.v.s. sedelcirkulationen den 31 mars är omkring 4 % större än den 28 februari.

Mars—april differensen har med undantag för åren omkring 1874 varit negativ med ett flackt minimum omkring år 1905 på c:a —3,5 % för att därefter långsamt växa. Den är nu c:a —1 %.

April—maj differensen har varit växande från c:a —2,5 % omkring år 1875 till c:a 2,5 % omkring år 1914 för att under krigsåren göra ett språng ned till c:a —2,5 %. Trenden har därefter varit fallande för att nu befinna sig i närheten av —4,5 %.

Maj—juni differensen har under perioden långsamt minskat från c:a 1,5 % till c:a —2,5 %.

Juni—juli differensen har från att vid periodens början hava varit svagt positiv minskat till omkring —2,5 %.

Juli—augusti differensens trend har från c:a 3 % småningom minskat till c:a 0.5 %.

Augusti—september differensens trend har från c:a 3,5 % minskat till ungefär 2 %.

September—oktober differensen, som före år 1900 var c:a 2,5 %, har därefter sjunkit till c:a —2 %.

Oktober—november differensens trend har varit i det närmaste horisontal på c:a —1 %. Avvikelserna från trenden ha dock varit rätt kraftiga omkring år 1900 och under krigsåren.

November—december differensens trend har varit stigande från c:a —3,5 % till c:a 3 %. Avvikelserna från trenden äro jämförelsevis stora. Tillväxtshastigheten synes ha minskat efter år 1925.

December—januari differensen har hela tiden förblivit negativ. Från att i periodens början ha varit c:a —2,5 %

har den efter ett flackt maximum i periodens mitt på c:a —2 % ånyo minskat till c:a —3 %.

De rörelser de olika månadernas säsongindextal utfört, ha naturligtvis även påverkat säsongfluktuationernas amplitud. Denna amplitud ha vi närmast av räknetekniska skäl mätt medelst summan av de olika säsongindextalens logaritmers talvärden ($\Sigma\,|\ln x|$). Motsvarande summa av de månatliga förändringarna ($\Sigma\,|d|$) ha vi även beräknat. Variationerna i säsongfluktuationernas amplitud, sådana de framgå av dessa tidsserier, ha vi åskådliggjort grafiskt i diagram 4.

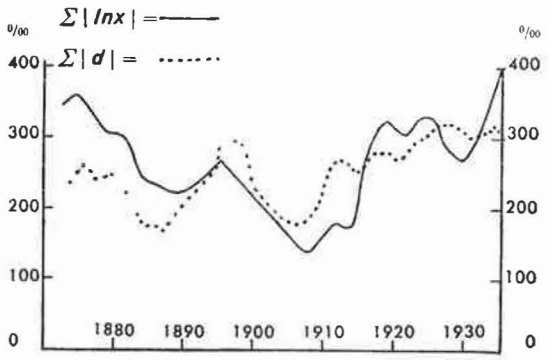


Diagram 4. Variationerna i sedelcirkulationens säsongfluktuationers amplitud åren 1874—1934. Siffrorna återfinnas i tabellerna II, 1 och II, 2 samt III, 1 och III, 2.

Vi observera, att amplituden omkring år 1908 nådde ett minimum, som utgjorde mindre än hälften av amplituden vid periodens början och vid dess slut. Vid minimitillfället var höst- och vårmaximet ungefär lika stora. I periodens början dominerades amplituden av höstmaximet, efter 1920 av vårmaximet.

Genom att jämföra de i diagrammen 2 och 3 inritade små cirklarna, vilka åskådliggöra de enligt den modifierade Harvardmetoden för tioårsperioder beräknade indextalen med de löpande sjuårs säsongindextalens kurvor och trendkurvorna, observerar man, att dessa cirklar i regel ligga nära kurvorna, oftast något närmare långtidstrenden än de sju-

års säsongindextal, som ha samma periodmitt. De största avvikelserna mellan de olika beräkningssättens resultat förekomma, såsom naturligt är, under inflationsperioden 1910—1920. Att cirklarna i allmänhet ligga närmare långtidstrenden beror naturligtvis främst på, att de höra till en längre period. I de fall, då den modifierade Harvardmetodens säsongindextal icke synas väl förenliga med de övriga metodernas, äro indextalen alla osäkra, beroende på de stora oregelbundenheterna i månad-till-månad-logaritmdifferenserna för sedelcirkulationen under den motsvarande perioden. I fråga om november månads säsongindextal för perioderna 1910—1919 och 1911—1920 är det tydligt, att de löpande sjuårs periodernas indextal äro tillförlitligare än de enligt den modifierade Harvardmetoden beräknade. I detta fall inverkar den metodtekniska finess, som den dubbla trendjusteringen innebär, även märkbart förbättrande på resultaten. Av den spridning, som existerar mellan de olika beräkningarnas resultat, får man intrycket, att indextalens osäkerhetsmoment är c:a ½ %. I ogynnsamma fall växer osäkerheten till c:a 1 %. För de trendrensade januari—februari-differenserna ha vi beräknat medelkvadratavvikelsen och erhållit för hela perioden 1870—1937 resultatet 2,5 %. I begreppet normalsäsongindex ingår dock en viss utjämning över en längre period. Utjämningen minskar spridningen (från trenden) hos normalsäsongindextalen jämförd med de ettåriga säsongindextalens spridning. Den på grundval av grafisk iakttagelse erhållna uppfattningen synes därför vara förenlig med stickprovet över spridningen i de ettåriga säsongindextalen. De grafiskt bestämda trendkurvorna ha, att döma av den spridning, som vid ett företaget experiment kunnat iakttagas vid olika personers försök att rita trender på kopior av diagrammen 2 och 3, en osäkerhet på ett par promille. Man kunde lätt förleda sig till att tro, att denna osäkerhet kunde minskas genom att beräkna trendlinjerna analytiskt, men i själva verket är den ökade precisionen blott skenbar. Beräknas trenden som ett löpande medeltal, kan detta icke anpassas till de variationer i sväng-

ningsperiodernas längd, som kunna iakttagas hos kurvorna, ej heller kan hänsyn tagas till de språngartade förändringarna i trenden under krigsåren. Dessutom får man icke trenden beräknad för de senaste åren i serierna. Dessa tilldraga sig vid prognoshärledningen det största intresset. Begagnar man åter analytiska funktioner (polynom av olika gradtal) kunna de icke lika smidigt som den grafiska trenden ta hänsyn till de speciella egendomligheterna i trendutvecklingen. Vid periodens slut bli dessa trender, speciellt om man begagnar polynom av högre gradtal, lätt missvisande. Vad man vinner i jämn krökning hos trendkurvorna, förlorar man i fråga om anpassningsförmåga till de empiriska siffrorna. Den omständigheten, att de grafiskt bestämda trendlinjerna i kanske högre grad än de medelst »mekaniska» medeltalsberäkningar eller anpassning av analytiska funktioner till den givna tidsserien erhållna äro beroende av statistikerns subjektiva omdöme, uppväges för det mesta av inbesparingen av arbete vid tillämpandet av den grafiska metoden [1]).

På grundval av de grafiskt bestämda trendlinjernas sannolika framtida utveckling och de senaste årens säsongindextal ha vi sökt numeriskt fixera vår matematiska förväntan beträffande säsongindextalen för perioden (1938 ± 5). Primärt ha vi härvid ställt prognosen på grundval av diagram 3 över de månatliga differenserna (d) och de därtillhörande siffrorna. Därpå ha de av dessa beroende säsongindextalen ln x och x beräknats (tabell V). I diagrammen 2 och 3 ha vi åskådliggjort vår prognos medelst ett kryss. I diagram 5 ha vi jämfört prognosen med den till tioårsperioden 1928—1937 hörande säsongindexen och den till perioden (1934 ± 3) hörande indexen samt Finlands Banks säsongindex för perioden 1925—1934. Prognosen skiljer sig blott obetydligt från de andra säsongindextalen, vilket tyder på att normalsäsongindexen efter kriget blott långsamt förändras. Den osäkerhet, vi känna oss böjda att tillskriva prognosen, är blott obetydligt större än osäkerhets-

[1]) Jfr *Hans Cl. Nybølle, Interpolation i statistiken,* Nordisk Statistisk Tidskrift, 1929, s. 115.

momenten för de övriga i diagram 5 inritade säsongindextalen. Ett fel i prognosen på mer än 1 % förefaller oss vara osannolikt. Prognosen synes oss ha en större chans än de till perioderna 1928—1937 och 1934 ± 3 hörande säsongindexarna att träffa nära de »ex post» beräknade säsongindextalen för den ifrågavarande perioden. Gäller det att

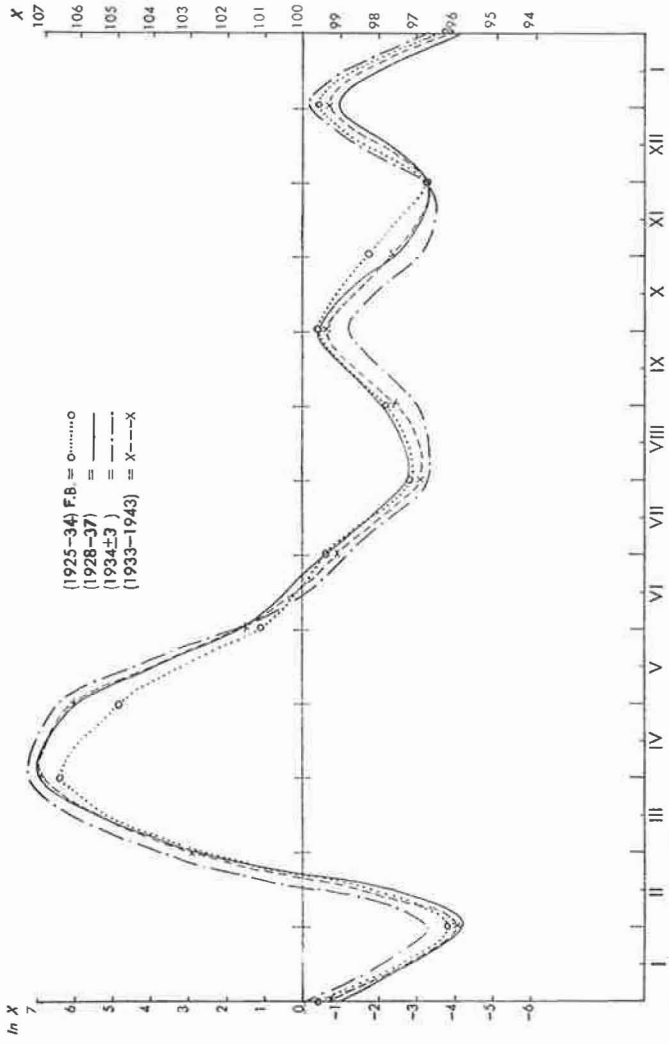


Diagram 5. Prognosen för sedelcirkulationens säsongindextal 1933—43, jämförd med säsongindexar för perioderna 1925—34, 1928—37 och 1931—37. Siffrorna finnas i tabell I, 3.

eliminera säsongrörelserna i en tidsserie för det löpande året, är det skäl, att vid de tillfällen, då man ej kan göra troligt, att det existerar en ideal normalsäsongindex, som är oföränderlig i tiden, så vitt möjligt begagna en prognostiskt bestämd säsongindex.

Varpå bero de ettåriga säsongindextalens avvikelser från trenden?

Om de krafter, som åstadkomma de säsongmässiga fluktuationerna i sedelcirkulationen ensamma för sig finge påverka sedelomloppet, vore det naturligt att vänta sig, att de ettåriga säsongindextalen för en viss månad skulle bilda en tidsserie med en jämförelsevis trögrörlig utveckling. Då detta i själva verket icke är fallet, frågar man sig, vilka omständigheter, som kunna tänkas förklara de avvikelser från säsongindextalens långtidstrend, som man kan observera.

En förklaring är, att sedelcirkulationens ökning eller minskning ej sker kontinuerligt utan i poster av ofta relativt till säsongrörelsernas storlek betydande belopp. Om Finlands Bank beviljar en kund ett större lån eller erhåller en större återbetalning på ett sådant, medför detta, att säsongrörelserna under den ifrågavarande månaden icke kunna iakttagas. I den mån störningen av de månatliga säsongmässiga variationerna är så stor, att den ifrågavarande månad-till-månad-differensen »dansar ur serien», påverkar den icke mycket de utjämnade säsongindextalen, emedan differensen i detta fall vid medeltalsbildningen tilldelas vikten 0. De stora störningar, som förekommo under inflationsperioden 1914—1920 äro i regel av denna karaktär, i undantagsfall inträffar det dock flere störningar inom mindre avstånd än 3 år från varandra, i vilket fall de även påverka de utjämnade säsongindextalen.

Intressantare äro de avvikelser från säsongindextalens trend, som kunna ställas i samband med konjunkturrörelserna i sedelcirkulationen. Det förefaller ju i och för sig mycket

naturligt att vänta sig, att de variationer i efterfrågan på sedlar, som åstadkomma de säsongmässiga fluktuationerna, skola variera med konjunkturrörelserna. Om t.ex. förklaringen, att det nuvarande vår-vintermaximet utgör en följd av skogsarbetenas ökade anspråk på sedlar, är riktig, ligger det nära till hands att vänta sig, att detta maximum skall vara kraftigare utvecklat under högkonjunkturen än under lågkonjunkturen.

Vi ha undersökt korrelationssambandet mellan säsongfluktuationerna och konjunkturrörelserna enligt den kanske något grova, men arbetsbesparande metoden att grafiskt jämföra avvikelserna från långtidstrenden i diagrammen 2,3 och 4 med en på genomskinligt papper ritad kopia av diagram 6, som åskådliggör löpande femårs medeltal (K_5) av avvikelserna från långtidstrenden hos sedelcirkulationens årsmedeltals logaritmer ($K_1 = ln\ S - ln\ T$). Diagrammet ha vi ritat i en skala, som gör att konjunkturkomponentens variationer, grafiskt sett, förefalla att vara ungefär lika stora som säsongindextalens avvikelser från trenden. Genom att ett flertal gånger söka fastställa, huruvida säsongindexkurvorna likna mera konjunkturkomponenten K_5 eller spegelbilden — K_5, har korrelationens tecken kunnat avläsas.

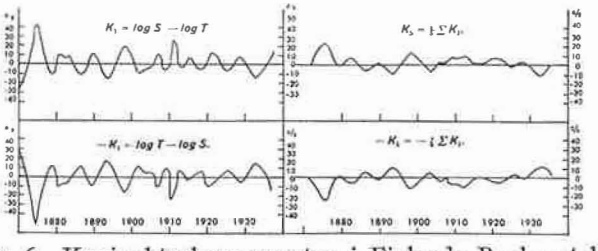


Diagram 6. Konjunkturkomponenten i Finlands Banks utelöpande sedlar och bronsmynt. K_1 = logaritmens för sedelcirkulationens årsmedeltal avvikelser från en grafiskt bestämd trend. K_5 = fem års löpande medeltal av K_1. Siffrorna finnas i tabellerna VIII, 1 och VIII, 2.

Under den studerade perioden växer i regel det säsongmässiga vår-vintermaximet (28/2—31/3) relativt till trenden vid stigande sedelomlopp och avtager, då sedelcirkulatio-

Tabell 2. *Korrelationen mellan säsongindextalens avvikelser från långtidstrenden och konjunkturkomponenten i sedelcirkulationen.*

Period	Jan.	Febr.	Mars	April	Maj	Juni	Juli	Aug.	Sept.	Okt.	Nov.	Dec.	$\Sigma \lvert ln\, x \rvert$	Period
1874—1905	$< o$	$\approx o$	$> o$	$> o$	$\approx o$	$\lessapprox o$	$\approx o$	$\approx o$	$> o$	$\approx o$	$\approx o$	$\lessapprox o$	$> o$	1874—1905
1906—1934	$\lessapprox o$	$\approx o$	$\gtrapprox o$	$\lessapprox o$	$\lessapprox o$	$\lessapprox o$	$\approx o$	$\approx o$	$\gtrapprox o$	$\approx o$	$< o$	$\approx o$	$\lessapprox o$	1906—1934
1874—1934	$< o$	$\approx o$	$> o$	$\approx o$	$\approx o$	$\lessapprox o$	$\approx o$	$\approx o$	$> o$	$\approx o$	$\approx o$	$\lessapprox o$	$\lessapprox o$	1874—1934

Korrelationen mellan säsongindextalens månatliga förändringars avvikelser från trenden och konjunkturkomponenten i sedelcirkulationen.

Period	Jan.-Febr.	Febr.-Mars	Mars-April	April-Maj	Maj-Juni	Juni-Juli	Juli-Aug.	Aug.-Sept.	Sept.-Okt.	Okt.-Nov.	Nov.-Dec.	Dec.-Jan.	$\Sigma \lvert d \rvert$	Period
1874—1905	$> o$	$\gtrapprox o$	$> o$	$\lessapprox o$	$\lessapprox o$	$> o$	$\approx o$	$\gtrapprox o$	$\lessapprox o$	$\lessapprox o$	$\lessapprox o$	$\gtrapprox o$	$> o$	1874—1905
1906—1934	$> o$	$\lessapprox o$	$< o$	$\approx o$	$\approx o$	$\lessapprox o$	$> o$	$\approx o$	$\approx o$	$\lessapprox o$	$\approx o$	$\approx o$	$\approx o$	1906—1934
1874—1934	$> o$	$\approx o$	$\approx o$	$\approx o$	$\lessapprox o$	$\approx o$	$\gtrapprox o$	$\approx o$	$\lessapprox o$	$\lessapprox o$	$\approx o$	$\approx o$	$\gtrapprox o$	1874—1934

Förklaringar: $> o$ betecknar att korrelationen är positiv: stigande sedelcirkulation höjer i detta fall säsongindextalen. $\gtrapprox o$ anger att korrelationen sannolikt är svagt positiv. $\approx o$ betecknar att korrelationskoefficientens tecken icke kan avläsas ur diagrammen. $\lessapprox o$ anger att korrelationen sannolikt är svagt negativ, $< o$ att korrelationen är negativ: stigande sedelomlopp minskar i detta fall säsongindextalet.

nen minskar, vilket tyder på att förklaringen, att vintermaximet står i samband med skogsarbetena, är riktig.

September månads säsongindextal visa sig även vara konjunkturkänsliga. En stigande sedelcirkulation har en tendens att höja denna månads säsongindextal. Närmare detaljer beträffande korrelationssambandet framgå av tabell 2.

Till en del kunde säsongindextalens konjunkturkänslighet tolkas så, att kraftiga konjunkturmässiga förändringar äro sannolikare under den tid på året, då sedelcirkulationen säsongmässigt är stigande än under övriga månader. Det är sålunda icke utan vidare klart, huruvida man skall betrakta de empiriskt iakttagna konjunkturmässiga fluktuationerna i säsongindextalen såsom variationer i säsongkomponenten; de kunna även uppfattas som variationer i konjunkturkomponenten, vilka på grund av säsongfluktuationernas existens ha en tendens att bliva kraftigare under vissa tidpunkter på året. F. Burjam [1]) har visat att diskontoförhöjningarna företrädesvis inträffa under höstmånaderna. Det synes därför vara sannolikt, att en del av september månads säsongindextals variationer böra tillskrivas konjunkturkomponenten och icke säsongkomponenten.

Sambandet mellan konjunkturen och säsongindextalens variationer är dock ganska löst. Den till synes största korrelationen existerar mellan jan.-febr.-differensens avvikelser från trenden ($d^{(3)}$) och konjunkturkomponenten K_5. I tabell VIII i bihanget ha vi sammanställt siffror för K_1, K_5 och $d^{(3)}$ för jan.—febr. På grundval av dessa närmevärden ha vi beräknat korrelationskoefficienten

$$r = \Sigma K_5 d^{(3)} : \sqrt{\Sigma (K_5)^2 \times \Sigma (d^{(3)})^2}$$

för perioderna 1874—1905, 1906—1934 och 1874—1934 och erhållit de i följande tabell, kolumn r, sammanställda resultaten:

[1]) *F. Burjam. Diskontoförskjutningar och säsongfluktuationer.* Ekon. Samf. Tidskr., Ny serie, H. XXVII, 1934.

Tabell 3.

Period	r	$\sigma(K_5)$	$\sigma(d^{(3)})$	b	$\sigma(d^{(4)})$	$r\,\sigma(d^{(3)})$
1874—1905	0,54	80	10,9	0,075	8,8	5,9
1906—1934	0,18	63	7,9	0,022	7,8	1,4
1874—1934	0,41	73	9,6	0,055	8,6	3,9

$\sigma\,(K_5) = \sqrt{\frac{1}{n}\Sigma\,(K_5)^2}$, $\sigma\,(d^{(3)}) = \sqrt{\frac{1}{n}\Sigma\,(d^{(3)})^2}$, $d^{(4)} = d^{(3)} - b \cdot K_5$, $b = \Sigma\,K_5\,d^{(3)} : \Sigma\,(K_5)^2 = r\,\sigma\,(d^{(3)}) : \sigma\,(K_5)$

n = antalet termer vid summationen (Σ).

Konjunkturkomponenten K_5 är under inflationsåren mycket svår att fastställa. Det är t.o.m. möjligt, att det vore riktigare att rita trenden så, att K_5 för dessa år bytte tecken, i vilket fall korrelationen för perioden 1906—1934 skulle något stiga. Under krigsåren var vår-vintermaximet mindre än vad man skulle väntat sig med hänsyn till trendutvecklingen.

Tidsserien $b \cdot K_5$ kan betraktas som den konjunkturmässiga komponenten i $d^{(3)}$, resttermen $d^{(4)}$ domineras av den slumpmässiga variationen. Spridningen i $d^{(4)}$ har beräknats enligt formeln $\sigma\,(d^{(4)}) = \sqrt{1 - r^2}\,\sigma\,(d^{(3)})$. Spridningen i den konjunkturmässiga komponenten är $\sigma\,(b\,K_5) = b\,\sigma(K_5) = r\,\sigma\,(d^{(3)})$. Anmärkningsvärd är den stabilitet i tiden, som den slumpmässiga variationen att döma av talen i kolumnen $\sigma\,(d^{(4)})$ har. Den konjunkturmässiga komponentens spridning varierar däremot kraftigt, beroende på vilken tidsperiod som betraktas.

Det skulle ju ligga nära till hands att antaga, att man skulle få en större korrelation mellan de ettåriga säsongindextalens månatliga förändringars (d—δ) avvikelser från trenden ($d_1 = d - \delta - tr$) och konjunkturkomponenten K_1 ($= ln\ S - ln\ T$) än den mellan $d^{(3)}$ och K_5. I själva verket är korrelationen betydligt mindre beroende på, att

den slumpmässiga variationen blivit starkare dämpad vid beräkningen av $d^{(3)}$ än den konjunkturmässiga variationen.

I tabell 4 ha vi sammanställt de vid beräkningen av korrelationskoefficienter mellan d_1 och K_1 erhållna resultaten.

Tabell 4.

Period	r	$\sigma(K_1)$	$\sigma(d_1)$	b	$\sigma(d_1^{(2)})$	$r\,\sigma(d_1)$
1874—1905	0,15	138	24,6	0,027	24,4	3,7
1906—1934	0,34	99	25,7	0,088	24,4	8,7
1874—1934	0,22	121	25,1	0,046	24,5	5,5

$d_1^{(2)} = d_1 - b\,K_1$, $b = r\,\sigma\,(d_1) : \sigma\,(K_1)$.

En jämförelse mellan talen i tabell 3 och 4 visar, att K_5 för perioden 1874—1934 har en spridning, som är 0,6 ggr så stor som spridningen i K_1, och att $\sigma\,(d^{(3)}) = 0{,}38\ \sigma\,(d_1)$ samt $\sigma\,(d^{(4)}) = 0{,}35\ \sigma\,(d_1^{(2)})$, medan den konjunkturmässiga komponenten i d_1 vid medeltalsbildningarna blott dämpats i förhållandet 0,7 : 1. Under sådana förhållanden förstår man, att korrelationen r i tabell 3 är betydligt större än den i tabell 4. Dämpningsfaktorn för den slumpmässiga variabeln i d_1 är obetydligt större än den dämpningsfaktor, man erhåller vid jämförelse mellan spridningen i den serie löpande medeltal med sju termer, som man får ur en ren slumpvariabel (t.ex. en serie ögon vid tärningskast) med spridningen i slumpvariabeln själv. I detta fall blir nämligen denna faktors matematiska förväntan $1 : \sqrt{7} = 0{,}38$.

Säsongindextalens variationer synas sålunda väl kunna uppdelas i tre komponenter: 1) en trögrörlig långtidstrend, 2) en konjunkturmässig variation och 3) en slumpmässig variation. Den konjunkturmässiga variationen är i regel mycket liten. För de starkast konjunkturkänsliga säsongindextalen är den konjunkturmässiga komponentens sprid-

ning av storleksordningen ½ %, medan den slumpmässiga variationen i de ettåriga säsongindextalen är betydligt större med en spridning av storleksordningen 2—3 %.

De olika sedel- och myntvalörernas säsongfluktuationer åren 1926—1937.

En fullständigare bild av sedelcirkulationens säsongfluktuationer få vi genom att studera de olika sedel- och myntvalörernas säsongrörelser skilt för sig. I tabell VII,1 ha vi sammanställt logaritmerna av säsongindextal hänförande sig till kvartalssluten 31/3, 30/6, 30/9 och 31/12 för de olika sedel- och myntvalörerna. Betydande avvikelser existera mellan myntens och sedlarnas fluktuationer. Mynten visa genomgående en betydligt lugnare utveckling än sedlarna. Detta är ju rätt naturligt på grund av besväret att transportera större mängder mynt till och från Finlands Bank. Anmärkningsvärd är den förändring, som ägde rum i 5 och 10 mks valörernas säsongfluktuationer efter utbytandet av sedlarna mot bronsmynt. Myntens säsongfluktuationer åskådliggöras grafiskt i diagram 7.

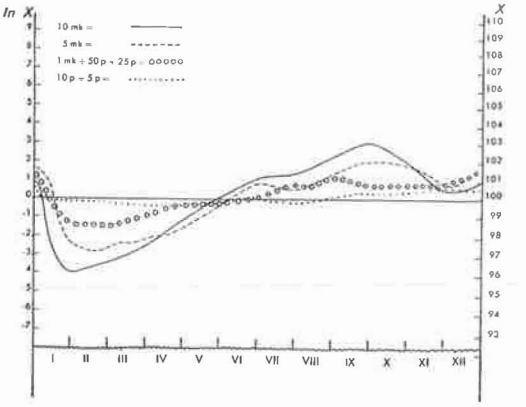


Diagram 7. Säsongfluktuationerna i olika myntslag. Siffrorna, vilka hänföra sig till perioden $^{31}/_{12}$ 1934—$^{31}/_{12}$ 1937, finnas i tabell VII,2.

Detta diagram har uppritats på grundval av säsongindexar för treårs perioden 31/12 1934—31/12 1937 (tabell VII,2) beräknade enligt en starkt förenklad aritmetisk metod beskriven i förklaringarna till tabell VII,3. Säsongindexarna äro visserligen ganska osäkra, men torde dock ge en i huvudsak riktig bild av säsongfluktuationerna i de olika sedel- och myntvalörerna. Myntcirkulationen visar en liten stegring i juni månad. I augusti och september är den utelöpande myntmängden relativt stor likaså vid jultiden. Vårvintermaximet saknas helt. Under januari eller februari månad når myntcirkulationen sitt minsta värde under året.

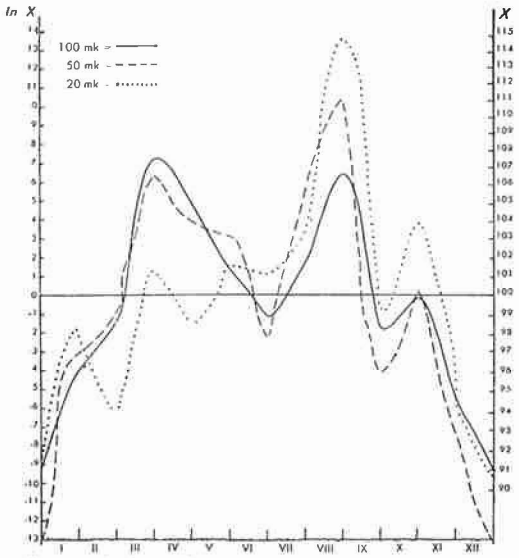


Diagram 8. Säsongfluktuationerna i 20 mks, 50 mks och 100 mks sedlarna. Siffrorna återfinnas i tabell VII,2.

De olika sedelvalörernas säsongindexar skilja sig även rätt mycket från varandra. Säsongindexkurvorna för 20 mks, 50 mks och 100 mks sedlarna likna dock varandra rätt mycket. Dessa säsongindexar ha grafiskt åskådliggjorts i diagram 8.

Man observerar, att marsmaximet förefinnes i alla kur-

vorna. Detta maximum överträffas dock av ett kraftigt »skördemaximum» i augusti månad. I oktober finnes även en mindre stegring, medan minimivärden antagas i december, juni och september. December minimet är desto märkligare, då den totala sedelcirkulationen vid årsskiftet uppvisar ett relativt maximum [1]).

I diagram 9 ha vi ritat säsongkurvorna för de två största sedelvalörerna 1000 och 500 mk. Vår-vintermaximet är i dem mycket kraftigt, medan skördemaximet är obetydligt. Minimipunkterna finnas för 1000 mks sedlarna i juli och november, medan 500 mks sedlarnas minima befinna sig i juni, september och december. 500 mks sedlarnas säsongindex är mindre tillförlitlig än de övrigas, emedan en icke säsongmässig variation störde 1936 års säsong, i det att beloppet utelöpande 500 mks sedlar plötsligt nedgick från 288,9 milj. mk den 30 juni till 207,7 milj. mk den 31 juli. Det onormalt låga värdet för de utelöpande 500 mks sed-

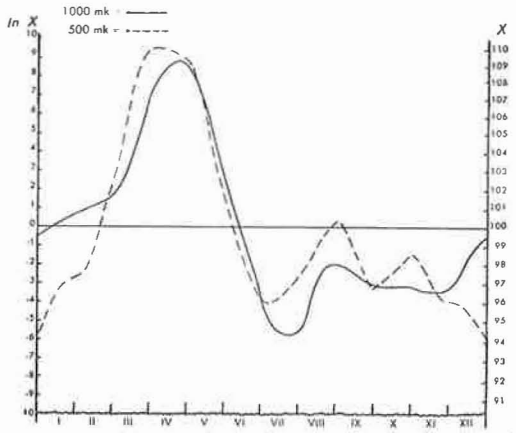


Diagram 9. Säsongfluktuationerna i 500 mks och 1,000 mks sedlarna.

[1]) På grund av de säsongmässigt låga siffrorna för 20, 50 och 100 mks sedlarna vid utgången av året utgöra siffrorna i Finlands Banks årsbok över sedelcirkulationens procentuella sammansättning vid årets slut icke några goda närmevärden för de procenttal, som man skulle få genom att beräkna de olika valörernas andel i cirkulationen på grundval av årsmedeltalen.

larna »korrigerades» i november, i det att det steg från 214,5 milj. mk den 31 oktober till 309,6 milj. mk den 30 november 1936. De två andra årens serier visa ej alls samma utveckling under perioden juni-november. Vi ha vid säsongindexberäkningen sökt eliminera störningen genom att öka beloppen 500 mks sedlar under augusti, september och oktober månad 1936 med 90 milj. mk.

Anmärkningsvärt är, att alla andra sedelvalörer förutom 1000 mks sedlarna säsongmässigt minska från den 30/11 till den 31/12. I bronsmyntens och nickelmyntens säsong kan man dock spåra denna stigningstendens.

En förklaring till de stora sedelvalörernas vårvintermaximum ligger kanske i, att löneutbetalningarna till skogsarbetarna för det mesta ske i poster på mer än 500 mk per likvid och arbetare. Arbetarna bära sedlarna på sig till dess de återvända från de ofta långt från banker och handelscentra belägna avverkningsplatserna till sina ordinarie boningsorter. Stegringen i 1000 mks sedlarna tyder samtidigt på, att en icke obetydlig del av skogsköpen betalas medelst sedlar. Stegringen i sedelomloppet under vårvintern påverkar ej mycket detaljhandelns omsättning, emedan sedlarnas omloppshastighet är lägre än normalt. De små sedel- och myntvalörerna, vilka endast i ringa mån behövas för löneutbetalningarna, visa en helt annan utveckling under vårvintern. Däremot synas skördearbetena i augusti i mindre mån taga de större sedlarna i anspråk; lönelikviderna fördelas i smärre poster, vilket framgår av att 20, 50 och 100 mks sedlarna visa den kraftigaste augustistegringen.

Ett säsongmässigt maximum i detaljhandelsomsättningen, synes icke alls erfordra en stegring av sedelomloppet. Den ökade livligheten i handeln möjliggöres av en ökning av sedlarnas omloppshastighet. En högsäsong inom detaljhandeln är snarare ägnad att minska sedelomloppet, emedan affärsmännen vanligen regelbundet deponera de under dagen erhållna penningmedlen i banker, varifrån de därpå strömma till Finlands Bank.

Det torde sålunda existera en negativ korrelation mellan de säsongmässiga fluktuationerna i sedelmängden och säsongrörelserna i sedlarnas omloppshastighet. Väl känt är däremot, att det existerar en stor positiv korrelation mellan de konjunkturmässiga variationerna i sedelmängden och i sedlarnas omloppshastighet.

Jämförelse mellan säsongfluktuationerna i Finlands Banks utelöpande sedlar och bronsmynt och säsongfluktuationerna i ett antal andra ekonomiska tidsserier.

I diagram 10 ha vi grafiskt åskådliggjort säsongindexar för följande tidsserier[1]):

1. = Finlands Banks utelöpande sedlar och bronsmynt åren 1922—37.
2. = Utelöpande sedlar utanför Finlands Bank och affärsbankerna 1922—37.
3. = Innestående på löpande räkning i affärsbankerna åren 1922—33.
4. = Innestående på checkräkning åren 1934—37.
5. = Olyftade kassakreditiv hos affärsbankerna 1922—33.
6. = Olyftad checkkredit hos affärsbankerna 1934—37.
7. = Finlands Banks clearingrörelses medelbelopp per dag för de olika månaderna åren 1925—34.
8. = Totalomsättningen per dag i Kansallis-Osake-Pankki i medeltal för de olika månaderna dividerad med månadsbalansens slutsumma.

Säsongindexkurvorna för tidsserierna 1. och 2. skilja sig från varandra blott däri, att de utanför bankerna befintliga sedlarna visa ett minimum vid årsskiftet, medan de utanför Finlands Bank befintliga sedlarna ha ett relativt maximivärde vid denna tidpunkt. Affärsbankernas kassor nå sitt maximum vid årsskiftet.

[1]) Säsongindexarna 2.—6. ha beräknats av min broder pol. kand. *Erik Törnqvist* enligt den modifierade Harvardmetoden. Vid beräkningen av säsongindexarna 7. och 8. ha indexar beräknade av Finlands Banks statistiska avdelning utnyttjats.

Säsongkurvorna för tidsserierna 3., 4., 5. och 6. avvika väsentligt från sedelcirkulationens säsongkurvor.

Säsongfluktuationerna i omsättningssiffrorna representerade av tidsserierna 7. och 8. synas icke utöva något iakttagbart inflytande på betalningsmedlens säsongfluk-

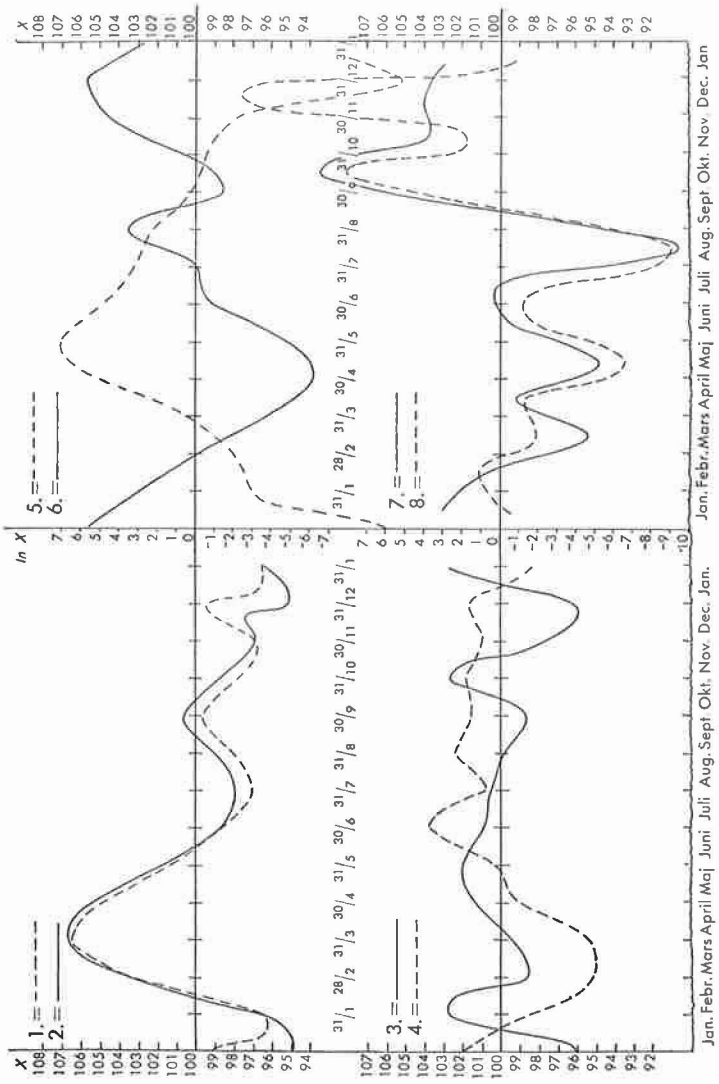


Diagram 10.

3

tuationer. Korttidsfluktuationerna i betalningsströmmarnas storlek och i betalningsmedelsvolymen synas sålunda ske oberoende av varandra. Om ett samband existerar är detta av en komplicerad karaktär. På grund av de relativt stora slumpmässiga variationerna speciellt i indexarna 3.—6. måste en viss försiktighet iakttagas vid dragandet av slutsatser.

Elimineringen av säsongfluktuationerna.

Så snart man erhållit en serie normalsäsongindexar, har man möjlighet att söka eliminera de normala säsongfluktuationerna ur den givna tidsserien. Ha räkningarna utförts logaritmiskt, har man härvid blott att från logaritmen för utgångstalen subtrahera det motsvarande normalsäsongindextalets logaritm. Vi ha säsongrensat sedelcirkulationens logaritmer ln S så, att vi från dem subtraherat ln x = logaritmen av säsongindextalet för den löpande sjuårs period, som har det ifrågavarande året till mitt. Vid periodens slut ha vi medelst lineär interpolation mellan de prognostiskt beräknade säsongindextalens logaritmer för perioden (1938 $\pm$ 5) och säsongindexlogaritmerna för perioden (1934 $\pm$ 3) beräknat talen ln x för åren 1935—37. Vill man uttrycka de säsongrensade serierna ln S—ln x i milj. mk, har man blott att slå upp numeruslogaritmerna för dessa tal.

I diagram 11 ha vi åskådliggjort den säsongrensade serien ln S—ln x. Kurvan har ritats med stöd av talen för den 31/3, 30/6, 30/9 och 31/12. I diagram 11 ha vi i större skala ritat den motsvarande månatliga serien för perioden 1920—37.

Diagram 11 och 12 äro värda ett grundligt studium. Sammanställda med andra ekonomiska tidsserier kunna de giva en god inblick i de konjunkturmässiga variationerna i Finlands näringsliv.

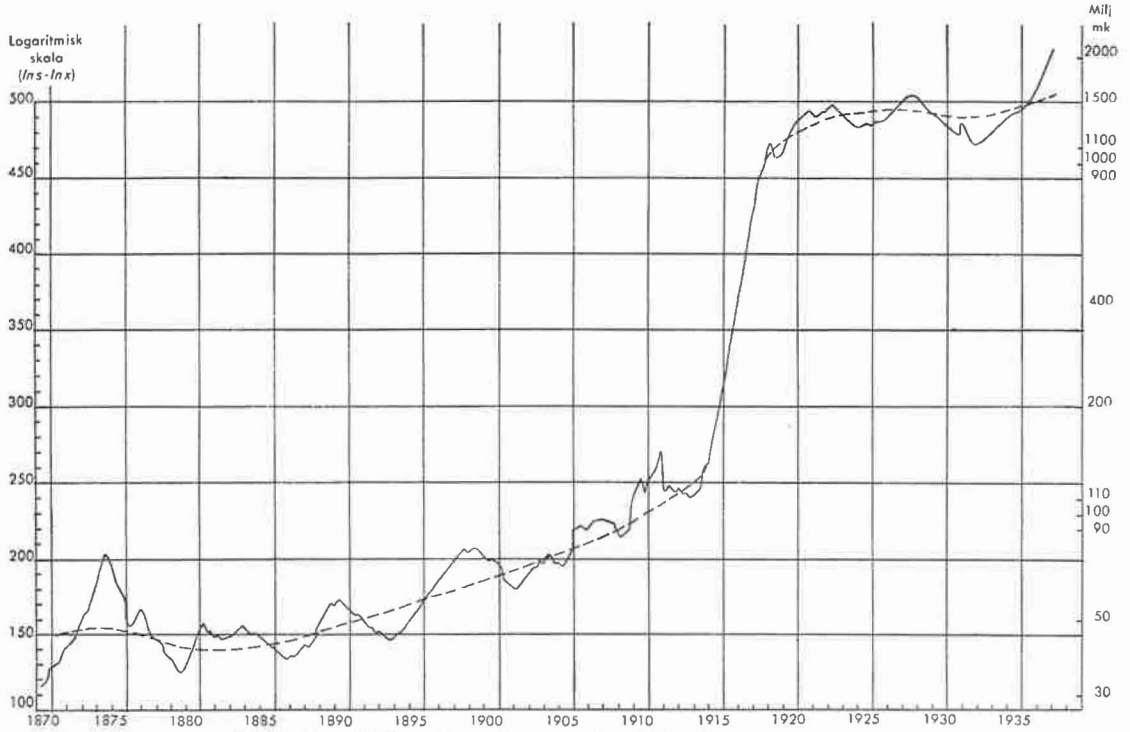


Diagram 11. Finlands Banks utelöpande sedlar och bronsmynt, säsongrensad serie. Diagrammet har uppritats med stöd av siffror hänförande sig till ultimo i varje kvartal åren 1870—1937. Den jämnt krökta kurvan utgör en grafiskt bestämd trend.

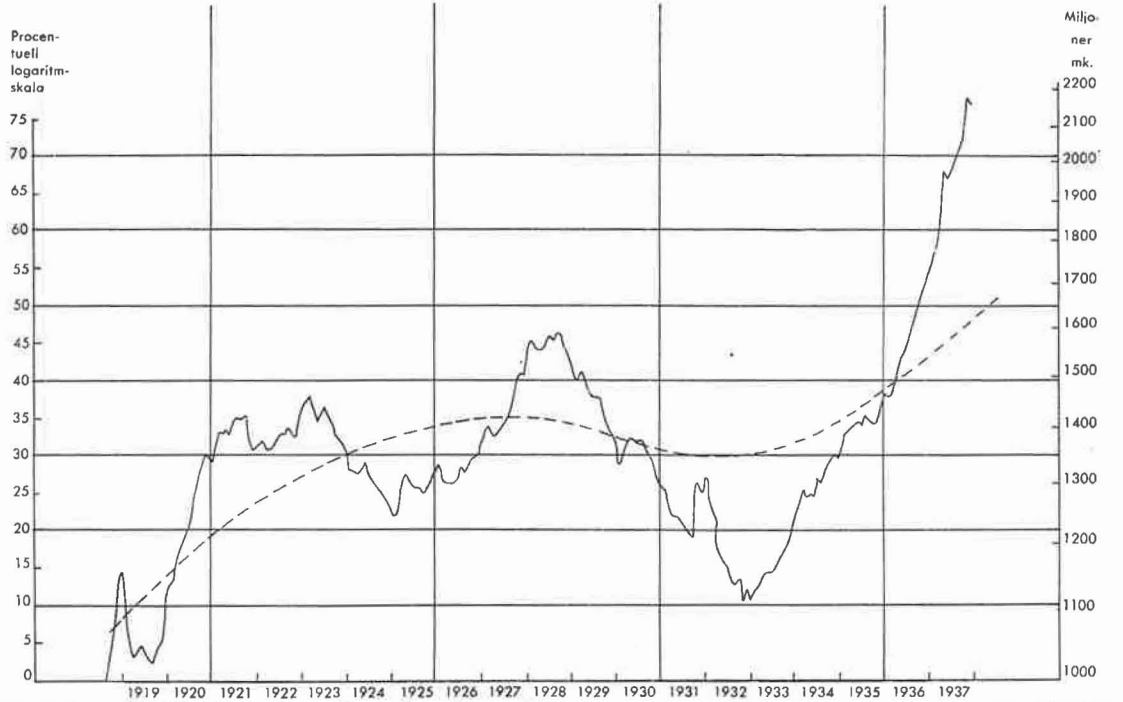


Diagram 12. Finlands Banks utelöpande sedlar och bronsmynt, säsongrensad serie hänförande sig till ultimo i varje månad åren 1919—1937 (tabell IX, 4). Den jämnt krökta kurvan utgör en grafiskt bestämd trend.

EKONOMISK STATISTIK

Efterfrågan på jordbruksprodukter och dess känslighet för pris- och inkomstförändringar av Herman Wold. Sthlm 1940. (SOU. 1940: 16. Jo.)

På uppdrag av 1938 års jordbruksutredning har fil. dr Herman Wold utfört en värdefull undersökning av pris- och inkomstelasticiteten för olika slags konsumtionsvaror, varvid huvudintresset koncentrerats på de animala jordbruksprodukterna mjölk, smör, ägg samt kött och fläsk. Då denna undersökning är av stort intresse såväl ur metodologisk som prispolitisk synpunkt, skola vi i det följande försöka giva en sammanfattning av de viktigaste resultaten, varvid vi skola underkasta några av de teoretiska resonemangen en kritisk granskning samt jämföra en del av Wolds resultat beträffande inkomstelasticiteterna med motsvarande elasticiteter beräknade på grundval av det material, som föreligger i den finska levnadskostnadsundersökningen[1] av år 1928.

Wold lyckas ge en även ur pedagogisk synpunkt god framställning av elasticitetsbegreppet och de statistiska metoderna för beräkning av elasticitetstal. Det problem, som han söker angripa, formulerar han populärt på följande sätt:

»Om man höjer priset på smör med 10 %, huru kommer efterfrågan att reagera? Kommer smörkonsumtionen att minskas med 10 %, med

[1] *F. O. S. XXXII. Sociala specialundersökningar 14.* Levnadskostnadsundersökningen 1928. A. Bokföringsfamiljerna i städer och övriga bosättningscentra. Helsingfors 1936.

5 %, med 0 %? Och om inkomstnivån stiger med 10 %, hur mycket kommer då smörkonsumtionen att ökas?»

Efterfrågeanalysens grundtanke är att studera konsumtionens faktiska storlek vid olika pris- och inkomstlägen i syfte att spåra upp regelbundenheter i det statistiska materialet, vilka kunna tolkas som ett samband mellan å ena sidan priset och inkomsten, å andra sidan konsumtionen. Med ledning av det erhållna sambandet kan man sedan uppskattningsvis beräkna, hur konsumtionen reagerar vid en tilltänkt pris- eller inkomständring.

Sambandet mellan konsumtionsändringen och pris- resp. inkomständringen uttryckes på ett enkelt sätt medelst s. k. elasticitetstal. Elasticitetens talvärde anger graden av konsumtionens känslighet för pris- eller inkomständringar. En priselasticitet på 0,4 betyder sålunda, att en prishöjning på 10 % medför en minskning av konsumtionen med c:a $0{,}4 \times 10\ \% = 4\ \%$. En inkomstelasticitet på 0,35 betyder, att en inkomstökning på 10 % ökar konsumtionen med c:a 3,5 %.

Wold redogör först för det matematiska elasticitetsbegreppet och för olika statistiska metoder för beräkningen av elasticitetstal. Med begagnande av det material, som står till buds i de svenska levnadskostnadsundersökningarna av åren 1913, 1923 och 1933, beräknar han inkomstelasticitetstal för levnadskostnadsbudgetens olika huvudgrupper samt för olika födoämnen. Slutligen studerar han ingående inkomst- och priselasticiteten för mjölk, smör, ägg samt kött och fläsk. Härvid begagnar han sig av konsumtionsföreningarnas marknadsstatistik.

Matematiskt uttryckt betecknar elasticiteten förhållandet mellan en infinitesimalt liten relativ tillväxt i den beroende variabeln och den motsvarande infinitesimala relativa tillväxten i den oberoende variabeln:

$$(1) \qquad E(x) = -e(x) = \frac{\delta k}{k} : \frac{dx}{x} = \frac{\delta \log k}{d \log x}.$$

Elasticiteten överensstämmer, såsom i formel (1) antytts, med förhållandet mellan tillväxten i logaritmen för den beroende variabeln och tillväxten i logaritmen för den oberoende variabeln, under förutsättning att tillväxterna äro infinitesimalt små. Om elasticiteten är konstant, har man

$$(2) \qquad E = -e = \frac{\log k_1 - \log k_0}{\log x_1 - \log x_0}$$

oberoende av läget och storleken av tillväxten $\log x_1 - \log x_0$. Då elasticiteten ej är konstant, ger formel (2) ett medelvärde av elasticiteten $E(x)$ inom intervallen (x_0, x_1). På grund av dessa egenskaper hos formel (2) är det ändamålsenligt att utföra efterfrågeanalysen i logaritmisk skala och studera, huru logaritmen för konsumtionen beror av logaritmen för priset och logaritmen för inkomsten. I den (dubbel)logaritmiska skalan åskådliggöras konsumtionskurvor med konstant elasticitet av räta linjer, elasticiteten anger kurvans vinkelkoefficient. Wold begagnar sig konsekvent av övergången till logaritmisk skala.

Då det gäller att beräkna elasticitetstal ur det statistiska materialet

bestående av en serie sammanhörande logaritmer för pris, inkomster och konsumtionskvantiteter av en vara, har man att välja mellan ett flertal olika utjämningsformler. Problemet att välja den lämpligaste formeln för den statistiska bearbetningen av det föreliggande primärmaterialet löser Wold med tillhjälp av en välkänd matematisk sats. Elasticiteten för en summa utgör det med respektive addender som vikter vägda medeltalet av de enskilda addendernas elasticitet i avseende å den betraktade variabeln. Hans beräkningar visa, att av de för elasticitetskoefficientberäkningar föreslagna regressionsformlerna blott den elementära enligt den s. k. minsta kvadratmetoden härledda regressionsformeln ger till resultat elasticitetstal, som approximativt uppfylla det teoretiska sambandet mellan elasticiteten för en summa och de däri ingående addendernas elasticiteter. Detta samband utnyttjar Wold även som en formel för beräkning av elasticitetstal fria från ett systematiskt fel härrörande från en s. k. skiktningseffekt. Skiktningseffekten uppkommer, när primärmaterialet är inhomogent, d. v. s. om spridningsdiagrammets form påverkats av andra med den oberoende variabeln (inkomsten) korrelerade variabler (barnens antal och ålder i olika hushåll). Med tillhjälp av skiktningseffekten undersöker Wold den amerikanska och den tyska skalans användbarhet för beräkningen av antalet konsumtionsenheter i ett hushåll. Han finner, att den amerikanska skalan (barn 0—3 år 0,15, 4—6 år 0,4, 7—10 år 0,75, 11—14 år 0,9, män över 14 år 1,0, kvinnor över 14 år 0,9 konsumtionsenheter) överskattar barnens betydelse för storleken av utgifterna för föda och bostad, men ger ett relativt gott resultat beträffande skodon och kläder. Den tyska skalan, som ger barnen en mindre vikt (barn 0—3 år 0,1, 4—6 år 0,2, 7—9 år 0,3, 10—12 år 0,4, 13—14 år 0,5, 15—16 år: män 0,7, kvinnor 0,6, 17—18 år: män 0,9, kvinnor 0,7, 19 år och däröver: män 1,0, kvinnor 0,8 konsumtionsenheter), ger en riktigare bild av födautgifternas stegring med barnantalet vid konstant realinkomst, än den amerikanska skalan. För bostadsutgifternas vidkommande måste man använda en skala, som ger en ännu mindre vikt än den tyska åt de yngre barnen, för att få skiktningseffekten att försvinna. Han finner sålunda, att man strängt taget borde använda olika skalor beroende på vilken utgiftspost, som skall reduceras med hänsyn till antalet konsumtionsenheter i hushållet. Om man önskar använda en enda skala, bör denna närmelsevis överensstämma med den tyska skalan.

Av stort intresse äro även Wolds anmärkningar angående sambandet mellan inkomstelasticiteten E_1 och priselasticiteten e_{11} för samma vara. Med utgångspunkt från det nära till hands liggande antagandet att konsumtionen av en vara förblir oförändrad, om samtliga varupris stiga i samma proportion som inkomsterna, härleder han formeln:

$$E_1 = e_{11} + e_{12} + \cdots + e_{1n}. \tag{3}$$

Formeln säger, att en varas inkomstelasticitet är lika med summan av varans priselasticitet i avseende å de olika förekommande varuprisen. Om varorna äro inbördes oberoende i efterfrågehänseende äro »kryss-elasticiteterna» $e_{1j}=0$, då j är olika 1, varför i detta fall $E_1 = e_{11}$. Priselasticiteten för en i efterfrågehänseende fristående vara är sålunda till-

närmelsevis lika med dess inkomstelasticitet. Om varuslagen V_1, V_2, V_3 bilda en fristående substitutionsgrupp är $e_{12} < 0$, $e_{13} < 0$, $e_{1j} = 0$ för $j > 3$, varför i så fall $e_{11} = E_1 - e_{12} - e_{13} > E_1$. Antagandet om oberoende i efterfrågehänseende utgör ett idealfall, som knappast fullständigt kan realiseras. Av hänsyn härtill kan man förvänta, att priselasticiteten e_{11} alltid skall visa sig vara större än inkomstelasticiteten.

På grundval av Kooperativa Förbundets marknadsstatistik för Konsumtionsföreningen Stockholm med Omnejd åren 1925—38 har Wold funnit de i följande tabell sammanställda pris- och inkomstelasticiteterna:

Tab. 1. *Pris- och inkomstelasticitet för angivna jordbruksprodukter.*

	Inkomst-elasticitet	Priselasticitet
Mjölk	0,25	0,2 à 0,3
Kött och fläsk.....	0,43	ca 0,7
Ägg	0,53	ca 1,0
Smör och margarin	0,25	0,4 à 0,5
Smör	0,42	a) 0,7 à 0,8 b) 1,4 à 1,5

a) Då förhållandet mellan smörpriset och margarinpriset är konstant.
b) Då margarinpriset är konstant.

Resultaten synas sålunda bekräfta riktigheten av satsen, att priselasticiteten är större än inkomstelasticiteten.

Detta resonemang, så bestickande det än kan förefalla, är likväl, såsom vi i det följande skola bevisa, i någon mån i behov av komplettering och korrigering.

Vad formel (3) beträffar förbiser Wold vid formelns härledning villkoret, att även kassahållningen bör utveckla sig proportionellt med inkomsterna eller närmare bestämt, att utgifterna böra utveckla sig proportionellt med inkomsterna. Vid snabba generalprisnivåförändringar är detta i regel ej fallet, varför formel (3) strängt taget icke gäller under det dynamiska händelseförloppet vid en prisnivåförskjutning. Ett beaktande av variationerna i normalhushållets bruttoförmögenhet och dess sammansättning samt i utgifterna för reklamändamål skulle vara nödvändigt för en uttömmande formelmässig beskrivning av efterfrågans beroende av olika ekonomiska variabler. Beaktas variationerna i konsumenternas lagerhållning, bli köparnas framtidsförväntningar beträffande prisutvecklingen av väsentlig betydelse för efterfrågans storlek. Det lönar sig nämligen för konsumenterna att öka sina lager, ifall priset på en vara vid den tidpunkt, då lagret konsumerats, är större än nupriset ökat med lagringskostnaderna intill denna tidpunkt. Om en generalprisstegring anteciperas av konsumenterna, ökas därför efterfrågan på lagringsbara produkter, även om samtliga pris utveckla sig proportionellt med inkomsterna. Ifall man i stället för penninginkomsterna betraktar penningvärdet av de under en viss period faktiskt konsumerade nyttigheterna, medelprisen för dessa varor enligt anskaffningspris plus lagerkostnader och i stället för de under den

ifrågavarande perioden inköpta varorna betraktar konsumtionens storlek, torde likheten (3) vara i det närmaste korrekt även under själva prisnivåförskjutningstiden. Om den betraktade perioden är tillräckligt lång, kan man bortse från variationerna i hushållets bruttoförmögenhet och dess sammansättning, varför formeln (3) närmelsevis gäller för långtidselasticiteterna.

Vad beträffar frågan, huruvida priselasticiteten e_{11} alltid är större än inkomstelasticiteten E_1, ha vi funnit, att denna relation i regel ej gäller för varuslag, vilkas inkomstelasticitet är större än 1, d. v. s. för de lyxbetonade varorna. Om V_1 betecknar en kvantitetsindex för en grupp nödvändighetsvaror, p_1 motsvarande prisindex, V_2 samtliga övriga varors kvantitetsindex och p_2 deras prisindex, så gäller, när normalhushållets utgifter och inkomster äro konstanta:

$$U = p_1 V_1 + p_2 V_2 = \text{konstant},$$

varav genom elasticitetsberäkning i avseende å priset p_1 fås

$$\frac{d \log U}{d \log p_1} = (1 - e_{11}) c_1 - e_{21} c_2 = 0; \; c_1 = p_1 V_1 / U, \; c_2 = p_2 V_2 / U$$

d. v. s.

$$(4) \qquad e_{21} = \frac{c_1}{c_2}(1 - e_{11}) \text{ och analogt } e_{12} = \frac{c_2}{c_1}(1 - e_{22}).$$

Ur formel (3) och (4) fås i detta fall

$$(5) \qquad E_2 = e_{22} + e_{21} = e_{22} + \frac{c_1}{c_2}(1 - e_{11})$$

$$(6) \qquad E_1 = e_{11} + e_{12} = e_{11} + \frac{c_2}{c_1}(1 - e_{22}).$$

Priselasticiteten e_{11} för en varugrupp är således alltid mindre än inkomstelasticiteten E_1 för samma varugrupp, om priselasticiteten för de till den betraktade gruppen icke hörande varuslagen är mindre än 1. Då priselasticiteten för nödvändighetsvarorna är mindre än 1, är således priselasticiteten för lyxvarorna i regel mindre än inkomstelasticiteten för dessa varor.

Ännu en omständighet kan vara värd omnämnande i detta sammanhang. Priselasticiteten

$e_{11} = \frac{\Delta \log V_1}{\Delta \log p_1}$ (vid konstant utgift U och konstant pris p_2) är *icke* identisk med realpriselasticiteten $\bar{e}_{11}$

$\bar{e}_{11} = \frac{\Delta \log V_1}{\Delta \log p_1}$ (vid konstant utgift U och konstant generalprisnivå P).

Vi ha nämligen, som man lätt kan bevisa: $\bar{e}_{11} = e_{11} - \frac{c_1}{c_2} e_{12}$,

$$(7) \qquad \bar{e}_{11} = e_{11} - \frac{c_1}{c_2}(E_1 - e_{11}) = e_{11} + e_{22} - 1 = \bar{e}_{22}.$$

Ur formel (6) och (7) kan man bl. a. utläsa följande olikheter:

Om e_{22} är större än 1, är $\bar{e}_{11} > e_{11} > E_1$.

Vid en statistisk analys är det lätt att förväxla priselasticiteten e_{11} med realpriselasticiteten $\bar{e}_{11}$. Härigenom komme man att i någon mån överskatta differensen mellan priselasticiteten och inkomstelasticiteten.

Av stort intresse är även frågan angående skillnaden mellan korttids- och långtidselasticiteten hos konsumtionen av en vara i avseende å inkomstförändringar. Wold finner, att för de icke kontraktbundna utgifterna korttidselasticiteten är större än långtidselasticiteten. Olika varuslag visa sig besitta olika tröghet vid den impuls till konsumtionsförändring, som en inkomstförändring innebär. Ett liknande fenomen inträffar vid prisförändringar. På grund av att prisförändringarna i regel anteciperas av konsumenterna, uppkommer vid en serie isolerade prisstegringar ett ganska komplicerat svängningsfenomen i efterfrågan på konsumtionsvaror. Sistnämnda problemkomplex förbigår Wold likväl med tystnad.

Den klassiska efterfrågeanalysen bygger på antagandet att elasticitetstalen skola visa sig vara närmelsevis konstanta vid olika inkomst- och prislägen. Wold varnar visserligen läsaren på ett par ställen för att draga för långt gående slutsatser på grundval av de erhållna elasticitetstalen, som egentligen blott gälla för inkomst- och prislägen i närheten av medelvärdena för de i det utnyttjade primärmaterialet förekommande inkomst- och prislägena, men han kunde gärna ha ägnat frågan om elasticitetstalens variation med inkomsterna och prisläget en större uppmärksamhet. I själva verket är antagandet om konstanta inkomstelasticiteter för samtliga varuslag en matematisk orimlighet, emedan i så fall utgifterna för de varuslag, vilkas inkomstelasticitet är större än 1 med växande inkomst, skulle bli större än själva inkomsten. Konsumtionen av nödvändighetsvaror ökar vid mycket små inkomster per konsumtionsenhet ungefär proportionellt med inkomsten samt närmar sig vid växande inkomst ett övre gränsvärde. Detta innebär, att nödvändighetsvarornas inkomstelasticitet varierar från ett tal i närheten av talet 1 mot talet 0, då inkomsten varierar från mycket små till höga värden. De lyxbetonade utgifterna förekomma i budgeten först från och med en viss minimi-inkomst, för att med stigande inkomstnivå växa till dess de omfatta största delen av utgifterna. De lyxbetonade varornas inkomstelasticitet avtager sålunda från ett mycket högt värde vid det inkomstläge, då de först uppträda i budgeten, mot värdet 0 eller 1 beroende på, om det ifrågavarande lyxbetonade behovet kan tillfredsställas ända till mättnad eller icke. Priselasticiteterna äro ej heller några konstanter, utan växa från värdet 0, då priset är = 0, mot gränsvärdet 1 eller oändligt stora värden beroende på, om varan är en absolut nödvändighetsvara eller icke. Vill man taga hänsyn till dessa omständigheter, har man att lägga efterfrågeanalysen på en annan grundval än den klassiska elasticitetskoefficientens.

Av det material, som Wold analyserat angående äggkonsumtionens variationer, framgår klart, att priselasticiteten ej är någon konstant vid olika realpris. Hans spridningsdiagram antyder, att priselasticiteten växer t. o. m. snabbare än realpriset.

Att döma av det statistiska material, som föreligger i den finska levnadskostnadsundersökningen av år 1928, torde man relativt väl kunna framställa sambandet mellan utgifterna och inkomsterna per normalhushåll för en nödvändighetsvara medelst algebraiska uttryck av den enkla formen

$$\text{(I)} \qquad u(x) = u(\infty)\frac{x}{x+a}; \quad E(x) = \frac{a}{x+a} = 1 - \frac{u(x)}{u(\infty)}.$$

Talet $u\,(\infty)$ betecknar det gränsvärde, utgiften för den betraktade varugruppen närmar sig, då hushållets inkomst x växer över alla gränser. Talet a, som bestämmer den hastighet, med vilken utgiften $u\,(x)$ närmar sig mättningsgränsen $u\,(\infty)$ och elasticiteten $E\,(x)$ gränsvärdet 0, beror av varugruppens prisindex p, denna prisindex' relation till övriga varors prisindexer, hushållets sammansättning m. fl. omständigheter. $k\,(x) = u\,(x) / p$ utgör en index över den konsumerade kvantiteten av det betraktade varuslaget eller varuslagen.

De s. k. relativa lyxvarorna ha en efterfrågefunktion, som i regel väl kan framställas i formen

$$\text{(II)} \qquad u(x) = u(\infty)\frac{x - x_0}{x+a}; \quad E(x) = \frac{a}{x+a} + \frac{x_0}{x - x_0}.$$

Talet x_0 betecknar den minimi-inkomst, som är nödvändig för att hushållet skall efterfråga de betraktade relativa lyxvarorna. De relativa lyxvarorna likna nödvändighetsvarorna däri, att hushållets utgifter för dem närmar sig en mättningsgräns $u\,(\infty)$, då inkomsten x växer över alla gränser.

De specifika lyxvarorna betalas medelst den del av inkomsten, som blir över, när de mättningsbara behoven först tillfredsställts. Det är därför naturligt att vänta sig, att efterfrågefunktionen för de specifika lyxvarorna kan skrivas i formen

$$\text{(III)} \quad u(x) = c\left(x - \frac{x_0 + a}{x+a}x\right) = cx\frac{x - x_0}{x+a}; \quad E(x) = 1 + \frac{a}{x+a} + \frac{x_0}{x - x_0}.$$

c betecknar den bråkdel av totalutgifterna, som de betraktade specifika lyxutgifterna närma sig, då totalutgifterna växa över alla gränser.

Om flera olika varuslag tillfredsställa ett likartat behov, såsom smör och margarin, har hela substitutionsgruppen i regel en efterfrågefunktion av typen I, medan de i gruppen ingående relativa lyxvarorna äro av typen II. Surrogatvarornas efterfrågefunktion torde därför i regel kunna väl framställas som skillnaden mellan en efterfrågefunktion av typen I och en efterfrågefunktion av typen II.

Tab. 2. *Närmeformler för sambandet mellan utgifterna och inkomsterna per normalhushåll om 3,4 amerikanska konsumtionsenheter. Arbetare och lägre tjänstemannahushåll.*

Utgiftsgrupp	Utgifter per normalhushåll				Inkomstelasticitet			
	Finland år 1928				Finland 1928			Sverige 1933
	Formel $u(x)$	$u(20)$	$u(40)$	$u(60)$	$E(20)$	$E(40)$	$E(60)$	H. Wold
Föda, animalisk och vegetabilisk	$\frac{26\,x}{x+31}$	10 200	14 600	17 140	0,60	0,43	0,34	0,37
Animalisk föda	$\frac{16\,x}{x+37}$	5 600	8 600	9 900	0,65	0,48	0,38	0,39
Vegetabilisk föda	$\frac{26\,x}{x+31} - \frac{16\,x}{x+37}$	4 600	6 000	7 240	0,54	0,36	0,28	0,33
Mjölk och grädde	$\frac{4{,}9\,x}{x+27}$	2 080	2 900	3 380	0,57	0,40	0,31	0,27
Grädde	$\frac{2(x-10)}{x+100}$	200	500	625	1,83	1,05	0,83	1,08
Mjölk	$\frac{4{,}9\,x}{x+27} - \frac{2(x-10)}{x+100}$	1 880	2 400	2 755	0,44	0,26	0,19	0,08
Smör, margarin o. a. fett	$\frac{3{,}1\,x}{x+17}$	1 670	2 180	2 416	0,46	0,30	0,22	0,34
Smör	$\frac{3{,}1\,x}{x+30}$	1 240	1 770	2 067	0,60	0,43	0,33	0,67
Margarin och fett	$\frac{3{,}1\,x}{x+17} - \frac{3{,}1\,x}{x+30}$	430	410	349	0,06	−0,03	−0,44	−0,33
Ost	$\frac{0{,}28\,(x-10)}{x+20}$	70	140	175	1,50	0,67	0,45	0,30
Ägg	$\frac{0{,}80\,(x-10)}{x+20}$	200	400	500	1,50	0,67	0,45	0,51
Kött, färskt	$\frac{2{,}75\,(x-9)}{x+10}$	1 000	1 700	2 000	1,15	0,49	0,29	0,49
Tobak	$\frac{1{,}8\,(x-5)}{x+45}$	416	741	944	1,03	0,67	0,52	0,94
Beklädnad	$\frac{21\,(x-10)}{x+160}$	1 500	3 700	5 500	1,89	1,13	0,93	0,96
Skodon	$\frac{3{,}5\,x}{x+68}$	795	1 296	1 640	0,77	0,63	0,53	
Bostad	$\frac{54\,(x-5)}{x+290}$	2 610	5 730	8 485	1,27	1,02	0,92	1,25
Inventarier	$\frac{0{,}2\,x\,(x-10)}{(x+80)}$	400	2 000	4 286	2,80	2,00	1,76	1,41
Totalutgifter	x	20 000	40 000	60 000	1,00	1,00	1,00	0,97

Beräkningen av de obestämda parametrarna $u(\infty)$, a, x_0 och c ur det föreliggande statistiska materialet med tillhjälp av en formel härledd enligt den s. k. minsta kvadratmetoden är ganska arbetskrävande. Ett praktiskt taget lika gott resultat erhålles på följande enkla sätt. Man utjämnar först grafiskt den kurva, som sammanbinder de punkter u_x, vilka representera medelutgiften för det betraktade varuslaget för en grupp likartade hushåll med medelinkomsten x. Man avläser från denna utjämnade kurva tre funktionsvärden (u_1, x_1), (u_2, x_2), (u_3, x_3), samt bestämmer de obestämda parametrarna så, att kurvan $u(x)$ går genom dessa tre punkter. Enligt teorin för de lineära substitutionerna veta vi, att, då $u(x)$ lyder formeln (II), är:

$$\frac{u(x)-u_1}{u(x)-u_2} : \frac{u_3-u_1}{u_3-u_2} = \frac{x-x_1}{x-x_2} : \frac{x_3-x_1}{x_3-x_2}$$

Genom att hyfsa denna likhet så, att $u(x)$ blir en explicit funktion av x, får man de sökta värdena $u(\infty)$, a och x_0. Formeln I är blott ett specialfall av II med $x_0 = 0$. Ifall $u(x)$ är av typen III, är $u(x) / x$ av typen II.

I tabell 2 ha vi sammanställt resultaten av ett försök att beräkna närmeformler för sambandet mellan utgifterna och inkomsterna per normalhushåll inom olika välståndsgrupper. Materialet hänför sig till arbetare och befattningshavare i Finland med en medelutgift år 1928 på 27 700 mk. Normalhushållet bestod av 3,40 amerikanska konsumtionsenheter. Som penningenhet användes i formlerna 1 000 mk av 1928 års penningvärde. Vi ha vidare anfört utgiftsfunktionernas och de motsvarande elasticitetsfunktionernas värden, då inkomsten $x = 20$, 40 och 60 tusen mk. I den sista kolumnen ha vi antecknat de av H. Wold beräknade motsvarande inkomstelasticiteterna (tab. 12, sid. 105, arbetare och lägre tjänstemän 1933). Hans elasticitetstal ligga ofta emellan elasticitetstalen $E(40)$ och $E(60)$, antydande att den svenska medelinkomstnivån (4 200 kr.) skulle motsvara en inkomst på c:a 50 tusen mark per normalhushåll, d. v. s. en väsentligt högre inkomstnivå än medelinkomstnivån i det finska materialet ($x = 28$). Även olikheterna i varornas relativa dyrhet i Finland och Sverige ha förskjutit elasticitetstalen i förhållandet till varandra.

Förutom på grund av inkomstnivåns höjd och prisläget påverkas inkomstelasticiteten av sysselsättningens art, bosättningsorten m. fl. omständigheter. För att belysa sysselsättningens inflytande ha vi i tabell 3 antecknat formler, som beskriva födautgifternas beroende av inkomsterna för normalhushåll hänförda till grupperna 1) arbetare, 2) befattningshavare, 3) arbetare och befattningshavare samt 4) tjänstemän vid den finska levnadskostnadsundersökningen av år 1928.

Födautgifterna stiga således kännbart, då hushållets huvudman har en sysselsättning, som kräver större fysiska ansträngningar. Ur metodologisk synpunkt är det av intresse att konstatera, att sammanförandet av arbetare och befattningshavare till en enda hushållsgrupp ger upphov till ett skiktningsfel, som åstadkommer, att inkomstelasticiteten för hela gruppen beräknad på grundval av spridningsdiagrammet blir lägre

Tab. 3. *Födautgifterna per normalhushåll vid varierande välstånd.*

Hushållsgrupp	Utgifter i 1 000 mk				Inkomstelasticitet			
	$u(x)$	$u(20)$	$u(40)$	$u(60)$	$E(x)$	$E(20)$	$E(40)$	$E(60)$
Arbetare	$\frac{28x}{x+34}$	10,5	15,3	17,9	$\frac{34}{x+34}$	0,63	0,46	0,36
Befattningshavare	$\frac{27x}{x+36}$	9,6	14,2	16,8	$\frac{36}{x+36}$	0,64	0,47	0,37
Arbetare och befattningshavare	$\frac{26x}{x+31}$	10,2	14,6	17,1	$\frac{31}{x+31}$	0,61	0,48	0,34
Tjänstemän	$\frac{26x}{x+40}$	8,7	13,1	15,6	$\frac{40}{x+40}$	0,67	0,50	0,40

än för arbetare och befattningshavare skilt för sig. Wold har ej beaktat denna felkälla vid sina beräkningar.

Den uppgift dr Wold haft att lösa, har blott varit att beräkna efterelasticiteten för en del viktiga jordbruksprodukter och icke att undersöka efterfrågeproblemet i hela sin vidd. De här framställda anmärkningarna äro därför närmast att uppfatta som en komplettering av hans mycket värdefulla och intressanta undersökning.

Leo Törnqvist.

RÄTTELSE.

A sid. 147 rad. 20—21 uppifrån *står* prisstegringsgenerande, *läs* prisstegringsgenererande.

EFTERFRÅGEANALYS OCH EKONOMETRISKA FORSKNINGSMETODER

HERMAN WOLD i samarbete med LARS JURÉEN: *Demand Analysis. A Study in Econometrics*. Almqvist & Wiksell. Stockholm 1952. John Wiley & Sons Inc. New York 1953. 358 sidor.

Professorn i statistik vid Uppsala universitet HERMAN WOLD har, efter ett ihärdigt arbete under tolv års tid, publicerat en bok, som »den omdömesgilla forskningen är tvungen att taga hänsyn till, därför att den bygger på en djup och långsamt förvärvad insikt inom det område den behandlar» (citat från Ragnar Granit, »En ung mans väg till Minerva», sid. 30, vilket förf. satt som motto över sitt i slutet av boken samlade fotnotsmaterial).

Boken utgör ett resultat av ett samspel mellan empirisk forskning och skarpsinniga teoretiska resonemang med ändamålet att finna nya metoder för att analysera statistiskt material av det slag man måste bearbeta vid empirisk efterfrågeanalys.

Till bokens teoretiska apparatur hör härvid förutom en fullt genomarbetad version av den klassiska pristeorin, teorin för regressionsanalys, teorin för sta-

tionära stochastiska processer och en inledning till en teori för rekursiva kausala utvecklingsförlopp.

Vid bearbetningen och analysen av det vidlyftiga empiriska iakttagelsematerialet har en av prof. Wolds förra elever, LARS JURÉEN, numera direktör för Svenska Statens Jordbruksnämnds marknadsforskningsbyrå, varit honom behjälplig.

Bokens teoretiska del utgör en syntes av alla viktiga forskningsresultat professor Wold och hans elever tidigare presterat, i regel publicerade som specialundersökningar för experter i matematisk statistik. Ställvis ha nya kompletterande analyser gjorts av de synpunkter och resultat som sedermera visat sig vara speciellt fruktbärande.

Wolds bok är väl lämpad som lärobok i ekonometrisk och statistisk forskningsteknik på det mera avancerade stadiet. Men också undervisningen i nationalekonomi kan utnyttja valda delar av boken. Hans analys av tillämpningsområdet för minstakvadratmetoden och den klassiska regressionsanalysen kan med fördel inarbetas i den elementära statistikundervisningen.

Vid ett grundligt studium av boken finner man en mängd detaljer, som kräva en ingående kunskap om den roll de behandlade frågorna spelat i den ekonometriska litteraturen för att man fullt skall kunna uppskatta den speciella finess varmed de behandlats eller såsom i många fall skett inordnats som ett övningsexempel till ett kapitel i boken.

På grund av den rika variationen av frågekomplex av olika svårighetsgrad och den mycket lättfattliga framställningen av frågor, som kunna behandlas utan en tung matematisk apparat kan en läsare, som icke låter sig förskräcka av de mera avancerade partierna, finna mycket han kan tillgodogöra sig, även om hans förkunskaper ej äro tillräckliga för att han skall begripa boken i dess helhet.

En intressant fråga inom området för den klassiska pristeorin är huruvida de empiriska efterfrågefunktionerna kunna antagas vara sådana, att de stå i strid med hypotesen att det existerar en preferensfunktion ur vilken alla de betraktade efterfrågefunktionerna kunna härledas med tillhjälp av principen att maximera preferensfunktionen under tilläggsvillkoret att budgetlikheten inkomst = utgift är uppfylld.

Wold anser sig ha lyckats bevisa, att ifall en konsuments efterfrågefunktioner bilda ett entydigt bestämt motsägelsefritt system, också en preferensfunktion definierande indifferensytorna för denna individ existerar. Andra författare (Samuelson, Houthakker) betvivla riktigheten av detta påstående. För marknaden i sin helhet existerar däremot enligt Wold i allmänhet icke någon preferensfunktion ur vilken marknadens efterfrågefunktioner kunna härledas.

Min ståndpunkt i denna kontrovers kan sammanfattas i följande förmodande. Det finnes icke någon praktisk möjlighet att motbevisa hypotesen om existensen av en preferensfunktion med de önskade egenskaperna, enär de empiriska efterfrågefunktionerna icke kunna bestämmas med en sådan exakthet att icke något system taget från det fantastiskt flexibla system av system av efterfrågefunktioner, som kunna härledas ur en generell preferensfunktion med tillhjälp av den betingade maximeringsprincipen, icke skulle kunna beskriva det observerade

statistiska materialet lika väl som de efterfrågefunktioner som fixerats utan avseende fäst vid frågan huruvida det erhållna systemet kan härledas ur en hypotetisk preferensfunktion eller icke.

Ifall man sålunda av förkärlek för en matematiskt elegant syntes önskar utgå ifrån att det existerar en preferensfunktion av det slag som den klassiska pristeorin antagit, kan det tillgängliga statistiska materialet icke »tvinga» forskaren att frångå denna hypotes såsom varande irrealistisk. Å andra sidan finnes det naturligtvis icke heller någon möjlighet att »bevisa» att efterfrågefunktionerna äro integrabla till en preferensfunktion.

Man har i detta fall att göra med en akademisk stridsfråga, som envar kan taga ställning till enligt sin personliga preferensfunktion.

Personligen är jag böjd för att pålägga det system av närmeformler som tilldelas uppgiften att giva en ideal statisk beskrivning av ett förefintligt statistiskt iakttagelsematerial angående efterfrågans variation vid variationer i inkomster och priser, kravet att detta system skall kunna integreras till en preferensfunktion, utan att härvid alls bekymra mig över den filosofiska frågan, huruvida efterfrågefunktionerna i det hypotetiska fall, då inga störningar av mer eller mindre slumpmässig karaktär skulle ha inverkat på iakttagelsematerialet, äro integrabla eller icke. Icke heller bekymrar mig frågan huruvida den erhållna preferensfunktionen för hela marknaden kan tolkas som en funktion de handlande subjekten medvetet försökt maximera under de tilläggsvillkor, som nödvändigtvis måste vara uppfyllda. Det erhållna systemets frihet från inkonsekvenser och dess matematiska elegans skulle vara en tillräcklig motivering för att genom detta extra krav i någon mån begränsa flexibiliteten i det använda systemet av närmeformler. I praktiken måste man i varje fall nöja sig med så grova uppskattningar av de obestämda parametrarna i formlerna, att det strängt taget icke blir aktuellt att göra ett verkligt val mellan integrabla och icke integrabla efterfrågefunktioner.

Ifall man åter icke vill nöja sig med den på sätt och vis statiska modell den klassiska pristeorin utarbetat, utan i stället önskar få en dynamisk modell som även beaktar de slumpartade elementen i den utvecklingsprocess som konsumtionen, inkomsterna och prisen beskriva, saknar frågan angående efterfrågefunktionernas integrabilitet allt intresse, emedan i detta fall själva existensen av efterfrågefunktioner av den matematiskt stränga karaktär, som den klassiska pristeorin opererar med blir ifrågasatt redan förrän själva analysen börjats. I en sådan modell måste man nämligen operera med sannolikhetsfördelningar beskrivande interkorrelerade stochastiska processer.

För att övergå till mindre högt flygande frågor skall jag göra några anmärkningar om de inkomst- och priselasticiteter Wold opererar med i sin undersökning.

Wolds analys av sambanden mellan olika inkomst- och priselasticiteter torde kunna väsentligt fördjupas genom att också studera de elasticiteter som beskriva efterfrågans variationer vid variationer i kvoten mellan prisindexen för en varugrupp och prisindexen samt utgiften för en större varugrupp till vilken den betraktade varugruppen på ett naturligt sätt kan hänföras. Genom att begagna sig av ett ändamålsenligt system av grupper och undergrupper samt genom att vid

detaljanalysen av efterfrågan låta totalutgiften för gruppen spela samma roll som inkomsten = totalutgiften spelar i Wolds analys, torde man få ett fastare grepp på efterfrågefunktionerna och deras elasticiteter än om man icke begagnar sig av några mellanled mellan den betraktade undergruppen och totalkonsumtionen.

I detta sammanhang vill jag uttrycka ett svagt missnöje över författarens beslut att låta den bekväma termen realpriselasticitet »degenerera» till en synonym till priselasticiteten med avseende å variationer i det nominella priset. Detta inträffar om man identifierar realpriselasticiteten med den *partiella* derivatan av logaritmen för efterfrågefunktionen i avseende å logaritmen för kvoten mellan den ifrågavarande varans pris och prisindexet för totalkonsumtionen (med ombytt förtecken).

I motsats till vad som är fallet då man beräknar elasticiteten med avseende å nominalpriset, kan man likväl icke tänka sig att endast ett realpris kan variera utan att samtidigt något annat realpris förändras, detta emedan en prisförändring på en vara även förändrar prisindexen för totalkonsumtionen. Detta faktum kan utnyttjas för att motivera att realpriselasticiteterna icke definieras på ett sätt som är direkt kopierat från definitionen av nominalpriselasticiteten.

Ett realpriselasticitetsbegrepp anpassat för det fall där man önskar beskriva variationen i efterfrågan som en funktion av *endast två* variabler, nämligen den ifrågavarande varans pris och realinkomsten, utgör kvoten mellan minskningen i efterfrågans logaritm vid konstant realinkomst och tillväxten i logaritmen för realpriset. Denna kvot har jag i en recension av ett tidigare arbete av Wold i Ekonomisk Tidskrift, årg. 43 (1941), betecknat som realpriselasticitet.

I samband med det teorem som fastslår att realpriselasticiteten enligt Wolds terminologi är en synonym till den nominella priselasticiteten anmärker Wold att jag kallat en storhet, som han kallar elasticiteten med avseende å nominalpriset vid en kompenserande variation i ett annat pris, för realpriselasticitet. Olyckligtvis har i den formel han i detta sammanhang anför insmugit sig en liten lapsus, som leder till det humoristiska resultatet att jag skulle ha kallat ett tal identiskt = 0 för realpriselasticitet. Den korrekta formeln lyder:

$$\overset{**}{e}_{p_i} Q_k = e_{p_i} Q_k - \frac{x_i}{x_j} e_{p_j} Q_k \; ; x_i = p_i Q_i \; , \; x_j = p_j Q_j \; .$$

Om man som Wold av ett förbiseende sätter $j = k$, d.v.s. identifierar den varugrupp Q_k vars priselasticitet studeras med den varugrupp Q_j vars prisförändring kompenserar prisförändring i priset p_i, fås resultatet $\overset{**}{e}_{p_k} Q_k = 0$.

Min »realpriselasticitet» hänförde sig till specialfallet $Q_k = Q_i \; , \; x_j = \mu - x_i$.

Om en prisstegring på andra varor har en tendens att öka konsumtionen av den betraktade varan (ifall dess pris förblir oförändrat) är $\overset{**}{e}_{p_i} Q_i > e_{p_i} Q_i$. Den praktiska betydelsen av detta resultat ligger i att man vid en regressionsanalys vid vilken endast realpriset $p_i = \acute{p}_i / P$ och realinkomsten $\mu = \acute{\mu} / P$ begagnats som förklarande variabler, i regel torde få ett närmevärde för $\overset{**}{e}_{p_i} Q_i$ och icke för $e_{p_i} Q_i$, som Wold synes förmoda. Om samtliga varors

realpriser ingå i regressionsformeln får man naturligtvis i princip ett närmevärde för $e_{pi}\ Q_i$, ehuru detta icke är säkert då tilläggsvillkoret $\Sigma\, x_i\, d \log p_i = 0$ kan väntas åstadkomma komplikationer för den statistiska regressionsanalysen (kollinearitet), som förhindrar konsistenta närmevärdena för $e_{p_i}\ Q_i$.

Wold begagnar sig på ett flertal ställen av det system för approximativ beskrivning av efterfrågefunktionernas form, som jag i min nämnda recension i Ekonomisk Tidskrift, sid. 216—225 av ett av hans tidigare arbeten om efterfrågeanalys utnyttjat. Detta system har han emellertid icke godtagit i sin helhet utan han har föredragit att utelämna den typ, som jag föreslagit för underlägsna varor, exempelvis margarin, vilka varor vid stigande välstånd undanträngas av »bättre» varor tillfredsställande ett liknande behov, t.ex. smör.

Mitt ursprungliga förslag utgjorde att man skulle beräkna efterfrågefunktioner för underlägsna varor såsom skillnaden mellan efterfrågefunktionen $d_1\,(\mu)$ för hela den betraktade varugruppen och efterfrågefunktionen $d_2\,(\mu)$ för den grupp som återstår då den underlägsna varan utelämnas.

Om hela gruppens efterfrågefunktion närmelsevis kan framställas medelst den brutna lineära funktionen av välståndsnivån μ:

$$d_1\,(\mu) = d_1\,(\infty)\ \frac{\mu - \mu_{01}}{\mu + \beta_1}\ ;\ \mu_{01} \lessapprox 0,\ \beta_1 > 0$$

typisk för nödvändighetsvaror, och efterfrågan på de »bättre» varorna i gruppen medelst funktionen:

$$d_2\,(\mu) = d_2\,(\infty)\ \frac{\mu - \mu_{02}}{\mu + \beta_2}\ ;\ \mu_{02} > 0,\ \beta_2 > 0$$

följer härav med logisk nödvändighet att efterfrågan på den underlägsna varugruppen $d_3\,(\mu) = d_1\,(\mu) - d_2\,(\mu)$ kan approximeras medelst funktionen:

$$d_3\,(\mu) = d_3\,(\infty) + \frac{\gamma_2}{\mu + \beta_2} - \frac{\gamma_1}{\mu + \beta_1}$$

där

$$\gamma_2 = d_2\,(\infty)\ (\mu_{02} + \beta_2),\ \gamma_1 = d_1\,(\infty)\ (\mu_{01} + \beta_1)$$

och

$$d_3\,(\infty) = d_1\,(\infty) - d_2\,(\infty).$$

Wold har antingen icke godtagit detta förslag eller förbisett detsamma, enär han i stället för detta resultat, begagnar funktioner av typen:

$$d_4\,(\mu) = \frac{d_4\,(\infty)\mu}{\mu - \beta_4}\ ,\ \beta_4 > 0$$

d.v.s. en modifikation av formlerna för nödvändighetsvaror, som närmeformel för efterfrågan på underlägsna varor, utan att han explicit påpekar att denna modifikation icke ingick i mitt ursprungliga system.

Formeln $d_4\,(\mu)$ har tvenne tråkiga svagheter. Den antager värdet ∞ för $\mu = \beta_4 > 0$ och saknar ett »välartat» maximum innanför det relevanta inkomst-

området. Ifall Wold ej velat godtaga närmeformler av typen $d_3(\mu)$ för underlägsna varor, något som är svårt att motivera, med hänvisning till svårigheter att beräkna alla parametrarna $d_3(\infty)$, γ_2, β_2, γ_1, β_1 direkt från det empiriska materialet, då detta icke alls är nödvändigt, då de kunna beräknas så snart parametrarna i de enklare funktionerna $d_1(\mu)$ och $d_2(\mu)$, vilka funktioner i och för sig äro minst lika intressanta som efterfrågefunktionen $d_3(\mu)$, beräknats, skulle han ha erhållit ett bättre resultat genom att utvidga användningsområdet för min närmeformel för efterfrågan för specifika lyxvaror:

$$d(\mu) = \varkappa \mu \frac{|\mu - \mu_0|}{\mu + \beta}$$

även till att omfatta »specifikt underlägsna varor» genom att i detta fall låta μ_0 vara större än μ. Denna funktion har för $\mu < \mu_0$ ett »välartat» maximum för $\mu = \sqrt{\mu_0 \beta}$. För $\mu > \mu_0$ har man att sätta $d(\mu) = 0$, emedan μ_0 för specifikt underlägsna varor anger icke såsom för lyxvarorna den lägsta inkomstnivå vid vilken de förekomma i budgeten, utan i stället den högsta inkomstnivå vid vilken de fortfarande efterfrågas.

Funktioner av typen $d_3(\mu) = d_1(\mu) - d_2(\mu) = \text{Min}\,(d_1(\mu), d_3(\infty) + \frac{\gamma_2}{\mu + \beta_2} - \frac{\gamma_1}{\mu + \beta_1}$ ha även ett välartat maximiställe för inkomstnivån:

$$\mu = \text{Max}\left(\mu_0, \frac{\beta_2 \sqrt{\gamma_1} - \beta_1 \sqrt{\gamma_2}}{\sqrt{\gamma_2} - \sqrt{\gamma_1}}\right).$$

De obestämda parametrarna i »Törnqvists efterfrågefunktioner» skola betraktas såsom funktioner av prisen på de varor som ingå i de ifrågavarande varugrupperna, och, om än i en relativt liten grad, av prisen på varor som tillhöra andra varugrupper. I första approximation är parametrarna i en efterfrågefunktion för en varugrupp av nödvändighetskaraktär proportionella med en prisindex för de varor som tillhöra gruppen. De relativa lyxvarornas parametrar växa med prisindexen för de relativa lyxvarorna, men avtaga om de underlägsna varornas pris stiga i förhållande till de bättre varornas prisindex. Parametern μ_{02} t.ex. är närmelsevis en lineär funktion av formen

$$\mu_{02} = P_2 \mu_{022} - P_1 \mu_{021}.$$

Wold har icke utnyttjat möjligheten att härleda efterfrågefunktionernas egenskaper vid variationer i prisen direkt från det statistiska materialet från en serie konsumtionsundersökningar. Härtill hade det blott krävts att de erhållna närmevärdeserierna för parametrarna β_1, γ och μ_0 etc. korrelerats med prisindextal för de ifrågavarande varugrupperna för de olika undersökningsåren. För att få en föreställning om efterfrågans beroende av prisen har han i stället utnyttjat tidsserier av marknadsstatistisk karaktär.

De ovan anförda kritiska anmärkningarna och kompletteringarna väga emellertid mycket lätt då man jämför dem med de obestridliga förtjänster Wolds arbete har.

Leo Törnqvist

EN BEFOLKNINGSSTATISTISK JÄMFÖRELSE MELLAN FINLAND OCH SVERIGE[1])

Av fil. doktor LEO TÖRNQVIST

En jämförelse mellan de statistiska uppgifter, som beskriva befolkningsförhållandena och befolkningsutvecklingen i Finland och Sverige, underlättas väsentligt av att det statistiska underlaget för en sådan jämförelse insamlats och bearbetats på ett mycket likartat sätt i de båda länderna. Under äldre tider var enhetligheten vid bearbetningen ännu större än nu, enär befolkningsstatistiken för Finland och Sverige för åren 1750—1808 utarbetades under samma ämbetsverks, »Tabellverkets», ledning. Senare har visserligen befolkningsstatistiken i Sverige gjort större framsteg än i Finland, varför man numera på grundval av publicerad officiell statistik får en detaljrikare och bättre analyserad bild av befolkningsutvecklingen och befolkningsförhållandena i Sverige än i Finland. Det enhetliga utgångsläget samt det nära samarbete, som existerat mellan befolkningsstatistikerna i Finland och Sverige, är en relativt god garanti för att man kan betrakta de befolkningsstatistiska siffrorna för Finland och Sverige såsom sinsemellan jämförbara. Vid vår jämförelse mellan de befolkningsstatistiska förhållandena i Finland och Sverige åren 1936—1940, visar det sig dock att det svenska materialet för vårt ändamål är bättre än det motsvarande materialet

[1]) För den hjälp jag erhållit av prof. Herman Wold, under den månad i januari—februari 1948 jag på inbjudan av Uppsala Universitet föreläste och arbetade vid Statistiska Institutionen i Uppsala, varvid jag blev i tillfälle att i ett seminarieföredrag framföra en del preliminära undersökningsresultat och diskutera dem, får jag här framföra mitt varma tack. Assistenten vid Helsingfors universitets Statistiska Institut, överaktuarien vid Statistiska Centralbyrån, magister Eino H. Laurila ber jag tillika få tacka för värdefull hjälp vid genomläsning och granskning av manuskriptet. För den hjälp jag haft av prof. Carl-Eric Quensels befolkningsstatistiska arbeten, samt de impulser till fortsatta undersökningar på detta område han givit mig är jag också skyldig ett tack.

för Finland. För Finlands vidkommande måste man sålunda på ett tidigare stadium än för Sveriges del begagna sig av uppskattningsmetoder för att få fram de tal som behövas.

Befolkningsutvecklingen sedan 1750.

Vi skola först kasta en blick på befolkningsutvecklingen i Finland och Sverige sedan år 1750.

För att få en möjligast åskådlig föreställning om folkmängdens förändringar i Finland och Sverige alltsedan vår befolkningsstatistiks barndom ha vi ritat diagram 1. Kurvorna ha ritats i logaritmisk skala varigenom relativt sett lika stora förändringar förskjuta kurvorna med samma sträcka. I diagrammet finnas inalles tio kurvor, nämligen för såväl Finlands som Sveriges del, kurvor, som beskriva följande tidsserier: I = Folkmängden. Kurvorna ha ritats så breda att kurvans övre kant närmelsevis beskriver den legala bokförda (kyrkskrivna och i civilregistret antecknade) befolkningen, medan den undre kanten åskådliggör den s.k. närvarande befolkningen i landet. Sistnämnda befolkning är c:a 5—8 % mindre än den bokförda befolkningen, som även upptager i utlandet bosatta medborgare och sådana, sannolikt döda, personer om vilka icke erhållits någon dödsanmälan, trots att det icke kommit något livstecken från dem på en längre tid. II = Giftermål, d.v.s. tio gånger antalet giftermål under det decennium vars sista år anger den tidpunkt som betraktas, då ifrågavarande punkt inritas i diagrammet. Genom multiplikationen med 10 förskjutes kurvan uppåt med en sträcka motsvarande avståndet mellan punkterna 1 och 10 i den logaritmiska skalan. Denna förskjutning är ändamålsenlig för att kurvan lättare skall kunna jämföras med kurvan för folkmängden. III = Antalet levande födda barn under den tioårsperiod, som föregår den tidpunkt, som bestämmer den ifrågavarande kurvpunktens läge på tidsaxeln. Man lägger märke till den snabba nedgången i antalet födda barn efter år 1910. En omsvängning i denna nedgångstendens inträffade under tioårsperioden 1930—40. IV = Antalet dödsfall under den senast förflutna tioårsperioden. Omkring år 1870 upphörde den relativt snabba stegringen av antalet dödsfall per decennium för att därefter hålla sig på närmelsevis samma nivå trots att folkmängden under samma tid vuxit väsentligt. Detta fenomen kommer till uttryck på ett annat sätt i kurvorna V, som beskriva de befolkningar, som skulle erhållas, ifall det årligen skulle födas 100.000 barn,

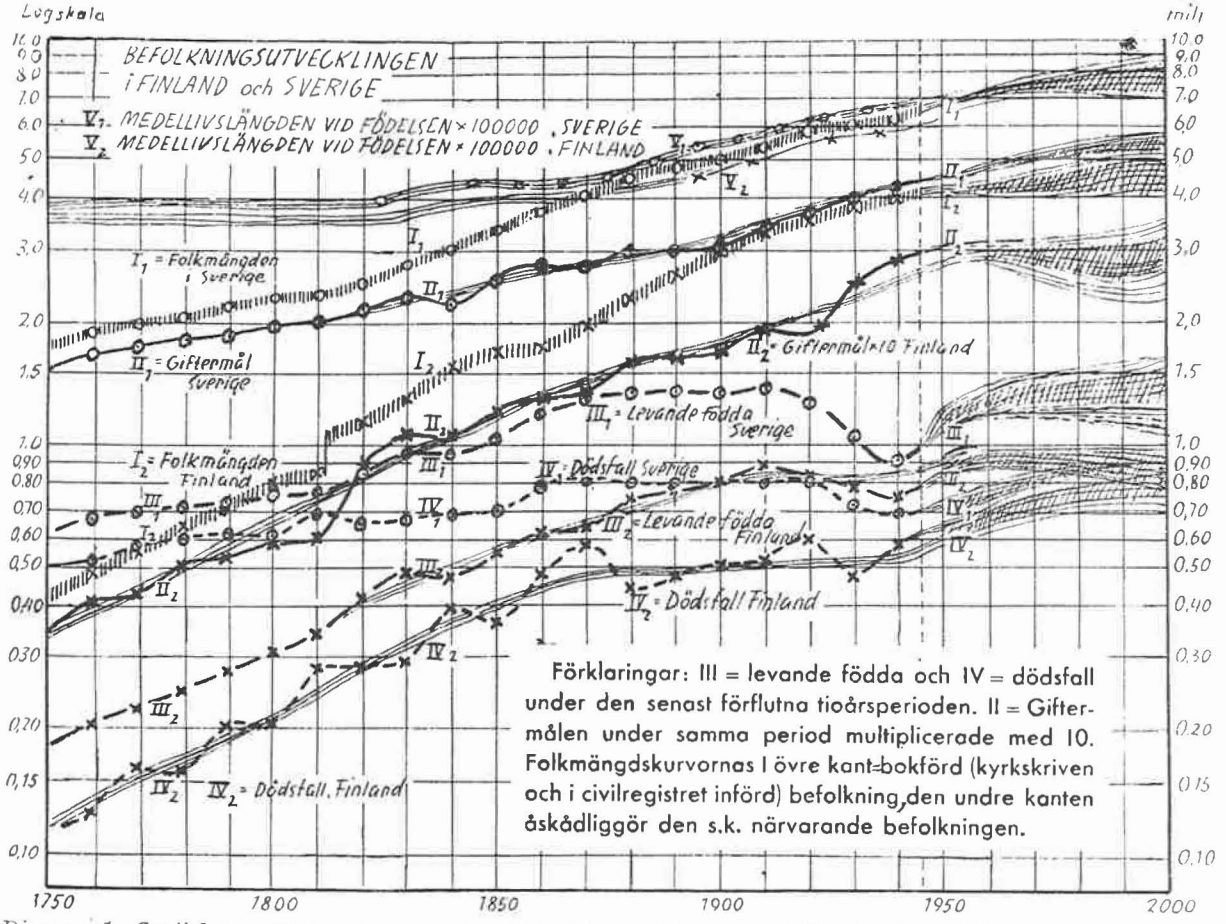


Diagram 1. Grafisk översikt över folkmängdens förändringar i Finland och Sverige åren 1750—2000. Det fält, inom vilket man kan vänta sig att välmotiverade prognoser för dessa tidsserier för åren 1950—2000 skola befinna sig, ha vi streckat. Denna streckning kan närmast tolkas som en bild av ett slags subjektivt bestämd sannolikhetstäthet för att olika slags prognoser skola visa sig vara lyckade prognoser.

vilkas medellivslängd skulle utgöra densamma, som i den officiella statistiken beräknats såsom den återstående medellivslängden vid födelsen i enlighet med överlevelsetabellerna för olika årtionden för båda könen gemensamt (se även diagram 2). Såsom av en jämförelse mellan kurvorna I och V framgår svarar nedgången i dödligheten, åtminstone för Sveriges vidkommande, för huvudparten av befolkningstillväxten efter 1880. Om denna mycket märkliga ökning av medellivslängden skulle ha uteblivit, torde Sveriges befolkningstillväxt ha avstannat redan omkring år 1925, för att efter att ha nått ett befolkningsmaximum på något över 5 milj. invånare minska på grund av nedgången i födelsetalen under perioden 1910—1940. Tack vare den fortsatta ökningen i medellivslängden är det möjligt att Sveriges befolkning överhuvud icke kommer att genomgå någon minskningsperiod under det innevarande århundradet, förutsatt att födelsekurvan icke efter den stegring som begynte omkring år 1935 ånyo snabbt vänder nedåt. Nedgången i dödligheten torde också för Finlands vidkommande vara i stånd till att förhindra, att den nedgång som inträffade i födelsetalen i Finland åren 1910—1940 skall såsom man befarat[1]) resultera i att befolkningskurvan för Finland vänder neråt före detta sekels slut. Av diagrammet framgår tydligt, att befolkningstillväxten i Finland under hela den betraktade perioden försiggått relativt sett mycket snabbare än i Sverige. Finlands folkmängd har 9-dubblats[2]) sedan år 1750, medan Sveriges befolkning blott 4-dubblats.

Efter detta »längdsnitt» belysande den historiska utvecklingen sedan år 1750 skola vi med varandra jämföra tvenne »tvärsnitt», nämligen de befolkningsstatistiska uppgifterna för Finland och Sverige för femårsperioden 1936—40. För att lättare kunna göra detta skola vi först redogöra för innebörden av vissa befolkningsstatistiska begreppsbildningar och undersökningsmetoder, som vi härvid få användning för.

Om begreppen stationär och stabil befolkning.

Då man vill få en möjligast klar bild av det sätt varpå äktenskapsfrekvensen, fruktsamheten, skilsmässofrekvensen, dödligheten samt

[1]) MODEEN och FOUGSTEDT, *Beräkningar om Finlands framtida befolkningsutveckling*, Helsingfors 1939. F. O. S. VI: 89.

[2]) Härvid bör dock beaktas att Finlands folkmängd ökades med 20 %, då det s.k. Gamla Finland (Karelen) den 11. 12. 1811 återförenades med det övriga Finland.

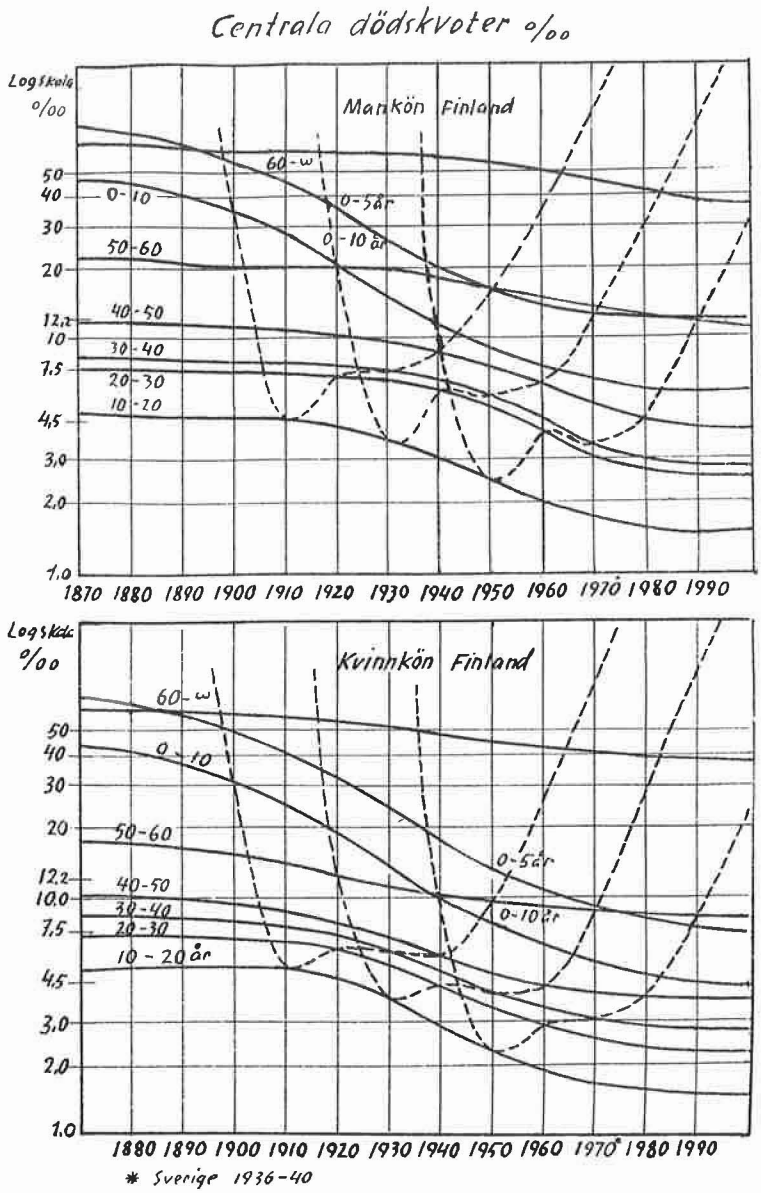


Diagram 2. Dödlighetens nedgång i Finland sedan år 1870. Motsvarande kurvor för Sverige fås närmelsevis genom att från årtalet subtrahera 30 år. Denna omständighet ha vi utnyttjat för att ställa en prognos för dödlighetens utveckling i Finland.

ut- och invandringen bestämma befolkningsutvecklingen, är det ändamålsenligt, att beräkna tal som belysa sådana hypotetiska befolkningar, vilkas statistiska egenskaper delvis äro desamma som den iakttagna befolkningens egenskaper i medeltal under en viss tidsperiod, delvis skilja sig från dessa genom att man ersatt en eller flere av de tidsserier som beskriva faktorer, vilka påverkat befolkningens struktur och egenskaper, med lämpligt valda konstanter. Av speciellt intresse äro de hypotetiska befolkningar vid vilka man ersatt de tidsserier, som angiva antalet levande födda barn under olika år och deras fördelning på kön, med konstanter, samt antager att någon utvandring eller invandring icke skulle förekomma, medan övriga befolkningsstatistiska faktorer, i den mån de äro förenliga med dessa antaganden, antagas stabilisera sig på den nivå de i medeltal hade under den studerade perioden. Om vi sålunda förutsätta t.ex., att det årligen födes 1.000 levande barn, varav 514 gossebarn och 486 flickebarn, samt antaga att dödlighets-, giftermåls- och skilsmässofrekvensen stabiliseras på den nivå de i medeltal hade åren 1936—1940 i Finland, kan man åtminstone approximativt beräkna huru stor befolkningen under dessa faktorers inverkan skulle bliva, fördelad enligt ålder, kön och civilstånd. Ifall vi ytterligare antaga att också de tal, som beskriva fruktsamheten i olika åldersklasser för gifta och ogifta kvinnor, skulle förbliva desamma, som i den betraktade faktiska befolkningen under den studerade perioden, skulle detta i regel resultera i, att i den sålunda beräknade *stationära* befolkningen skulle komma att födas (levande) ett annat barnantal, t.ex. 1.050, än det antal 1.000, som lagts till grund för beräkningen av den stationära befolkningens storlek. Om det sålunda erhållna barnantalet är större än 1.000, tyder detta på att befolkningen, för den händelse att de befolkningsstatistiska faktorerna skulle stabiliseras på den nivå de hade under den studerade perioden, skulle ha en tendens att på lång sikt växa. Om det beräknade antalet levande födda barn i den stationära befolkningen blir mindre än 1.000, t.ex. 960, tyder detta på att befolkningen på lång sikt har en tendens att minska; fruktsamheten är med andra ord för liten för att befolkningen skulle kunna förbliva stationär. Blott i det fall, då det beräknade antalet nyfödda barn (i enlighet med fruktsamhetsförhållandena i den faktiska befolkningen samt ålders-, köns-, och civilståndsfördelningen i den stationära befolkningen), överenstämmer med det antal (1.000), som man utgått ifrån vid beräkningen av den stationära befolkningen, befinna sig de befolkningsstatistiska grundfaktorerna i ett slags dynamisk jämvikt.

En intressant hypotetisk befolkning erhålles även, om man stabiliserar alla andra befolkningsstatistiska grundfaktorer förutom antalet årligen födda barn. Om vi t.ex. starta med en stationär befolkning, samt förutsätta att fruktsamhetstalen förbliva konstanta, kommer befolkningen i regel att antingen växa eller minska. Vi erhålla på detta sätt en befolkning, vars storlek och åldersfördelning förändras i tiden. Vid en matematisk analys av den utvecklingsprocess, som en dylik befolkning kommer att genomlöpa, visar det sig, att befolkningens procentuella fördelning, enligt ålder, kön och civilstånd med tiden kommer att asymtotiskt närma sig ett stabilt gränsläge. Detta stabila gränsläge skiljer sig desto mera från den stationära befolkningens motsvarande läge, ju mera fruktsamhetstalen skilja sig från dem, vilka skulle resultera i en stationär befolkning. En befolkning, som kan vara antingen växande eller avtagande sålunda att kvoten mellan befolkningens storlek under tvenne på varandra följande år samt befolkningens procentuella fördelning enligt ålder, kön och civilstånd förblir konstant och överenstämmande med det gränsläge, som betingas av de fruktsamhetstal som karakteriserar en faktisk befolkning under en viss tidsperiod, kallas för en *stabil* befolkning. I en stabil befolkning är befolkningens sammansättning enligt ålder och kön densamma, som skulle erhållas, om man diskonterar antalet personer i olika åldersklasser till födelseåret, med begagnande av den diskonteringsfaktor $v = (1/\ (1+p/100)$, som svarar mot befolkningens årliga tillväxtprocent p. Om den stabila befolkningen är avtagande, är denna tillväxtprocent p naturligtvis negativ. Dylika stabila befolkningars egenskaper ha studerats av Dublin och Lotka [1]). I det följande skola vi dock nöja oss med att studera de grundläggande demografiska faktorerna i Finland och Sverige med tillhjälp av beräkningar, som beskriva egenskaperna hos de stationära befolkningar, som svara mot befolkningsförhållandena i dessa länder åren 1936—1940. Det hade kanske varit önskvärt att utsträcka undersökningen till även en senare tidsperiod, men de uppgifter som stå till buds äro ej tillräckliga för att möjliggöra en lika fullständig analys som för perioden 1936—1940. Vi skola dock något beröra utvecklingen av födelsefrekvensen också efter år 1940.

En jämförelse mellan två länders eller två tidsperioders befolkningsstatistiska grundfaktorer belyses väl genom en jämförelse mellan de

[1]) DUBLIN och LOTKA: *On the true rate of natural increase.* Journal of the American Statistical Association Vol. 20 (1925) s. 305—399 och LOTKA: *Some recent results in population analysis,* ibid. Vol. 33 (1938) s. 164—178.

motsvarande stationära befolkningarna. En jämförelse mellan de stabila befolkningar, som svara mot de iakttagna befolkningarnas demografiska faktorer, skulle naturligtvis också ha sitt stora intresse. Beräkningar som gå ut på att beskriva dessa stabila befolkningar äro dock mycket arbetsdryga. De förutsätta i själva verket att det statistiska materialet först studerats med tillhjälp av de analysinstrument, som de stationära befolkningarna utgöra.

Om svårigheterna att i praktiken genomföra beräkningar, som gå ut på att fastställa den stationära befolkningens struktur, med utgångspunkt från statistiska uppgifter, som belysa dynamiska befolkningars egenskaper.

Då man gör ett försök att beräkna den stationära befolkning, som svarar mot de befolkningsstatistiska förhållandena i ett visst land under en viss period, visar det sig att man råkar ut för en mängd vanskligheter av såväl metodteknisk som principiell karaktär. Man blir ställd inför problemet om en stationär befolkning överhuvud kan beräknas på ett entydigt sätt med utgångspunkt från de befolkningsstatistiska data, som beskriva dynamiska befolkningars egenskaper, d.v.s. befolkningar med ofta en väsentligt annan fördelning enligt ålder, kön och civilstånd, än den som är kännetecknande för den stationära befolkningen (se diagram 3.). Vid beräkningarna inträffar det att de tal, som erhållas ur de grundläggande antagandena såsom närmevärden för tal som logiskt sett borde vara lika stora, t.ex. antalet gifta män och antalet gifta kvinnor, avvika systematiskt från varandra. En eliminering av dessa logiska skönhetsfel skulle förutsätta, att de grundläggande antagandena, varpå beräkningarna grunda sig, justeras så, att utgångstalen komma att avvika från de tal, som beskriva den faktiska dynamiska befolkningen under den ifrågavarande perioden. En dylik justering skulle åter förutsätta precisa kunskaper om det sätt, varpå de dynamiska element, som åstadkommit att den faktiska befolkningens fördelning enligt ålder, kön och civilstånd avviker från den stationära befolkningens motsvarande tal, påverkat de data, som valts till utgångspunkt för beräkningarna. Man måste därför i praktiken nöja sig med mer eller mindre grova approximationer, när man vill numeriskt bestämma de tal, som beskriva den stationära befolkning, som motsvarar de befolkningsstatistiska grundfaktorerna i ett visst land under en viss period. Trots detta giva dylika beräkningar en intressant bild av sambandet mellan dessa faktorer, befolkningsför-

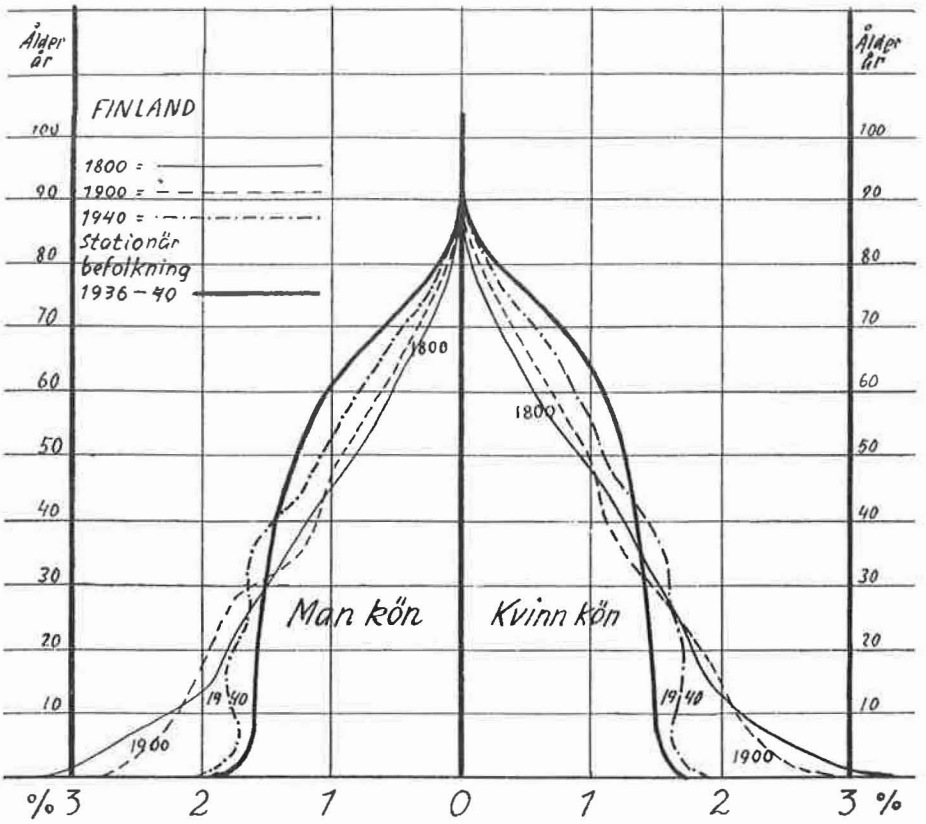


Diagram 3. Befolkningens i Finland procentuella sammansättning enligt ålder åren 1800, 1900 och 1940 jämförd med åldersfördelningen i 1936—40 års stationära befolkning.

hållandena och den sannolika utvecklingen i de länder man vill jämföra med varandra. Också de komplikationer beräkningarna möta äro ägnade att belysa samspelet mellan de faktorer, som styra befolkningsutvecklingen. Närmare detaljer angående den räkneteknik vi använt framgå av bilagan: Teorin för befolkningsstatistiska kalkyler.

En redogörelse för de erhållna närmevärdena till de tal, som beskriva den stationära befolkningen i Finland och Sverige i enlighet med förhållandena åren 1936—1940.

Vid våra beräkningar ha vi utgått från de tabeller, som angiva antalet kvarlevande av 100.000 levande födda gossar och av 100.000 levande

födda flickor enligt dödlighetsförhållandena åren 1936—1940. Under denna period fördelade sig de levande födda barnen i Finland och även i Sverige sålunda, att på 1.000 levande födda barn kom 514 gossar och 486 flickor. Genom att multiplicera antalet kvarlevande gossar av 100.000 födda med 0,00514 och antalet kvarlevande flickor av 100.000 födda med 0,00486 fås antalet kvarlevande gossebarn och flickebarn av 1.000 födda barn, vid de tidpunkter då de fylla t.ex. 5, 10, 15, 20, o.s.v. år. För att inbespara räknearbete och för att de erhållna tabellerna icke skulle bliva svåröverskådligt stora, ha vi nöjt oss med att såsom grundläggande tidsenhet begagna 5 år. Antalet kvarlevande gossar och flickor av 1.000 födda barn vid varje heltalig multipel av fem år ha vi angivit i tabell 1 (s. 175). Såsom vi längre fram komma att inse, äro icke ens dessa tal fullt tillförlitliga uttryck för antalet kvarlevande gossar och flickor i en stationär befolkning påverkad av de dödsrisker, som karakteriserade åren 1936—1940. De överlevelsesannolikheter, som ligga till grund för beräkningarna av de i denna tabell ingående talen, påverkas nämligen av civilståndsfördelningen för den faktiska befolkningen åren 1936—1940. Denna faktiska fördelning avviker från civilståndsfördelningen i den stationära befolkning, som har egenskapen att äktenskapsfrekvensen, skilsmässofrekvensen och dödsriskerna överensstämma med medelvärdet för åren 1936—1940. En ökning av äktenskapsfrekvensen torde på grund av gifta personers relativt lägre dödsrisker, även om detta till stor del utgör ett resultat av att giftermålssannolikheten stiger med hälsotillståndet, ha en tendens att öka antalet kvarlevande i åldersklasserna över 30 år. Talen i tabell 1 böra därför uppfattas endast som en första approximation till den tabell, som angiver antalet kvarlevande i den motsvarande stationära befolkningen. Vid ett studium av tabell 1 fäster man sig främst vid verkningarna av den relativt sett mycket höga barnadödligheten i Finland.

Av tabell 2 framgår på vilket sätt de stationära befolkningar, som svara mot 1936—1940 års officiella överlevelsetabeller i Finland och Sverige, fördela sig enligt kön och enligt ålder. Dessa tal angiva strängt taget det sammanlagda antal år, som de överlevande av 514 samtidigt födda gossar och 486 flickor kunna väntas leva i de angivna femårsåldersklasserna, för den händelse dödlighetsförhållandena i olika åldersklasser under hela deras livstid förbleve desamma som under den studerade perioden. De kunna dock även tolkas såsom angivande antalet personer, som kunna väntas befinna sig i respektive femårs åldersklass vid en på må få vald tidpunkt i den stationära befolkning,

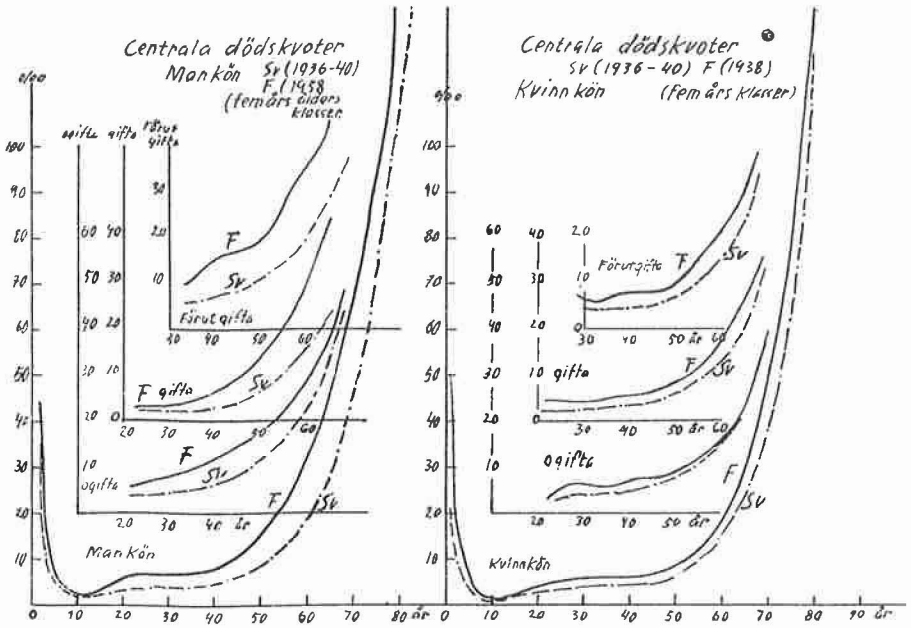
som skulle uppkomma, om det årligen skulle födas 514 gossar och 486 flickor, vilka vore utsatta för dödsrisker av den styrka överlevelsetabellerna beskriva.

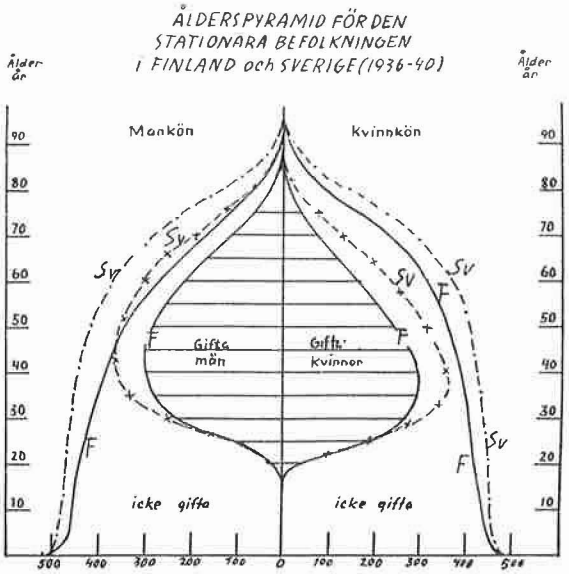
Tabell 3 innehåller resultaten av vårt försök att uppskatta den närvarande medelbefolkningen i Finland åren 1936—40 med fördelning enligt civilstånd, jämte en jämförelse mellan det relativa antalet förutgifta av summan av gifta och förutgifta i Finland 1936—40 och i Sverige 31/12 1940. Tabell 3 bildar grundvalen för beräkningen av dödsfallsfrekvenstalen enligt civilstånd (se tabell 4.1 och 4.2). Genom att till de ogiftas dödsfallsfrekvenser addera giftermålsfrekvenstalen fås de ogiftas avgångsfrekvenstal (se tabell 4.3), som möjliggöra en beräkning av antalet kvarvarande ogifta vid olika åldersår (tabell 5.1). Genom att utnyttja en tabell över antalet bestående äktenskap i Norge år 1900 med fördelning enligt hustruns och mannens ålder (tabell 10.8) blev det möjligt att uppskatta frekvenstalen för övergången mellan det gifta och förutgifta ståndet. Sedan vi på detta sätt kompletterat frekvenstalen för civilståndsväxlingarna kunde vi på i bilagan om teorin för befolkningsstatistiska kalkyler närmare beskrivet sätt beräkna den första approximationen till den mot 1936—40 års förhållanden svarande stationära befolkningen (tabellerna 5 och 6). Vid ett studium av resultaten framgår att antalet genomlevda år såsom gifta blev signifikativt större för männen än för kvinnorna. För att undersöka möjligheterna att eliminera detta logiska skönhetsfel från närmevärdena för den stationära befolkningens sammansättning enligt ålder, kön och civilstånd, ha vi gjort ett försök att kalkylera antalet ingångna äktenskap och antalet såsom gifta genomlevda år fördelade enligt mannens och hustruns ålder (tabellerna 10.1—10.9). På detta sätt kan man undgå de logiska självmotsägelser, som erhållas i den stationära befolkningens sammansättning enligt kön och civilstånd. De erhållna resultaten (approximation III) äro vi dock icke helt nöjda med, emedan de beräknats på grundval av en arbetshypotes (formel 19.2), som vi numera skulle vara böjda att ersätta med en annan (19.2'). Vi antogo nämligen, att de frekvenstal förbliva konstanta, som angiva antalet giftermål i en viss åldersklasskombination, per 10.000 genomlevda år för det kön, som har det mindre antalet till äktenskap lediga kandidater (de förutgifta »omräknades» härvid till ogifta med användning av en omräkningsfaktor, som tog hänsyn till de förutgiftas relativt sett större giftermålsfrekvens). Resultaten enligt denna metod giva en undre gräns för antalet ingångna giftermål i den stationära befolkningen. Detta antagande synes likväl lägga för stor vikt vid »minimi-

faktorn» bland de till äktenskap lediga i den ifrågavarande ålderskombinationen. Ett bättre antagande torde vara att utgå ifrån konstans hos kvoten mellan antalet parkombinationer, som kunde erhållas, om de personer, vilka per tidsenhet ingå giftermål i den ifrågavarande åldersklassen (x, y), skulle byta partner på alla möjliga sätt,och totalantalet parkombinationer, som kunde erhållas genom att på olika möjliga sätt sammankoppla äktenskapskandidaterna. Detta antagande skulle i praktiken leda till att giftermålen i den ifrågavarande ålderskombinationen skulle sättas i relation till det harmoniska medeltalet av antalet genomlevda år för de möjliga äktenskapskandidaterna. Den för denna undersökning anslagna tiden och utrymmet medgav likväl icke en beräkning av den stationära befolkningens i Finland och Sverige fördelning enligt civilstånd med utgångspunkt från detta grundläggande antagande. Vi måste nöja oss med att justera den första approximationens resultat, så att antalet genomlevda år för de gifta männen och kvinnorna blir lika stort genom en procentuell förhöjning av de gifta och förutgifta kvinnornas antal i Finland med 2.3 %, samt minska de ogiftas antal så mycket, att totalantalet genomlevda år för kvinnorna i de olika femårsklasserna förblir detsamma som enligt den första approximationen. De i den I approximationen för Finland erhållna siffrorna för antalet genomlevda år för de gifta och förutgifta männen ha vi åter minskat med 2.8 %, varigenom totalantalet genomlevda år för männen och kvinnorna blir lika stort, som enligt den tredje approximationen för Finland (11.500 år). »Ex analogia» ha vi höjt siffrorna för Sverige för de gifta och förutgifta kvinnornas del med 3 % och sänkt de motsvarande siffrorna för männen med 3.1 %. Den sålunda erhållna II approximationen för den stationära befolkningen i Finland och Sverige (tabell 8) kunde i sin tur begagnas för att beräkna en tredje, fjärde o.s.v. approximation med tillhjälp t.ex. av det senast diskuterade antagandet angående giftermålsfrekvensen i olika åldersklasskombinationer av män och kvinnor. Ett fullföljande av uppslaget får dock anstå till ett annat tillfälle.

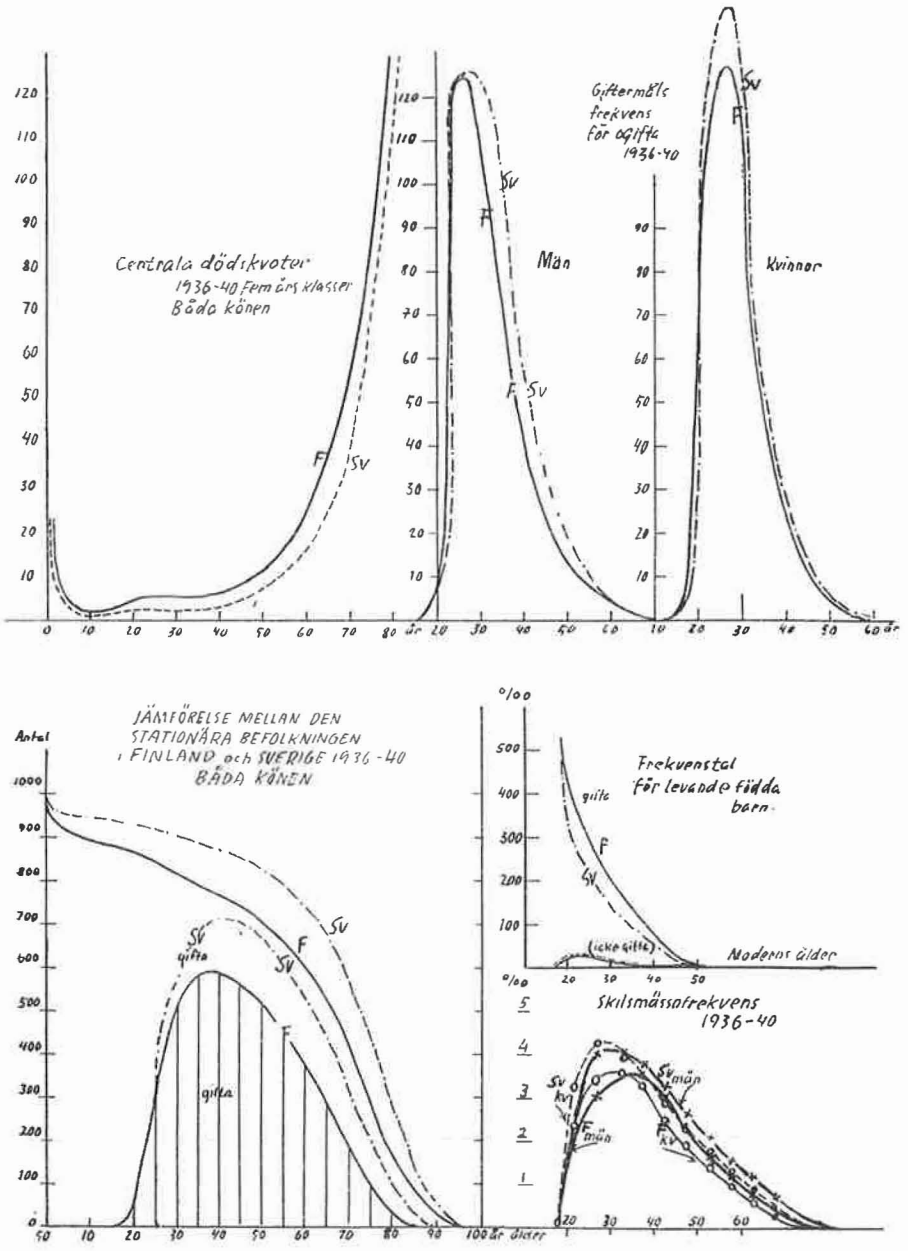
Diagramsammanställningarna 4, 5 och 6, vilka bygga på tabellerna (4, 5, 6 och 9) åskådliggöra de viktigaste resultaten av undersökningen.

Med utgångspunkt från den i tabell 8 angivna II approximationen till den stationära befolkningen i Finland och Sverige, som inom en felgräns av storleksordningen (2—3) % torde giva en korrekt bild av den stationära befolkning, som svarar mot dödlighets-, giftermåls-, och skilsmässofrekvenstalen för Finland och Sverige åren 1936—40, ha vi försökt beräkna antalet barn, som i enlighet med födelsefrekvenstalen

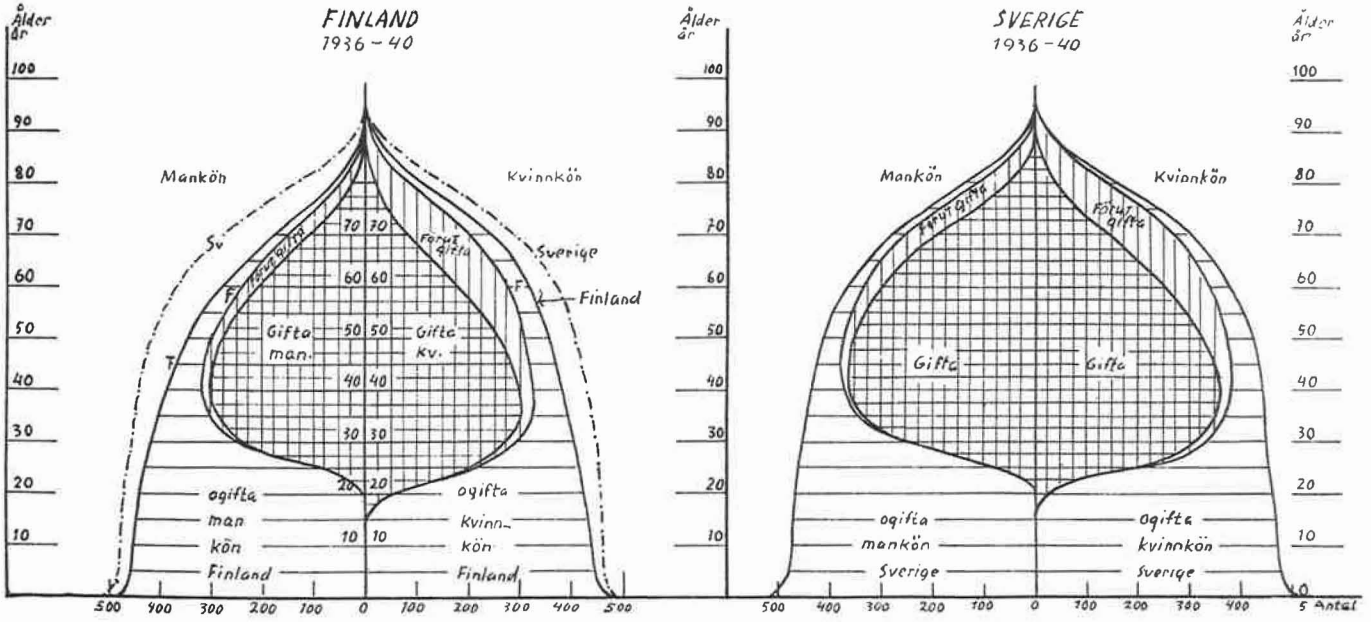




Diagramsammanställning 4.



Diagramsammanställning 5.



Diagramsammanställning 6. Den stationära befolkningens i Finland och Sverige 1936—40 fördelning enligt ålder och civilstånd.

för Finland och Sverige åren 1936—40 och år 1944 kunde väntas bliva födda levande i den stationära befolkningen. De erhållna resultaten för åren 1936—40 framgå av tabell I (s. 156).

Om vi skulle ha använt oss av den första approximationen i stället för den andra skulle vi ha erhållit »nettoreproduktionstalet» 1.066 för Finland och 871 för Sverige. Enligt den tredje approximationen för antalet genomlevda år, en uppskattning som dock torde underskatta antalet gifta kvinnor i de yngre åldersklasserna, erhölls enligt tabell I ett nettoreproduktionstal för Finland på 1.053. Resultaten för Sverige utvisa en tendens till *folkminskning* (med en hastighet på 3 à 4 promille per år) och resultatet för Finland på en svag tendens till folkökning (med en hastighet på c:a 2 $^0/_{00}$ per år). Bilden förändras för Sveriges del väsentligt, om vi i stället för födelsefrekvenserna för åren 1936—40 begagna födelsefrekvenstalen för år 1944. För år 1944 ha vi erhållit de i tabell II angivna räkneresultaten.

År 1944 voro således de specifika födelsetalen såväl i Sverige som Finland så stora, att den stationära befolkningen av åren 1936—40 skulle ha fått en tendens att övergå i en växande befolkning med en tillväxthastighet på 4—5 $^0/_{00}$ per år, ifall den skulle hava samma specifika födelsefrekvenstal, som den faktiska befolkningen hade år 1944. Härvid är att observera, att vi vid denna kalkyl icke räknat med någon fortsatt nedgång av dödligheten, en faktor som kan väntas spela en större roll för befolkningens utveckling under de närmaste årtiondena än nettoreproduktionstalets avvikelser från jämviktsläget 1.000 (se diagram 2, s. 143).

Den ovan använda metoden för beräkning av nettoreproduktionstalet har den teoretiska svagheten, att den jämförelsevis känsligt inregistrerar icke enbart variationerna i fruktsamheten utan även korttidsfluktuationerna i äktenskapsfrekvensen, då ju fruktsamheten normalt är väsentligt större i början av ett äktenskap än för jämnåriga gifta kvinnor i äktenskap, som varat en längre tid. I detta avseende fås ett bättre resultat genom att sätta födelserna i relation till det antal levnadsår, som fås, då man omräknar de icke giftas och nygiftas levnadsår till ett slags normalår för gifta kvinnor (se formlerna 20. 3,20.4 och 20.5 i bilaga I).

Resultatet av en beräkning av nettoreproduktionstalet enligt denna metod för åren 1936—40 framgår av tabell III.

Talen för Finland för åren 1936—40 tyda sålunda även enligt denna beräkningsmetod på en svag tendens till en naturlig folkökning (med

ca 1.5‰ per år), medan nettoreproduktionstalet för Sverige likaledes minskar något och angiver en tendens för en naturlig folkminskning på c:a 4.6‰ per år. År 1944 hade såväl Finland som Sverige en befolkning, vars födelsetal var tillräckligt stor för att folkmängden från en generation till den närmast följande (d.v.s. efter förloppet av en tidrymd motsvarande det geometriska medeltalet av barnaföderskornas ålder, d.v.s. för Finland 29.6 år och för Sverige 29.0 år) skulle komma att växa för Finland med c:a 12 % och i Sverige med c:a 9 %, d.v.s. med omkr. 2.2 respektive 2. 9‰ per år. Sedan år 1944 har nativiteten i Finland fortsatt att stiga,[1]) medan nativiteten i Sverige obetydligt nedgått.

Det kan kanske ännu ha sitt intresse att påpeka, att under de åren 1936—40 rådande förhållandena skulle en 15 års flicka i Finland ha en sannolikhet på omkring (355/432) = 0.82 för att bliva gift åtminstone en gång medan motsvarande sannolikhet för en 15 års flicka i Sverige kan uppskattas till (398/458) = 0.87. För gossarna äro motsvarande tal för Finland 356/447 = 0.80 och för Sverige 409/477 = 0.86.

De olika civilståndens och livets varaktighet i den stationära befolkningen 1936—40 framgår av tabell IV.

Våra beräkningar giva sålunda en något längre medellivslängd vid födelsen än de officiella beräkningarna, vilka icke beakta den omständigheten, att de giftas dödlighetsfrekvenser äro väsentligt lägre än de icke giftas. Förklaringen till att vi erhållit högre medellivslängd torde i främsta rummet bero på att äktenskapsfrekvensen åren 1936—40 var större än i medeltal under den nulevande generationens livstid. I tabellerna 9.1 och 9.2 ha vi jämfört mera i detalj resultaten enligt vår beräkningsmetod och resultaten enligt de officiella beräkningarna. Avvikelserna mellan resultaten äro jämförelsevis små. De avvikelser, som för Finland kunna observeras i de högre åldersklasserna, torde i huvudsak bero på bristfälligheter i utgångsmaterialet. Av denna orsak ha vi nämligen uppskattat dödsfallsfrekvenserna för de ogifta och förutgifta, samt för åldrarna över 80 år även för de gifta med stöd av de motsvarande siffrorna för Sverige och utvecklingstendensen i de yngre åldersklasserna. För åldersklasserna över 80 år ha vi, emedan officiella beräkningar icke publicerats, beräknat dödsfallsfrekvenserna för de olika civilstånden med utgångspunkt från de publicerade siffrorna över antalet dödsfall och medelbefolkning enligt civilstånd åren 1930—40 i Sverige. (S.O.S. Befolkningsrörelsen 1930—40).)

[1]) Se t.ex. Statistiska Översikter, 1947, 11—12

Tabell I.

Beräkning av antalet levande födda barn i 1936—40 års stationära befolkning i Finland och Sverige enligt födelsefrekvenstalen för samma period.

Första metoden.

Moderns ålder	II approximationen 1936—1940														III approximationen				
	Finland							Sverige							Finland. 1936—40				
																	Levande födda		
	Levande födda på 1 000 levnadsår för gifta kvinnor	Levnadsår för gifta kvinnor i stationär bef.	Inom äktenskap levande födda	Levande födda på 1 000 levnadsår för icke gifta kvinnor.	Levnadsår för icke gifta kv. i stationär bef.	Levande födda utom äktenskap	Levande födda inalles	Levande födda på 1 000 levnadsår för gifta kvinnor	Levnadsår för gifta kvinnor i stationär bef.	Inom äktenskap födas	Levande födda på 1 000 levnadsår för icke gifta kvinnor.	Levnadsår för icke gifta kv. i stationär bef.	Levande födas utom äktenskap	Levande födda inalles.	Levnadsår för gifta kvinnor i station. bef.	Levnadsår för icke gifta kv. i stationär befolkning	Inom äktenskap	Utom äktenskap	Inalles
15—20	482.4	50	24	4.3	2 089	9.0	33	511.1	44	23	11.95	2 238	26.7	49	49	2 090	24	9.0	33
20—25	332.7	573	191	17.3	1 521	26.3	217	257.7	596	153	24.1	1 663	40.1	243	533	1 561	177	27.0	204
25—30	235.0	1 274	299	18.7	772	14.4	314	167.2	1 438	240	17.0	788	13.4	254	1 200	846	282	15.8	298
30—35	158.4	1 515	240	14.3	476	6.8	247	116.0	1 778	206	12.6	418	5.3	212	1 492	499	236	7.1	243
35—40	114.6	1 522	174	10.8	429	4.6	179	71.7	1 818	130	8.5	348	3.0	133	1 514	437	174	4.7	179
40—45	57.4	1 468	84	5.1	444	2.3	87	28.2	1 778	50	3.8	351	1.3	51	1 465	447	84	2.3	86
45—50	7.6	1 344	10.3	0.6	512	0.3	11	3.2	1 678	6	0.4	392	0.1	6	1 352	504	10	0.3	10
15—50	126.6	7 746	1 022	10.6	6 243	66.0	1 088	88.6	9 130	808	12.9	6 198	90	898	7 605	6 386	987	66.2	1 053

Tabell II.

Beräkning av antalet levande födda barn enligt II approximationens stationära befolkning för 1936—40 och 1944 års födelsefrekvenstal. (Första metoden).

Moderns ålder	Finland							Sverige						
	Levande födda på 1 000 levnadsår för gifta kvinnor	Levnadsår för gifta kvinnor i station. bef.	Inom äktenskap levande födda	Levande födda på 1 000 levnadsår för icke gifta kv.	Levnadsår för icke gifta kvinnor i stationär befolkn.	Levande födda utom äktenskap	Levande födda inalles	Levande födda på 1 000 levnadsår för gifta kvinnor	Levnadsår för gifta kvinnor i stationär bef.	Inom äktenskap födda	Levande födda på 1 000 levnadsår för icke gifta kv.	Levnadsår för icke gifta kv. i stationär bef.	Levande födas utom äktenskap	Levande födda inalles
15—20	392	50	20	3.5	2 089	7.3	27	587.7	44	26	14.7	2 238	32.9	59
20—25	303	573	174	18.4	1 521	28.0	202	302.0	596	180	28.6	1 663	47.5	228
25—30	228	1 274	290	20.4	772	15.7	306	209.5	1 438	301	21.4	788	16.9	318
30—35	179	1 515	271	16.1	476	7.7	279	151.7	1 778	270	10.1	418	4.2	274
35—40	124	1 522	189	11.0	429	4.7	194	92.5	1 818	168	9.6	348	3.3	171
40—45	60	1 468	89	4.2	444	1.9	91	32.7	1 778	58	3.2	351	1.1	59
45—50	8	1 344	11	0.7	512	0.4	11	2.6	1 678	5	0.4	392	0.1	5
15—50	135	7 746	1 044	10.5	6 243	65.7	1 110	117.9	9 130	1 008	15.5	6 198	96.0	1 104

Tabell III.

Kalkyl över antalet barn, som kan beräknas bliva födda i en stationär befolkning av den storlek och sammansättning som anges av II approximationen, Tabell 8, bilaga II. (Andra beräkningsmetoden).

Moderns ålder	Levnadsår för gifta kvinnor	5 % av levnadsåren för icke gifta kvinnor	3 ggr antalet ingångna äktenskap	Normala äktenskapsår för kvinnor	Levande födda per 1 000 normalår (1936—40)	Levande födda (1936—40) i stationär befolkning	1944		Levnadsår för gifta kvinnor	5 % av levnadsåren för icke gifta kvinnor	3 ggr antalet ingångna äktenskap	Normala äktenskapsår för kvinnor	Levnade födda per 1 000 normalår (1936—40)	Levande födda (1936—40) i stationär befolkning	1944	
							Levande födda per tusen normalår	Levande födda i stationär befolkning							Levande födda per tusen normalår	Levande födda i stationär befolkning av åren (1936—40)
	L(g)	0.05 x L(o+f)	3N(→g)	L(→o)	λ(→o)	N(→o)	λ(→o)	N(→o)	L(g)	0.05 x L(o+f)	3N(→g)	L(→o)	λ(→o)	N(o)	λ(→o)	N(→o)
	Finland								Sverige[1])							
15—20	50	104	129	283	115.6	32.7	86.5	24.5	44	112	108	264	175.3	46.3	190.4	50.3
20—25	573	76	510	1 159	172.1	198.3	191.7	222.2	596	83	482	1 261	153.8	193.8	177.2	223.4
25—30	1 274	38	312	1 624	186.8	303.4	174.7	283.7	1 438	39	366	1 843	128.4	236.8	170.0	313.3
30—35	1 515	24	105	1 644	145.5	239.0	156.6	257.4	1 778	21	131	1 930	106.4	203.4	140.7	271.6
35—40	1 522	21	60	1 603	110.6	177.3	112.0	179.5	1 818	17	55	1 890	70.3	132.8	89.6	169.3
40—45	1 468	22	33	1 523	57.0	86.8	56.7	86.4	1 778	18	30	1 825	29.3	53.5	32.5	53.5
45—50	1 344	26	21	1 391	7.0	9.7	8.7	11.1	1 678	20	16	1 713	3.4	5.8	2.6	4.5
15—50	7 746	311	1 170	9 227	113.7	1 047	115.4	1 064.8	9 130	310	1 088	12 428	70.0	874	87.0	1 086

[1]) Beträffande de officiella beräkningarna över nettoreproduktionstalets storlek i Sverige 1936—40 och 1944 se t.ex. S. O. S. Befolkningsrörelsen 1944. De giva för åren 1936—40 ett lägre nettoreproduktionstal (0,796) och för år 1944 ett högre nettoreproduktionstal (1.140) än det av oss beräknade. De officiella beräkningarna utföras enligt en annan metod än de av oss begagnade.

Tabell IV.

De olika civilståndens och livets varaktighet i den stationära befolkningen i Finland och Sverige (1936—40) enl. II. approximationen.

	Mankön		Kvinnkön		Båda könen	
	Finland	Sverige	Finland	Sverige	Finland	Sverige
	Varaktighet i år					
Såsom ogift leves i medeltal efter födelsen	29.3	31.5	27.2	27.6	28.3	29.6
Äktenskapens varaktighet	29.0	33.8	29.0	33.8	29.0	33.8
De förutgifta leva i medeltal såsom förut-gifta	9.2	10.4	16.5	15.8	13.8	13.5
Medellivslängd enl. våra beräkningar	54.57	64.53	60.28	66.86	57.25	65.66
Medellivslängd enl. officiella beräkningar	54.31	64.30	59.48	66.91	56.83	65.57

I övrigt få vi hänvisa till tabellbilagans mera detaljerade uppgifter samt med avseende å beräkningsmetoderna till den teoretiska bilagan.

Bilaga I. *Teorin för befolkningsstatistiska kalkyler.*[1])

Befolkningsstatistiska kalkyler utnyttja i stor utsträckning räkningar med frekvenstal. Med ett frekvenstal $\lambda\ (a_i \rightarrow a_j)$ förstå vi kvoten mellan antalet överflyttningar $N\ (a_i \rightarrow a_j)$ av personer från klassen a_i till klassen a_j och den sammanlagda tid $L\ (a_i)$, som de personer, vilka befunnit sig i klassen a_i genomlevt under den betraktade tidsintervallen: $(x < t < x+h)$. Vid våra kalkyler ha vi i regel mätt tiden i år och valt $h = 5$ år.

[1]) En mängd formler, som kunna användas för befolkningsstatistiska kalkyler angivas t.ex. i L. v. BORTKIEWITCZ: *Die Sterbeziffer und der Frauenüberschuss in der stationären und in der progressiven Bevölkerung.* Bulletin de l' Institute International de Statistique, Vol. 19, och A. J. LOTKA: *Théorie analytique des associations biologique,* Deuxième Partie — Analyse démographique. Hermann et Cie. Paris (1939). I tidskrifterna Population, Population Studies, Human Biology, Journal of American Statistical Association, Skandinavisk Aktuarietidskrift, Statsvetenskaplig tidskrift, m.fl. tidskrifter ingå artiklar, i vilka finnas fomler för befolkningsstatistiska kalkyler. Den följande framställningen har utarbetats helt självständigt. Den tager endast sikte på att giva det teoretiska underlaget för de vid beräkningen av de i tabellbilagan angivna resultaten utförda kalkylerna.

$$\lambda(a \rightarrow a_j) = \frac{N(a_i \rightarrow a_j)}{L(a_i)}. \tag{1}$$

Vid tabelleringen uttryckes frekvenstalen $\lambda(a_i \rightarrow a_j)$ oftast i promille, d.v.s. man antecknar i tabellen talen $10^3 \cdot \lambda(a_i \rightarrow a_j)$.

Av (1) följer

$$N(a_i \rightarrow a_j) = \lambda(a_i \rightarrow a_j) L(a_i). \tag{2}$$

Om klasserna a_j; $j = 1, \ldots n$, äro varandra inbördes uteslutande, d.v.s. om ingen person samtidigt kan tillhöra flere än en klass är:

$$\lambda\left(a_i \rightarrow \sum_{j=1}^{n} a_j\right) = \sum_{j=1}^{n} \lambda(a_i \rightarrow a_j). \tag{3}$$

I detta fall är nämligen

$$N\left(a_i \rightarrow \sum_{j=1}^{n} a_j\right) = \sum_{j=1}^{n} N(a_i \rightarrow a_j) = \sum_{j=1}^{n} \lambda(a_i \rightarrow a_j) L(a_i). \tag{4}$$

Ur (4) fås genom division med den för termerna i vänstra membrum gemensamma faktorn $L(a_i)$ räkneregeln (3). Formeln (3) kan utnyttjas speciellt för beräkning av totala avgångsfrekvensen $\lambda(a_i \rightarrow)$ från klassen a_i då delfrekvenstalen $\lambda(a_i \rightarrow a_j)$ för alla möjliga klasser a_j äro givna

$$\lambda(a_i \rightarrow) = \sum_{j \neq i} \lambda(a_i \rightarrow a_j). \tag{5,1}$$

Då $\lambda(a_i \rightarrow)$ och $L(a_i)$ äro givna, fås $N(a_i \rightarrow)$ medels formeln

$$N(a_i \rightarrow) = \lambda(a_i \rightarrow) L(a_i). \tag{5,2}$$

Antalet inträden till klassen a_i kan åter beräknas medelst formeln

$$N(\rightarrow a_i) = \sum_j N(a_j \rightarrow a_i) = \sum \lambda(a_j \rightarrow a_i) L(a_j) \tag{5,3}$$

som övergår i

$$N(\rightarrow a_i) = \lambda(\rightarrow a_i) L\left(\sum_{j \neq i} a_j\right), \tag{6}$$

om vi definiera

$$\lambda(\rightarrow a_i) = \sum_j \lambda(a_j \rightarrow a_i) \cdot \frac{L(a_j)}{L\left(\sum_{j \neq i} a_j\right)}, \tag{7}$$

$$L\left(\sum_{j\neq i} a_j\right) = \sum_{j\neq i} L(a_j). \tag{8}$$

Den totala ingångsfrekvensen $\lambda(\rightarrow a_i)$ är således ett vägt medeltal av övergångsfrekvenserna $\lambda(a_j \rightarrow a_i)$ för $j \neq i$.

Den motsvarande genomlevda tiden $L\left(\sum_{i\neq j} a_j\right)$ lyder däremot en additionsregel av samma karaktär (jfr (3)), som de »utåtriktade» frekvenstalen.

För den händelse frekvenstalen $\lambda(a_j \rightarrow a_i)$ äro av väsentligt olika storleksordning kan man uppnå bättre stabilitet i tiden för frekvenstalet $\lambda(\rightarrow a_i)$ genom att använda en annan definition för $\lambda(\rightarrow a_i)$ än (7), d.v.s. en definition av typen:

$$\lambda(\rightarrow a_i) = \frac{N(\rightarrow a_i)}{L(\rightarrow a_i)}. \tag{5,1}$$

där

$$L(\rightarrow a_i) = \sum_{j\neq i} k_j L(a_j)\,; \qquad k_j \approx \frac{\lambda(a_j \rightarrow a_i)}{\lambda(a_s \rightarrow a_i)}. \tag{8}$$

Talen k_j kunna lämpligen väljas sålunda, att $k_j = 1$ för den klass $j = s$ för vilken $(N(a_j \rightarrow a_i))$ i medeltal är störst, medan övriga k_j väljas så, att variansen i trendavvikelserna för den tidsserie, som fås, då $\lambda(\rightarrow a_i)$ bestämmes under en hel serie av undersökningsperioder, blir ett minimum.

Då man vill konstruera den befolkning, som skulle uppkomma ur en viss utgångsbefolkning, under förutsättningen att växlingsfrekvenstalen $\lambda(a_i \rightarrow a_j)$ för samtliga klasser äro givna, behöver man ännu formler, som angiva sambandet mellan talen $L(a_i)$ och talen $N(a_i \rightarrow a_j)$. Vi beteckna med $N_x(a_i)$ antalet personer som överhuvud kunde övergå från klassen a_i efter tidpunkten $t = x$ för den händelse $N(\rightarrow a_i)$ vore $= 0$, d.v.s. om något inflöde ej skulle ske efter denna tidpunkt. Då $N(\rightarrow a_i) \neq 0$, förändras $N(a_i)$ från periodens början till dess slut med

$$\triangle N(a_i) = N_{x+h}(a_i) - N_x(a_i) = N(\rightarrow a_i) - N(a_i \rightarrow)\,. \tag{9}$$

Formeln (9) kunde vi kalla för »lagertillväxtformeln». Om »lagerbehållningen» vid tidpunkten $t = x$ är bekant $= N_x(a_i)$ kan man beräkna $N_{x+h}(a_i)$ medelst formeln

$$N_{x+h}(a_i) = N_x(a_i) + \triangle N(a_i)\,. \tag{10}$$

(Obs. Vi ha i regel utelämnat tidsintervallbeteckningen $\overline{(x,\ x+h)}$

från talen $N\,(a_i)$, $N\,(a_i \rightarrow a_j)$ och $\lambda\,(a_i \rightarrow a_j)$ för att icke i onödan överbelasta symbolerna. I de fall, då tidsattributen växla, tillfoga vi deras symboler som undre tecken till de motsvarande symbolerna).

Tillväxten i lagerbehållningen under den närmast föregående perioden $(x-h,\ x)$ beteckna vi medelst en upp och nedvänd liten triangel framför den symbol, som betecknar lagervolymen vid tidpunkten x. Vi ha användning speciellt för tillväxten:

$$(11) \qquad \bigtriangledown N\,(a_i) = N_x\,(a_i) - N_{x-h}\,(a_i) = \bigtriangleup N_{x-h}\,(a_i)\,.$$

Sambandet mellan talen $N_{x+\tau}\,(a_i)$ och $L\,(a_i)$ är matematiskt sett rätt komplicerat i det, att vi måste begagna ett integraltecken för att angiva sambandet exakt:

$$(12) \qquad L\,(a_i) = \int_{\tau=0}^{h} N_{x+\tau}\,(a_i)\,d\,\tau = \sum_{\nu} \tau_\nu\,(x, x+h, a_i)\,.$$

I den sist angivna formeln betecknar $\tau_\nu\,(x, x+h, a_i)$ längden av den tidsintervall, under vilken individen nummer ν tillhört klassen a_i under perioden $(x,\ x+h)$. Om antalet individer är stort, kan man jämförelsevis väl approximera $N_{x+\tau}\,(a_i)$ inom intervallen $(x,\ x+h)$ medelst ett polynom av andra graden av formen:

$$(13) \quad N_{x+\tau}\,(a_i) \approx N_x(a_i) + \tfrac{1}{2}(\bigtriangleup N\,(a_i) + \bigtriangledown N(a_i))\,\frac{\tau}{h} + \tfrac{1}{2}\,(\bigtriangleup N\,(a_i) - \bigtriangledown N\,(a_i))\,\frac{\tau^2}{h^2}\,.$$

(Regeln för interpolation med tillhjälp av andra grads polynom).

Man konstaterar lätt med tillhjälp av (9) och (11), att interpolationsformeln (13) stämmer exakt för $\tau = -h$, 0 och h. Då uttrycket (13) införes i (12) och integrationen utföres fås

$$(14) \quad L\,(a_i) = \int_{\tau=0}^{h} N_{x+\tau}\,(a_i)\,d\,\tau \approx h\,\{N_x\,(a_i) + \tfrac{1}{12}(\bigtriangleup N\,(a_i) + \bigtriangledown N\,(a_i)) + \\ + \tfrac{1}{6}\,(\bigtriangleup N\,(a_i) - \bigtriangledown N\,(a_i))\}$$

d.v.s.

$$(15) \qquad L\,(a_i) \approx h\,(N_x\,(a_i) + \tfrac{1}{12}\bigtriangledown N\,(a_i)) + \tfrac{5}{12}\,h \bigtriangleup N\,(a_i)$$

som då $h = \frac{10}{2} = 5$ kan skrivas i den för praktiska beräkningar bekväma formen:

$$(15;\, h=5) \qquad L\,(a_i) \approx \tfrac{1}{24}\,(12\,N_x\,(a_i) + \bigtriangledown N\,(a_i) + 5 \bigtriangleup N\,(a_i))\,.$$

Av (15; $h = 5$) följer, att ett relativt gott närmevärde för tillväxten $\bigtriangledown N(a_i)$ i »lagerbeståndet» $N(a_i)$, (då $L(a_i)$, $N_x(a_i)$ och $\bigtriangledown N(a_i)$ äro givna), kan beräknas medelst formeln

$$(16; h=5) \qquad \begin{aligned} \bigtriangleup N(a_i) &\approx \tfrac{12}{25} L(a_i) - \tfrac{60}{25} N_x(a_i) - \tfrac{5}{25} \bigtriangledown N(a_i) \\ &= 0{,}48\, L(a_i) - 2{,}4\, N_x(a_i) - 0{,}2 \bigtriangledown N(a_i). \end{aligned}$$

I regel är dock icke antalet genomlevda år känt för den befolkning man vill studera, utan endast de relativa frekvenstalen $\lambda(a_i \rightarrow a_j)$ för antalet växlingar $N(a_i \rightarrow a_j)$. Genom att i enlighet med formel (9) ersätta $\bigtriangleup N(a_i)$ med $N(\rightarrow a_i) - N(a_i \rightarrow)$ fås en formel för beräkning av $L(a_i)$. Vi omforma först (9) med tillhjälp av (5,2) till

$$(17; h=5) \qquad \bigtriangleup N(a_i) = N(\rightarrow a_i) - N(a_i \rightarrow) = N(\rightarrow a_i) - \lambda(a_i \rightarrow) L(a_i).$$

Då vi sätta detta uttryck lika med högra membrum i formeln (16: $h = 5$) fås

$$(18{,}1; h=5) \qquad \begin{aligned} \bigtriangleup N(a_i) &= N(\rightarrow a_i) - \lambda(a_i \rightarrow) L(a_i) \\ &= 0.48\, L(a_i) - 2.4\, N_x(a_i) - 0.2\, N_x(a_i), \end{aligned}$$

varav följer

$$(18{,}2; h=5) \qquad L(a_i) = \frac{N(\rightarrow a_i) + 2{,}4\, N_x(a_i) + 0{,}2 \bigtriangledown N(a_i)}{0{,}48 + \lambda(a_i \rightarrow)}.$$

Denna formel möjliggör en successiv beräkning av $L(a_i)$, $\bigtriangleup N(a_i)$ och $N_{x+h}(a_i)$, om ingångstalen $N(\rightarrow a_i)$ samt talen $N_x(a_i)$ och $\bigtriangledown N(a_i) = N_x(a_i) - N_{x-h}(a_i)$ äro kända. I specialfallet $N(\rightarrow a_i) = 0$ ger (18,2) en exaktare och mindre arbetskrävande metod för beräkning av s.k. överlevelsetabeller än de metoder, som använts vid de officiella beräkningarna i Finland och Sverige.

Vid de kalkyler vi utfört över den stationära befolkningens fördelning enligt civilstånd ha vi begagnat formeln (18,2). För de ogifta ($a_1 = 0$) ha vi härvid det enkla specialfallet $N(\rightarrow o) = 0$ för varje åldersklass $x, x+5$; $x > 0$. ($N_0(0) = 514$ för gossarna och 486 för flickorna.) Avgångsfrekvenstalet för de ogifta fås enligt formeln

$$(5{,}1;\ 0) \qquad \lambda(0 \rightarrow) = \lambda(o \rightarrow g) + \lambda(o \rightarrow d) \quad \text{(jfr (5,1))}.$$

I fråga om de gifta är situationen mera komplicerad. Om vi skulle uppdela de gifta i klasserna första, andra, tredje o.s.v. gången gifta män

resp. kvinnor och antaga, att övergångsfrekvenserna för alla dessa klasser äro kända, kunde man beräkna först $L\,(g_1)$, (där g_1 = första gången gifta), genom att i formeln (18,2) införa $N\,(\rightarrow g_1) = \lambda\,(0\rightarrow g)\,L\,(0)$. Därpå kunde man beräkna antalet nytillkomna förutgifta, vilka varit gifta en gång tidigare medelst formeln $N\,(g_1\rightarrow f_1)) = \lambda\,(g\rightarrow f)\cdot L\,(g_1)$, varpå man kan konstruera en tabell över $N\,(f_1)$ och $L\,(f_1)$ med tillhjälp av (18,2) genom att sätta $N\,(\rightarrow f_1) = N\,(g_1\rightarrow f_1)$ och $\lambda\,(f_1\rightarrow) = \lambda\,(f\rightarrow)$. Härpå kan man beräkna antalet personer, som gifta sig för andra gången medelst formeln $N\,(\rightarrow g_2) = \lambda\,(f\rightarrow g)\,L\,(f_1)$, vilka tal ånyo införas i formeln (18,2) för beräkning av $L\,(g_2)$ — o.s.v., varpå man slutligen beräknar $L\,(g) = \sum_i L\,(g_i)$ och $L\,(f) = \sum_i L\,(f_i)$). Vid våra beräkningar av den stationära kvinnliga befolkningens storlek och fördelning enligt civilstånd ha vi dock icke begagnat denna teoretiskt intressanta metod utan ett slags »regula di falsi» förfarande, som visserligen bygger på samma princip, men där vi i stället startade med det närmevärde för $N\,(\rightarrow f) = \lambda\,(g\rightarrow f)\,L\,(g)$, som vi erhöllo, då vi för $L\,(g)$ begagnade det närmevärde vi erhöllo genom att uppdela skillnaden $L\,(o+g+f) - L\,(o) = L\,(g+f)$ i gifta och förutgifta med tillhjälp av den faktiska befolkningens procentuella fördelning enligt civilstånd i Sverge 1940 och i Finland omkr. år 1938. (Tabell 3). Om man vill, kan man helt undgå »regula di falsi» respektive den fullständiga beräkningen av den gifta befolkningen i första, andra och tredje giftet genom att begagna den räknemetod, som erhålles genom att i enlighet med formeln (5,3) införa i (18,2)

$$N\,(\rightarrow a_i) = \sum_{j\neq i} \lambda\,(a_j - a_i)\,L\,(a_j) \tag{5,3}$$

samt lösa det sålunda erhållna lineära ekvationssystemet

$$(0{,}48 + \lambda\,(a_i\rightarrow))\,L\,(a_i) = \sum_{i\neq j} \lambda\,(a_j - a_i)\,L\,(a_j) + 2{,}4\,N_x(a_i) + 0{,}2\bigtriangledown N(a_i) \tag{18,3}$$

där a_i; $i = 1, 2, 3, 4$, i tur och ordning betecknar de »fyra» civilstånden a_1 = ogifta (o), a_2 = gifta (g), a_3 = förutgifta (f) och a_4 = döda (d). Man finner på detta sätt för $L\,(g)$ formeln:

$$L(g) = \frac{(2{,}4\,N\,(g) + 0{,}2\bigtriangledown N(g) + N\,(0\rightarrow g)) + \lambda\,(f\rightarrow g)L(f\,;\,N\,(f\rightarrow g) = 0)}{(0{,}48 + \lambda\,(g\rightarrow)) - \lambda\,(f\rightarrow g)\left(\dfrac{\lambda\,(g\rightarrow f)}{0{,}48 + \lambda\,(f\rightarrow)}\right)}, \tag{18,4}$$

där

$$L\,(f\,;\,N\,(f\rightarrow g) = 0)) = \frac{2{,}4\,N\,(f) + 0{,}2\bigtriangledown N\,(f)}{0{,}48 + \lambda\,(f\rightarrow)}. \tag{18,5}$$

Då $L\ (g)$ beräknats enligt (18,4) erhålles $L\ (f)$ för de förutgifta lätt medelst formeln (18,2) genom att välja $a_i = f$ och införa $N\ (\rightarrow f) = \lambda\ (g \rightarrow f) \cdot L(g)$

$$(18{,}2\,;\, a_i = f) \qquad L(f) = \frac{\lambda(g \rightarrow f)\, L(g) + 2{,}4\, N(f) + N(f)}{0{,}48 + \lambda(f \rightarrow)}\,.$$

Formeln (18,4) och (18,2; $a_i = f$) ha vi begagnat vid beräkningen av antalet genomlevda år för gifta och förut gifta män i Finland och Sverige i 1936—40 års stationära befolkning, samt för stickprovskontroll av att det för kvinnornas vidkommande begagnade regula di falsi förfarandet lett till samma resultat, som skulle ha erhållits genom en direkt tillämpning av formeln (18,4), (vilken i själva verket härleddes först sedan tabellen för kvinnornas vidkommande redan beräknats). Räknetekniskt äro de båda förfarandena närmelsevis likvärdiga, regula di falsi förfarandet kan, om man startar från ett mindre gott närmevärde för $N\ (\rightarrow a_i)$, erfordra ett något större antal räkneoperationer förrän man erhållit resultaten med två eller tre riktiga siffror, men är på sätt och viss riskfriare, i det att ett vid den första kalkylen av $L\ (g)$ och $L\ (f)$ begånget räknefel upptäckes och korrigeras, då kalkylen upprepas med utgångspunkt från de på grundval av den första kalkylen erhållna förbättrade närmevärdena för N $(\rightarrow g)$ och $N\ (\rightarrow f)$. Om talen $L\ (f)$ och $L\ (g)$ önskas med fyra eller fem siffror, är formeln (18,4) överlägsen »regula di falsi» metoden.

Beträffande frågan om huru övergångsfrekvenstalen $\lambda\ (a_i \rightarrow a_j)$ rent statistiskt skall kunna bestämmas nämna vi att vi i de fall, då officiella beräkningar icke stått till buds, uppskattat dem genom att sätta antalet under den studerade perioden 1936—40 (eller ibland endast 1938) inträffade i den officiella statistiken redovisade antalet övergångar $N\ (a_i \rightarrow a_j)$ i relation till den h-dubbla medelfolkmängden $\overline{N}\ (a_i)$ i den faktiska befolkningen i den betraktade klassen under den studerade perioden, (vars längd $= h$) i enlighet med formeln:

$$(19{,}1) \qquad \lambda(a_i \rightarrow a_j) = \frac{N(a_i \rightarrow a_j)}{h \cdot \overline{N}(a_i)}\,;\ \overline{N}(a_i) \approx \frac{L(a_i)}{h}\,.$$

För Finlands vidkommande ha vi varit tvungna att uppskatta $\overline{N}\ (a_i)$ för de olika civilstånden med utgångspunkt från de officiella talen över den närvarande befolkningens fördelning på civilstånd vid slutet av år 1930 och år 1940. Vid uppskattningen begagnades lineär interpolation till mitten av år 1938 mellan talen för 31/12 1930 och de tal, som kunde

ha väntats vid slutet av 1940 om den överdödlighet, som kriget åstadkommit, icke skulle ha förekommit. Speciella svårigheter uppkom för uppskattningen av medelantalet änkor och frånskilda i de yngre årsklasserna. Vi löste detta genom att antaga att deras procentuella andel »normalt» för perioden 1936—40 skulle ha utgjort samma procenttal av summan av gifta och förutgifta, som i den närvarande befolkningen 1930.

Frekvenstalen av typen $\lambda\,(a_i \rightarrow)$ beräknades genom addition av delfrekvenstalen $\lambda\,(a_i \rightarrow a_j)$ över samtliga $j \neq i$. Den största svårigheten uppkom vid försöken att bestämma frekvenstalen $\lambda\,(g \rightarrow f)$, då i detta fall talen $N\,(g \rightarrow f)$ ej äro bekanta från den officiella statistiken för olika femårsklasser av gifta. En lösning på problemet kunde finnas genom att först konstruera fördelningen av de bestående äktenskapen enligt mannens åldersklass x, $x+5$ och hustruns åldersklass y, $y+5$ med utgångspunkt från inträdesfrekvenstal av typen

$$\lambda(\rightarrow g_{xy}) = \frac{N(\rightarrow g_{xy})}{L(\rightarrow g_{xy})}, \tag{19,1}$$

där

$$L\,(\rightarrow g_{xy}) = \text{Min}\ ((L_x\,(\rightarrow g;\ \text{män}),\ L_y\,(\rightarrow g;\ \text{kvinnor})\,^{1}) \tag{19,2}$$

i vilket uttryck (t.ex. för männens vidkommande)

$$L_x(\rightarrow g,\ \text{män}) = L(0) + \frac{\lambda(f \rightarrow g)}{\lambda(0 \rightarrow g)} \cdot L(f), \tag{19,3}$$

samt avgångsfrekvenstal beräknade enligt formeln

$$\lambda(g_{xy} \rightarrow) = \lambda(g_x \rightarrow d;\ \text{män}) + \lambda(g_y \rightarrow d;\ \text{kvinnor}) + \lambda(g_{xy} \rightarrow s_{xy}), \tag{19,4}$$

där s_{xy} betecknar klassen av frånskilda par i åldersklassen (x, y). Räkningarna stöta dock på det nya hindret, att en exakt beräkning av skilsmässofrekvensen $\lambda\,(g_{xy} \rightarrow s_{xy})$ skulle förutsätta, *att man skulle ha kännedom om det sätt, varpå de bestående äktenskapen fördela sig enligt mannens och hustruns ålder*, så att man skulle kunna uppskatta talen $L\,(g_{xy})$ för den faktiska befolkningen under den studerade perioden. Tyvärr *saknas*

[1]) Ett i avseende å variationer i L_x $(\rightarrow g,$ män) och L_y $(\rightarrow g,$ *kv*) måhända något stabilare frekvenstal, som dock är väsentligt mera arbetskrävande är

$$\lambda(\rightarrow g_{xy}) = \frac{N(\rightarrow g_{xy})\,(L_x(\rightarrow g,\ \text{män}) + L_y(\rightarrow g,\ kv))}{2\,(L_x(\rightarrow g,\ \text{män}) \cdot L_y(\rightarrow g,\ kv)}. \tag{19,2'}$$

dessa för en mera djupgående analys av de befolkningsstatistiska problemen grundläggande uppgifter i den officiella statistiken i Finland och med undantag för åren 1920 och 1930 även i Sverige, varjämte man för Finlands vidkommande ytterligare stöter på svårigheten, att man icke har uppgifter om de ingångna första giftermålens fördelning enligt mannens och hustruns ålder, endast angående samtliga ingångna äktenskaps fördelning. Denna svårighet kan nödtorftigt övervinnas genom att i enlighet med formel (19,3) »omräkna» de förutgifta till ogifta genom att uppfatta de förutgifta såsom ekvivalenta med det antal, som fås, då de multipliceras med förhållandet $\lambda\,(f \rightarrow g) : \lambda\,(0 \rightarrow g)$. Vid de kalkyler vi utfört i avsikt att beräkna giftermålens och de bestående äktenskapens fördelning i den stationära befolkningen enligt mannens och hustruns ålder ha vi utnyttjat denna möjlighet. Antalet skilsmässor i den stationära befolkningen ha vi uppskattat genom att begagna formeln

$$(19{,}5) \qquad N\,(g_{xy} \rightarrow s_{xy}) = (\lambda(g_y \rightarrow s_y) \cdot L\,(g_y) \cdot \frac{N\,(g_{xy} \rightarrow s_{xy})}{N\,(g_y \rightarrow s_y)},$$

där

$N\,(g_{xy} \rightarrow s_{xy})$ och $N\,(g_y \rightarrow s_y)$ äro empiriska tal för åren 1936—40.

De på detta sätt beräknade resultaten, vilka måste uppfattas blott som en jämförelsevis grov första approximation till de sökta talen för Finland, ha vi återgivit i tabellerna (10,1—10,9). Då beräkningarna äro mycket arbetskrävande ha vi icke-fullföljt uppslaget genom att fortsätta kalkylerna enligt det begagnade »regula di falsi» förfarandet, så att vi skulle fått dessa resultat korrigerade med tillhjälp av de erhållna preliminära resultaten o.s.v., utan avbrutit räkningarna tillsvidare. Ej heller ha vi beräknat de motsvarande talen för Sverige. De preliminära resultaten för Finland visade nämligen att fördelningen av de bestående äktenskapen enligt mannens och hustruns ålder mycket liknade en i befolkningsstatistisk litteratur mycket sällsynt tabell över de bestående äktenskapens fördelning enligt mannens och hustruns ålder hänförande sig till Norge år 1900. Genom att utnyttja denna i Jahns lärobok »Statistikkens teknikk og metode» återgivna tabell får man nämligen relativt tillfredsställande närmevärden för övergångsfrekvenserna $\lambda\,(g \rightarrow f)$ mellan gifta och förutgifta medelst formeln

$$(19{,}6) \qquad \lambda_x(g \rightarrow f) = \lambda_x\,(g \rightarrow s) + \sum_y \left(\frac{L\,(g_{xy})}{\Sigma L\,(g_{xy})} \right) \lambda_y\,(g \rightarrow d),$$

genom att ersätta antalet genomlevda år $L\ (g_{xy})$ i åldersklasskombinationen $(x, x+5),\ y, y+5))$ med antalet bestående äktenskap i denna kombination i Norge år 1900. Denna lösning av det intrikata problemet ha vi utnyttjat vid våra kalkyler av den första approximationen till den stationära befolkningens fördelning enligt ålder och civilstånd.

Resultaten av beräkningarna enligt den första approximationen bekräftar en teoretiskt intressant iakttagelse, som redan beräkningarna enligt metoden att uppdela de gifta och förutgifta i samma proportion som i den faktiska befolkningen tydde på. De visa nämligen, att antalet ingångna äktenskap samt antalet bestående äktenskap skulle bliva c:a 5—6 % större, om männens lust att ingå äktenskap åren 1936—40, mätt medelst frekvenstalen $\lambda\ (0 \rightarrow g$; män) och $\lambda\ (f \rightarrow g$; män) (vilka äro karakteristiska för en dynamisk växande befolkning, som under den tidrymd, då de nu i äktenskapsåldern befintliga personer föddes, växte med c:a 1 % årligen), skulle bliva utslagsgivande för äktenskapsfrekvensen i den motsvarande stationära befolkningen, än om kvinnornas äktenskapsfrekvens skulle bestämma äktenskapsfrekvensen i den stationära befolkningen. Den utförda preliminära kalkylen över antalet ingångna äktenskaps fördelning enligt mannens och hustruns ålder i den stationära befolkningen för Finland 1936—40 gav till resultat, att antalet ingångna äktenskap per 1.000 födda skulle uppgå till omkr. 397, medan man enligt männens äktenskapsfrekvens skulle ha erhållit 408 äktenskap och kvinnornas äktenskapsfrekvens blott skulle resultera i omkr. 389 äktenskap. Förklaringen till detta resultat utgör, att männen i medeltal ingå äktenskap vid en högre ålder än kvinnorna. I en växande befolkning finnes det relativt sett flere kvinnor i de yngre åldersklasserna än i en stationär befolkning. För att man i en stationär befolkning skall få lika många ingångna äktenskap av män och kvinnor samt uppnå, att totalantalet genomlevda år av gifta män och kvinnor överensstämma, måste därför kvinnornas äktenskapsfrekvens stiga (eventuellt även skillnaden mellan männens och kvinnornas medelålder vid äktenskapets ingående avtaga) för att männens tendens att ingå äktenskap skall kunna realiseras. Jämvikten kan naturligtvis även uppnås genom en minskning i männens äktenskapsfrekvens på grund av de större svårigheterna för dem att finna en äktenskapspartner av det slag, de skulle ha valt, om de yngre åldersklasserna bland kvinnorna vore relativt sett lika stora som i en växande befolkning. Vår kalkyl över antalet äktenskap, som skulle ingås, ifall frekvenstalen $\lambda\ (\rightarrow g_{xy})$ definierade (enl. formel 19,1--19,3) såsom kvoten mellan antalet ingångna

äktenskap i en viss åldersklasskombination och maximiantalet äktenskapsår för möjliga nya äktenskap av detta slag skulle bliva konstanta, tyder på att de i en stationär befolkning ingångna äktenskapens antal närmelsevis torde kunna beräknas såsom ett medelvärde av de resultat, som erhålles enligt männens och kvinnornas giftermålsfrekvens tagna skilt för sig. En bättre lösning på problemet kunde erhållas genom att fullfölja uppslaget att beräkna den stationära befolkningens sammansättning med utgångspunkt för frekvenstal, som beskriva giftermålsfrekvensen och äktenskapens upplösningsfrekvens i olika åldersklasskombinationer. Denna tankes realiserande är det kanske dock bäst att uppskjuta till dess man, såsom vi hoppas efter slutförandet av 1950 års folkräkning, har tillgång till uppgifter om de bestående äktenskapens fördelning enligt makarnas ålderskombinationer. Vi ha vid kalkylerna över antalet barn, som komma att födas i den stationära befolkningen, använt oss av antagandet, att det i den stationära befolkning, som svarar mot förhållandena åren 1936—40 i Finland, kan antagas finnas omkr. 2,3 % flere och i Sverige 3 % flere gifta kvinnor än som kvinnornas giftermålsfrekvens under denna period skulle förutsätta. Vid beräkningen av antalet barn, som man väntar sig att skola födas i en stationär befolkning, utgår man i regel ifrån att de s.k. specifika födelsefrekvens-

$$\lambda(\rightarrow 0\,;\, g, y) = \frac{N(\rightarrow 0\,;\, g, y)}{L_y(g\,;\, kv)} \tag{20,1}$$

talen för varje åldersklass av gifta kvinnor och för varje åldersklass av icke gifta kvinnor

$$\lambda(\rightarrow 0\,;\, f+0, y) = \frac{N(\rightarrow 0\,;\, f+0, y)}{L_y(f+0)} \tag{20 2}$$

stabiliseras på den nivå de hade under den period, vars födelsefrekvenser studerats.

I formeln (20,1) betecknar $N(\rightarrow 0;\ g,\ y)$ det antal barn, som födas (levande) av gifta kvinnor i åldersklassen $(y,\ y+5)$, medan de i motsvarande åldersklass av gifta kvinnor genomlevda åren uppgå till L_y (g; kv). I formeln (20,2) ingå de motsvarande talen för icke gifta kvinnor, d.v.s. ogifta och förutgifta. Totalantalet levande födda barn N $(\rightarrow 0)$ erhålles, om $\lambda(\rightarrow 0;\ g,\ y)$ och $\lambda(\rightarrow 0;\ 0+f,\ y)$ äro kända enligt formeln:

$$N_0(\rightarrow 0) = \sum_y (\lambda(\rightarrow 0\,;\, g, y)\, L_y(g\,;\, kv)) + \sum_{\ddot{u}} (\lambda(\rightarrow 0{,}0+f, y)\, L_y(o+f\,;\, kv))\,. \tag{20,3}$$

Då födelsefrekvenstalen $\lambda\,(\rightarrow 0;\, g,\, y)$ respektive $\lambda\,(\rightarrow 0;\, f+0,\, y)$ beräknas på grundval av förhållandena under en period av stigande äktenskapsfrekvens, har denna beräkningsmetod en tendens att överskatta fruktsamheten i den stationära befolkningen beroende på att födelsefrekvenstalen icke allenast bero av kvinnans ålder och civilstånd, utan även av den tid äktenskapet varat. Under de första åren av ett nytt äktenskap är födelsefrekvensen mer än dubbelt högre än för kvinnor i samma ålder, som 5—10 år levat i äktenskap. Detta leder till en överskattning av födelsefrekvenstalen $\lambda\,(\rightarrow 0)$ under perioder då äktenskapsfrekvensen är större än normalt och en underskattning av födelsefrekvensens storlek, när äktenskapsfrekvensen är mindre än normalt. Ett sätt att minska dessa risker för feluppskattningar (som räknetekniskt ej är alltför betungande) är, att i stället för att operera med födelsefrekvenstal av det slag, som definierats i (20,1) och (20,2), begagna frekvenstal, som även beakta antalet nygifta kvinnor i åldersklassen $y,\ y+5$. Härvid är det skäl att beakta, att nära 50 % av det barnantal $N\,(\rightarrow 0;\, f+0,\, y)$, som födes av icke gifta kvinnor, utgöra föräktenskapliga barn och sålunda äro beroende av giftermålsfrekvenstalen $Ny(\rightarrow g;$ kv). Genom ett studium av den i Sveriges Officiella statistik »Befolkningsrörelsen» ingående tabellen över födelsefrekvensens beroende av äktenskapets varaktighet ha vi kommit till uppfattningen, att ett relativt gott resultat beträffande det antal barn, som kan väntas bliva födda i den stationära befolkning, som beskriver de demografiska grundfaktorernas styrka i medeltal under en viss period, torde kunna erhållas, om man i stället för födelsefrekvenstal av typen (20,1) och 20,2) begagnar frekvenstal av typen:

$$(20,3) \qquad \lambda\,(\rightarrow 0\,;y) = \frac{N\,(\rightarrow 0\,;y)}{L_y\,(\rightarrow 0)}\,.$$

Talet $Ly\,(\rightarrow 0)$ betecknar i denna formel ett närmevärde för det antal »normala äktenskapsår», som kvinnor tillhörande åldersklassen $y,\ y+5$ leva tillsammans med män, uppskattad enligt en formel av typen

$$(20,4) \quad L_y\,(\rightarrow 0\,;kv) = L_y\,(g\,;kv) + k_1 N_y\,(0\rightarrow g\,;kv) + k_2\ N_y\,(0+f\,;kv)\,.$$

Härvid uppskatta vi talet k_1 till c:a 3 år och $k_2 = \frac{1}{2}\cdot\frac{\lambda(0\,;0+f)}{\lambda(0\,;9)} \approx 0{,}05\,.$

Om goda närmevärden för frekvenstalen $\lambda(\rightarrow 0;\, y)$ äro kända, fås antalet förväntade levande födda barn enligt formeln:

$$(20,5) \qquad N\,(\rightarrow 0) = \sum_{\mu} \lambda\,(\rightarrow 0\,,\, y)\, L_y\,(\rightarrow 0)\,.$$

De kalkyler med frekvenstal vi utfört grunda sig alla på antagandet, att frekvenstalen förbli närmelsevis konstanta under den 100-årsperiod, som förflyter förrän den sista av 1.000 samtidigt födda barn dött. Om man vill göra realistiska framtidsprognoser på grundval av dylika kalkyler, får man de tillförlitligaste resultaten genom att uppgiva denna förutsättning och i stället operera med prognoser för frekvenstalen $\lambda\ (a_i \rightarrow a_j)$ beräknade genom ett studium av trendutvecklingen i dessa tal under en längre tidrymd. I fråga om speciellt dödsfallsfrekvenserna $\lambda\ (a_i \rightarrow d)$ har trendutvecklingen under det senast förflutna århundradet tenderat så kraftigt nedåt, att ett försummande av den sannolika framtida nedgången under prognosperioden leder till en väsentlig underskattning av den framtida befolkningens sannolika storlek (se diagram 2.) En vidare utveckling av teorin för befolkningsprognoser i denna riktning jämte ställandet av möjligast realistiska befolkningsprognoser för Finland och Sverige faller dock utanför ramen för denna undersökning.

Till slut kan det kanske vara skäl att påpeka, att de ovan angivna formlerna i teorin för utnyttjande av frekvenstal för beräkning av stationära befolkningar och deras statistiska egenskaper och sannolika utvecklingsförlopp kan utnyttjas även då det icke är fråga om befolkningar av levande människor. Teorin kan lika väl utnyttjas för kalkyler beträffande varaktiga objekt överhuvudtaget och de växlingar mellan olika statistiska massor, som de kunna undergå, då dessa objekts statistiska attribut förändras i tiden. I ekonometrin intresserar man sig främst för de attributväxlingar, som försiggå med de värdeobjekt, vilka människan intresserar sig för under sin livstid. Befolkningstatistiken riktar sitt intresse på de attributväxlingar, som de levande personerna själva undergå. Den statistiska teorin för attributväxlingar hos objekt med en ändlig »livstid» är i själva verket en abstrakt teori, som i princip kan tillämpas på såväl i fysiologisk mening levande objekt, som enbart i fysikalisk mening existerande objekt.

The post-war population development of Finland compared with predictions made after the war

LEO TÖRNQVIST

1. As a result of the war, and of post-war developments, all earlier Finnish population forecasts became clearly misleading. Therefore in 1947 a new series of predictions was made under the direction of the author. In these predictions extending to the year 2000 the future development was calculated by means of graphically predicted survival rates for each sex and five-year age group, and corresponding birth-rates for women. About each of them three different predictions were made called "high", "low", and "medium". The two first-mentioned were aimed at giving an 80 per cent confidence interval for the actual development. The names of those predictions tell if a change from "medium" to "high" or "low" will cause an increase resp. a decrease in the population. Because migration was not believed to have any notable effects on the results of the calculations, and the effects of new antibiotics were not yet predictable, these factors were not taken into account.

2. When predicting the future development of the age specific fertility rates of Finnish females the observed high numbers for the year 1947 were supposed to descend rapidly to the level prevailing before the war. The number of births for the period 1946-1950 was, however, badly underestimated. The total 24 per cent decrease in the number of births from the maximum level to a "normal" level almost agrees with the "high" prediction according to which the corresponding decrease was supposed to be 24.7 per cent. But as a matter of fact this "normalization" of the birth-rates, which was expected to be completed by the beginning of the 1950's, did not end before the second half of the decade. This was due to a remarkable increase in the proportional number of marriages in all fertile age groups. This increase has been especially striking in the female age group of 20-25, in which the population of married women has increased from 27.6 per cent to 45.4 per cent during the period 1940-1960. As a result of the underestimation of the birth-rate the actual number of births during the period 1946-1960 was 9.8 per cent higher than the "high" prediction.

3. New antibiotics like penicillin, streptomycin and aureomycin were already known in the 1940's, but because their demographic effects could not then be determined the steady increase in the survival rates taking place in the 1930's was supposed to continue. In reality, however, the development in this respect has been so much stronger that the level, which according to the forecast could not be reached until the year 2,000, was approximately reached at the end of the 1950's. During the period 1936-1940 new-born boys were expected to live 54.3 years and new-born girls 59.5 years. During 1956-1960 the same figures were 64.9 and 71.6 years, while the corresponding "high" predictions were about eight years lower. For comparison it may be mentioned that according to the "high" prediction the expected length of life in the year 2000 was supposed to be 64.0 years for new-born boys and 73.9 years for new-born girls.

4. The exceptionally rapid increase in the expected length of life is, I think, mainly due to new antibiotics, but also the increase in the number of marriages may have caused a slight decrease in mortality.

5. Curves describing the post-war development of total population, number of births, deaths and the total deductions including the number of deaths, net emigration and statistical corrections are presented in figure I. Actual and predicted population growth are illustrated by the shadowed areas between the curves describing the number of births and deductions.

6. The actual state of the Finnish population in 1960 and the three predictions are presented by the age pyramid in figure II. In the scale used, the part of the pyramid describing the actual and the predicted population aged over 55 years agree with each other very well. In the age group of 20-54 years, on the contrary, the actual number of persons is 0.8 per cent smaller than what was expected according to the "low" prediction. This is due to the fact that migration has caused an un-

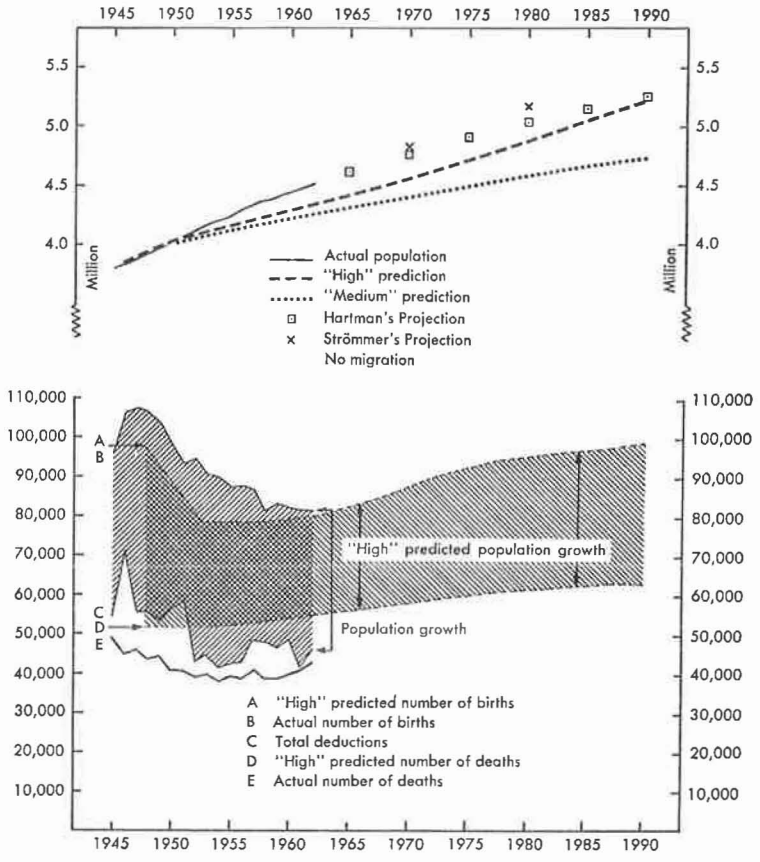


Figure I

Post-war population development of Finland

expectedly great decrease in this age group, overcompensating the error caused by the underestimation of the increase in survival rates. The number of persons born before 1945 aged over 15 years in the year 1960, however, practically agrees with the corresponding number of the "low" prediction. The effect of migration has been relatively small on the population aged 15-19 years. The underestimation of the decrease in mortality explains why the actual number of persons in this age group is 4.4 per cent higher than the "high" prediction. Taken altogether, the population in the year 1960 was 5.9 per cent higher than "high", 7.6 per cent higher than "medium" and 11.3 per cent higher than "low".

7. In the years 1950 and 1960 complete censuses of population were made in Finland. The information collected from them forms the basis of our new population projections that were made under the direction of Tor Hartman in the year 1963. These projections for the years 1970, 1980 and 1990 are described in the publication "Regional projections of population and households to the year 1990". The projections are based on the following assumptions:

(*a*) The net emigration remains at the level registered in the 1950's, namely, 7,000 per year;

(*b*) The migration from one region to another or from rural communes to urban communes or *vice versa* will, regarding direction and intensity, remain as calculated in the later half of the 1950's;

(*c*) The life tables remain invariable during the period of projection as calculated on the basis of the mortality for the year 1961;

(*d*) The birth rates and their differences between different parts of the country decrease.

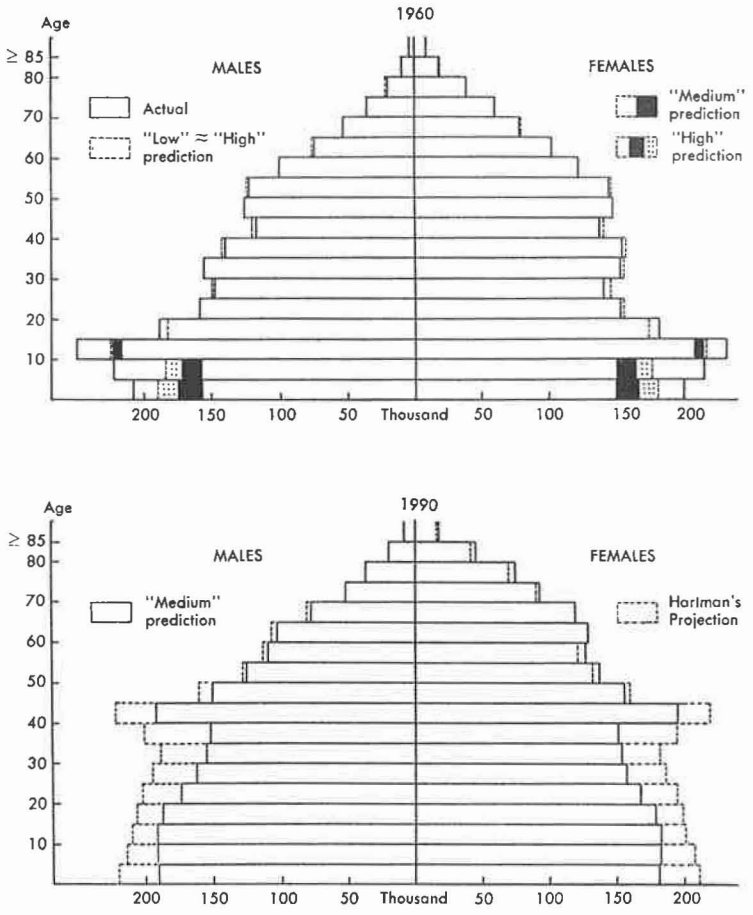


Figure II
Population of Finland

8. Results of these projections until the year 1990 are also described in figures I and II, comparing them graphically with the "high" or the "medium" prediction of our first post-war forecast. It is seen from the figures that according to the calculations the number of births will remain continuously at a higher level than expected according to the "medium" prediction though, contrary to the "medium" prediction, any clear repetition of the post-war "baby boom" is not envisaged in this projection. As to the part of the population aged over 45 years in 1990, who had already been born in 1945, the number of persons in this age group will be practically the same according to both the calculations. So the larger number of our "medium" prediction in the female age group is fully compensated by the smaller number of males, and the decrease in mortality by the assumptions made in the later projections about migration.

9. If the welfare of the Finnish people continues to increase without disturbances until 1990, Hartman's projected number for 1990 of 5.25 million could, I think, be considered as a "low" prediction. The latest 1964 projections directed by A. Strömmer, "Calculations concerning the population development in communes of Finland during 1960-1990" (publ. The National Planning Offices, series A:16) result in the corresponding number 5.19 million, according to the calculations taking migration into account, while the higher number, 5.48 million, is obtained according to calculations taking no account of migration. If besides migration we also take into account the further increase in the expected length of life during the period 1960-1990, and the fact that all earlier plausible predictions and projections made have been surpassed by the actual developments, it would not surprise me at all if the actual number for 1990 were above 5.5 million.

MODELS FOR A NON-EXPLOSIVE WORLD POPULATION DEVELOPMENT

by

Leo Törnqvist

We are used to talk and read about the population »explosion», when the world population development of the last centuries is considered. Often the writer speculates about a peculiar future when the population densities would reach ridiculous values assuming that the rapid increase of the last 100 years would go on for thousands of years. Relatively few writers have thought seriously about how the population of the world will most likely develop during the next hundred years.

One basic fact, which has to be incorporated in seriously meant population models, is the existence of an upper limit of the world population. For this reason, any such model ought to describe a bounded or at least a non-explosive process. Our purpose is to present some models of this type.

We consider here the interdependence of the world population process P_t and a process X_t which we call the deterioration of conditions for human lives. This process X_t, where t refers to historical time, measures — roughly speaking — the accumulated pollution caused by human population. Unfortunately the past historical path of X_t is unknown, even if a lot of separate details have been observed.

The available data of the P_t-process consists mainly of estimates of the world population during the 20th century. The »guesstimates» for 1650 of about 500 millions, for 1750 of about 750 millions, for 1800 of about 960 millions and for 1850 of 1240 millions are not used in this paper. The data used for this century is presented in table 1.

The basic idea about the interdependence between P_t and X_t is formulated in the following model:

$$(1)\quad \log P_t - \log P_{t-5} = A_0 + Y_{t-5}A_1 - X^2_{t-5} - A_2 P^2_{t-5} + R_1(t)$$

$$(2)\quad X_t - X_{t-5} = \frac{-B_1 X_{t-5}}{1 + C_1 X_{t-5} + C_2 X^2_{t-5}} + \frac{B_2 P_t^3}{C_3 + C_4 P_t + P_t^2} + R_2(t)$$

$$(3)\quad Y_t = D_0 + D_1 X_t + D_2 X^2_t + R_3(t),$$

where Y_t describes the positive effects of the socio-economical, scientific and technological development on the population growth. The coefficients A, B, C and D are unknown constants, $R_1(t)$, $R_2(t)$ and $R_3(t)$ are (stochastic) rest terms, t is measured in years and P_t is measured in 10^9 persons.

The term

$$(4)\quad X_t(-) = \frac{B_1 X_{t-5}}{1 + C_1 X_{t-5} + C_2 X^2_{t-5}}$$

describes nature's ability to recover from the accumulated pollution, and the term

$$(5) \qquad X_t(+) = \frac{B_2 P_t^3}{C_3 + C_4 P_t + P_t^2}$$

represents the amount of new pollution during the time period (t—5, t). Some specific forms of these terms are illustrated in the diagrams 1 and 2.

The Y_t -process can be eliminated from the model by substituting formula (3) into formula (1). The model can thus be written in the form

$$(1') \quad \log P_t - \log P_{t-5} = A_0' + A_1' X_{t-5} - A_2' X^2_{t-5} - A_3' P^2_{t-5} + R_1'(t)$$

$$(2') \quad X_t - X_{t-5} = X_t(+) - X_t(-) + R_2(t).$$

To simplify the estimation procedures we assume that a sufficiently good approximation for the $X_t(-)$-term can be obtained by putting $C_1 = C_2 = 0$ and for $X_t(+)$ by putting $C_3 = 0$ and $C_4 = P_{t_0}$, where $t_0 = 1900$ In our trials to generate the X_t-values we will put $X_{t_0} = P_{t_0} = 1.650$ and neglect the error term ($R_2(t)$).

The parameters of equations (1') and (2') were estimated by minimizing the sum of squared errors of equation (1') with respect to A_0', A_1', A_2' and A_3' for given B_1 and B_2, and assigning a sequence of different values for B_1 and B_2. The values $B_1 = 0.033$ and $B_2 = 0.15$ were found to correspond approximately to the overall minimum of the sum of squared errors.

The data used (see tables 1 and 3) consisted of the UN estimates of the world population for the years 1900 and 1920—1965, the UN (»medium») predictions for the years 1970—2000 and graphical interpolations for the remaining years. The parameters A_0', A_1', A_2' and A_3' were estimated separately using the UN »high» and »low» predictions as alternative data, but keeping B_1 and B_2 fixed ($B_1 = 0.033$; $B_2 = 0.15$). Using these data we obtained the estimates:

	Low	Medium	High
A_0'	—0.10903	—0.09426	—0.104876
A_1'	0.10306	0.08762	0.092072
A_2'	0.00353	0.004256	0.006977
A_3'	0.01057	0.005505	0.002325
$\sigma^2 R'_1$	$1.133 \cdot 10^{-4}$	$0.8793 \cdot 10^{-4}$	$1.190 \cdot 10^{-4}$

In table 1, the »medium» data are compared with generated values of this model, which also can be written in the form:

$$P_t = P_{t-5} \exp [A_0' + A_1' X_{t-5} - A_2' X^2_{t-5} - A_3' P^2_{t-5}]$$

$$X_t = (1 - B_1) X_{t-5} + \frac{B_2 P_t^2}{P_{t_0} + P_t}$$

The generated values of P_t and X_t, corresponding to different data, are presented up to the year 2500 in table 3 and in diagram 3.

A temptation to make use of the general forms (4) and (5) of $X_t(+)$ and $X_t(-)$ was countered by the fact that the available data contains very little information about the pollution structure. However, it might be possible to construct an indicator describing $X_t(+)$ as a weighted mean of production time series of the world. The estimate $B_1 = 0.033 \approx {}^1/_{30}$ is strikingly small, because according to this simplified model, it would take about 100 years before one half of the accumulated pollution would be removed if no new pollution would be produced. This estimate is, however, not very accurate, because it can be doubled without significantly decreasing the fit of the model to the data.

It would be of great importance to know more about the form of the $X_t(-)$-function, because there are heavy reasons to fear that the $X_t(-)$-function actually decreases with increasing X_{t-5} when X_{t-5} is sufficiently large, as indicated in diagram 2. In such a case the survival of mankind would be on stake. Such a risk exists for instance if the amount of plankton, green plants and trees starts to decrease.

It might be possible to construct models of this kind for different regions of the world. Then, of course, the models have to be extended to take into account the interaction between regional populations, regional pollution and world pollution processes.

Studying the results obtained we see that the »low» variant generates the highest asymptotic value ($6.3 \cdot 10^9$) of the world population, and the »high» variant the lowest asymptotic value ($3.7 \cdot 10^9$). The latter value almost coincides with the number of the present population.

According to these results, the world population would most likely stay between $2.5 \cdot 10^9$ and $8 \cdot 10^9$ persons.

I am aware of only two other opinions about this question, one due to Bogue *(1)* and one due to Forrester *(3)*. My results agree well with Forrester's preliminary results, but Bogue seems to believe in a range of 10^{10}—$5 \cdot 10^{10}$ persons for the future population.

I like to thank Mr. Markku Rahiala and Miss Mirja Ollila for valuable assistance in preparation of this paper.

Diagram 1

Two examples of the form of the "new deterioration"-function $X_t^{(-)}$

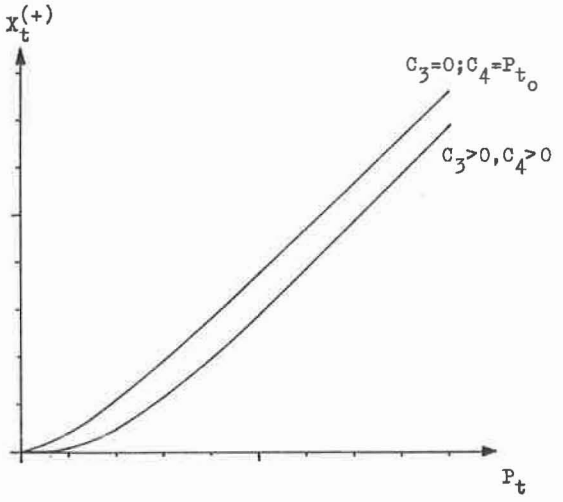
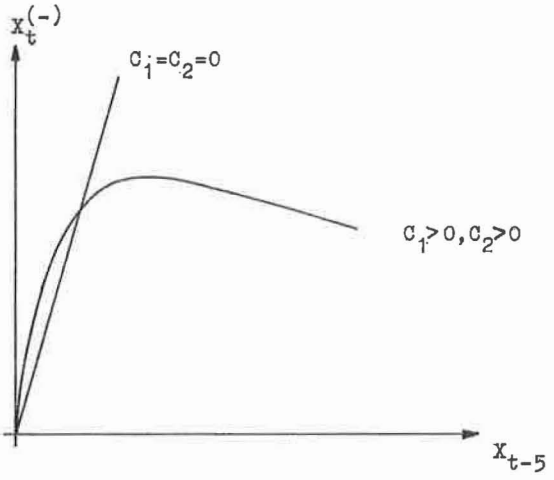


Diagram 2

Two examples of the form of the $X_t^{(-)}$-function describing natures recovery from deterioration



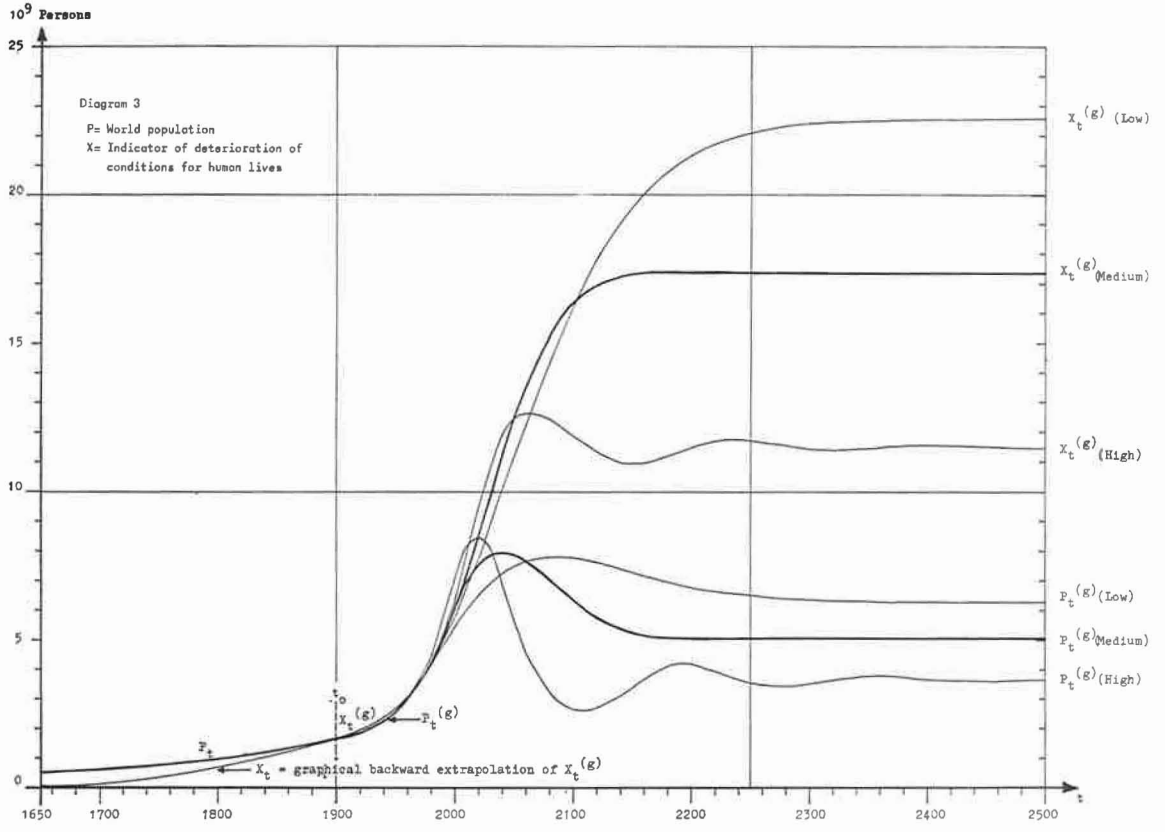

10^9 Persons
25
20
15
10
5
0
Diagram 3
P= World population
X= Indicator of deterioration of conditions for humon lives
$X_t^{(g)}$ (Low)
$X_t^{(g)}$ (Medium)
$X_t^{(g)}$ (High)
$P_t^{(g)}$ (Low)
$P_t^{(g)}$ (Medium)
$P_t^{(g)}$ (High)
t_o
$X_t^{(g)}$
$P_t^{(g)}$
P_t
X_t = graphical backward extrapolation of $X_t^{(g)}$
1650
1700
1800
1900
2000
2100
2200
2300
2400
2500
t

Table 1

UN estimates of world population complemented by graphical estimates and UN »medium» predicted values, compared with generated values ($P_t^{(g)}$) according to the model (1')—(2') with estimated parameters.

Year	P_t	$P_t^{(g)}$	$100 \cdot (P_t^{(g)}/P_t - 1)$
1900	1.650	1.650	0
*1905	1.700	1.689	—0.65
*1910	1.754	1.737	—0.70
*1915	1.814	1.795	—1.05
1920	1.862	1.864	0.11
*1925	1.960	1.944	—0.82
1930	2.070	2.038	—1.55
*1935	2.182	2.147	—1.61
1940	2.295	2.273	—0.96
*1945	2.408	2.419	0.46
1950	2.517	2.588	2.82
1955	2.730	2.784	1.98
1960	2.998	3.010	0.40
1965	3.281	3.270	—0.34
1970	3.592	3.570	—0.61
1975	3.944	3.910	—0.86
1980	4.300	4.294	—0.84
1985	4.746	4.718	—0.59
1990	5.188	5.174	—0.27
1995	5.648	5.647	—0.02
2000	6.130	6.118	—0.20

Graphically interpolated values are indicated by *.

Table 2

UN predictions for the world population for the years 1975—2000, »high» and »low» variants.

Year	Low	High
1975	3.480	4.070
1980	4.147	4.551
1985	4.463	5.096
1990	4.783	5.690
1995	5.109	6.326
2000	5.449	6.994

Table 3

Generated values of P_t—and X_t -series for the years 1900—2500, based on different data variants.

Year	P_t			X_t		
	Medium	Low	High	Medium	High	Low
1920	1.864	1.864	1.852	1.968	1.967	1.964
1940	2.273	2.283	2.254	2.395	2.397	2.384
1960	3.010	3.008	3.015	3.056	3.061	3.042
1980	4.294	4.127	4.485	4.173	4.134	4.214
2000	6.118	5.437	6.952	6.030	5.742	6.370
2020	7.547	6.504	8.426	8.563	7.786	9.426
2040	7.970	7.222	6.767	11.21	10.034	11.773
2060	7.691	7.632	4.518	13.496	12.287	12.598
2080	7.064	7.789	3.185	15.204	14.391	12.423
2100	6.373	7.758	2.684	16.321	16.248	11.877
2120	5.797	7.609	2.720	16.955	17.811	11.329
2140	5.402	7.400	3.111	17.258	19.072	10.985
2160	5.197	7.176	3.671	17.370	20.053	10.946
2180	5.080	6.967	4.108	17.393	20.791	11.172
2200	5.052	6.790	4.187	17.386	21.332	11.484
2220	5.052	6.647	3.969	17.376	21.72	11.692
2240	5.059	6.539	3.684	17.370	21.993	11.728
2260	5.064	6.459	3.497	17.368	22.183	11.641
2280	5.066	6.402	3.453	17.367	22.313	11.516
2300	5.066	6.361	3.521	17.368	22.402	11.422
2320	5.066	6.334	3.637	17.368	22.462	11.393
2340	5.066	6.315	3.733	17.368	22.502	11.424
2360	5.066	6.302	3.767	17.368	22.530	11.481
2380	5.066	6.293	3.742	17.368	22.548	11.528
2400	5.066	6.287	3.689	17.368	22.560	11.546
2420	5.066	6.283	3.644	17.368	22.569	11.536
2440	5.066	6.281	3.625	17.368	22.574	11.511
2460	5.066	6.279	3.633	17.368	22.578	11.490
2480	5.066	6.278	3.654	17.368	22.580	11.479
2500	5.066	6.277	3.676	17.368	22.582	11.482

(1) Bogue, D. J.: »The End of the Population Explosion»; The Public Interest No 7, 1967

(2) Ehrlich, P. R. — Ehrlich, A. H.: »Population Resources Environment, Issues in Human Ecology»; W. H. Freeman & Co., San Fransisco, 1970

(3) Forrester, J. W.: Discussion paper, M.I.T., nov. 6. 1970

On the economic theory of lottery-gambles.

By **Leo Törnqvist** (Helsingfors).

Let us survey an economic subject A_0 who at the point t_0 is planning to offer for sale a number q of lots during a space of time

$$t_a = \overset{n_a}{(+)}_{\alpha=1} t_\alpha \tag{1}$$

= selling period of lottery ticket.

At the point $t_\alpha = t_0 + \tau_\alpha$ another economic subject A_α may buy $q_\alpha \geq 0$ lots. We suppose that A_α pays for the bought lots an amount of money

$$x_{\alpha\alpha} = p_{\alpha\alpha}\, q_\alpha \tag{2}$$

= stake of the lot-buyer A_α in the lottery.

The lottery arranger may estimate this stake, on discounting it to the point t_0, at *the present value*

$$x_\alpha = e^{-\varrho_0 \tau_\alpha} x_{\alpha\alpha} = p_\alpha \cdot q_\alpha \tag{3}$$

= present value of A_α's stake.

The total number of lots sold during the selling period may be: $q_a = \sum_{\alpha=1}^{n_a} q_\alpha$ pieces of sold lots. The present value of the gross income, derived from the selling of lots, A_0 may estimate at the amount

$$x_a = \sum_{\alpha=1}^{n_a} x_\alpha = p_a\, q_a. \tag{4}$$

As a rule, the lottery arranger does not succeed in selling all the lottery tickets, but he is compelled to take to himself q_0 pieces of lots:

(5) $$q_0 = q - q_a$$

= gambling-lots of the lottery arranger.

The prizes in the lottery may be disbursed during the space of time t_b,

(6) $$t_b = \overset{n_b}{\underset{\alpha=1}{(+)}} t_{b\alpha}$$

= period of disbursing the prizes.

The lottery arranger may calculate the present value of the prizes to be disbursed at the point $t_{b\alpha}$ to the lot-buyer A_α at $x_{b\alpha}$ kr.

(7) $$x_{b\alpha} = p_{b\alpha} q_a$$

= present value of A_α's gambling-gain.

The present value of the disbursed prizes amounts, in all, to x_b

(8) $$x_b = \sum_{\alpha=0}^{n_a} x_{b\alpha} = p_b q = p_{b0} q_0 + p_{b\alpha} q_a = p_b q_0 + p_b q_a.$$

The present value of the prizes the lottery arranger himself draws as a gain on his q_0 pieces of gambling-lots makes, in this connexion $x_{b0} = p_{b0} q_0$, while the other buyers of lottery tickets obtain prizes with a present value estimated by the lottery arranger at $x_{b\alpha}$,

(9) $$x_{b\alpha} = \sum_{\alpha=1}^{n_a} x_{b\alpha} = p_{b\alpha} q_a;\ x_{b\alpha} = p_{b\alpha} q_\alpha = \sum_{i=0}^{n_i} p_{b\alpha i} q_{\alpha i}$$

$q_{\alpha i}$ = number of lots falling due with a prize of the size $v_i = p_{b\alpha i} e^{\varrho_0 \tau_{b\alpha}}$.

The costs of the lottery arranger A_0 for the procuring and selling of the lottery tickets, and the puff for the sale of the tickets form a stream of expenses, the present value of which discounted to the point t_0 we designate with x_c kr. This amount might with a good approximation be conceived as a linear function of q_0 and q_a of the form:

$$x_c = p_{c0}\, q_0 + p_{ca}\, q_a \tag{10}$$

= present value of the lottery costs.

With regard to that $q_0 = q - q_a$, the formula (10) may also be written in the form

$$x_c = p_{c0}\, q + \acute{p}_c\, q_a \text{ where } \acute{p}_c = p_{ca} - p_{c0}. \tag{11}$$

In this formula the first term may be interpreted as an expression for the so-called firm costs, which do not depend on the scope of the sale, only on the number of lots procured for sale during the selling period t_a. The second term may be interpreted as an expression for the so-called moving costs, i.e. commissions of sale etc. If on the other hand we eliminate $q_a = q - q_0$ from the formula (10), we get

$$x_c = p_{ca}\, q - \acute{p}_c\, q_0. \tag{12}$$

In this formula the first term may be interpreted as the present value of the costs which A_0 would have, if all q lots would find buyers during the selling period t_a. The expression $\acute{p}_c \cdot q_0$ may be interpreted as the decrease from this maximum value which takes place in x_c owing to the lottery arranger's taking to himself q_0 pieces of lots.

Making use of the formulas (4), (9), and (11) we may write the present value of the gain made by A_0 in the lottery in the form

$$\begin{aligned} x_v &= x_a - x_{b\,a} - x_c = p_a\, q_a - p_{b\,a}\, q_a - \acute{p}_c\, q_a - p_{c0}\, q = \\ &= (p_a - p_{b\,a} - \acute{p}_c)\, q_a - p_{c0}\, q = p_v\, q_a - p_{c0}\, q. \end{aligned} \tag{13}$$

In this formula the symbol p_v

$$p_v = p_a - p_{b\,a} - \acute{p}_c \tag{14}$$

indicates the moving gain for A_0 per lot sold. If p denotes the mathematical expectation for p_v, the quantity $\acute{p}_v = p_v - p = 0$, when all lots have found other buyers than the lottery arranger A_0 himself.

Ex ante, i.e. when planning the lottery, A_0 does not know how it will turn out. The quantities q_a, p_v, and all variables

influencing them are for him stochastic variables, the probability distribution functions of which depend on the proceeding-parameters he may have at his disposal. The economic-political problem of the lottery arranger consists in choosing, out of all possible complexes $\boldsymbol{p}$ indicating the possible values of the proceeding-parameters, a number complex $\boldsymbol{p}_0$, which has the quality that the probability distribution function $F(x; \boldsymbol{p}_0)$ indicating the probability, conditioned by $\boldsymbol{p}_0$, that x_v will be $\leq x$, will attract A_0 more than all other probability distribution functions $F(x; \boldsymbol{p})$ he subjects to valuation. We suppose this procedure of valuation has been rationalized in the manner that A_0, for the distribution function $F(x; \boldsymbol{p})$ or properly for his conception of this distribution function, will appoint an economic-political value $x_{\boldsymbol{p}A_0}$; the size of which depends on the form of the distribution function. Further it seems to me reasonable to presume that a good approximate value for $x_{\boldsymbol{p}A_0}$ may be calculated, as a medium value of a very general type obtained by help of a medium-value formation function $U(x)$. This medium-value formation function might be interpreted as a measure of the feeling respectively of pleasure and displeasure, connected by the subject A_0 with the thought of obtaining, with a small probability $dF(x; \boldsymbol{p})$, a prize of the size x, while as a unity of measure is used the feeling of pleasure connected by the subject with the thought of obtaining, with the same probability, a prize of the value of one money-unity. The general medium value x_{pu} we presume to approximate the economic-political value $x_{\boldsymbol{p}A_0}$ is defined through the formula

$$U(x_{\boldsymbol{p}u}) = \int_{x=-\infty}^{\infty} U(x)\, dF(x; \boldsymbol{p}) \tag{15}$$

The proceeding-plan $\boldsymbol{p}$, which maximizes $x_{\boldsymbol{p}A_0} \approx x_{\boldsymbol{p}u}$, gives the solution of A_0's problem of option.

In practice, the acting and valuing subject limits his appreciation of the different proceeding-plans to a very small number of alternatives and chooses, more or less at random, his proceeding-plan $\boldsymbol{p}_0$ among these alternatives, guided only by a more or less reasonably motivated feeling that the chosen

alternative of proceeding is to be preferred to the other alternatives which have been subject to his valuation. The probability of A_0's choosing the proceeding-plan $\boldsymbol{p}_z$, which we shall designate $P(\boldsymbol{p}_z = \boldsymbol{p}_0)$, may be written in the form

$$(16) \qquad P(\boldsymbol{p}_z = \boldsymbol{p}_0) = P(p_z A_0 \neq 0) \cdot P\Big(\underset{\boldsymbol{p} A_0 \neq 0}{\&} (x_{\boldsymbol{p}_z A_0} \geq x_{\boldsymbol{p} A_0})\Big)$$

i.e. the probability of A_0's choosing the proceeding-plan $\boldsymbol{p}_z$ is equal to the product of the probability of A_0's taking the plan $\boldsymbol{p}_z$ into consideration, and of the probability of his forming the conception that the economic-political value of this proceeding-plan is greater than that of any other valued proceeding-plan. For the proceeding-plans that will not be subject to any consideration, the symbol combination $\boldsymbol{p} A_0 = 0$ is put down. Often enough it happens, that A_0 is not able to decide about the succession of size between the economic-political values of two alternatives. Already on the very occasion of valuing, cases may thus appear when A_0 only tries to estimate which alternative of proceeding has the greatest probability of being of the greatest economic-political value for the valuing subject.

A critical reader may perhaps remark that the above-mentioned probabilities cannot be made object of any attempts of measurement, and therefore cannot be considered to be any probabilities in the mathematically statistic sense. Naturally, this objection is just in some degree. Yet one may imagine that about one thousand thoroughly congenerous subjects might be within the sphere of action of a lot-seller, and that all these subjects might have the same possibilities to come under the influence of the selling activity the lot-seller develops. In this case one may study which per mille number of the surveyed subjects will, on the whole, take the question of buying lots of the surveyed buyer into consideration, and which per mille number of these latter subjects will thereupon decide on buying 0, 1, 2 etc. pieces of lots. The relative frequency numbers, obtained in this manner, we regard as the approximate values for the above-examined probabilities.

From the fact that people buy lottery tickets, notwithstanding the mathematical expectation of gain on the purchase is smaller than the price of a lot, and that fire-insurances are made though the mathematical value of the fire-risk is smaller than the premium of fire-insurance, one may draw certain general conclusions as to the qualities of that medium-value formation function $U(x)$, which has the quality of giving such a medium value x_{pu} that a reasonable explanation for the proceeding-manner of the subject will be obtained. This medium-value formation function must evidently give more weight to the relatively great gains and losses than to the medium-value formation function $U(x) = x$, which is used when calculating the mathematical expectation.[1] The medium-value formation function $U(x)$ is not symmetrical in point of losses and gains, for an expected loss weighs more heavy in the balance than an equally great gain, anticipated with the same probability. At the calculation, the anticipated gains and losses are likely to be put in relation to that maximum stake, which the subject might be able to do after the lottery-gamble has been concluded. Under such circumstances it seems to be possible, at least in respect of such amounts of gain and loss that occur at lottery-gambles, thing-insurances, and the like, to well approximate $U(x)$ by help of a mathematical function of the form

$$U(x) = x\left(1 + k\left(\frac{x}{K + x}\right)^2\right). \tag{17}$$

This medium-value formation function gives an over-normally great importance to the extremely great gains $x > 0$ and the losses $x < 0$. Losses approaching to the figure K, that depends on the wealth of the subject, have a powerfully deterring effect on the subject. The idea of obtaining a very great gain, in proportion to the wealth of the subject, creates with most people a feeling of pleasure, which grows faster

[1] In INGVAR SVENNILSSON'S dissertation: Ekonomisk planering, Uppsala 1938, among other things these over- and under-valuation phenomena are discussed, and some quantities, corresponding to the quotients $U(x) : x$, studied. The discussion does not lead to any quantitative results.

than the size of the gain. It seems to me right to presume that the relative over-valuation $(U(x) - x) : x$ does not grow over all limits, x going towards the infinite, but approaches a limit value, which we have designated with the letter k.

An estimation of the figure k may be obtained through surveying the simple lottery where only one great prize e.g. 1 million kr. is dealt out. The number of tickets offered for sale may be 500,000 and the price of a lott $p_a = 5$ kr. The mathematical value of a lot is in this case somewhat less than 2 kr., i.e. $p_b = 2$ kr. The experience at the money-lotteries shows that there are many people who would be interested in buying lots in such a lottery. If we presume, that the medium-value formation function for a buyer has the form (17), the economic-political value of for inst. g pieces of lots may be calculated by means of the formula

$$(18) \quad U(x_g) = \left(1 - \frac{g}{q}\right)\left(-p_a g - k\frac{(p_a g)^3}{(K - p_a g)^2}\right) + \\ + \frac{g}{q}\left(p_b q - p_a g + k\frac{(p_b q - p_a g)^3}{(K + p_b q - p_a g)^2}\right).$$

In case the net profit for a buyer, obtaining the great prize $p_b q = 10^6$ kr., is much greater than the maximum stake K the surveyed lot-buyer would be able to make, the quantity $U(x_g)$, calculated by the formula (18), is approximately equal to the limit value

$$(19) \quad U(x_g;\ q = \infty) = ((1 + k)p_b - p_a)g - k\frac{(p_a g)^3}{(K - p_a g)^2},$$

that $U(x_g)$ approaches to, when q is going towards the infinite. If we put in as a new variable

$$(20) \quad z = \frac{p_a g}{K - p_a g}$$

and designate shorter the expression

$$(21) \quad \frac{(1 + k)p_b - p_a}{k p_a} = h^2,$$

an expression, which must be positive, i.e. $k > \frac{p_a - p_b}{p_b}$, in order that it shall not be more attractive for the buyer in question to choose $g = 0$ than to buy a positive number of lots, the expression (19) may be written in the mathematically more easily overlooked form

$$U(x_g;\ q = \infty) = k\,K\,z\,(h^2 - z^2)(z + 1)^{-1}. \tag{22}$$

This function reaches its maximum value, when z is a root to the equation got by derivating the expression (22) in respect of z and by putting the derivate $= 0$. This equation may be written in the form $2z^3 + 3z^2 = h^2$, an equation, which also may be rendered in a form more suited for the calculation of the root determining the maximum value for the expression (22):

$$z = \frac{h}{\sqrt{3 + 2z}} = \frac{h}{h_0};\ h_0 = \sqrt{3 + 2\frac{h}{h_0}} = 1.8\ldots \tag{23}$$

If we suppose e.g. that $k = 2$, $p_a = 5$, $p_b = 2$, we find by the formula (21) $h^2 = 0.1$, and through a »regula di falsi» method by the formula (23) the root $z = 0.18$. If we solve the equation (20) in respect of $p_a \cdot g$, i.e. in respect of the stake of the lot-buyer, we find the expression

$$p_a g = K z (1 + z)^{-1}. \quad \text{If } z = \frac{h}{h_0} \text{ is } p_a g = \frac{K h}{h_0 + h} \approx \frac{K h}{1.8 + h}. \tag{24}$$

If in this formula we put $z = 0.18$, we find $p_a g = 0.15\ K$.

It may further interest to examine which ratio $p_b : p_a$ is thus conditioned, that the surveyed lot-buyer would come to give the greatest possible contribution,

$$v = p_v \cdot g = K\frac{p_v}{p_a} z\,(1 + z)^{-1} = K\left(1 - \frac{p'_c}{p_a} - \frac{p_b}{p_a}\right) z\,(1 + z)^{-1}, \tag{25}$$

to the moving gain of the lottery arranger. If we solve the equation (21) in respect of the quotient $p_b : p_a$, we find that this ratio may be written in the form

$$\frac{p_b}{p_a} = \frac{1 + k h^2}{1 + k};\ \therefore k = \left(\frac{1 - p_b : p_a}{p_b : p_a - h^2}\right). \tag{26}$$

In case we introduce the equation $h^2 = 2z^3 + 3z^2$, derived from the maximizing condition for $U(x_g;\ q = \infty)$, into the expression (26), and determine the value of z maximizing v, we find that this quantity forms the root to the equation

$$(27) \qquad (z+1)^2(6z^2+2z-1) + H = 0;\ H = \frac{k+1}{k} \cdot \frac{\acute{p}_c}{p_a}$$

which equation may also be written

$$(28) \quad z = \frac{\sqrt{7}-1}{6} - \frac{H}{6}(z+1)^{-2}\left(z + \frac{\sqrt{7}+1}{6}\right)^{-1} =$$

$$= 0.274 - \frac{H}{6}(z+1)^{-2}(z+0.61)^{-1}.$$

If we suppose e.g. that $k = 2$, and that the moving costs for the lottery arranger amount to 20 % of the price of a lot, i.e. that $\acute{p}_c : p_a = 0.2$, we find the value 0.3 for the figure H and the value $z = 0.237$ for the only positive root to (28). If this value of z is introduced into the formula $h^2 = 2z^3 + 3z^2 = (1+z)^2(2z-1) + 1$, one finds that $h^2 = 0.21$, which introduced into the formula (26) gives the value 0.47 for the sought quotient $p_b : p_a$. The stake $p_a g = 0.191\,K$.

In order that the ratio 0.4, chosen in the example, between p_b and p_a, will give a maximum value to the gain v, the figure k must have the value $k = 3.2$, $z = \frac{\sqrt{10}-1}{9} =$ $= 0.240$, $h^2 = 0.21$. In that case the stake of the surveyed buyer will be $p_a g = 0.194\,K$.

In case the lottery, forming the object of our study, is arranged for some charitable purpose, the buyers' aim-setting is altered. The buyers who from feeling of social responsibility wish to support the good purpose through buying lots, do not try to maximize $U(x_g)$, but choose, instead, the number of bought lots in the manner that the economic-political value they assign to the bought lots agrees with the economic-political value of the sacrifice, which they are willing to make for the purpose in question. If this amount is $= x_f$, the buyer chooses g in such manner that $U(x_g) = U(-x_f)$. As however

not all buyers have a social cast of mind, we may as the normal case regard the case where the number of lots is determined in the manner that $U(x_g) = 0$. If on the other hand we survey the example discussed earlier, when q was supposed to be so great that $U(x_g) \approx U(x;\, q = \infty) = k\,K\dfrac{z(h^2 - z^2)}{1 + z}$, we perceive that to this value of g corresponds the value h for z; viz. $z = h = \sqrt{\dfrac{(k+1)\,p_b - p_a}{p_a\,k}}$. This value of z is, according to the formula (24), adequate to a stake of the size $x_{\alpha\alpha} = p_a\,g = \dfrac{h}{1+h}K$. In case $k = 2$, $p_b = 2$, $p_a = 5$, we get $h = \dfrac{\sqrt{10}}{10} = 0.316$ and, consequently, the stake $= 0.24\,K$, i.e. nearly 60 % greater than in the corresponding case when the buyer tries to maximize $U(x_g)$. In this case we got $x_{\alpha\alpha} = 0.15\,K$.

In order that a buyer, who adjusts his action in the manner that $U(x_g)$ becomes equal to zero, will give a greatest possible contribution to the charitable purpose, for which the lottery is arranged, the quotient $p_b : p_a$ must be chosen so that the quantity h becomes a root to the equation

$$(29) \qquad (h+1)^2(2h-1) + H = 0;\; H = \frac{k+1}{k}\frac{p_c'}{p_a}.$$

If $p_c' : p_a = 0.2$ and $k = 2$, H becomes 0.3 and $h = 0.425$, a value of h corresponding to the quotient $p_b : p_a = 0.46$, i.e. practically the same ratio as in the case when the buyers only try to regard their prospects of gain. In this case, namely, we obtained the quotient $p_b : p_a = 0.47$. In order that the ratio $p_b : p_a = 2 : 5 = 0.4$ shall be the rightest one from the point of view that the buyer in question will give a greatest possible contribution to the charitable purpose, we must according to (26) and (29) presume that $k = 2.84$, $h^2 = 0.189$, $z = h = (\sqrt{13} - 1) : 6 = 0.434$, $p_a\,g = 0.30\,K$. In the corresponding case, when the buyer tries to maximate $U(x_g)$, we got $k = 3.2$, $h^2 = 0.210$, $z = 0.240$, $h = 0.458$, $p_a\,g = 0.194\,K$.

16—45690 *Skandinavisk Aktuarietidskrift 1945*

The above examples seems to me to give support to the conception that the quantity k for the bulk of people shall be of the order of size $k = 3$.

The reason why there in practice mostly occur lotteries in which, besides a few great prizes, also a great number of small prizes are put in, seems to me to be that the risk for the lottery arranger, that the lottery will go with loss, will hereby diminish. To this comes that the buyer of several lots will have a small, but still a probability of obtaining several prize-draws, which in all may amount to so great a sum that the desire to buy more than one lot increases. The circumstance that the probability, on the whole, of getting a prize becomes relatively great, has also proved to be a stimulation for the desire of buying. The fact that the probability of obtaining an economically noticeable prize is very small, appears less conspicuous, if we, in the puff for the lots, can state that e.g. each fourth lot will win, or, as is the case with respect to the modern kinds of lotteries called A, B, C, D lots, colour lots, gold purse lots etc., that a buyer who buys of all kinds of lots will certainly win one lot. In order to explain why such lots attract a greater interest with the buyers than common lots, we must presume that the buyers apportion a relatively great value to the thought of obtaining a prize, even if this prize afterwards proves to be so small that the buyer will not even trouble himself to call for the prize. The last-mentioned circumstance may also be a contributive motive for the lottery arranger to distribute, in addition to the chief prizes, many small prizes.

It has further appeared that the desire of buying increases perceptibly, if part of lot-drawings have taken place already before the selling period of lots, as the very buying of lots and not only the later drawings will then become an exciting act. To the increase of the desire of buying contributes, however, also the circumstance that the period passing between the distribution of prizes and the byuing of lots is shortened, whereby the factor of discounting, with which the buyers multiply their expectations of gain, increases.

These observations may be theoretically interpreted as a confirmation of the fact that the buyers vigorously simplify the probability distribution function $F(x; g)$, the value of which they try to estimate when they decide on buying a certain number of lots $g = q_\alpha$. They direct their attention chiefly to the size of the chief prizes, and to the probability of obtaining on the whole a positive gain. Moreover, such psychical income posts are taken notice of, as the excitement at the lot-drawings and the pleasure of contributing to the purpose, most frequently worthy of having compassion on, for which the lottery is arranged. As a negative psychical moment comes the trouble of going to the lot-seller's. If on the other hand the lottery-agents put a direct question to the buyer if he wishes to buy lots in a well-advertised lottery, a refusal of buying is most frequently associated with a feeling of displeasure, which circumstance may influence the resolution of buying. On account of this, the lottery arranger may, through varying the quotient $p_c : p_a$, influence the demand for the lots offered for sale. Theoretically seen, the advertising costs and the size of the agents' commissions influence in the first place the probability that a person, chosen at random within the district for lot-selling, will take the question of buying a lot into earnest consideration. These costs may, moreover, increase the desire of the buyers to contribute to the purpose, worthy of having compassion on, for which the lottery is arranged. It is further possible, that a well-conducted advertising is able to influence the size of the figure k, which determines the lot-buyers' over-valuation of the chance of obtaining a great prize.

It may be proper to underline, in this connection, that the demand for lots cannot be understood as a function univocally determined by the lottery conditions. The number of lots q_α, bought by a certain buyer A_α, as well as the total number of bought lots q_a, must be understood as stochastic variables, and the form of the probability distribution function for the total number of bought lots is, moreover, influenced by the lottery conditions $\boldsymbol{p}$. We conceive the probability $f_\alpha(g; \boldsymbol{p})$ that the buyer A_α will buy g pieces of lots, as a

product of the probability of A_α's taking the lot-purchase into consideration, a probability $f_a\left(L_0; \frac{p_c'}{p_a}\right)$, which depends on the moving costs of the lottery arranger, and of the probability $f_a(g; L_0 \& \boldsymbol{p})$ that A_α will decide on buying g pieces of lots in case the question of lot-buying is subjected to his consideration. Because the number of lot-buyers is generally very great, and the resolutions of the different buyers might be considered to be approximately independent of each other, we may, conformably to the fundamental proposition of probability calculation, expect that the distribution function for the total number of sold lots $q_a = \sum_{\alpha=1}^{n_a} q_\alpha$ will be an approximately normally distributed statistic varable with a medium value $E(q_a) = \sum_{\alpha=1}^{n_\alpha} E[q_\alpha] = g_a$ and a variance $\sigma_a^2 = E(q_a - g_a)^2$ which increases rather proportionally with the medium quantity of sold lots g_a. The number g_a is proportional with a medium $f_a\left(L_0; \frac{p_c'}{p_a}\right)$ of the probabilities that the subjects A_α will take the question of buying lots of A_0 into reconsideration. Calculations we have done, based on different plausible distribution functions $f_a(g; L_0 \& \boldsymbol{p})$, point at $\sigma_a^2 : g_a = \sigma_{a_1}^2 \approx 1$.

The probability $f_a(g; \boldsymbol{p})$, that A_0 will succeed in selling $q_a = g$ pieces of lots, might therefore be approximately calculated by the formula

$$(30) \quad f_a(g; \boldsymbol{p}) = \boldsymbol{\Phi}\left(\frac{g + \frac{1}{2} - g_a}{\sigma_a}\right) - \boldsymbol{\Phi}\left(\frac{g - \frac{1}{2} - g_a}{\sigma_a}\right); \quad g \neq q$$

$$f_a(q; \boldsymbol{p}) = 1 - \boldsymbol{\Phi}\left(\frac{q - \frac{1}{2} - g_a}{\sigma_a}\right)$$

where

$$(31) \quad \boldsymbol{\Phi}(g) = \frac{1}{\sqrt{2\pi}} \int^{g} e^{-\frac{1}{2}x^2}\, dx; \; g_a = \frac{K_a}{p_a} \frac{h}{h + h_0} = \sum_{g=0}^{q} f_a(g; \boldsymbol{p}) \cdot g,$$

$$\sigma_a = \sqrt{g_a} \cdot \sigma_{a_1}; \; \sigma_{a_1} \approx 1, \; K_a = \sum_{\alpha=1}^{n_a} K_\alpha, \; h \approx \sqrt{(4 p_b - p_a) : 3 p_a},$$

$$h_0 = 1.5 \pm 0.5.$$

If $F(x;\ g\ \&\ \boldsymbol{p})$ denotes the conditioned probability that A_0 will obtain a gain $x_v \leq x$, in case the number of sold lots $q_a = g$, we may, by Bayes' proposition, calculate the probability that A_0 will obtain a gain $x_v \leq x$, by means of the formula:

$$F(x;\ \boldsymbol{p}) = \sum_{g=0}^{q} f_a(g;\ \boldsymbol{p})\, F(x;\ g\ \&\ \boldsymbol{p}). \tag{32}$$

The conditioned probability $F(x;\ g\ \&\ \boldsymbol{p})$ is a function that may, approximately at least, be calculated as soon as the number q_i of lots of different kinds i are known, and the mathematical value p_i of the contribution to x_v one lot of the kind i produces has been calculated. When designating the medium of the numbers p_i weighed with the number of lots of respective kinds $p = \sum_i p_i \frac{q_i}{q}$, the mathematical expectation of gain, in the case $q_a = g$, is a quantity which may be written in the form

$$\underset{q_a=g}{E(x_v)} = p\,g - p_{c0}\,q. \tag{33}$$

We designate further the variance in x_v, in the case $q_a = g = 1$, with

$$\sigma_p^2 = \sum_i (p_i - p)^2 \frac{q_i}{q}, \tag{34}$$

and the third moment, determining the skewness, round p with

$$\varrho_p^3 = \sum_i (p_i - p)^3 \frac{q_i}{q} \tag{35}$$

and the fourth moment round p with

$$\mu_p^4 = \sum_i (p_i - p)^4 \cdot \frac{q_i}{q}. \tag{36}$$

If we limit ourselves to survey a lottery in which every bought lot may only win a single prize, the probability, that the sold g pieces of lots will consist of such a lot number

combination $\boldsymbol{g}$ as at the lot-drawings will fall due with g_i pieces of prizes of the kind i, is a quantity which may be calculated by the formula

$$f(\boldsymbol{g};\, g) = \prod_i \binom{q_i}{g_i} \binom{q}{g}^{-1}. \tag{37}$$

In this formula the product of binomial coefficients $\prod_i \binom{q_i}{g_i}$ denotes, namely, the quantity of different lot number combinations, which are so conditioned that they consist of g_i pieces of lots of the kind i, while the denominator $\binom{q}{g}$ denotes the total quantity of different lot number combinations $\boldsymbol{g}$ having the quality of consisting of $g = \sum_i g_i$ pieces of lots. The distribution function $F(x;\, g\,\&\,\boldsymbol{p})$ may be calculated by the formula:

$$F(x;\, g\,\&\,\boldsymbol{p}) = \sum_{x(\boldsymbol{g}) \leq x} f(\boldsymbol{g};\, g);\;\; x(\boldsymbol{g}) = \sum_i g_i p_i - p_{c0}\, q. \tag{38}$$

If the number of prize-draws is relatively small, $F(x;\, g\,\&\,\boldsymbol{p})$ may easily be calculated by help of these formulas. If on the other hand the number of prize-draws is great, it becomes troublesome to calculate this function.

A good insight into the qualities of the distribution function $F(x;\, g\,\&\,\boldsymbol{p})$ is however possible, because we can write the so-called probability functional

$$G(\boldsymbol{p},\, \uparrow_x) = \int_{z=-\infty}^{\infty} (d\, F(z;\, p))\, \uparrow_x^z = \sum_{\boldsymbol{g}} (f(\boldsymbol{g};\, g)\, \uparrow_x^{x(\boldsymbol{g})}) f_a(g;\, \boldsymbol{p}) \tag{39}$$

in the form

$$G(\boldsymbol{p},\, \uparrow_x) = \overset{0}{\underset{y}{\uparrow}} \left(\uparrow_x^{-p_{c0}\, q} \prod (1 + \uparrow_x^{p_i} \uparrow_y)^{q_i} \binom{q}{y}^{-1} f_a(y;\, \boldsymbol{p}) \right). \tag{40}$$

In this formulas the operator $\uparrow_x$ has the quality $\uparrow_x^z U(x) = U(x + z)$ and $\uparrow_y^{g_i}$ the quality $\uparrow_y^{g_i} \binom{q}{y}^{-1} = \binom{q}{y + g_i}^{-1}$, $\uparrow_y^g = \overset{y+g}{\underset{y}{\uparrow}}$.

Replacing, in this functional, the variable increase symbol $\uparrow_x$ by e^D, and successively derivating the obtained function of D, we find rather easily the first three moments round the medium $m_v = E(x_v)$:

$$m_v = p\, g_a - p_{c0}\, q, \tag{41}$$

$$\sigma_v^2 = g_a \frac{(q - g_a - \sigma_{a1}^2)}{q-1} \sigma_p^2 = g_a (1-\theta)\, \sigma_p^2;\ \theta = \frac{g_a + \sigma_{a1}^2 - 1}{q-1}, \tag{42}$$

$$\varrho_v^3 = E\left(\frac{g(q-g)(q-2g)}{(q-1)(q-2)} \cdot \varrho_p^3\right) = \tag{43}$$

$$= \frac{g_a (q - g_a - 3\sigma_{a1}^2)(q - 2g_a) + 2\varrho_a^3}{(q-1)(q-2)} \varrho_p^3.$$

The exact expression for the fourth moment μ_v^4 is rather complicated

$$\mu_v^4 = 3\, E(\psi_1(g))\, \sigma_p^4 + E(\psi_2(g))\, \mu_p^4; \tag{44}$$

where

$$\psi_1(g) = \frac{g(g-1)(q-g)(q-g-1)q}{(q-1)(q-2)(q-3)},$$

$$\psi_2(g) = g\left(\frac{q-g}{q-1}\right)\left(1 - 6\frac{q(g-1)}{(q-2)(q-3)}\right),$$

and

$$E(\psi_1(g)) = \sum_{g=0}^{q} f_a(g; p)\, \psi_1(g) =$$

$$= \psi_1(g_a) + \frac{1}{2}\psi_1^{(2)}(g_a)\sigma_a^2 + \frac{1}{3!}\psi_1^{(3)}(g_a)\varrho_a^3 + \frac{1}{4!}\psi_1^{(4)}(g_a)\mu_a^4.$$

Thus

$$E(\psi_1(g)) = \psi_1(g_a) + \frac{qQ}{(q-1)(q-2)(q-3)};$$

$$Q = \{3[q^2 - (2(q-g_a)+1)g_a - 1]\sigma_a^2 - (2q - 4g_a - 1)\varrho_a^3 + \mu_a^4\},$$

and

$$E(\psi_2(g)) = \psi_2(g_a) - \frac{\{(q^2 + \ldots q - 18 g_a + 12)\sigma_a^2 - 6\varrho_a^3\}}{(q-1)(q-2)(q-3)}$$

When g_a and $q - g_a$ both are great, we make a relatively little error when putting

$$(44') \qquad \mu_v^4 \approx 3\,\sigma_v^4\left(1 + \frac{3\,\sigma_{a1}^2 - 1 + \theta}{g_a(1-\theta)} + \frac{1}{3}\left(\frac{1-6\,\theta}{1-\theta}\right)\frac{\mu_p^4}{\varrho_p^4\, g_a}\right) \approx 3\,\sigma_v^4.$$

The formulas (43) and (44) have not been published earlier, to my knowledge not even in the special case when $\sigma_a^2 = 0$. In this case the formula (42) gives a well known formula for the mean square deviation of the prize obtained by buying g_a pieces of lots.

If we approximate the medium-value formation function $U(x)$ by a polynom of fourth degree:

$$(45) \qquad U_4(x) = x + k_3\,x^3 - k_4\,x^4$$

the medium value of $U(x)$ becomes approximately equal to the medium value of $U_4(x)$, i.e.

$$(46) \qquad U(x_{\boldsymbol{p}u}) \approx \int_{\chi=-\infty}^{\infty} U_4(x)\,d\,F(x;\boldsymbol{p}) =$$

$$= U_4(m_v) + k_3(3\,m_v\,\sigma_v^2 + \varrho_v^3) - k_4(4\,m_v\,\varrho_v^3 + 6\,m_v^2\,\sigma_v^2 + \mu_v^4).$$

In order that $U(x_{\boldsymbol{p}u})$ will become a positive number in the case when m_v is negativ, the third moment ϱ_v^3 must be a relatively great positive number. If g_a becomes greater than $\frac{1}{2}\,q$, the third moment changes the sign, as appears from the formula (43).

As to the medium-value formation function $U(x) = x\left(1 + k\left(\frac{x}{K+x}\right)^2\right)$, it meets with mathematical difficulties to find simple expressions for the medium $E(U(x)) = U(x_{\boldsymbol{p}u}) = G(\boldsymbol{p}, \uparrow_x)\,U(x)$. If on the other hand we consider the following medium-value formation function, may be still better in theoretical respect

$$(47) \qquad U(x) = x(1 + k(1 - e^{-\frac{x}{K}})^2) =$$

$$= x(1 + k) - 2\,k\,x\,e^{-\frac{x}{K}} + k\,x\,e^{-\frac{2x}{K}}$$

we may calculate $U(x_{pu})$ by means of the formula

$$(48) \qquad U(x_{pu}) = (1+k)\,m_v - k\left(2\,\dot{G}(\boldsymbol{p}, e^{-\frac{1}{K}}) - \dot{G}(\boldsymbol{p}, e^{-\frac{2}{K}})\right).$$

In this formula the designation $\dot{G}$ denotes the derivate, in respect of u to the function $G(\boldsymbol{p}, e^u)$, that is

$$(49) \qquad \dot{G}(\boldsymbol{p}, e^u) = \frac{d}{d\,u}\left(\int_{-\infty}^{\infty} e^{ux}\,d\,F(x;\boldsymbol{p})\right) = \int_{x=-\infty}^{\infty} x\,e^{ux}\,d\,F(x;\boldsymbol{p}).$$

The formula (48) is specially usable, when the distribution function $F(x;\boldsymbol{p})$ is approximately normal, in which case

$$(50) \qquad G(\boldsymbol{p}, e^u) = e^{m_v u + \frac{1}{2}\sigma_v^2 u^2 + \frac{1}{6}\varrho_v^3 u^3},$$

and the derivate

$$(51) \qquad \dot{G}(\boldsymbol{p}, e^u) = (m_v + \sigma_v^2\,u + \tfrac{1}{2}\,\varrho_v^3\,u^2)\,G(\boldsymbol{p}, e^u).$$

We can then write $U(x_{pu})$ in the form

$$(52) \qquad U(x_{pu}) = (1+k)\,m_v - k\left(2\,m_v - \frac{2\,\sigma_v^2}{K} + \frac{\varrho_v^3}{K^2}\right)G(\boldsymbol{p}, e^{-\frac{1}{K}}) + \\ + k\left(m_v - \frac{2\,\sigma_v^2}{K} + \frac{2\,\varrho_v^3}{K^2}\right)G(\boldsymbol{p}, e^{-\frac{2}{K}}).$$

If we in this formula put the expressions for the medium value m_v of x_v, the second moment σ_v^2 and third moment ϱ_v^3 from the formulas (41), (42) and (43) we can express $U(x_{pu})$ as a function of p_a, p_b, p_c, p_{c0}, q, σ_p^2, ϱ_p^3, σ_{a1}^2, ϱ_{a1}^3 and g_a. If as usual g_a is a great number, we can without considerable change in $U(x_{pu})$ put $\sigma_{a1}^2 = 1$ and $\varrho_{a1}^3 = 0$. The number g_a is a function of the proceeding parameters that approximately may be written in the form

$$(53) \qquad g_a = \frac{K_a}{p_a} \cdot \frac{\gamma_c}{\gamma_0 + \gamma_c} \frac{\theta\,h_a}{1 + \theta\,h_a}$$

where γ_c is a function of the selling cost of the lots perhaps of the form

$$\gamma_c = \left(\frac{p'_c}{p_a} - \gamma\right)^{\frac{1}{2}}. \tag{54}$$

The symbol h_a denotes in accordance with formula (21) and (23) a function of $p_b : p_a$ of the form

$$h_a = \left(\frac{k_a + 1}{k_a}\frac{p_b}{p_a} - \frac{1}{k_a}\right)^{\frac{1}{2}}. \tag{55}$$

The symbols γ_0, γ, k_a, K_a and θ are so slowly changing functions of the proceeding parameters of the lottery arranger that they may be approximated by constants. If p_b/p_a and p'_c/p_a both increase above all limits $p_a g_a$ approach the limit K_a. If p_b is smaller than $p_a : (k_a + 1)$ nobody is interested of the lots. If $p'_c : p_a$ is smaller than γ nobody will take the job of selling lot for A_0. The number k_a is of the order of size 3 ± 1, the number $\gamma \approx 0.1 \pm 0.005$ and $\theta \approx 0.7 \pm 0.3$.

If A_0 tries to find the optimal values of the proceeding parameters q, p_b, p_a, p'_c, σ_p^2, ϱ_p^3, which maximate the function $U(x_{pu})$ he must solve the equation system

$$\frac{\delta U(x_{pu})}{\delta q} = \frac{\delta U(x_{pu})}{\delta p_b} = \frac{\delta U(x_{pu})}{\delta p_a} = \frac{\delta U(x_{pu})}{\delta p'_c} = \frac{\delta U(x_{pu})}{\delta \sigma_p^2} = \frac{\delta U(x_{pu})}{\delta \varrho_p^3} = 0. \tag{56}$$

If he approximate $U(x_{pu})$ by a function of the form (52) he can calculate these partial derivates. Using hypothetical values for the figures K_a, K, p_{c0}, k, k_a, γ, γ_0 and θ he can determine these derivates for a number of interesting examples. In this way he can get some information about the order of size of the optimal proceeding parameters.

It may be of interest to note that the above discussed formulas can be used in a general theory of economic activity. The classic price theory can not be used for a theory of lottery gambles.

THE INTERPLAY BETWEEN DECISION MAKERS

Leo Törnqvist
Helsinki

We consider a set B of decision makers b,b', etc. Every b has available a set A_b of alternative actions, which is extended taking into account the possibility of using mixed strategies. The decision to use a certain mixed strategy

$$s_b = \{s_{ba};\ a \in A_b\} \tag{1}$$

means using random sampling numbers for choosing an alternative $a \in A_b$ with the probability s_{ba}.

We assume that every decision maker $b \in B$ tries to maximize his own preference function

$$U_b = U_b(s_B) = U_b(s_b, s_{B-b}) \tag{2}$$

of the combinations $s_B = \{s_b;\ b \in B\}$ of mixed strategies available to the decision makers in B. The decision making problem of b can then be formulated as the problem of finding such a strategy $s_b = s_b(s_{B-b})$ for which this preference function reaches its maximum:

$$U_b = U_b(s_b(s_{B-b}), s_{B-b}) = \operatorname*{Max}_{s_b} U_b(s_b, s_{B-b}) \tag{3}$$

If the preference functions U_b, $b \in B$ have continuous partial derivatives $\partial U_b/\partial s_{ba}$ for every s_B, we define

$$\varepsilon_b(s_B) = \frac{\sum_{a \in A_b} s_{ba}\left(\frac{\partial U_b}{\partial s_{ba}} - \operatorname*{Max}_{a \in A_b} \frac{\partial U_b}{\partial s_{ba}}\right)}{\operatorname*{Min}_{a \in A_b} \frac{\partial U_b}{\partial s_{ba}} - \operatorname*{Max}_{a \in A_b} \frac{\partial U_b}{\partial s_{ba}}} \tag{4}$$

as a measure of non-optimality of the strategy s_b for given s_{B-b}. If for $s_B = \hat{s}_B$ we have

$$\varepsilon_b(s_B) = 0 \quad \text{for every } b \in B, \tag{5}$$

we say that $\hat{s}_B$ is a combination of mutually optimal strategies $\hat{s}_b$ for the decision makers b in B.

An approximate solution to the problem of finding mutually optimal strategies for decision makers in B might in principle be obtainable by means of the "stratify-sample-maximize-algoritm": stratify (preferably so that the variation in $\partial U_b/\partial s_{ba}$ is relatively smaller between representative alternatives from the same stratum than between representative alternatives from different strata) the set of admissible alternatives A_b for every $b \in B$ in at least two and at most for instance ten strata. Choose then at random a representative alternative from each stratum. Adjungate to each representative $a \in A_b$ a tentative probability

$$s_{ba}^{(\nu)} = n_{ba}^{(\nu)}/n_b^{(\nu)}, \quad \nu = 1,2,... \tag{6}$$

where $n_{ba}^{(\nu)}$ denotes the number of strata in A_b represented by $a \in A_b$, and $n_b^{(\nu)}$ the total number of strata in A_b after the νth stratification. Determine the winning representative in the contest

$$\frac{\partial U_b}{\partial s_{ba}} = \underset{a \in A_b}{\text{Max}} \frac{\partial U_b}{\partial s_{ba}}; \; s_{ba} = s_{ba}^{(\nu)} \neq 0, \; b \in B, \; \Sigma s_{ba} = 1. \tag{7}$$

Stratify anew (step $\nu + 1$) those strata to which the winning representatives belong and repeat the procedure until

$$\varepsilon_b(s_B^{(\nu)}) \leq \varepsilon \text{ for every } b \in B \tag{8}$$

where ε is a number that determines when a sufficiently good approximation to $\hat{s}_B$ has been reached. If some strata containing only one "indivisible" alternative have to be further subdivided according to the "stratify-sample-maximize-algoritm", we choose $n_{ba}^{(\nu+1)} = n_{ba}^{(\nu)} + 10$ if a is a winner, $n_{ba}^{(\nu+1)} = n_{ba}^{(\nu)}$ otherwise. Put then $n_b^{(\nu+1)} = \Sigma_{a \in A_b} n_{ba}^{(\nu+1)}$ and let the alternative a represent $n_{ba}^{(\nu+1)}$ fictional strata. If the partial derivatives $\partial U_b/\partial s_{ba}$ vary suffiently slowly by small variations in s_B, the procedure will converge towards

a solution $\hat{s}_B$ for the problem stated.

Supplementary ideas for attacking the decision making problem.

If the reaction of the other decision makers $b' \in B-b$ to some strategy of s_b can be considered as a known function $s_{B-b}(s_b)$, the optimal strategy of b will make

$$\bar{U}_b(s_b) = U_b(s_b, s_{B-b}(s_b)) \tag{9}$$

a maximum with respect to variations in s_b, and this strategy $\bar{s}_b$ will have the property

$$s_{ba}\left(\frac{\partial \bar{U}_b}{\partial s_{ba}} - \operatorname*{Max}_{a \in A_b} \frac{\partial \bar{U}_b}{\partial s_{ba}}\right) = 0 \tag{10}$$

for $s_{ba} = \bar{s}_{ba}$.

It might often be possible to consider $\bar{U}_b$ as a function of the form

$$\bar{U}_b(s_b) = U_b(s_b, y_b(s_b)) \tag{11}$$

where $y_b(s_b)$ can somehow directly be observed by b and is approximately a function dependent only on the reactions of other decision makers to the strategy used by b.

To make use of the rule $s_b = s_b(s_{B-b})$ for finding an optimal strategy for b as a function of s_{B-b}, the decision maker b will need some method for estimating the probabilities determining the strategies of $b' \in B-b$. One such method would be to gather information about how often the other decision makers have used the different alternatives of actions. Another method for getting some insight about the probabilities s_{B-b} is to use a method we could call the method of playing roles. The principal idea can be formalized as follows. Let $U_{b'b}$ denote the preference function of b when he plays the role of b'. When he plays his own role, we put $U_{bb} = U_b$. If he is able to obtain a reasonably good approximation $U_{b'b}$ to $U_{b'}$, he can play the whole game according to the hypothesis that $U_{b'b} =$

= $U_{b'}$ and try to find an approximately optimal solution $\hat{s}_{Bb}$. In practice he might by this role playing method find a better strategy $\hat{s}_{bb}$ for himself than if he tries to use only the very poor data about the strategy s_{B-b} really used he might have at his disposition. If he later on gets information which gives a better estimate $s_{b'b}$ for $s_{b'}$ than $\hat{s}_{b'b}$ he can adjust his strategy from $\hat{s}_{bb}$ to $s_{bb} = s_{bb}(s_{b'b}, \hat{s}_{(B-b')b})$ and thus make use of whatever new information he can get. In principle $U_{b'b}$ is a synthesis of the information available for b about the preferences and situation of b' and his own preferences when b thinks he would have to make the decisions b' in fact has to make. Thus $U_{b'b}$ is a more or less crude estimate of $U_{b'}$.

The question under which conditions the attempts of the decision maker b in B to maximize U_b will lead to the optimal set $\hat{s}_B$ seems to be quite intricate. It seems a priori to be possible that the attempts to maximize U_b lead to a process of changes in s_B that does not converge towards any stable solution of the problem at all.

The reasons for the existence of complex interplay.

If U_b is very sensitive to changes in the strategies of b' but $U_{b'}$ changes very little by some changes in $s_{b'}$ which favour b, then the decision maker b will have a great interest in finding methods for influencing b' to make such changes in $s_{b'}$ favourable to b. By making use of side-payments, conditional promises etc. b might be able to make it favourable also for b' to make such changes in his strategies. We see that the decision makers in B very easily will become involved in a very complicated game of finding methods for influencing the choice of strategies of other players in a direction which is favourable for all or some subset of the decision makers

in B. This interplay can be rationalized by making use of organized decision making. The decision makers $b' \in B'$ may decide to form a new decision making unit $C_{B'}$ for some types of decisions to be made. By means of somehow organized decision making it is possible to change the set of available alternatives of action for the members of the set and more easily than otherwise obtain high probabilities for certain combinations $a_{B'}$ of co-operations which are considered to be especially good for all $b' \in B'$. The problem of incentives, sanctions, communication, common resources and inner hierarchy of the organization will require decision making activity of a kind typical for organized decision making. The interplay between the decision makers may also influence the form of the preference functions U_b of the individuals and thus have very deep going consequences for all decision makers.

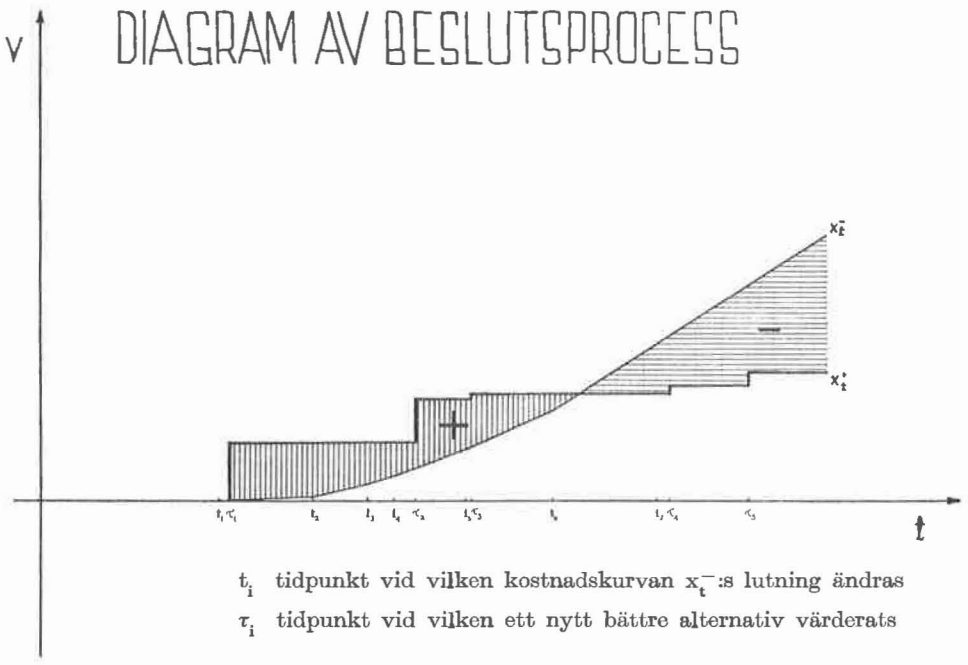
BESLUTSVETENSKAPEN OCH MÄNNISKANS STRÄVAN ATT STYRA SKEENDET

Människan försöker i sin dagliga gärning styra eller åtminstone påverka händelsernas gång medelst beslutsprocesser. Dessa mångskiftande och ofta komplexa processer synes mig värda ett intensivt och systematiskt studium.

Våren 1968 utgav pol.lic. Nordberg och jag boken »Päätäntätieteen keskeisiä ongelmia», dvs. beslutsvetenskapens centrala problem, i vilken vi bl.a. behandlar två särskilt intressanta egenskaper hos beslutsprocesser, nämligen deras stokastiska och deras tidskrävande karaktär.

En genuin beslutsprocess karakteriseras av att beslutsfattaren i en viss beslutssituation kan välja mellan flera alternativa handlingar, vilka förknippas med en föreställning om en av alternativet betingad sannolikhetsfördelning över möjliga förlopp för skeendet i framtiden — en föreställning, som beslutsfattaren tilldelar ett visst värde relativt till andra konkurrerande alternativs motsvarande föreställningar. Hans uppgift$\hat{?}$ är att i en sådan situation $h_{\to t}$ finna det bästa alternativet a bland en ofta oöverskådligt stor mängd A av alternativ. Tyvärr är själva letandet tidskrävande varför beslutssituationen och därmed alternativens relativa värde förändras under beslutsprocessens gång. Uppskovet med att fatta ett aktivt beslut kan lätt komma att kosta mer än värdedifferenserna mellan alternativen. Frågan uppstår hur länge man skall fortsätta letandet innan man anser sig ha nått ett tillfredsställande handlingsalternativ.

Situationen kan åskådliggöras med en bild. Horisontalaxeln anger tiden och den vertikala det subjektiva värdet. Beslutsfattaren värderar olika alternativ $a_1, a_2, \ldots$. Alternativens subjektiva värde är en stokastisk variabel. Den aldrig avtagande trappstegsfunktionen x_τ^+ är det största värdet bland de alternativs väntevärden, som värderats före tidpunkten τ. Det under processens gång tillämpade alternativet betecknas med a_0. Behovet att fatta ett beslut annat än a_0 blir brännande först då kostnaderna för att förbli passiv börjar stiga. Enligt diagrammet åstadkommer tillämpandet av a_0 ingen kostnad x_t^- före t_1, då beslutskostnaderna slumpmässigt börjar stiga. Medan beslutsfattaren är upptagen med att behandla det först inkomna ärendet tänker vi oss att nya ärenden anmäler sig. Höjden på kostnadskurvan x_t^- anger alltså vid var tidpunkt t de stigande kostnader beslutsfattaren pådrar sig ifall han framhärdar med a_0.



t_i tidpunkt vid vilken kostnadskurvan x_t^-:s lutning ändras

τ_i tidpunkt vid vilken ett nytt bättre alternativ värderats

Bild 1. Stokastisk realisation av beslutsprocess

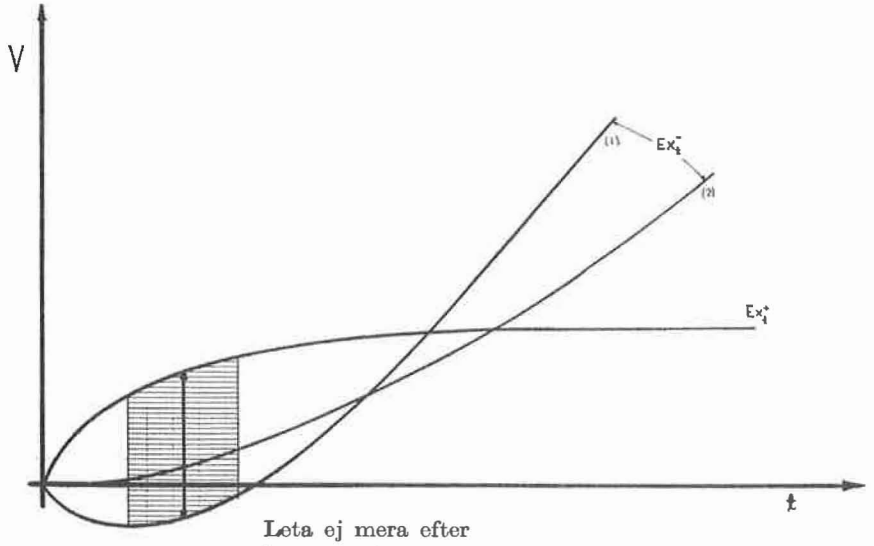


Bild 2. Väntevärdeutveckling av beslutsprocess

Trots att beslutsfattaren under processens gång finner allt bättre alternativ kan hans handlande bli dumt. Låter han sig tillräckligt länge gäckas av lusten att finna det bästa beslutet $\hat{a}$, befinner han sig snart till höger om skärningspunkten mellan x_i^+ och kostnadskurvan x_i^-. Man kan isåfall paradoxalt säga att det bästa är det godas fiende.

Situationen kompliceras av att en god värdering av ett alternativ är ett väntevärde för en subjektiv sannolikhetsfördelning och i regel avviker det faktiskt använda jämförelsevärdet slumpmässigt från detta väntevärde. Man vet alltså inte med någon större säkerhet vad beslutsalternativets värde är relativt till a_0. Man kan tänka sig att spridningen för dessa sannolikhetsfördelningar avtar med värderingstiden, vilket gör att beslutsfattaren känner sig mest osäker i början av beslutsprocessen, då fjäsk lätt resulterar i förhastade beslut.

Beslutsprocessen är alltså till sin grundkaraktär stokastisk. Vår bild analyserar dock blott en enda beslutsfattare i en specifik beslutssituation. En allmän behandling förutsätter att man betraktar samspelet mellan beslutsfattare, vilka fattar sekvenser av beslut under ett manipulerbart socialt tryck. Även i det fall att den enskilde individen skulle ha klart för sig de samband vi illustrerat i diagrammet och skulle kunna fatta goda beslut är steget ännu långt till förnuftiga beslutssekvenser och en klok koordinering av flera individers aktioner.

Mänskans strävan mot en förnuftig styrning av skeendet i vidare bemärkelse behöver större resurser än hittills. Framförallt krävs det att vi frigör oss från en mängd illusioner om och idealiseringar av beslutsprocessernas natur och att vi accepterar slumpens ofrånkomlighet. Jag är övertygad om att beslutsvetenskapen kan göra en värdefull insats i mänskans strävan att påverka skeendet i en förnuftig riktning.

A Method for Calculating Changes in Regression Coefficients and Inverse Matrices Corresponding to Changes in the Set of Available Data

By **Leo Törnqvist** (Helsingfors)

1. Introduction

When we have made a regression analysis, for instance on the basis of time series, it is often of interest to know how the results would change if we take into account observations made later on. Because it seems that the whole work of solving the normal equations must be made over again, we seldom continue the calculations by taking into account later information. It is, however, easy to find the adjustments required by a method developed in this article. It is possible to get time series showing the development of the regression coefficients without formidable work. We can in this way get a deeper insight in the problem to be studied than by making the regression analysis only once for all. If the purpose of the regression analysis is to obtain formulas to be used for forecasting, time series of regression coefficients give a better starting point than if we only have regression coefficients for a certain period.

The basic idea can also be utilized for attacking other problems of interest.

2. A Method of Adjusting Regression Coefficients when Adding New Data

If we use matrix notations, the regression coefficients $\alpha_1, \ldots, \alpha_n$ are the solutions

$$\alpha = A^{-1} c \tag{1}$$

of the equations

$$A\alpha = c, \tag{2}$$

where $A = \{a_{ij}\}$ denotes the coefficient matrix of the equations and c the constant vector.

We assume that the elements α_{ij} of the matrix A^{-1} have been calculated for a certain set of data. We then take a new observation $x_1, \ldots, x_n$ of the independent variables. If this observation is also included in the set of data used for determining the regression coefficients, the elements a_{ij} of A would be increased by $x_i x_j$. If the new matrix obtained in this way is denoted by B, we have

$$B = A + x\,x' = \{b_{ij}\} = \{a_{ij} + x_i\,x_j\}, \tag{3}$$

where x' denotes the row vector corresponding to the column vector x.

Instead of the normal equations (2) we obtain the new normal equations

$$B\beta = c + x\,y', \tag{4}$$

where the scalar y' denotes the new observation of the variable to be explained by the regression formula

$$y' = x'\beta + \varepsilon_\beta = x'\alpha + \varepsilon_\alpha, \tag{5}$$

where ε_β and ε_α denote the unexplained error terms obtained when using the regression coefficients β and α respectively.

To find the inverse matrix B^{-1} giving

$$\beta = B^{-1}(c + x\,y'), \tag{6}$$

the following idea leads to a surprisingly simple formula.

We start from the identity

$$B^{-1} = A^{-1} - A^{-1}(B - A)\,B^{-1} = A^{-1} - A^{-1}\,B\,B^{-1} + A^{-1}\,A\,B^{-1}. \tag{7}$$

Because according to (3) we have $B - A = x\,x'$, we can write (7) in the form

$$B^{-1} = A^{-1} - (A^{-1}x)\,(x'B^{-1}), \tag{8}$$

from which follows

$$x'B^{-1} = x'A^{-1} - (x'A^{-1}x)\,(x'B^{-1}), \tag{9}$$

and thus

$$x'B^{-1} = \frac{x'A^{-1}}{1 + x'A^{-1}x}, \tag{10}$$

where the denominator is a scalar. By inserting (10) in (8) we get

$$B^{-1} = A^{-1} - \frac{(A^{-1}x)(x'A^{-1})}{1 + x'A^{-1}x}, \tag{11}$$

and finally by inserting (11) in (6)

$$\beta = \alpha + \frac{A^{-1}x(y' - x'\alpha)}{1 + x'A^{-1}x} = \alpha + \frac{A^{-1}x\varepsilon_\alpha}{1 + x'A^{-1}x}. \tag{12}$$

To the matrix formula (11) corresponds the following formula (13) for the elements β_{ji} of the new inverse matrix B^{-1}:

$$\beta_{ji} = \alpha_{ji} - \frac{\sum_i \alpha_{ji}x_i \sum_j x_j \alpha_{ji}}{1 + \sum_{j,i} x_j \alpha_{ji} x_i} = \alpha_{ji} - \alpha_{jx}\gamma\,\alpha_{xi}, \tag{13}$$

where

$$\gamma = \frac{1}{1 + \sum_{j,i} x_j \alpha_{ji} x_i} = \frac{1}{1 + \sum_i \alpha_{xi}x_i} = \frac{1}{1 + \alpha_{xx'}}, \tag{14}$$

$$\alpha_{jx} = \sum_i \alpha_{ji}x_i, \quad \alpha_{xi} = \sum_j x_j \alpha_{ji}. \tag{15}$$

The new regression coefficients β_j can in correspondence to (12) be calculated according to the formula

$$\beta_j = \alpha_j + \alpha_{jx}\gamma\,\varepsilon_\alpha, \tag{16}$$

where

$$\varepsilon_\alpha = y - x'\alpha$$

denotes the unexplained part of y when we use the old regression coefficients α.

The calculation of the new regression coefficients by means of the formula (16) requires very little work. The formula (13) gives the elements of the new inverse matrix. If we get more data we repeat the procedure and can by this method get a whole sequence of regression coefficients. This sequence will refer to a successively increasing period of time covered by the data used. The coefficients will therefore not be strictly comparable with each other. The following method will give us a sequence of regression coefficients, which can be considered as comparable to each other.

3. A Method for Getting a Sequence of Comparable Regression Coefficients

If we in the regression analysis successively take into account new data, the mean lag between different data used and the latest

data will increase with increasing length of the time period covered. This tendency could be counterbalanced by leaving out the oldest data and correcting the regression coefficients by means of a formula corresponding to (16). It seems, however, to be better to subtract from the old moments such a fraction that after adding the new terms $x_i x_j$ the mean lag will remain unchanged. If the mean lag is T and the time interval between successive observations is τ, the fraction ϱA, where $\varrho = \tau/(T+\tau)$, will have this property. The new moment matrix and the new normal equations will then be

$$B = (1-\varrho) A + x x', \tag{3,ϱ}$$

$$B\beta = (1-\varrho) c + x y', \tag{4,ϱ}$$

giving instead of (13) and (16) the formulas

$$\beta_{ji}(\varrho) = \frac{1}{1-\varrho}(\alpha_{ji} - \alpha_{jx}\gamma_\varrho \alpha_{xi}), \tag{13,ϱ}$$

$$\beta_j(\varrho) = \alpha_j + \alpha_{jx}\gamma_\varrho \varepsilon_\alpha, \tag{16,ϱ}$$

where

$$\gamma_\varrho = \frac{1}{1-\varrho+\alpha_{xx'}}. \tag{14,ϱ}$$

The solution (16,ϱ) will minimize the discounted sum of squares

$$Q^2_{t,\beta_t} = \varepsilon^2_{t,\beta_t} + (1-\varrho_t) Q^2_{t-\tau,\beta_t}, \tag{17}$$

where ε_{t,β_t} denotes the unexplained part of y_t when we use the regression coefficients β_t for the last time point t, and τ denotes the equidistant time difference between the observations and $(1-\varrho_t)$ the last discounting factor used. In the classical approach we have $\varrho_t = 0$ and

$$Q^2_{t,\beta_t} = \sum_\nu \varepsilon^2_{t-\nu\tau,\beta_t}. \tag{18}$$

In principle we could try, when we have plenty of data, to determine $\varrho_t = \varrho$ so that

$$s^2_{t,\alpha_t} = Q^2_{t,\alpha_t}/Q^{(0)}_t, \tag{19}$$

where

$$Q^2_{t,\alpha_t} = \varepsilon^2_{t,\alpha_t} + (1-\varrho_t) Q^2_{t-\tau,\alpha_{t-\tau}} \quad \text{and} \quad Q^{(0)}_t = 1 + (1-\varrho_t) Q^{(0)}_{t-\tau},$$

will become as small as possible. Here $\varepsilon^2_{t,\alpha_t}$ denotes the square of the unexplained part of y_t when the regression coefficients $\alpha_{t,j} =$

$= \beta_{t-\tau,j}$ are used. In practice we must, I think, calculate s^2_{t,α_t} for instance for $\varrho = 0$, 0.05, 0.1 and 0.2, and by means of interpolation methods determine an approximation to the optimal value of ϱ.

It can be mentioned that

$$\varepsilon_{t,\beta_t} = (1-\varrho_t)\gamma_{\varrho t}\varepsilon_{t,\alpha_t} \tag{20}$$

while

$$Q^2_{t,\beta_t} = \varepsilon_{t,\alpha_t}\varepsilon_{t,\beta_t} + (1-\varrho_t)Q^2_{t-\tau,\beta_{t-\tau}}. \tag{21}$$

An analogue to the classical standard error of the regression coefficients would be

$$\varepsilon(\beta_{jt}) = s_{t,\alpha}\sqrt{\beta_{jjt}}. \tag{22}$$

It is, however, clear that the assumptions made for the derivation of this formula are not in general valid for autocorrelated time series. It is therefore safer to estimate the standard error of β_{jt} directly from the time series obtained for β_{jt}.

4. Generalizations and Other Applications of the Basic Results

We have earlier assumed that the difference $B - A$ between the new and old matrix is of the form xx'. If we instead assume that

$$B - A = zx', \tag{23}$$

the corresponding formula for B^{-1} becomes

$$B^{-1} = A^{-1} - \frac{A^{-1}z \cdot x'A^{-1}}{1 + x'A^{-1}z}. \tag{24}$$

This result can be used for calculating the changes $\beta_{ji} - \alpha_{ji}$ corresponding to the changes in only one row or column of A. If we for instance put $x_j = \Delta a_{hj}$ while $z_i = 0$, $i \neq h$, $z_h = 1$, we get the new elements β_{ji} according to the formulas

$$\begin{aligned} \beta_{ji} &= \alpha_{ji} - \alpha_{jh}\, v_{hi}/(1+v_{hh}), \\ \beta_{jh} &= \alpha_{jh}/(1+v_{hh}), \end{aligned} \tag{25}$$

where $v_{hi} = \sum_j \Delta a_{hj}\alpha_{ji}$.

If we like to change both the elements in the row $i = h$ and the column $j = k$, this can be done in one single step by means of the formulas

$$\left.\begin{aligned}
\beta_{kh} &= \frac{\alpha_{kh}}{u_{kk}v_{hh} - \alpha_{kh}w_{hk}} = \gamma_{hk}\alpha_{kh},\\
\beta_{ki} &= \gamma_{hk}(\alpha_{ki}v_{hh} - \alpha_{kh}v_{hi}) = \gamma_{hk}\lambda_{hki}, \quad i \neq h,\\
\beta_{jh} &= \gamma_{hk}(\alpha_{jh}u_{kk} - \alpha_{kh}u_{jk}) = \gamma_{hk}\lambda_{jhk}, \quad j \neq k,\\
\beta_{ji} &= \alpha_{ji} + \frac{\beta_{jh}\beta_{ki}}{\beta_{kh}} - \frac{\alpha_{jh}\alpha_{ki}}{\alpha_{kh}} = \alpha_{ji} + \frac{\gamma_{hk}\lambda_{hki}\lambda_{jhk}}{\alpha_{kh}} - \frac{\alpha_{jh}\alpha_{ki}}{\alpha_{kh}}\\
&= \alpha_{ji} - u_{jk}\beta_{ki} - \gamma_{hk}\alpha_{jh}(v_{hi}u_{kk} - \alpha_{ki}w_{hk})\\
&= \alpha_{ji} - v_{hi}\beta_{jh} - \gamma_{hk}\alpha_{ki}(u_{jk}v_{hh} - \alpha_{jh}w_{hk}), \quad i \neq h, j \neq k,
\end{aligned}\right\} \tag{26}$$

where

$$\begin{aligned}
u_{jk} &= \sum_{i \neq h} \alpha_{ji}\Delta a_{ik}, \quad j \neq k, \qquad u_{kk} = \sum_{i \neq h} \alpha_{ki}\Delta a_{ik} - 1,\\
v_{hi} &= \sum_{j \neq k} \Delta a_{hj}\alpha_{ji}, \quad i \neq h, \qquad v_{hh} = \sum_{j \neq k} \Delta a_{hj}\alpha_{jh} - 1,\\
w_{hk} &= \sum_{i \neq h} v_{hi}\Delta a_{ik} - \Delta a_{hk}.
\end{aligned}$$

The formulas (26) remain valid even if we throughout use b_{ik} instead of Δa_{ik} and b_{hj} instead of Δa_{hj} at calculating u_{jk}, u_{kk}, v_{hi}, v_{hh} and w_{hk}, and at inserting in (26) use $u'_{kk} = u_{kk} - 1 = \sum_{i \neq h} \alpha_{ki} b_{ik}$ instead of u_{kk} and $v'_{hh} = v_{hh} - 1 = \sum_{j \neq k} b_{hj}\alpha_{jh}$ instead of v_{hh}.

In the special case when $b_{hj} = 0$, $j \neq k$ and $b_{ik} = 0$, $i \neq h$ $b_{hk} = 1$ we obtain the classical "simplex" formulas

$$\beta_{ji} = \alpha_{ji} - \frac{\alpha_{jh}\alpha_{ki}}{\alpha_{kh}}, \qquad \beta_{kh} = 1 \tag{27}$$

for leaving out the row $i = h$ and the column $j = k$.

In the case where $\alpha_{ki} = 0$, $i \neq h$ and $\alpha_{jh} = 0$, $j \neq k$, $\alpha_{kh} = 1$, and thus $a_{ik} = 0$, $i \neq h$, $a_{hj} = 0$, $j \neq k$, $a_{hk} = 1$, we obtain the special formulas

$$\left.\begin{aligned}
\beta_{ji} &= \alpha_{ji} - u_{jk}\beta_{ki}, \quad i \neq h, \ j \neq k,\\
\beta_{ki} &= -\beta_{kh}v_{hi}, \quad i \neq h,\\
\beta_{jh} &= -\beta_{kh}u_{jk}, \quad j \neq k,\\
\beta_{kh} &= 1/(1 - w_{hk})
\end{aligned}\right\} \tag{28}$$

for adding the new row $i = h$ and the new column $j = k$.

The inversion of a matrix B can be done by means of (28) by calculating a sequence of inverse matrices B_ν^{-1}. When this method is used in solving linear programming problems, the primarily used

identification numbers i and j do not necessarily have to denote the ordinal numbers for rows and columns. Let the subscripts of the somehow rearranged rows and columns be $(i_1, \ldots, i_n)$ and $(j_1, \ldots, j_n)$ respectively. By ordering the rows and columns it may be a good principle to use such an ordering that $|b_{i_\nu j_\nu}| \geqslant |b_{i_\varrho j_\mu}|$ for $\varrho \geqslant \nu$, $\mu \geqslant \nu$. We may then define the sequence of matrices B_ν to be inverted by means of the rule

$$B_\nu = \{b_{\nu_{i_\varrho} j_\mu} = b_{i_\varrho j_\mu} \text{ for } \varrho \geqslant \nu, \mu \geqslant \nu,\ b_{\nu_{i_\varrho} j_\mu} = I_{\varrho\mu} \text{ for } \varrho > \nu, \mu > \nu\}, \quad (29)$$

where $I_{\varrho\mu} = 0$ if $\varrho \neq \mu$, $I_{\varrho\varrho} = 1$ and $\nu = 1, 2, \ldots, n$. At each step $\nu \to \nu + 1$ one new row with $\nu + 1$ elements differing from $I_{\nu+1,\mu}$ and one new column differing from $I_{\varrho,\nu+1}$ will be added. By this method we obtain the last inverse matrix $B_n^{-1} = B^{-1}$ after in all n^3 multiplications or divisions. In the symmetric case we can obtain B^{-1} by making only $(n^3 + n)/2$ such operations. In regression analysis all the matrices B_ν^{-1} obtained as byproducts during the calculations can be of direct interest.

The basic formulas (26) can be used for obtaining many different sequences of inverse matrices B_ν^{-1} of interest in a regression analysis.

If we have performed the preliminary calculations with only a few number of decimals. we may check and increase the accuracy of the results $\{\alpha_{ji}\}$ by means of the formula

$$\beta_{ji} \approx \alpha_{ji} - \frac{d_{ji}}{1+z} + \varkappa (d_{ji} - z\,\alpha_{ji}), \quad (30)$$

where

$$d_{ji} = \sum_q \sum_p \alpha_{jp}\,\Delta a_{pq}\,\alpha_{qi} = \sum_q z_{jq}\,\alpha_{qi},$$

$$z = \sum_q \sum_p \alpha_{qp}\,\Delta a_{pq} = \sum_q z_{qq},$$

$$z_{jq} = \sum_p \alpha_{jp}\,\Delta a_{pq} = \sum_p \alpha_{jp} b_{pq} - I_{jq},$$

$$\varkappa = \frac{z^{(2)} - z^2}{z^{(2)} - z^2 + (n-1)z},$$

$$z^{(2)} = \sum_q \sum_p z_{qp}\,z_{pq} = \sum_q z_{qq}^{(2)}.$$

This approximation formula is exact if $\Delta a_{pq} = -\varrho\,a_{pq} + z_p x_q$. Even if this is not exactly true, the formula (30) gives a very good approximation to β_{ji} whenever the discrepancies from the exact case are not very large.

5. Remarks

I have to thank Mr. Aarni Nyberg for checking the computations. He has at the same time studied a part of the vast literature on the subject and discovered that Bartlett [1] has earlier found the facts expressed by the formula (24). Further it seems that many authors have developed different variants of the submatrix inversion method mentioned by Frazer, Duncan and Collar [2]. These variants seem to be applications of the mathematical facts expressed in the formula (28). As all articles dealing with similar problems have not been available, other details of our results may have been published earlier. I hope that at least some of the results will be really new contributions to the subject. I suppose that the formulas containing the discounting factor $(1-\varrho)$ are new. I think that the presentation of the results given will be helpful for those who, like myself, sometimes make extensive use of regresssion analysis. The methods described have been very efficient in an investigation of the interdependence between a large set of time series describing the economic development of the Post and Telegraph Office in Finland. When this investigation has been completed, it will be possible to demonstrate the usefulness of my approach numerically.

References

[1] M. S. Bartlett, An inverse matrix adjustment arising in discriminant analysis. *Annals of Math. Stat.*, Vol. 22 (1951), pp. 107–111.

[2] R. A. Fraser, W. J, Duncan, and A. R. Collar, Elementary Matrices. Cambridge University Press, 1938.

[3] P. S. Dwyer and F. V. Waugh, On errors in matrix inversion. *Journ. Amer. Stat. Ass.*, Vol. 48 (1953), pp. 289–319.

[4] V. I. West, Replacing variables in correlation problems. *Journ. Amer. Stat. Ass.*, Vol. 47 (1952), pp. 185–190.

Numerous other references can be found in [3].

(Received December 1957.)

INDUSTRIPRODUKTIONENS UTVECKLING I FINLAND ÅREN 1925—56

Av professor LEO TÖRNQVIST

Den 27 maj 1958 granskades vid Helsingfors universitet en ekonometrisk doktorsavhandling av OLAVI NIITAMO[1]), behandlande produktivitetens utveckling i Finlands industri under åren 1925—52. I min egenskap av ex officio opponent framförde jag bland annat följande synpunkter. Författaren har med berömvärd flit och seghet angripit ett krävande ekonometriskt problem och utfört ett grundläggande arbete för att beskriva och förklara produktivitetens utveckling i Finlands industri. Då undersökningen är den första i sitt slag i Finland, har författaren haft att utföra ett röjningsarbete på det vidsträckta arbetsfält, som består i att förklara utvecklingen av viktiga ekonomiska tidsserier. Doktor Niitamo har omsorgsfullt konstruerat en mängd tidsserier, som belysa den betraktande tidsperiodens kvantitativa ekonomiska historia. Ekonometriska forskningsresultat äro i allmänhet inte slutgiltiga. Detta gäller naturligtvis också Niitamos resultat. En relativt framgångsrik attack på ett intressant ekonometriskt problem har en förmåga att väcka intresse och locka till nya försök att angripa det redan angripna problemet eller liknande problem. Ett sådant försök gjorde jag redan till disputationstillfället, då jag omnämnde några nya modeller, som bättre än Niitamos kunde förklara utvecklingen åren 1925—52.

Vid ifrågavarande tillfälle framhöll jag att en ekonometrisk undersöknings kvalitetsnivå väsentligt beror på hur väl forskaren lyckas lösa följande grundproblem:

1. Formulering av forskningsuppgiften, d.v.s. en precisering av syftemålet med undersökningen. Härvid måste man angiva vilka statistiska variabler, som man önskar förklara med tillhjälp av

[1]) OLAVI NIITAMO: *Tuottavuuden kehitys Suomen teollisuudessa vuosina 1925—52.* (Produktivitetens utveckling i Finlands industri åren 1925—52.) Kansantaloudellisia Tutkimuksia XX; även som akad. avh. English Summary. 192 sid. Helsingfors 1958. Mk 900:—.

kvantitativa uppgifter beskrivande en viss grupp av statistiska variabler, vilkas variationer måhända kan förklara variationerna i de variabler, som undersökningen syftar till att förklara.

2. En omsorgsfullt utförd kvalitetskontroll av det tillbudstående materialet. Härvid måste man undersöka, om det tillgängliga siffermaterialet med tillfredsställande noggrannhet under hela undersökningsperioden kan anses beskriva variationerna i de statistiska variabler, som man har behov av för att uppnå undersökningens syftemål.
3. Byggandet av en modell, som anger det förmodade matematiskt-statistiska sambandet mellan de studerade variablerna, varvid man har att bedöma, huruvida sambandet kan bättre uttryckas medelst additiva eller multiplikativa kopplingar mellan de ifrågavarande variablerna och de vid detta stadium av undersökningen ännu obekanta parametrar, som ingå i modellen.
4. Problemet, att med utgångspunkt från det tillbudstående siffermaterialet skatta de obekanta parametrarnas talvärden under antagandet att de förblivit konstanta vid de variationer i de studerade variablerna, som beskrivas av det tillbudstående siffermaterialet. Studiet av variansen och andra statistiska egenskaper hos avvikelserna mellan de enligt modellen förväntade och observerade talen.
5. En prövning av huru väl de tal, som kunna beräknas med tillhjälp av modellen, överensstämma med nytt material, som icke utnyttjats vid modellens härledning.
6. Nyanpassning av modellen med utnyttjandet av nytt material.
7. Frågan angående utsträckningen av modellens approximativa giltighetsområde.
8. Bedömningen av värdet av den information, som sinsemellan tävlande modeller ger för uppnåendet av undersökningens syfte.
9. Redogörelsen för resultaten och forskningstekniken på ett sätt, som är möjligast lättfattligt, intresseväckande och verkningsfullt för dem, som möjligen kunde utnyttja dessa resultat för praktiskt bruk.
10. Den vid uppgifternas lösning använda forskningsteknikens elegans och effektivitet.

Syftemålet med denna artikel är att förklara utvecklingen av volymindexen för Finlands industriproduktion sedan år 1925 medelst variationerna i användningen av de viktigaste produktionsfaktorerna och i utrikeshandeln samt skolbildningen hos Finlands folk genom att utgå ifrån de serier och modeller, som Niitamo studerat. Genom att något förbättra de serier, som Niitamo använder för energiinsatsen och utrikeshandeln, samt utelämna en tilläggsbetingelse för parametrarna i hans modeller — som möjligen kan anses vara saklogiskt motiverbar, om man accepterar tankegången att ifall enbart antalet sysselsatta och energiförbrukningen fördubblades, så skulle även produktionen fördubblas — visar det sig att man kan erhålla väsentligt bättre resultat. Vidare är det avsikten, att undersöka huru väl de förbättrade modellerna för åren 1925—52 förmå förklara industriproduktionens utveckling under åren 1953—56. En sådan beräkning fungerar som en god test på modellens effektivitet. Av intresse är även att se, huru modellens parametrar förändras då det nya materialet ett år i sänder beaktas vid härledningen av modellen.

Med tanke på att göra en mycket effektiv beräkningsmetod, som jag utarbetat för dylika ändamål[1]), tillgänglig för andra forskare, synes det vara behövligt att i korthet angiva, huru beräkningarna genomförts vid härledningen av de nya modellerna.

Efter att ha utfört en del preliminära försök har jag kommit fram till följande modell

$$\text{Log } Q = a_0 + a_1 \log L_m + a_2 \log H_r + a_3 \log E + a_4 \log I_q + a_5 \log W_q + u = \log Q^* + u$$

Q Industriproduktionens volymindex.

L_m Antalet årsarbetare i industrin.

H_r Den nulevande mellanskolegångna befolkningens förhållande till befolkningen i arbetsduglig ålder.

E Energiserien, som har beräknats så att till industrins totala elektricitetsförbrukning har adderats den energimängd, som uppskattats ha använts av andra kraftmaskiner än elektriska motorer. Den sistnämnda ha vi uppskattat genom att med konstanten 2.300 multiplicera antalet hästkrafter, som dessa maskiner kunna prestera, så att vi erhållit ett närmevärde för den motsvarande mängden

[1]) *Regressioanalyysin laskelmien käytännöllinen suorittaminen esimerkin valossa.* Helsingin Yliopiston Tilastotieteen laitos, moniste n:o 2 (april 1958).

kilowatt-timmar. Industrins elektricitetsförbrukning har för åren 1925—29 uppskattats till 91% av elektricitetsproduktionen under dessa år.

I_q En obetydligt utjämnad serie för importvolymen $\acute{I}_q$ för varor för produktionsändamål. Utjämningen har skett enligt formlerna

$$I_{qt} = (\bar{I}_{qt} + I'_{qt})/2 \text{ där } I_{qt} = (\bar{I}_{qt-1} + I'_{qt})/2$$

Härigenom beaktas i någon mån lagerhållningen av importerade varor och de tjänster, som under tidigare år importerade varor gjort.

W_q Exportvolymen

u Den oförklarade resttermen.

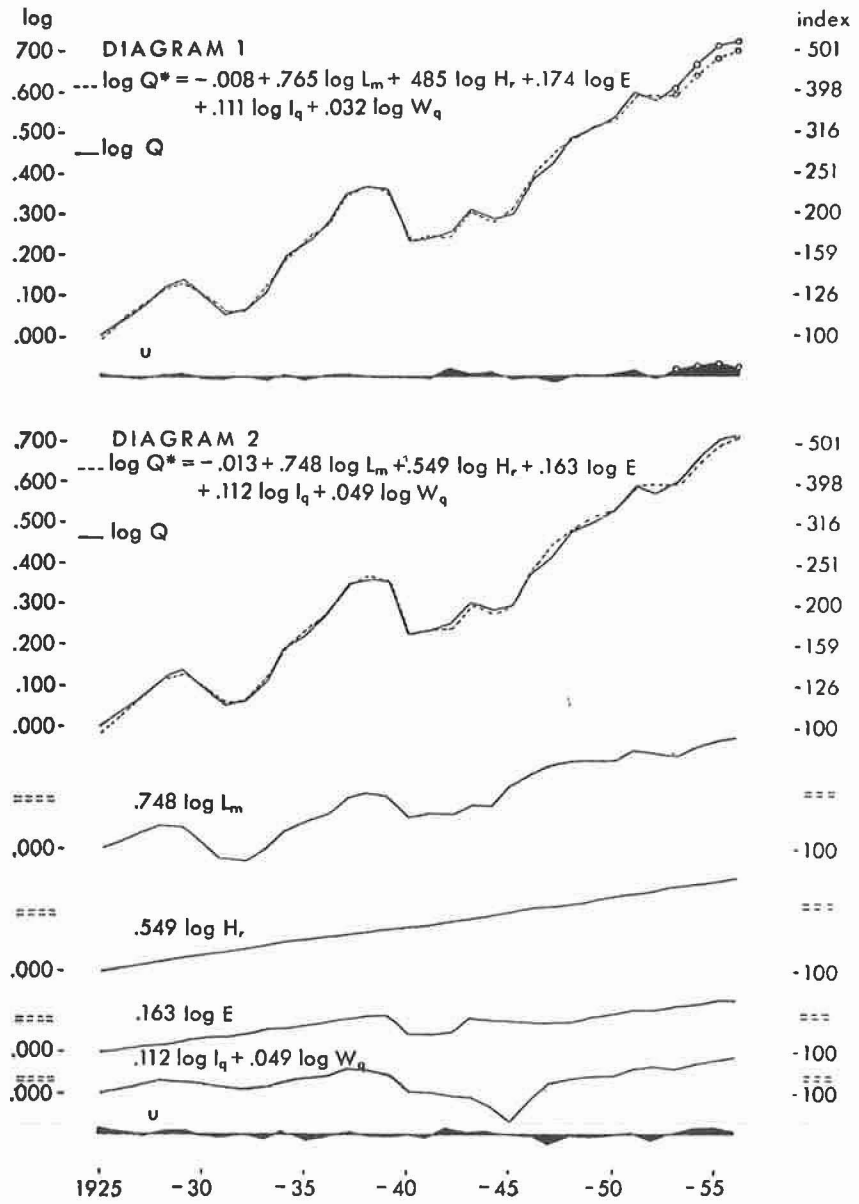
Diagram 1 åskådliggör den efter skattningen av de obekanta parametrarna a_0, a_1, etc. erhållna förklaringen $\log Q^*$ och resttermen u för industriproduktionens volym. De parametervärden, som erhållits genom att beakta materialet för åren 1925—52, leder till en visserligen liten men systematisk underskattning av industriproduktionen för åren 1953—56. Denna iakttagelse kan möjligen bero på, att industriproduktionens volymindex för krigsskadeståndsåren 1945—52 är relativt svårberäknad eller på att industriproduktionens effektivitet steg märkbart då de speciella svårigheter, som hörde ihop med de omställningar i produktionen, som krävdes under skadeståndsåren, upphörde. Då det statistiska material, som föreligger för hela perioden 1925—56, lägges till grund för härledningen av modellens parametrar fås den i *diagram 2* åskådliggjorda modellen.

I tabell 4 ha vi sammanställt de indexserier, som ingå i modellen. Där ingå tillika indexserier som beskriva produktionen per sysselsatt (L_m) och per arbetstimme (L_h). I den sista kolumnen finns produktionen per arbetstimme i USA. Produktionen per arbetstimme inom industrin har utvecklats på ett rätt likartat sätt i Finland och i USA. Utvecklingstakten i Finland har t.o.m. varit något snabbare än i USA.

För produktionen per sysselsatt person kan man ur den i diagram 2 åskådliggjorda modellen härleda följande modell

$$\log Q^*/L_m = -.013 - .252 \log L_m + .112 \log I_q + .049 \log W_q + .163 \log E + .549 \log H_r$$

De multiplikativa modeller, som vi här i likhet med Niitamo begagnat, torde vara approximativt giltiga även för relativt kraftiga variationer i L_m och H_r. Däremot är det tvivel underkastat, huruvida goda approximationer kunna erhållas till eventuella statistiska data, i den händelse att indexen för energimängden och utrikeshandeln varierar kraftigt relativt till L_m.


log
index
DIAGRAM 1
--- log Q* = − .008 + .765 log Lm + 485 log Hr + .174 log E
+ .111 log Iq + .032 log Wq
— log Q
700 -
.600 -
.500 -
.400 -
.300 -
.200 -
.100 -
.000 -
501
398
316
251
200
159
126
100
U
DIAGRAM 2
--- log Q* = − .013 + .748 log Lm + .549 log Hr + .163 log E
+ .112 log Iq + .049 log Wq
— log Q
.748 log Lm
.549 log Hr
.163 log E
.112 log Iq + .049 log Wq
1925
30
35
40
45
50
55

Resultaten beträffande utbildningsfaktorns betydelse äro anmärkningsvärda. De visa tydligt, att utbildningen spelar en mycket stor roll för industriproduktionens effektivitet. Beaktas bör dock, att den term i modellen, som tar hänsyn till utbildningsfaktorn, även implicit inbegriper forskningsverksamhetens och uppfinnarnas ansträngningar. Importvolymfaktorns (I_q) verkan kan ha förstärkts av variationer i inflödet av värdefulla tekniska informationer från utlandet. Denna faktor inbegriper även en liten energivariationsfaktor, beroende på variationen i importen av drivmedel för industrimaskiner. Det kan vara skäl att påpeka att utvecklingen av industrins fasta kapital torde vara starkt korrelerad med H_r, I_q och E.

Redogörelse för den begagnade beräkningstekniken.

Vid beräkningarna ha vi använt oss av minsta-kvadrat-metoden, d.v.s. minimerat summan av resttermernas (u) kvadrater. Först ha vi logaritmerat de ursprungliga tidsserierna, och — på vanligt sätt — ur dem beräknat följande momentmatris.

	log Q	1.000	log L_m	log H_r	log E	log I_q	log W_q
log Q	2.826 616	7.530 000	1.520 282	1.849 013	3.572 602	1.786 455	.249 088
1.000	7.530 000	28.000 000	3.799 000	5.126 000	10.331 000	5.003 000	.460 000
log L_m	1.520 282	3.799 000	.846 565	.985 812	1.857 729	.903 320	.079 489
log H_r	1.849 013	5.126 000	.985 812	1.253 364	2.373 131	1.048 075	.005 042
log E	3.572 602	10.331 000	1.857 729	2.373 131	4.761 303	2.227 705	.316 846
log I	1.786 455	5.003 000	.903 320	1.048 075	2.227 705	1.908 313	.905 656
log W_q	.249 088	.460 000	.079 489	.005 042	.316 846	.905 656	1.059 202

Vid lösandet av normalekvationerna utgå vi från att först utvälja en variabel, — taga så det inversa värdet av motsvarande momentmatris diagonalelement — samt beräkna variabelns koefficient och restkvadratsumman. Därefter välja vi en annan variabel och beräkna den inversionsmatris samt de koefficienter och den restkvadratsumma (q^2), som svara mot dessa två variablers momentmatris. Beräkningarna fortsättas genom att taga med en ny variabel i gången — tills man medtagit alla i modellen ingående variabler. I anslutning till tabellerna 1a och 1b visas hur det sista steget har utförts. På rad log Q finns den första radens värden för momentmatrisen. I kolumn a och på rad a finnas parameter-

värdena för det nästsista steget. Innanför det inramade fältet finns inversionsmatrisen till motsvarande momentmatris. I kolumn log W_q finnas talen .249088, . . . , 1.059202, den nya variabelns (log W_q) momentmatrisvärden. Från denna 5×5 inversionsmatris övergå vi till den slutliga 6×6 inversionsmatrisen på följande sätt. Vi beräkna de tal, som höra till rad d, så, att vi först multiplicera talen i kolumn log W_q med motsvarande tal i kolumn (1.000) och addera produkterna, således .212605 × .460000 + .787811 × .079489 + (—.707834) × .005042 + (—.358102) × .316846 + (—.123513) × .905656 = —.068472. På samma sätt räknas för varje inversionsmatris' kolumn talen d. Från talet på rad log W_q och i kolumn log W_q (log W_q:s diagonalelement i momentmatrisen) subtraheras summan av produkterna av de tal som finnas i den nya kolumnen (log W_q) och på rad d; d.v.s. 1.059202 — (—.068472) × .460000 — .052865 × .079489 — (—1.551071) × .005042 — .605258 × .316846 — .774388 × .905656 = .20121472. Det erhållna talet antecknas på rad $1/k$ i kolumn log W_q. Såsom lätt inses, är d-talen värdena för de parametrar, som man skulle ha erhållit, om log W_q hade varit den serie, som skulle förklaras, och de variabler, som ingå i inversionsmatrisen, varit förklarande variabler. Talet på $1/k$-raden är restkvadratsumman för en dylik modell.

Inversionsmatrisens diagonalelement för den nya variabeln log W_q är det inversa värdet k av talet i kolumn $1/k$; $1/.20121472 = 4.969815$. De övriga talen, som komma på samma rad, ha motsatt tecken mot detta tals och d-talens produkter ($= -kd$). Således är t.ex. 4.969815 × (—.068472) = .340293, 4.969815 × .052865 = —.262729 o.s.v. Eftersom både inversions- och momentmatrisen äro symmetriska i förhållande till sin huvuddiagonal, placeras även de sålunda erhållna talen i kolumn log W_q.

Det tal, som finns på den nya inversionsmatrisens rad (1.000) och i dess kolumn (1.000), d.v.s. talet .235906, få vi genom att från motsvarande tal i den föregående inversionsmatrisen .212605 subtrahera produkten $(-kd_{(1.000)})\ d_{(1.000)} = .340293 \times (-.068472)$. I enlighet härmed är talet på rad log L_m och i kolumn log L_m 19.174584 = 19.160695 — (—.262729) × .052865. Talet på rad (1.000) i kolumn log L_m och på rad log L_m i kolumn (1.000), d.v.s. .76921, få vi, då vi från motsvarande tal i den föregående inversionsmatrisen subtrahera produkten $(-kd_{(1.000)})\ d_L$ eller $(-kd_L)\ d_{(1.000)}$, således .787811 — .340293 × .052865 = .787811 —(—.262729 × (—.068472) = .769821. Då hela inversionsmatrisen har beräknats på detta sätt, kunna vi beräkna parametrarnas värden. Varje

variabels koefficient beräknas genom att multiplicera talen på rad log Q med talen i motsvarande variabels inversionsmatrisrad. T.ex. $a_{(1.000)}$ = .235906 × 7.530000 + .769821 × 1.520282 + (—.180015) × 1.849013 + (—.564067) × 3.572602 + (—.387032) × 1.786455 + .340293 × .249088 = —.007972.

När man erhållit värdena för alla koefficienter, beräknas restkvadratsumman genom att från talet på rad log Q och i kolumnen log Q (diagonalelementet i motsvarande momentmatris), 2.826616, subtrahera summan av produkterna av koefficienterna och motsvarande tal på rad log Q; 2.826616 — (—.007975) × 7.530000 — .764990 × 1.520282 — .484717 × 1.849013 — .173647 × 3.572602 — .110668 × 1.786455 — .032342 × .249088 = .001288, som antecknas på raden q^2 i kolumn log Q. Man kan kontrollera talen i inversionsmatrisen genom att beakta att produkternas summa för talen i en rad och i motsvarande momentmatrisrad = 1; om talen i en rad i inversionsmatrisen multipliceras med talen i en annan än i motsvarande momentmatrisrad, är produkternas summa=0.

Koefficienterna kunna även beräknas enligt en kortare metod än den ovan skildrade — och med lika stor noggrannhet. Härvid räknar man först på vanligt sätt den nya variabelns koefficient (i tabellen a_W = .032342). De övriga koefficienterna får man genom att från de gamla koefficienterna subtrahera produkten av den erhållna koefficientens värde och motsvarande d-tal; $a_{(1.000)}$ = —.010190—.032342 × (—.068472) = —.007975.

Restkvadratsumman q^2 kan också beräknas på ett annat sätt. Om man från den gamla restkvadratsumman subtraherar $(a_W)^2 \times (1/k)$, får man en ny restkvadratsumma; .001504—(.032342)² × .201215 = .001294 (de två sista siffrorna äro osäkra i exemplet).

Restvariansen s^2 fås genom att dividera modellens restkvadratsumma q^2 med antalet frihetsgrader, d.v.s. med antalet observerade år (28) minskat med antalet beräknade parametrar (5 i 1a och 6 i 1b).

Koefficienternas medelfel beräknas genom att taga kvadratroten ur produkten av den inversionsmatris' diagonalelement, som svarar mot ifrågavarande koefficient och restvariansen. Ex. medelfelet

$$\varepsilon_a = \sqrt{\frac{.001504}{(28-5)} \times 19.160695} = .035400$$

för koefficienten a = .766700 för log L_m.

I tabell 2 se vi hur värdena för de nya parametrarna och den nya inversionsmatrisen beräknats, då man beaktat de olika variablernas

värden för år 1956, och då man på motsvarande sätt kommit till inversionsmatriserna och koefficienterna t.o.m. år 1955. Variablerna för år 1956 finnas på rad x_{56}. Först beräknas d_x-talen på samma sätt som d-talen beräknats, då man utökat modellen med en variabel i sänder. T.ex. $1.000 \times .218544 + .373 \times .701322 + .422 \times .067871 + .776 \times (-.600553) + .643 \times (-.390241) + .319 \times .409982 = -.077391$. Man bestämmer så den nya iakttagelsens, $Q = .720$, oförklarade del z_{55} genom att använda regressionskoefficienternas a_{55}-värden; $.720 - (-.012145) \times 1.000 - .749774 \times .373 - .543327 \times .422 - .164589 \times .776 - .110485 \times .643 - .048856 \times .319 = .008847$ .Därefter beräknas $1/k' = 1 + 1.000 \times (-.077391) + .373 \times (-.227953) + .422 \times .831684 + .776 \times (-.200870) + .643 \times .260854 + .319 \times .046099 = 1.215112762$. Till talet 1 måste man således addera produkternas summa för d_x- och x_{56}-talen. Det sistnämnda talets inversa värde, $k' = .822969$, placeras på rad x, i kolumn x, i tabell 2b. De andra talen på raden x och i kolumn x äro produkter kd_x, sålunda t.ex. $-.063690 = .822969 \times (-.077391)$ och $-.187598 = .822969 \times (-.227953)$ o.s.v. Talen i inversionsmatrisen i tabell 2b få vi från talen i den föregående tabellen 2 a genom att från dem subtrahera produkten av kd_x, och det tal d_x, som finns i samma kolumn. På inversionsmatrisens rad (1.000) i kolumn (1.000) få vi således talet $.213615 = .218544 - (-.063690) \times (-.077391)$. På motsvarande sätt få vi på rad (1.000) i kolumnen log L_m och på rad log L_m i kolumn (1.000) $.686804 = .701322 - (-.063690) \times (-.227953) = .701322 - (-.187598) \times (-.077391)$. På rad log Q finnas de nya momenten $4.129853 + .720 \times .720, = 4\,648253, 9.503000 + 1.000 \times .720 = 10.223000$, $2.181601 + .373 \times .720 = 2.450161$ o.s.v. De nya regressionskoefficienterna a_{56} kunna nu beräknas såsom tidigare visats, med tillhjälp av inversionsmatrisen. Ett annat sätt är att beräkna dem ur de gamla koefficienterna a_{55} genom att till dem addera produkten av talet (kd_x) på raden x i motsvarande kolumn och talet z_{55}. Sålunda få vi t.ex. $a_{(1.000);56} = -.012145 + (-.063690) \times .008849 = -.012709$ (koefficienterna beräknade på detta sätt skiljer sig från de tidigare först fr.o.m. den 5:te decimalen).

Vi får restkvadratsumman $q^2 = 4.648253 - (-.012709) \times 10.223000 - .748114 \times 2.450161 - .549384 \times 2.929063 - .163126 \times 5.599736 - .112385 \times 3.284053 - .049192 \times 1.037402 = .002424$.

Räkningarna kunna kontrolleras t.ex. genom att bestämma regressionskoefficienterna på två skilda sätt, och dessutom genom att konstatera att produkternas summa för en inversionsmatrisrad och motsvarande momentmatrisrad=1. Då man beräknar korrelationskoefficienten subtraherar man först från talet 4.648253 på rad log Q i kolumn

log Q den del som konstanten (1.000) lämnar oförklarad; $q_1^2=(4.648253 - (10.223000)^2/32.000000)=1.328324$. Talet $1.328324/(32-1)=.042849$ är variansen $s^2_{\log Q}$ för den serie som skall förklaras. Restvariansen för den slutliga modellen s_u^2 är q_{56}^2/f eller $.002458/(32-6)=.0000945$, där f är antalet frihetsgrader. Själva korrelationskoefficienten R beräknas ur ekvationen

$$R=\sqrt{1-s_u^2/s^2_{\log Q}} \quad =\sqrt{1-.002205}=.999$$

I tabell 3 ha vi angivit parametervärdena i vår modell vid en successiv förlängning av den period som beaktats vid härledningen av parametervärdena. I den nästsista kolumnen anges restvariansen och i den sista korrelationskoefficienten mellan den observerade log Q-serien och de beräknade modelltalen log Q^*. Det kan nämnas, att för Niitamos bästa modeller för perioden 1925—52 korrelationskoefficienten $R=0.997$. I hans modell (11) har regressionskoefficienter för log L_m värdet .779 och för log H_r värdet .545. De andra två förklarande variablerna ha en något annan innebörd än log W_q och log E varför någon direkt jämförelse mellan regressionskoefficienterna för dem ej kan göras.

Till slut har jag den angenäma plikten att tacka pol.mag. Reino Helppi för att han bistått mig med att genomföra de praktiska räkningarna.

Tabell 1. *Övergången från en modell till en annan modell med en ny förklarande variabel* ($\log W_q$).

(1 a)	$\log Q$	a	1.000	$\log L_m$	$\log H_r$	$\log E$	$\log I_q$	$\log W_q$
$\log Q$	2.826 616		7.530 000	1.520 282	1.849 013	3.572 602	1.786 455	.249 088
1.000	7.530 000	−.010 190	.212 605	.787 811	− .707 834	− .358 102	− .123 513	.460 000
$\log L_m$	1.520 282	.766 700	.787 811	19.160 695	−17.946 099	.788 367	−2.199 399	.079 489
$\log H_r$	1.849 013	.434 552	− .707 834	−17.946 099	31.948 778	−8.856 293	3.142 523	.005 042
$\log E$	3.572 602	.193 222	− .358 102	.788 367	− 8.856 293	5.626 017	−1.137 970	.316 846
$\log I_q$	1.786 455	.135 713	− .123 513	− 2.199 399	3.142 523	−1.137 970	1.491 459	.905 656
$\log W_q$								1.059 202
a			− .010 190	.766 700	.434 552	.193 222	.135 713	
ε_a			± .003 729	± .035 400	± .045 711	± .019 182	± .009 876	
q^2	.001 504							
d			− .068 472	.052 865	− 1.551 071	.605 258	.774 388	
$1/k$								.20 121 472
(1 b)								
$\log Q$	2.826 616		7.530 000	1.520 282	1.849 013	3.572 602	1.786 455	.249 088
1.000	7.530 000	−.007 972	.235 906	.769 821	− .180 015	− .564 067	− .387 032	.340 293
$\log L_m$	1.520 282	.764 990	.769 821	19.174 584	−18.353 611	.947 386	−1.995 945	− .262 729
$\log H_r$	1.849 013	.484 717	− .180 015	−18.353 611	43.905 265	−13.521 946	−2.826 875	7.708 536
$\log E$	3.572 602	.173 647	− .564 067	.947 386	−13.521 946	7.446 645	1.191 405	−3.008 020
$\log I_q$	1.786 455	.110 668	− .387 032	− 1.995 945	− 2.826 875	1.191 405	4.471 742	−3.848 565
$\log W_q$	.249 088	.032 342	.340 293	− .262 729	7.708 536	−3.008 020	−3.848 565	4.969 815
a			− .007 972	.764 990	.484 717	.173 647	.110 668	.032 342
ε_a			± .003 715	± .033 492	± .050 680	± .020 871	± .016 174	± .017 050
q^2	.001 288							

Tabell 2. *Övergången från en modell* (2a) *byggande på material för åren 1925–55 till en modell* (2b) *byggande på material för åren 1925–56.*

(2 a)	log Q	a	1.000	log L_m	log H_r	log E	log I_q	log W_q	x
log Q	4.129 853		9.503 000	2.181 601	2.625 223	5.041 016	2.821 093	.807 722	
1.000	9.503 000	−.012 145	.218 544	.701 322	.067 871	− .600 553	− .390 241	.409 982	−.073 685
log L_m	2.181 601	.749 774	.701 322	18.779 019	−17.056 168	.713 953	−2.044 658	.076 933	−.067 630
log H_r	2.625 223	.543 327	.067 871	−17.056 168	39.457 061	−12.722 889	−2.699 059	6.530 850	.381 196
log E	5.041 016	.164 589	− .600 553	.713 953	−12.722 889	7.279 642	1.164 148	−2.805 992	.058 450
log I_q	2.821 093	.110 485	− .390 241	− 2.044 658	2.699 059	1.164 148	4.459 740	−3.818 922	.039 885
log W_q	.807 722	.048 856	.409 982	.076 933	6.530 850	− 2.805 992	−3.818 922	4.653 350	.161 430
			− .073 685	− .067 630	.381 196	.058 450	.039 885	.161 430	.820 244
a			− .012 145	.749 774	.543 327	.164 589	.110 485	.048 856	
$\pm\varepsilon_a$			± .004 540	± .042 082	± .060 999	± .026 201	± .020 507	± .020 948	
q^2	.002 358								
$-x_{56}$	.720		1.000	.373	.422	.776	.643	.319	
d_x			− .077 391	− .227 953	.831 684	− .200 870	.260 854	.046 099	
$1/k$									1.215 112 762
z_{55}									.008 847
(2 b)									
log Q	4.648 253		10.223 000	2.450 161	2.929 063	5.599 736	3.284 053	1.037 402	
1.000	10.223 000	−.012 709	.213 615	.686 804	.120 841	− .613 347	− .373 627	.412 918	−.063 690
log L_m	2.450 161	.748 114	.686 804	18.736 255	−16.900 146	.676 270	−1.995 722	.085 581	−.187 598
log H_r	2.929 063	.549 384	.120 841	−16.900 146	38.887 815	−12.585 403	−2.877 601	6.499 298	.684 450
log E	5.599 736	.163 126	− .613 347	.676 270	−12.585 403	7.246 436	1.207 270	−2.798 371	−.165 310
log I_q	3.284 053	.112 385	− .373 627	− 1.995 722	− 2.877 601	1.207 270	4.403 741	−3.828 818	.214 675
log W_q	1.037 402	.049 192	.412 918	.085 581	6.499 298	− 2.798 371	−3.828 818	4.651 601	.037 938
x			− .063 690	− .187 598	.684 450	− .165 310	.214 675	.037 938	.822 969
a			− .012 709	.748 114	.549 384	.163 126	.112 385	.049 192	
$\pm\varepsilon_a$			± .004 463	± .041 798	± .060 188	± .025 992	± .020 262	± .020 885	
q^2	.002 424								

Tabell. 3. *Modellens parametrar vid successiv utökning av det utnyttjade utgångsmaterialet med nya observationer för ett år i sänder.*

Period	1 a_0	log L_m a_1	log H_r a_2	log E a_3	log I_q a_4	log W_q a_5	$S^2_{(\varepsilon)}$	R
1925—52	—.0080	.7650	.4847	.1736	.1107	.0323	.0000 585	.9992
1925—53	—.0091	.7570	.5081	.1696	.1093	.0384	.0000 648	.9992
1925—54	—.0106	.7512	.5352	.1633	.1096	.0454	.0000 829	.9991
1925—55	—.0121	.7498	.5433	.1646	.1105	.0489	.0000 943	.9990
1925—56	—.0127	.7481	.5494	.1631	.1124	.0492	.0000 945	.9991

Koefficienterna $a_1, \ldots, a_5$ ange den procentuella ökningen i industriproduktionens volym Q^* vid en ökning på 1% i den frågavarande på produktionens storlek inverkande faktorn.

Tabell 4. *Utgångsserierna för analysen samt produktionen per sysselsatt och per arbetstimme jämfört med produktionen per arbetstimme i U.S.A. samt den enligt modellen för tidsperioden 1925–56 beräknade produktionsvolymen Q^* och Q^*/L_m.*

År	Q	Q^*	L_m	H_r	E	I_q	W_q	Q/L_m	Q^*/L_m	Q/L_h	Q/L_h (USA) / Q/L_h (USA)
1925	100	97	100	100	100	100	100	100	97	100	100
1926	108	107	106	103	109	120	103	102	101	101	101
1927	118	119	113	107	118	142	116	105	105	105	103
1928	132	130	120	111	124	166	114	110	108	109	108
1929	136	134	117	115	149	156	121	116	115	117	114
1930	124	124	104	119	155	147	109	119	119	124	113
1931	112	114	93	123	158	122	110	121	123	128	114
1932	116	116	91	127	181	112	116	127	127	131	109
1933	128	129	100	131	206	128	134	129	129	130	119
1934	155	155	114	135	226	160	145	136	136	136	122
1935	169	173	123	139	256	184	155	137	141	138	127
1936	189	190	130	143	281	206	173	145	146	146	131
1937	222	220	146	146	322	260	186	152	151	152	132
1938	229	231	152	150	357	252	160	151	152	153	129
1939	224	226	148	153	348	224	152	151	153	157	136
1940	168	168	127	157	199	152	45	132	132	138	144
1941	172	174	131	162	189	139	58	132	133	139	147
1942	179	173	129	168	199	113	64	139	134	144	146
1943	203	200	138	172	328	105	71	147	145	150	145
1944	191	188	137	178	301	73	49	140	137	144	145
1945	199	199	162	184	293	42	28	122	123	128	145
1946	241	244	177	190	291	87	73	136	138	148	141
1947	269	283	188	196	287	164	99	143	151	155	148
1948	305	308	195	203	302	210	104	157	158	167	154
1949	320	327	196	210	355	212	123	164	167	177	163
1950	340	344	197	217	399	223	137	173	175	187	173
1951	395	394	214	225	443	283	170	185	184	199	173
1952	379	395	205	232	447	338	150	185	192	200	179
1953	400	396	201	239	503	277	168	199	197	...	186
1954	460	445	217	247	530	347	195	212	205	...	194
1955	511	492	229	255	631	380	212	223	215	...	201
1956	525	516	236	265	598	440	208	222	219	...	...

On the distribution function for a function of n statistic variables and the Central Limit Theorem in the Mathematical Theory of Probability.

By **Leo Törnqvist** (Helsingfors).

According to the Central Limit Theorem in the mathematical theory of probability, the sum i. e. a specific function of n mutually independent statistic variables — under very general conditions — converges towards the normal distribution function:

$$(0) \qquad \Phi\left(\frac{x-b}{a}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-b}{a}} e^{-\frac{u^2}{2}} du$$

when the number of the statistic variables entering into the sum (function) examined tends to infinity. This theorem has attracted the attention of many mathematically interested statisticians. A great number of famous mathematicians, as MOIVRE, LAPLACE, LJAPOUNOFF, LINDEBERG, KOLMOGOROFF, KHINTCHINE, PETROWSKY, LÉVY, FELLER, CRAMÉR etc. have dealt with the formulation of and the proof of the central limit theorem. As a result of these efforts, the space of validity of the central limit theorem has been found to be very extensive and very difficult to define, so that the limits of the space within which the theorem provably is valid have repeatedly been widened. FELLER and LÉVY have, however, found a formulation touching the limits for the space of validity of the theorem in the case that only considers the sum of n mutually independent variables, the relative signification of which approaches zero.

The criterion formulated by FELLER for the validity of the central limit theorem within the domain studied, runs as follows [1]:

FELLER's formulation of *the central limit theorem:*

A sufficient and necessary condition, in order that the distribution function

$$W_n(x) = \int_{-\infty}^{\infty} W_{n-1}(x-y)\, d\, V_k(y + b_k) \tag{1}$$

of a sum of n mutually independent statistic variables $x_1 \ldots x_n = x_n' - b_n$ *shall have the property that there exists such a succession of numbers* $\{a_n, b_n\}$ *that*

$$W_n(a_n x) \to \Phi(x) \tag{2}$$

while the distribution functions $V_k(x)$ *for* $x_k' = x_k + b_k$ *have the property*

$$(\overset{\infty}{\underset{n}{\uparrow}}\ V_k(a_n x + b_k) \to \Phi\left(\frac{x}{0}\right) = \begin{matrix} 0;\ x<0 \\ 1;\ x>0 \end{matrix};\ k = 1, \ldots, n, \tag{3}$$

when the function $V_k(x)$ has been normed so that

$$V_k(0-) < 1-\lambda,\ V_k(0+) > \lambda, \tag{4}$$

is that the distribution functions $V_k(x)$ *have the properties* (5) *and* (6)

$$\lim_{\substack{X\to\infty \\ n\to\infty}} \left(X^2 \sum_{n=1}^{n} \int_{|x|>X} d\, V_n(x) \right) : \left(\sum_{n=1}^{n} \int_{|x|<X} d\, V_k(x) \right) = 0 \tag{5}$$

$$\lim_{X\to\infty} \lim_{n\to\infty} \left(X^2 : \sum_{k=1}^{n} \int_{|x|<X} x^2\, d\, V_n(x) \right) = 0. \tag{6}$$

There exists then (and only then) a succession of numbers $X_n \to \infty$ with the properties (7) and (8)

$$\left(X_n^2 : \sum_{k=1}^{n} \int_{|x|<X_n} x^2\, d\, V_k(x) \right) \to 0 \tag{7}$$

$$\sum_{k=1}^{n} \int_{|x|>X_n} d\, V_k(x) \to 0. \tag{8}$$

We can in that case put

$$a_n^2 = \sum_{k=1}^{n} \left\{ \int_{|x|<X_n} x^2 \, d\, V_k(x) - \left(\int_{|x|<X_n} x \, d\, V_k(x) \right)^2 \right\} \tag{9}$$

and define the numbers b_n in the following manner

$$\int_{|x|<a_n} (x - b_n) \, d\, V_n(x) = 0. \tag{10}$$

Feller's theorem does not, however, define the whole space of validity of the central limit theorem. The reason is that the condition (3) does not represent an absolutely necessary condition, as the convergence will not be ventured if we in the formula (3) replace $\Phi\left(\frac{x}{0}\right)$ by $\Phi\left(\frac{x}{\alpha}\right)$. In the case $\alpha > 0$ the condition (7) is not fulfilled. A condition — although not necessary — consists in narrowing the field of study to a specific function, viz. a *sum* of n mutually *independent* statistic variables. It is possible essentially to extend the space of validity of the central limit theorem by examining a much more general problem, viz. that of defining under which circumstances the distribution function for a function of n statistic variables which need not be independent of each other converges towards the normal distribution function. In the following we shall try to elucidate this complex of problems and to derive convergent serial representations of the distribution function of a function of n statistic variables. This problem has previously been treated by R. v. MISES [2] and CRAMÉR [3] as for mutually independent variables and functions which have partial derivatives of first and second order.

Let us consider a function $u_n(\boldsymbol{x})$ of a system of n statistic variables $\boldsymbol{x} = (x_1, \ldots, x_n)$, and let the number system $\boldsymbol{x}$ have a n-dimensional probability distribution function $F_n(\boldsymbol{x})$. The probability that the function $u_n(\boldsymbol{x})$ will take a value $\leq au + b$, can in that case be written in the form:

$$W_n(au + b) = \int \Phi\left(\frac{au + b - u_n(\boldsymbol{x})}{0}\right) d\, F_n(\boldsymbol{x}) = \tag{11}$$

$$= \int \Phi\left(\frac{u - u_{n\,a\,b}(\boldsymbol{x})}{0}\right) d\, F_n(\boldsymbol{x}).$$

In the last formula we have used the notation

$$u_{nab}(\boldsymbol{x}) = \frac{u_n(\boldsymbol{x}) - b}{a}. \tag{12}$$

The integrations are done here and in the following, when not expressively otherwise stated over the whole n-dimensional space of variation. The function over which the integration is done, takes the value zero for $u_{nab}(\boldsymbol{x}) < u$ and the value 1 for $u_{nab}(\boldsymbol{x}) > u$. The sum obtained by integration thus indicates the probability that the function defined by formula (12), will take a value, at the most $= u$. For $u_{nab}(\boldsymbol{x}) = u$ the integrant $\Phi\left(\frac{u - u_{nab}(\boldsymbol{x})}{0}\right)$ remains indeterminate between the limits 0 and 1. It is, however, only in exceptional cases that $W_n(au + b)$ receives a not infinitesimally small contribution from the number systems $\boldsymbol{x}$, belonging to the surface $u_{nab}(\boldsymbol{x}) = u$. In these exceptional cases $W_n(au + b)$ possesses a finite spring-stead indicating the probability that $\boldsymbol{x}$ will be situated on the surface considered.

The expression (11) can be transformed in many different ways. Interesting results are obtained, if we e. g. develop the integrant in a sum of orthogonal functions of the normal distribution function $\Phi(u)$. Such a serial development is represented by the formula (13)

$$\Phi\left(\frac{u - u_{nab}(\boldsymbol{x})}{0}\right) = \sum_{\nu=-\infty}^{\infty} c_\nu e^{2\pi\nu i \Phi(u)};\ c_\nu = c_\nu(u_{nab}(\boldsymbol{x}));\ \boldsymbol{i} = \sqrt{-1}. \tag{13}$$

Multiplying each member by the factor $e^{-2\pi g i \Phi(u)} d\Phi(u)$ and integrating the expression thus obtained over all u's, we find that the coefficient c_g may be written in the form

$$c_g = \int_{u=-\infty}^{\infty} \Phi\left(\frac{u - u_{nab}(\boldsymbol{x})}{0}\right) e^{-2\pi i g \Phi(u)} d\Phi(u) = c_g(u_{nab}(\boldsymbol{x})), \tag{14}$$

because the integrals

$$\int_{u=-\infty}^{\infty} c_\nu e^{2\pi(\nu-g)i\Phi(u)} d\Phi(u) = c_\nu \frac{e^{2\pi(\nu-g)i} - e^{2\pi(\nu-g)i\cdot 0}}{2\pi(\nu-g)i} = 0, \tag{15}$$

ν and g being whole numbers and $\nu \neq g$. When $\nu = g$, the integral (15) becomes equal to (14). The expression (14) may, with regard to that $\Phi\left(\frac{u - u_{nab}(x)}{0}\right) = 0$ for $u < u_{nab}(x)$ and 1 for $u > u_{nab}(x)$, when $g \neq 0$, be written in the form

$$(16) \quad c_g = \int_{u=u_{nab}(x)}^{\infty} e^{-2\pi i g \Phi(u)}\, d\,\Phi(u) = -\frac{1}{2\pi i g} \Big|_{u=u_{nab}(x)}^{\infty} e^{-2\pi i g \Phi(u)},$$

that is

$$(17) \quad c_g = -\frac{1}{2\pi i g} + \frac{1}{2\pi i g} e^{-2\pi i g \Phi(u_{nab}(x))} = -\frac{1}{2\pi i g} + \\ + w_{nabg}(x).$$

In case $g = 0$, we find for c_0 the expression

$$(18) \quad c_0 = \int_{u=u_{nab}(x)}^{\infty} d\,\Phi(u) = 1 - \Phi(u_{nab}(x)) = \tfrac{1}{2} + \\ + \left(\tfrac{1}{2} - \Phi(u_{nab}(x)) = \tfrac{1}{2} + w_{nab0}(x)\right).$$

For $W_n(au + b)$ we find by means of medium value calculation term for term in the series (13)

$$(19) \quad W_n(au + b) = \left(\frac{1}{2} - \sum_{g \neq 0} \frac{1}{2\pi i g} e^{2\pi i g \Phi(u)}\right) + \sum w_{nabg}\, e^{2\pi i g \Phi(u)};$$

where

$$(20) \quad \omega_{nab0} = \tfrac{1}{2} - \int \Phi(u_{nab}(x))\, d\, F_n(x),$$

$$w_{nabg} = \frac{1}{2\pi i g} \int e^{-2\pi i g \Phi(u_{nab}(x))}\, d\, F_n(x); \quad g \neq 0.$$

The first parenthesis in the expression (19) is a serial development for $\Phi(u)$

$$(21) \quad \Phi(u) = \frac{1}{2} - \sum_{g \neq 0} \frac{1}{2\pi i g} e^{2\pi i g \Phi(u)},$$

which can be seen, if we replace $F_n(x)$ by $\Phi(x)$ and $u_n(x)$ by x and observe that $\int \Phi(x)\, d\,\Phi(x) = \frac{1}{2}$, $\int e^{2\pi g i \Phi(x)}\, d\,\Phi(x) = 0$. By

help of Euler's formula $e^{2\pi i g \Phi(u)} = \cos 2\pi g \Phi(u) + i \sin 2\pi g \Phi(u)$ we can transform (21) into

$$\Phi(u) = \frac{1}{2} + \sum_{g=1}^{\infty} \frac{1}{\pi g} \sin 2\pi g \Phi(u). \tag{22}$$

The second parenthesis in the expression (19) is a function of a and b varying from $-\frac{1}{2}$ to $\frac{1}{2}$, when $(a > 0)$ and b varies between $-\infty$ and $+\infty$. The expression

$$W_{na}(b) = \int \Phi\left(\frac{b-y}{a}\right) d\, W_n(y) = \int \Phi\left(\frac{b - u_n(x)}{a}\right) dF_n(x) = \tag{23}$$

$$= \tfrac{1}{2} + (\tfrac{1}{2} - \int \Phi(u_{nab}(x))\, dF_n(x))$$

actually represents a distribution function, converging towards $W_n(b)$ for $a = 0$ and towards $W_{n\infty}(b) = \Phi\left(\frac{b}{\infty}\right) = \frac{1}{2}$ for $a \to \infty$. Independent of the value of the number a, we can thus choose a number $b = b_{na}$ so that the expression $W_{na}(b) - \frac{1}{2} = (\frac{1}{2} - \int \Phi(u_{nab}(x)\, dF_n(x)) = 0$. Since this expression in any case must tend to zero, in order that $W_n(au + b)$ shall converge towards $\Phi(u)$, we can, without reducing thereby the space of convergence for the succession of functions $W_n(au + b)$, presume that $b = b_{na}$. The probability distribution function for the variable $u_{na}(x) = \frac{u_n(x) - b_{na}}{a}$ can thus be written in the form

$$W_n(au + b_{na}) = \Phi(u) + \sum_{g \neq 0} w_{nag}\, e^{2\pi i g \Phi(u)} = \tag{24}$$

$$= \Phi(u) + \sum_{g \neq 0} |\, \omega_{nag}\, |\, e^{2\pi i g (\Phi(u) - \tau_{nag})};$$

$$w_{nag} = w_{nabg},\ b = b_{na}.$$

By using Euler's formula this can be transformed into

$$W_n(au + b_{na}) = \Phi(u) + 2 \sum_{g=1}^{\infty} |\, w_{nag}\, | \cos 2\pi g (\Phi(u) - \tau_{nag}). \tag{25}$$

The difference

$$(26) \qquad w_{na}(u) = W_n(au + b_{na}) - \Phi(u) = \sum_{g \neq 0} w_{nag}\, e^{2\pi i g \Phi(u)},$$

where

$$(27) \qquad w_{nag} = \frac{1}{2\pi i g} \int e^{-2\pi i g \Phi(u_{na}(x))}\, d F_n(x) = |w_{nag}|\, e^{-2\pi i g \tau_{nag}},$$

may be conceived as a function of a statistic variable u, which is normally distributed. The medium value of this function is

$$(28) \qquad m = \int w_{na}(u)\, d\Phi(u) = \sum_{g \neq 0} w_{nag} \int e^{2\pi i g \Phi(u)}\, d\Phi(u) = 0.$$

On the other hand, the variance $w_n^2(a)$ of the function $w_{na}(u)$ is normally $\neq 0$. On account of the properties of orthogonality in the serial development (26) we find that $w_n^2(a)$ can be calculated by means of the formula

$$(29) \qquad w_n^2(a) = \int (w_{na}(u))^2\, d\Phi(u) = \sum_{g+\nu=0} w_{nag}\, w_{na\nu} = 2 \sum_{g=1}^{\infty} |w_{nag}|^2.$$

The expression (29) can, with regard to (27), be written in the form

$$(30) \qquad w_n^2(a) = \sum_{g=1}^{\infty} \frac{1}{2\pi^2 g^2} \left| \int e^{-2\pi i g \Phi(u_{na}(x))}\, d F_n(x) \right|^2.$$

This expression disappears identically, only if all numbers $w_{nag} = 0$. It reaches its greatest possible value when $a = 0$, in which case $u_{na}(x) = \infty$ whence

$$(31) \qquad w_n^2(0) = \sum_{g=1}^{\infty} \frac{1}{2\pi^2 g^2} = \frac{\pi^2}{6 \cdot 2\pi^2} = \frac{1}{12} = \operatorname*{Max}_{(a)} w_n^2(a).$$

When the number $a \to \infty$, we find that $w_n^2(a)$ again approaches the maximum[1] value $\frac{1}{12}$. If the distribution function

[1] We find, however, $\omega_{na\infty}^2 = \int \left(\Phi\left(\frac{u - u_{na}(x)}{0}\right) - \Phi(u)\right)^2 d\Phi(u) =$ $= (\Phi(u_{na}(x)) - \frac{1}{2})^2 + \frac{1}{12} > \frac{1}{12}$, when $\Phi(u_{na}(x)) \neq \frac{1}{2}$, $\int \left(\Phi\left(\frac{u - u_{na}(x)}{0}\right) - \right.$ $\left. - \Phi(u) - \sum_{|g| \leq N} \omega_{nag}(x) e^{2\pi i g \Phi(u)}\right)^2 d\Phi(u) = \left(\frac{1}{12} - \sum_{g=1}^{N} \frac{1}{2\pi^2 g^2}\right) \to 0$ when $N \to \infty$.

$W_n(a_n u + b_{na})$ is not quite deformed, that is to say if it cannot be represented by an expression of the form $\Phi\left(\frac{u}{0}\right) = \begin{cases} 0; & u < 0 \\ 1; & u > 0 \end{cases}$ or $\Phi\left(\frac{u}{\infty}\right) = \frac{1}{2}$ if $u \neq \pm \infty$ there will be a positive finite value a, which reduces $w_n^2(a)$ to a minimum $= w_n^2 \geq 0$. Let this minimum be reached when $a = a_n$. Then

$$(32) \qquad w_n^2 = w_n^2(a_n) = \operatorname*{Min}_{(a)} w_n^2(a) = \operatorname*{Min}_{(a)} \int (W_n(a u + b_{na}) - \Phi(u))^2 \, d\Phi(u) < \tfrac{1}{12}.$$

Definition. *That space in the functional expanse* $\overset{\infty}{\underset{n=1}{\{}} u_n(\boldsymbol{x}), F_n(\boldsymbol{x})\}$, *which is so conditioned that the limes* $w_n^2 \to 0$, *is here called the space of validity of the Central Limit Theorem. Those successions of pairs of functions* $\overset{\infty}{\underset{n=1}{\{}} u_n(\boldsymbol{x}), F_n(\boldsymbol{x})\}$, *for which* $w_n^2 \to 0$, *are said to belong to the space of convergence of the normal distribution function.*

Now, there exists a class of functions $\overset{\infty}{\underset{n=1}{\{}} u_n(\boldsymbol{x}), F_n(\boldsymbol{x})\}$ with the property $w_n^2 = 0$ for each n. To this elegant class belong all those pairs of functions which are obtained by transforming, by help of substitutions of some kind, a linear function

$$(33) \qquad l_n(\boldsymbol{y}) = \sum_{k=1}^{n} l_{nk} y_k; \quad \sum_{k=1}^{n} l_{nk}^2 = 1,$$

and a normal distribution function of the form

$$(34) \qquad \Phi_n(\boldsymbol{y}) = \prod_{k=1}^{n} \Phi(y_k); \quad \Phi(y_k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_k} e^{-\frac{u^2}{2}} du.$$

In fact, we can transform the distribution function $\Phi_n(\boldsymbol{y})$, using substitutions of the form

$$(35) \qquad \begin{aligned} \boldsymbol{x} &= \boldsymbol{x}(\boldsymbol{y}) = (x_1(\boldsymbol{y}), \; x_2(\boldsymbol{y}), \; \ldots, \; x_n(\boldsymbol{y})) \\ \boldsymbol{y} &= y(\boldsymbol{x}) = (y_1(\boldsymbol{x}), \; y_2(\boldsymbol{x}), \; \ldots, \; y_n(\boldsymbol{x})) \end{aligned}$$

where

$$(36) \qquad x_k(\boldsymbol{y}) = x_k(y_1, y_2 \ldots y_n), \quad y_k(\boldsymbol{x}) = y_k(x_1, \ldots, x_n),$$

into any distribution function $F_n(\boldsymbol{x})$ prescribed in advance. We only need to define $x_k(\boldsymbol{y})$ so that $F_n(\boldsymbol{x}(\boldsymbol{y})) = \Phi_n(\boldsymbol{y})$; $n = 1, 2, \ldots$ In the case of $dF_n(x)$ denoting a finite probability, the substitution function will be so conditioned that it collocates to every *not* infinitesimal space $\boldsymbol{x}(\boldsymbol{y}) = \boldsymbol{x}$ in the expanse $\boldsymbol{y}$ with a mass of probability $\int\limits_{\boldsymbol{x}(\boldsymbol{y})=\boldsymbol{x}} d\,\Phi_n(\boldsymbol{y}) > 0$ a space in the expanse $\boldsymbol{x}$ with infinitesimal dimensions. The substitution which has the property of transforming $\Phi_n(\boldsymbol{y})$ into $F_n(\boldsymbol{x})$, is therefore, as a rule, neither analytical nor continued, but in principle it has a freely definable form.

The distribution function $W_n(a\,u + b)$ is invariant with respect to all substitutions of the type $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{y})$ and, consequently also with respect to the substitution making $F_n(\boldsymbol{x}) = F_n(\boldsymbol{x}(\boldsymbol{y}))$ equal to $\Phi_n(\boldsymbol{y})$ and $u_n(\boldsymbol{x}) = u_n(\boldsymbol{x}(\boldsymbol{y})) = v_n(\boldsymbol{y}) = a\,v_{na}(\boldsymbol{y}) + b_{na}$. The problem of examining what kind of successions of functional pairs $\overset{\infty}{\underset{n=1}{\{}} u_n(\boldsymbol{x}), F_n(\boldsymbol{x})\}$ belong to the space of convergence of the normal distribution function, may therefore be brought back to the simpler problem of examining what kind of successions of functions $\overset{\infty}{\underset{n=1}{\{}} v_n(\boldsymbol{y}), \Phi_n(\boldsymbol{y})\}$ are such that

$$w_n^2 = \operatorname*{Min}_{(a)} w_n^2(a) \to 0. \tag{37}$$

For $w_n^2(a)$ we obtain, by specializing the formula (30), the expression

$$w_n^2(a) = \sum_{g=1}^{\infty} \frac{1}{2\,\pi^2\,g^2} \left| \int e^{-2\pi i g \Phi(v_{na}(y))}\, d\,\Phi_n(\boldsymbol{y}) \right|^2. \tag{38}$$

As the distribution function of the variable $l_n(\boldsymbol{y}) = \sum\limits_{\lambda=1}^{n} l_{n\lambda}\, y_\lambda$, if $\sum\limits_{\lambda=1}^{n} l_{n\lambda}^2 = 1$, is $= \Phi(u)$, the function $W_n(a\,u + b_{na}) = \Phi(u)$, and consequently $w_n^2 = 0$ *in case*

$$v_{na}(\boldsymbol{y}) = \frac{v_n(\boldsymbol{y}) - b_{na}}{a} = l_n(\boldsymbol{y}). \tag{39}$$

When $v_{na}(\boldsymbol{y}) = l_n(\boldsymbol{y})$, the expression

$$\varepsilon_{nal}(\boldsymbol{y}) = \boldsymbol{\Phi}(v_{na}(\boldsymbol{y})) - \boldsymbol{\Phi}(l_n(\boldsymbol{y})) \tag{40}$$

is equal to zero for each $\boldsymbol{y}$. In that case the sum

$$\psi_{nal}(\boldsymbol{y}) = \boldsymbol{\Phi}(v_{na}(\boldsymbol{y})) + \boldsymbol{\Phi}(l_n(\boldsymbol{y})) \tag{41}$$

is equal to $2\,\boldsymbol{\Phi}(l_n(\boldsymbol{y}))$.

We presume that the numbers a and $l_{n1}, \ldots, l_{nn}$ have been fixed so that

$$\varepsilon_{nal} = \int |\,\varepsilon_{nal}(\boldsymbol{y})\,|\,d\,\boldsymbol{\Phi}_n(\boldsymbol{y}) \tag{42}$$

becomes a minimum. Be this minimum

$$\varepsilon_n = \operatorname*{Min}_{(a\,l)} \varepsilon_{nal}. \tag{43}$$

Because $w_n^2 = 0$, $\varepsilon_n = 0$ when $v_{na}(\boldsymbol{y}) = l_n(\boldsymbol{y})$, we must, on account of the formula (38) for each integer g have

$$\int e^{-2\pi i g \Phi(l_n(y))}\,d\,\boldsymbol{\Phi}_n(\boldsymbol{y}) = 0 \tag{44}$$

and consequently

$$w_{nag} = \frac{1}{2\,\pi\,g\,\boldsymbol{i}} \int e^{-2\pi i g \Phi(v_{na}(y))}\,d\,\boldsymbol{\Phi}_n(\boldsymbol{y}) = \tag{45}$$

$$= \frac{1}{2\,\pi\,i\,g} \int \left(e^{-2\pi i g \Phi(v_{na}(y))} - e^{-2\pi i g \Phi(l_n(y))}\right) d\,\boldsymbol{\Phi}_n(\boldsymbol{y}).$$

With trigonometrical writing, we have

$$w_{nag} = -\frac{1}{\pi\,g} \int e^{-\pi i g \psi_{nal}(y)} \left(\frac{1}{2\,\boldsymbol{i}}\, e^{-\pi i g \varepsilon_{nal}(y)} - e^{\pi i g \varepsilon_{nal}(y)}\right) \cdot d\,\boldsymbol{\Phi}_n(\boldsymbol{y}) = \tag{46}$$

$$= -\frac{1}{\pi\,g} \int \cos \pi\,g\,\psi_{nal}(\boldsymbol{y}) \sin \pi\,g\,\varepsilon_{nal}(\boldsymbol{y})\,d\,\boldsymbol{\Phi}_n(\boldsymbol{y}) +$$

$$+ \frac{\boldsymbol{i}}{\pi\,g} \int \sin \pi\,g\,\psi_{nal}(\boldsymbol{y}) \sin \pi\,g\,\varepsilon_{nal}(\boldsymbol{y})\,d\,\boldsymbol{\Phi}(\boldsymbol{y}).$$

A necessary and sufficient condition that $w_n^2 \to 0$ *is that the number* a *can be chosen so that* $|w_{nag}| \to 0$ *for each* g.

The expression $w_n^2(a) = 2\sum_{g=1}^{\infty} |w_{nag}|^2$ can be written in the form

$$(47)\quad w_n^2(a) = \frac{2}{\pi^2}\sum_{g=1}^{\infty}\frac{1}{g^2}\left(\int \sin \pi g\, \varepsilon_{nal}(\boldsymbol{y}) \cos \pi g\, \psi_{nal}(\boldsymbol{y})\, d\, \boldsymbol{\Phi}_n(\boldsymbol{y})\right)^2 +$$

$$+ \frac{2}{\pi^2}\sum_{g=1}^{\infty}\frac{1}{g^2}\left(\int \sin \pi g\, \varepsilon_{nal}(\boldsymbol{y}) \sin \pi g\, \psi_{nal}(\boldsymbol{y})\, d\, \boldsymbol{\Phi}_n(\boldsymbol{y})\right)^2.$$

Considering that $\cos \pi g\, \psi_{nal}(\boldsymbol{y})$ and $\sin \pi g\, \psi_{nal}(\boldsymbol{y})$ as to their number value are always $\leqq 1$, while

$$(48)\quad \left|\frac{\sin \pi g\, \varepsilon_{nal}(\boldsymbol{y})}{\pi g}\right| \leq |\varepsilon_{nal}(\boldsymbol{y})|$$

we see that

$$(49)\quad w_n^2 \leq \operatorname*{Min}_{(a,l)} \sum_{g=1}^{\infty} 4\left(\int \left|\frac{\sin \pi g\, \varepsilon_{nal}(\boldsymbol{y})}{\pi g}\right| d\, \boldsymbol{\Phi}_n(\boldsymbol{y})\right)^2 \leqq$$

$$\leqq 4\, N\, \varepsilon_n^2 + \frac{4}{\pi^2}\sum_{g=N+1}^{\infty}\frac{1}{g^2}.$$

If $\varepsilon_n \to 0$, we can here let the numbers N grow so slowly towards ∞ that $N\varepsilon_n^2 \to 0$, while

$$(50)\quad R(N) = \frac{4}{\pi^2}\sum_{g=N+1}^{\infty}\frac{1}{g^2} \to 0,$$

for which reason also $w_n^2 \to 0$.

Thus we can formulate the theorem

Theorem 1. *A sufficient condition that $W_n(a\,u + b_{na})$ will converge towards $\boldsymbol{\Phi}(u)$ is that the number*

$$(51)\quad \varepsilon_n = \operatorname*{Min}_{a,l} \int \varepsilon_{nal}(\boldsymbol{y})\, d\, \boldsymbol{\Phi}_n(\boldsymbol{y}) =$$

$$= \operatorname*{Min}_{(a,l)} \int \left|\boldsymbol{\Phi}\left(\frac{v_n(\boldsymbol{y}) - b_{na}}{a}\right) - \boldsymbol{\Phi}(l_n(\boldsymbol{y}))\right| d\, \boldsymbol{\Phi}_n(\boldsymbol{y})$$

converges towards zero.

If, in stead of w_n^2 as measure for the abnormality of the distribution function $W_n(a\,u + b_n)$, we use the medium deviation

$$(52)\quad d_n = \operatorname*{Min}_{a} \int |W_n(a\,u + b_{na}) - \boldsymbol{\Phi}(u)|\, d\, \boldsymbol{\Phi}(u)$$

we can in a rather elementary manner show that $d_n \leq \varepsilon_n$.

We have namely

$$W_n(a\,u + b_{n\,a}) = \int \Phi\left(\frac{u - v_{n\,a}(\boldsymbol{y})}{0}\right) d\,\Phi_n(\boldsymbol{y}) \tag{53}$$

and

$$\Phi(u) = \int \Phi\left(\frac{u - l_n(\boldsymbol{y})}{0}\right) d\,\Phi_n(\boldsymbol{y}) \tag{54}$$

whence

$$d_n = \operatorname*{Min}_{(a,\,l)} \int_{u=-\infty}^{\infty} \left| \int \Phi\left(\frac{u - v_{n\,a}(\boldsymbol{y})}{0}\right) - \Phi\left(\frac{u - l_n(\boldsymbol{y})}{0}\right) d\,\Phi_n(\boldsymbol{y}) \right| d\,\Phi(u). \tag{55}$$

Considering that the number value of a sum is $\leq$ the sum of the number values of the terms, we obtain the formula

$$d_n \leq \operatorname*{Min}_{a,\,l} \int_{u=-\infty}^{\infty} \left| \Phi\left(\frac{u - v_{n\,a}(\boldsymbol{y})}{0}\right) - \Phi\left(\frac{u - l_n(\boldsymbol{y})}{0}\right) \right| (d\,\Phi(u))\, d\,\Phi_n(\boldsymbol{y}). \tag{56}$$

Integrating with respect to u, we find

$$d_n \leq \operatorname*{Min}_{(a,\,l)} \int |\,\Phi(v_{n\,a}(\boldsymbol{y})) - \Phi(l_n(\boldsymbol{y}))\,|\, d\,\Phi_n(\boldsymbol{y}) = \varepsilon_n \tag{57}$$

that is

$$d_n \leq \varepsilon_n. \tag{58}$$

Thus, if $\varepsilon_n \to 0$, also $d_n \to 0$. On the other hand, it is scarcely possible to prove that $\varepsilon_n \to 0$ be a necessary condition, in order that w_n^2 and $d_n \to 0$. For it is possible that there exists such functions $v_n(\boldsymbol{y})$ which have the property $w_n^2 \to 0$, without $\varepsilon_n \to 0$, because the coefficient of correlation $r^{(l)}_{n\,a\,l\,g}$ between $\sin \pi\,g\,\varepsilon_{n\,a\,l}(\boldsymbol{y})$ and $\cos \pi\,g\,\psi_{n\,a\,l}(\boldsymbol{y})$ and the coefficient of correlation $r^{(2)}_{n\,a\,l\,g}$ between $\sin \pi\,g\,\varepsilon_{n\,a\,l}(\boldsymbol{y})$ and $\sin \pi\,g\,\psi_{n\,a\,l}(\boldsymbol{y})$ may be thought to converge towards zero for each g when $n \to \infty$, whereby the convergence $w_n^2 \to 0$ would be secured even if ε_n would converge towards a positive number $\varepsilon_\infty > 0$. The condi-

tion $d_n \to 0$ is a necessary condition, in order that $\{ v_n(y), \Phi_n(y) \}$ shall belong to the space of convergence of the normal distribution function.

It is still of interest to put forward an easily demonstrable theorem, containing, however, no more general condition that $v_n(y)$ will belong to the space of convergence to the normal distribution function, than the condition $\varepsilon_n \to 0$.

Theorem 2. *If $v_n(y)$ can be approximated by means of such a linear expression $L_n(y) = a_n l_n(y) + b_n = \sum_{k=1}^{n} L_{nk} y_k + b_n$ that the probability $P_{L_n}(a_n \varepsilon)$ of the variable $|v_{L_n}(y)| = |v_n(y) - L_n(y)| : a_n = |v_{na}(y) - l_n(y)|$ being $> \varepsilon$, converges towards zero for each $\varepsilon > 0$, when $n \to \infty$, the distribution function $W_n(a_n u + b_n)$ will converge towards $\Phi(u)$.*

In order to prove this theorem, which is in fact only a special case of a general theorem of convergence demonstrated by H. CRAMÉR [4], we observe that $v_{nab}(y) = l_n(y) + v_{L_n}(y) = \left(\frac{v_n(y) - b_n}{a_n}\right)$

and

$$(59) \quad W_n(a_n u + b_n) = \int_{|v_{L_n}(y)| \leq \varepsilon} \Phi\left(\frac{(u - v_{L_n}(y) - l_n(y))}{0}\right) d\Phi_n(y) + \int_{|v_{L_n}(y)| > \varepsilon} \Phi\left(\frac{(u - v_{L_n}(y) - l_n(y))}{0}\right) d\Phi_n(y).$$

The integrant $\Phi\left(\frac{u - l_n(y) - v_{L_n}(y)}{0}\right)$ of the first integral will take the value 1, if $l_n(y) < u - v_{L_n}(y) < u + \varepsilon$, but the value 0 if $l_n(y) > u - v_{L_n}(y)$. Thus, the integrant is always $\leq \Phi\left(\frac{u + \varepsilon - l_n(y)}{0}\right)$ which takes the value 1 for each value y such that $l_n(y) < u + \varepsilon$. If we replace $\Phi\left(\frac{u - v_{L_n}(y) - l_n(y)}{0}\right)$

by $\Phi\left(\frac{u+\varepsilon-l_n(y)}{0}\right)$ and add together the not-negative differentials of probability $d\,\Phi_n(y)$ over all the space for which $\Phi\left(\frac{u+\varepsilon-l_n(y)}{0}\right)=1$ and not only over the part of it having the property $|v_{L_n}(y)|\leqq\varepsilon$, we are sure to obtain a result which is greater than or equal to the integral considered. Now $\int\Phi\left(\frac{u+\varepsilon-l_n(y)}{0}\right)d\,\Phi_n(y)=\Phi(u+\varepsilon)$ for which reason the first considered integral in (59) is $\leq\Phi(u+\varepsilon)$.

If we in the other integral replace the integrant by 1, we are sure to obtain a result which is not less than that obtained when the integration takes place over $\Phi\left(\frac{u-l_n(y)-v_n(y)}{0}\right)\leq 1$, which also can take the value 0 within some part of the space of integration $|v_{L_n}(y)|>\varepsilon$. The integral $\int\limits_{|v_{L_n}(y)|>\varepsilon}d\,\Phi_n(y)=P_{L_n}(a_n\varepsilon)$ indicates the probability that $|v_{L_n}(y)|>\varepsilon$. Thus we have for each ε

$$W_n(a_n u+b_n)\leq\Phi(u+\varepsilon)+P_{L_n}(a_n\varepsilon)<\Phi(u)+ \tag{60}$$
$$+P_{L_n}(a_n\varepsilon)+\frac{\varepsilon}{\sqrt{2\pi}}.$$

Estimating in the same way the size of the number

$$1-W_n(a_n u-b_n)=\int\Phi\left(\frac{v_{L_n}(y)+l_n(y)-u}{0}\right)d\,\Phi_n(y) \tag{61}$$

we obtain the formula

$$1-W_n(a_n u+b_n)\leq 1-\Phi(u-\varepsilon)+P_{L_n}(a_n\varepsilon) \tag{62}$$

i. e.

$$W_n(a_n u+b_n)\geq\Phi(u-\varepsilon)-P_{L_n}(a_n\varepsilon)\geq \tag{63}$$
$$\geq\Phi(u)-\frac{\varepsilon}{\sqrt{2\pi}}-P_{L_n}(a_n\varepsilon).$$

Thus we have

$$|W_n(a\,u+b_n)-\Phi(u)|\leq\frac{\varepsilon}{\sqrt{2\pi}}+P_{L_n}(a_n\varepsilon).$$

If therefore the linear expression $L_n(\boldsymbol{y})$ can be chosen in such a manner that the probability of $|v_{L_n}(\boldsymbol{y})| > \varepsilon$ converges towards zero when $n \to \infty$ for each positive number ε, $W_n(a_n u + b_n)$ accordingly converges towards $\Phi(u)$ for each value of u.

If $u_n(\boldsymbol{x}) = \sum_{k=1}^{n} x_k$, $F_n(\boldsymbol{x}) = \prod_{k=1}^{n} V_k(x_k + \beta_k)$, the corresponding function $v_n(\boldsymbol{y}) = \sum_{k=1}^{n} x_k(y_k)$

and

$$v_{L_n}(\boldsymbol{y}) = \sum_{k=1}^{n} \left(\frac{x_k(y_k) - L_{nk} y_k - \beta_k}{a_n} \right); \; b_n = \sum_{k=1}^{n} \beta_k.$$

We will further prove the following theorem.

Theorem 3. *If the number* $k_n = \operatorname{Min}_{\alpha > 0} k_{n\alpha} \to 0$,

where

$$k_{n\alpha}^{\alpha} = \operatorname{Min}_{L_n(y)} (E| \Phi(v_{L_n}(\boldsymbol{y})) - \tfrac{1}{2}|^{\alpha} = k_{n\alpha}^{\alpha}(L_n)) \tag{64}$$

and further

$$v_{L_n}(\boldsymbol{y}) = \frac{1}{a_n} \left(v_n(\boldsymbol{y}) - \sum_{\varrho=1}^{n} L_{n\varrho} y_\varrho - b_n \right), \tag{65}$$

$$P(y_1 \leq t_1, \ldots, y_n \leq t_n) = \prod_{\varrho=1}^{n} \Phi(t_\varrho),$$

$$a_n^2 = \sum_{\varrho=1}^{n} L_{n\varrho}^2, \tag{66}$$

the distribution function $W_n(a_n u + b_n)$ of

$$v_{n\alpha b} = (v_n(\boldsymbol{y}) - b_n) : a_n \tag{67}$$

converges towards

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt. \tag{68}$$

By the theorem of Taylor we have

$$\Phi(v_{L_n}(y)) = \Phi(0) + \Phi'(\theta\, v_{L_n}(y)) \cdot v_{L_n}(y); \quad 0 < \theta < 1. \tag{69}$$

Because $\Phi(0) = \frac{1}{2}$, $\frac{1}{\sqrt{2\pi}} e^{-\frac{v_{L_n}(y)^2}{2}} \leq \Phi'(\theta\, v_{L_n}(y)) =$

$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta v_{L_n}(y))^2} \leq \frac{1}{\sqrt{2\pi}}$, $|\Phi(x) - \frac{1}{2}| \leq \frac{1}{2}$ we thus find that when $\alpha > 0$

$$\left(\frac{|v_{L_n}(y)|}{\sqrt{2\pi}}\right)^\alpha e^{-\frac{\alpha}{2}(v_{L_n}(y))^2} \leq |\Phi(v_{L_n}(y)) - \tfrac{1}{2}|^\alpha \leq \tag{70}$$

$$\leq \operatorname{Min}\left((\tfrac{1}{2})^\alpha, \left|\frac{v_{L_n}(y)}{\sqrt{2\pi}}\right|\right)^\alpha.$$

We now define the two stochastic variables $\varepsilon_1(y;\, L_n)$ and $\varepsilon_2(y;\, L_n)$ in the following way:

$$\varepsilon_1(y;\, L_n) = \frac{\varepsilon}{\sqrt{2\pi}} e^{-\frac{1}{2}\varepsilon^2}; \qquad |v_{L_n}(y)| > \varepsilon, \tag{71}$$

$$\varepsilon_1(y;\, L_n) = 0; \qquad |v_{L_n}(y)| \leq \varepsilon,$$

and

$$\varepsilon_2(y;\, L_n) = \tfrac{1}{2}; \qquad |v_{L_n}(y)| > \varepsilon, \tag{72}$$

$$\varepsilon_2(y;\, L_n) = \frac{|v_{L_n}(y)|}{\sqrt{2\pi}}; \qquad |v_{L_n}(y)| \leq \varepsilon.$$

Then we have for every y and $\alpha > 0$

$$\varepsilon_1(y;\, L_n)^\alpha \leq |\Phi(v_{L_n}(y)) - \tfrac{1}{2}|^\alpha \leq \varepsilon_2(y,\, L_n)^\alpha. \tag{73}$$

When we calculate the mathematical expectation of the stochastic variables of (73), we find that

$$E(\varepsilon_1(y;\, L_n))^\alpha \leq k^\alpha_{n_\alpha}(L_n) \leq E(\varepsilon_2(y);\, L_n)^\alpha. \tag{74}$$

The lower limit in (74) may be written

$$E(\varepsilon_1(y;\, L_n))^\alpha = \frac{\varepsilon^\alpha}{(\sqrt{2\pi})^\alpha} e^{-\frac{\alpha}{2}\varepsilon^2} P_{L_n}(\varepsilon a_n) \tag{75}$$

where

$$(76)\quad P_{L_n}(\varepsilon a_n)=P(|v_{L_n}(\boldsymbol{y})|>\varepsilon)=P(|v_n(\boldsymbol{y})-\sum_{\varrho=1}^{n} L_{n\varrho}\, y_\varrho - b_n|>\varepsilon a_n).$$

The upper limit of $k_{n\alpha}^{\alpha}$ is a number that may be calculated by the formula

$$(77)\quad K_{L_n}^{\alpha}(\varepsilon,\alpha)=E(\varepsilon_2(\boldsymbol{y}, L_n))^{\alpha}=\frac{1}{2^{\alpha}}P_{L_n}(\varepsilon a_n)+$$

$$+\int_{z=0}^{\varepsilon}\frac{z^{\alpha}}{(\sqrt{2\pi})^{\alpha}}d(1-P_{L_n}(z a_n)).$$

If the number $k_n \to 0$, there exist such linear expressions $L_n(\boldsymbol{y})$, $n=1,\ldots,\infty$ that $P_{L_n}(\varepsilon a_n)\to 0$, which (as said in the Theorem 2) is a sufficient condition for the convergence of $W_n(a_n u + b_n)\to \Phi(u)$. If $P_{L_n}(\varepsilon a_n)\to 0$ for every $\varepsilon>0$, we may choose ε so small that $K_{L_n}(\varepsilon,\alpha)$ will be smaller than every fixed number ε'. Accordingly, if there exist such linear expressions $L_n(\boldsymbol{y})$ that $P_{L_n}(\varepsilon a_n)\to 0$ for every $\varepsilon>0$, the corresponding $k_{n\alpha}$ must converge towards zero.

If the upper limit K_n of k_n, that may be obtained by minimizing $K_{L_n}(\varepsilon,\alpha)$ with respect to variations in $L_n(\boldsymbol{y})$, ε and $\alpha>0$, converges towards zero, $k_n\to 0$. We can thus formulate a new theorem:

Theorem 4. *If the number*

$$(78)\quad K_n=\operatorname*{Min}_{\alpha>0,\ \varepsilon,\ L_n(y)} K_{L_n}(\varepsilon,\alpha)\to 0;$$

$$K_{L_n}(\varepsilon,\alpha)=\left(\frac{1}{2^{\alpha}}P_{L_n}(\varepsilon a_n)+\int_{z=0}^{\varepsilon}\left(\frac{z}{\sqrt{2\pi}}\right)^{\alpha}d(1-P_{L_n}(z a))\right)^{\frac{1}{\alpha}},$$

the distribution function of $(v_n(\boldsymbol{y})-b_n):a_n$ *converges towards* $\Phi(u)$.

This theorem is a special case of the following more general Central Limit Theorem.

Theorem 5. *If there exists a subspace* A_n *such that*

$$(79)\quad K_n(A_n,\alpha)=\operatorname*{Min}_{L_n(y)}\left(\frac{1}{2^\alpha}P(\boldsymbol{y}<A_n)+\int\limits_{y\not<A_n}\left(\frac{|v_{L_n}(y)|}{\sqrt{2\pi}}\right)^\alpha d\,\boldsymbol{\Phi}_n(\boldsymbol{y})\right)^{\frac{1}{\alpha}}$$

converges towards zero, for some (α,ε) *the function* $W_n(a_n u+b_n)\to\boldsymbol{\Phi}(u)$.

The probability that the point $\boldsymbol{y}$ will be a point of A_n we have denoted by

$$(80)\qquad P(\boldsymbol{y}<A_n)=\int\limits_{y<A_n} d\,\boldsymbol{\Phi}_n(\boldsymbol{y})=\int\limits_{y<A_n}\prod_{\varrho=1}^{n} d\,\boldsymbol{\Phi}(y_\varrho).$$

The symbol $\boldsymbol{y}\not<A_n$ is used to denote all the points $\boldsymbol{y}$ that *do not* belong to A_n.

By help of the inequality (70) we find that $K_n(A_n,\alpha)>k_n$. If $K_n(A_n,\alpha)\to 0$, the number $k_n\to 0$. By the Theorem 3 we then conclude that $W_n(a_n+b_n)\to\boldsymbol{\Phi}(u)$.

If $v_{L_n}(\boldsymbol{y})$ can be expressed as a sum of n independent variables $\Delta v_{L_n}(y_\varrho)$, that is if

$$(81)\qquad v_{L_n}(\boldsymbol{y})=\sum_{\varrho=1}^{n}\Delta v_{L_n}(y_\varrho)=\sum_{\varrho=1}^{n}\left(\frac{x_\varrho(y_\varrho)\,L_{n\varrho}\,y_\varrho-\beta_{n\varrho}}{a_n}\right)$$

we can for instance define the subspace A_n so that $\boldsymbol{y}$ will belong to A_n if $|\Delta v_{L_n}(y_\varrho)|>\varepsilon_n$ for some $\varrho=1,\ldots,n$. In this case

$$(82)\quad P(\boldsymbol{y}<A_n)=1-\prod_{\varrho=1}^{n}(1-P(|\Delta v_{L_n}(y_\varrho)|>\varepsilon_n))<$$

$$<\sum_{\varrho=1}^{n}P(|\Delta v_{L_n}(y_\varrho)|>\varepsilon_n)$$

that is

$$P(\boldsymbol{y}<A_n)<\sum_{\varrho=1}^{n}P_\varrho(\varepsilon_n a_n)$$

where

$$P_\varrho(\varepsilon\,a_n)=P(|x_\varrho(y_\varrho)-L_{n\varrho}\,y_\varrho-\beta_{n\varrho}|>\varepsilon_n a_n).$$

We will now study the special case $\alpha=2$. Here we have by (82) and Theorem 5:

$$(83)\quad K_n^2 \leqq \left(\frac{1}{4}\sum_{\varrho=1}^{n} P_\varrho(X_n) + \frac{1}{2\pi}\int\limits_{y \not\subset A_n(X_n)} (v_{L_n}(\boldsymbol{y}))^2\, d\,\boldsymbol{\Phi}_n(\boldsymbol{y})\right) = \\ = H^2(A_n(X_n);\, L_n)$$

where $\quad P_\varrho(X_n) = P(|x_\varrho(y_\varrho) - L_{n\varrho}\, y_\varrho - \beta_{n\varrho}| > X_n),$

$$A_n = A_n(X_n) = \sum_{\varrho=1}^{n} (|x_\varrho(y_\varrho) - L_n\, y_\varrho - \beta_{n\varrho}| > X_n).$$

If we choose $L_{n\varrho}$, and $\beta_{n\varrho}$ so that the »truncated» variance of $a_n\, v_{L_n}(\boldsymbol{y}) = \sum_{\varrho=1}^{n} (x_\varrho(y_\varrho) - L_{n\varrho}\, y_\varrho - \beta_{n\varrho})$,

$$(84)\quad \sigma^2(X_n;\, L_n(\boldsymbol{y})) = \int\limits_{y \not\subset A_n} (\Sigma(x_\varrho(y_\varrho) - L_{n\varrho}\, y_\varrho - \beta_{n\varrho}))^2\, d\,\boldsymbol{\Phi}_n(\boldsymbol{y})$$

will be a minimum in respect of variations in the coefficients of the linear function $L_n(\boldsymbol{y})$, we find that

$$(85)\quad L_{n\varrho} \approx L_{n\varrho}(X_n) = \frac{\int\limits_{y_\varrho \not\subset A_n} (x_\varrho(y_\varrho) - \beta_{n\varrho}(X_n))\, y_\varrho\, d\,\boldsymbol{\Phi}(y_\varrho)}{\int\limits_{y_\varrho \not\subset A_n} y_\varrho^2\, d\,\boldsymbol{\Phi}(y_\varrho) - \left(\int\limits_{y_\varrho \not\subset A_n} y_\varrho\, d\,\boldsymbol{\Phi}(y_\varrho)\right)^2}$$

$$L_{n\varrho}(X_n) \to E(x_\varrho\, y_\varrho) = E(x_\varrho\, \boldsymbol{\Phi}^{-1}(V_\varrho(x_\varrho)) \text{ when } X_n \to \infty.$$

$$(86)\quad \beta_{n\varrho} \approx \beta_{n\varrho}(X_n) = \frac{\int\limits_{y_\varrho \not\subset A_n} x_\varrho(y_\varrho)\, d\,\boldsymbol{\Phi}(y_\varrho)}{\int\limits_{y_\varrho \not\subset A_n} d\,\boldsymbol{\Phi}(y_\varrho)} \to E(x_\varrho), \text{ when } X_n \to \infty.$$

$$(87)\quad a_n^2 \approx (a_n(X_n))^2 = \sum_{\varrho=1}^{n} (L_{n\varrho}(X_n))^2 \to \\ \to \sum_{\varrho=1}^{n} (E(x_\varrho\, y_\varrho))^2 \text{ when } X_n \to \infty.$$

$$(88)\quad b_n \approx b_n(X_n) = \sum_{\varrho=1}^{n} \beta_{n\varrho}(X_n) \to \\ \to \sum_{\varrho=1}^{n} E(x_\varrho) = E\,u_n(\boldsymbol{x}) \text{ when } X_n \to \infty.$$

$$(89)\quad \sigma^2(X_n) = \sum_{\varrho=1}^{n} \sigma_\varrho^2(X_n) \to \sum_{\varrho=1}^{n} (E(x_\varrho^2) - (E\,x_\varrho\,y_\varrho)^2) \text{ when } X_n \to \infty.$$

$$(90)\quad \sigma_\varrho^2(X_n) = \int_{y_\varrho \notin A_n} (x_\varrho(y_\varrho))^2\, d\,\Phi(y_\varrho) - L_{n\varrho}(X_n)^2 \cdot \cdot \Big(\int_{y_\varrho \notin A_n} y_\varrho^2\, d\,\Phi(y_\varrho) - \Big(\int_{y_\varrho \notin A_n} y_\varrho\, d\,\Phi(y_\varrho)\Big)^2\Big).$$

Because $\int_{y \notin A_n} (v_{L_n}(\boldsymbol{y}))^2\, d\,\Phi_n(\boldsymbol{y}) = \sigma_n^2(X_n) : a_n^2(X_n)$, when $L_n(\boldsymbol{y}) = \sum_{\varrho=1}^{n} L_{n\varrho}(X_n)\, y_\varrho + b_n(X_n)$ and because $H^2(A_n(X_n);\ L_n) > k_n^2$, we find by Theorem 3.

Theorem 6. *If* $v_n(\boldsymbol{y}) = \sum_{\varrho=1}^{n} x_\varrho(y_\varrho)$, *where* $P(y_1 \leq t_1, \ldots, y_n \leq t_n) = \prod_{\varrho=1}^{n} \Phi(t_\varrho)$ *and*

$$(91)\quad H_n^2 = \operatorname*{Min}_{X_n} \left(\frac{1}{4} \sum_{\varrho=1}^{n} P_\varrho(X_n) + \frac{1}{2\pi} \frac{\sigma^2(X_n)}{a^2(X_n)} \right) \to 0,$$

the distribution function of $(v_n(y) - b_n) : a_n$ *converges towards* $\Phi(u)$ *for* $b_n = b_n(X_n)$ *and* $a_n = a_n(X_n)$, *if* $H^2(A_n(X_n);\ L_n) = H_n^2$.

This Theorem says a little more than the corresponding theorems of Feller [1] and the theorems of M. Loevy [5].

We now suppose that the numbers a_n and $b_n = b_{na}$ have been determined so that

$$(92)\quad E(\Phi(v_{na}(\boldsymbol{y})) - \tfrac{1}{2}) = E(\Phi(l_n(\boldsymbol{y})) - \tfrac{1}{2}) = 0$$

$$(93)\quad E(\Phi(v_{na}(\boldsymbol{y})) - \tfrac{1}{2})^2 = E(\Phi(l_n(\boldsymbol{y})) - \tfrac{1}{2})^2 = = \int (\Phi - \tfrac{1}{2})^2\, d\,\Phi = \tfrac{1}{12}.$$

The greatest coefficient of correlation that can be obtained between $\Phi(v_{na}(\boldsymbol{y}))$ and $\Phi(l_n(\boldsymbol{y}))$ by varying the coefficients of the linear equation

$$l_n(\boldsymbol{y}) = \sum_{\varrho=1}^{n} l_{n\varrho}\, y_\varrho;\ \sum_{\varrho=1}^{n} l_{n\varrho}^2 = 1,$$

can then be calculated by means of the formula

$$R_n = 12 \operatorname*{Max}_{l_n(y)} E\left(\Phi(v_{na}(\boldsymbol{y})) - \tfrac{1}{2}\right)\left(\Phi(l_n(\boldsymbol{y})) - \tfrac{1}{2}\right) \tag{94}$$

$$= 1 - 6 \operatorname*{Min}_{l_n(y)} E\left(\Phi(v_{na}(\boldsymbol{y})) - \Phi(l_n(\boldsymbol{y}))\right)^2 =$$

$$= 1 - 6 \operatorname*{Min}_{l_n(y)} \left(E\left(\Phi(v_{na}(\boldsymbol{y})) - \tfrac{1}{2}\right)^2 + \right.$$

$$\left. + E\left(\Phi(l_n(\boldsymbol{y})) - \tfrac{1}{2}\right)^2 - 2E\left(\Phi(v_{na}(\boldsymbol{y})) - \tfrac{1}{2}\right)\left(\Phi(l_n(\boldsymbol{y})) - \tfrac{1}{2}\right)\right)$$

$$= 1 - 6\left(\tfrac{1}{12} + \tfrac{1}{12} - \operatorname*{Max}_{l_n(y)} \tfrac{2}{12} \cdot \left(12 \cdot E\left(\Phi(v_{na}(\boldsymbol{y})) - \tfrac{1}{2}\right)\left(\Phi(l_n(\boldsymbol{y}) - \tfrac{1}{2}\right)\right).$$

If $R_n \to 1$, $\operatorname{Min} E\left(\Phi(v_{na}(\boldsymbol{y})) - \Phi(l_n(\boldsymbol{y}))^2\right) \geq \left(\operatorname{Min} E \left| \Phi(v_{na}(\boldsymbol{y})) - \Phi(l_n(\boldsymbol{y}))\right|\right)^2 = \varepsilon_n^2$ must converge towards zero. According to Theorem 1 we thus get

Theorem 7. *If the coefficient of correlation R_n between $\Phi(v_{na}(\boldsymbol{y}))$ and $\Phi(l_n(\boldsymbol{y}))$ converges towards 1, the distribution function of $v_{na}(\boldsymbol{y})$ converges towards $\Phi(u)$.*

The difference $w_{na}(u) = W_n(a_n u + b_{na}) - \Phi(u)$ can in a simple way be expressed by means of a sum of ortogonal polynoms $\alpha_n \acute{p}_k(s)$ of the variable $s = \Phi(u) - \frac{1}{2}$. These polynoms have the properties

$$\int \acute{p}_k(s) \cdot \acute{p}_g(s)\, ds = 0; \qquad k \neq g \tag{95}$$

$$\int \acute{p}_k(s)\ \acute{p}_k(s)\, ds = 1; \qquad k = g.$$

We denote the integrals of these ortogonal polynoms $p_k(s) = \int_0^s \acute{p}_k(s)\, ds$. The first four ortogonal polynoms of this kind are

$$\acute{p}_0(s) = 1, \quad \acute{p}_1(s) = \sqrt{3}\,(2s), \quad \acute{p}_2(s) = \sqrt{5}\,(6s^2 - \tfrac{1}{2}), \tag{96}$$

$$\acute{p}_3(s) = \sqrt{7}\,(20s^3 - 3s).$$

(97) $$p_0(s) = s, \quad p_1(s) = \sqrt{3}\, s^2, \quad p_2(s) = \sqrt{5}\,(2\,s^3 - \tfrac{1}{2}\,s),$$

$$p_3(s) = \sqrt{7}\,(5\,s^4 - \tfrac{3}{2}\,s^2)$$

generally:

(98) $$\acute{p}_k(s) = \sqrt{2\,k+1} \cdot \left((2\,s)^k - \sum_{g=0}^{k-1} c_{k\,g}\, \acute{p}_g(s) \right) : \sqrt{1 - \sum_{g=1}^{k-1} c_{k\,g}^2}$$

and

(99) $$p_k(s) = \sqrt{2\,k+1} \left(\frac{(2\,s)^{k+1}}{k+1} - \sum_{g=0}^{k-1} c_{k\,g}\, p_g(s) \right) : \sqrt{1 - \sum_{g=1}^{k-1} c_{k\,g}^2}$$

where

(100) $$c_{k\,g} = \sqrt{2\,k+1} \int (2\,s)^k \acute{p}_g(s)\, ds.$$

When we put

(101) $$w_{n\,a}(u) = E\left(\Phi\left(\frac{u_g - v_{n\,a}(\boldsymbol{y})}{0} \right) - \Phi\left(\frac{u_g - l_n(\boldsymbol{y})}{0} \right) \right) = \sum_{k=0}^{\infty} \alpha_{n\,g}\, \acute{p}_g(s)$$

and determine the coefficients so that the variances

(102) $$w_{n\,a\,\nu}^2 = \int \left(w_{n\,a}(u_s) - \sum_{g=1}^{\nu} \alpha_{n\,g} \acute{p}_g(s) \right)^2 ds$$

will be a minimum, we find for the coefficients $\alpha_{n\,g}$ the expressions

(103) $$\alpha_{n\,g} = \int w_{n\,a}(u_s)\, \acute{p}_g\ (s)\, ds =$$

$$= E\left(\int \Phi\left(\frac{u_s - v_{n\,a}(\boldsymbol{y})}{0} \right) \acute{p}_g(s)\, ds - \int \Phi\left(\frac{u_s - l_n(\boldsymbol{y})}{0} \right) \acute{p}_g(s)\, ds \right) =$$

$$= E(p\ \ (\Phi(v_{n\,a}(\boldsymbol{y})) - \tfrac{1}{2})) - E(p_g(\Phi(l_n(\boldsymbol{y})) - \tfrac{1}{2})).$$

The numbers a_n and $b_{n\,a}$ can be chosen so that

(104) $$\alpha_{n\,0} = E(\Phi(v_{n\,a}(\boldsymbol{y}) - \tfrac{1}{2}) - E(\Phi(l_n(\boldsymbol{y})) - \tfrac{1}{2}) =$$

$$= E(\Phi(v_{n\,a}(\boldsymbol{y}) - \tfrac{1}{2}) - 0 = 0.$$

$$(105)\quad \alpha_{n1} = \sqrt{3}\,(E(\Phi(v_{na}(y)) - \tfrac{1}{2})^2 - E(\Phi(l_n(y)) - \tfrac{1}{2})^2) =$$

$$= \sqrt{3}\,(E(\Phi(v_{na}(y)) - \tfrac{1}{2})^2 - \tfrac{1}{12}) = 0.$$

$$\sqrt{3}\,(\tfrac{1}{12} - \tfrac{1}{12}) = 0.$$

For $g > 1$ we have further $E p_g(\Phi(l_n(y)) - \tfrac{1}{2})) = \int p_g(s)\,ds =$ $= \tfrac{1}{2}(p_g(\tfrac{1}{2}) + p_g(-\tfrac{1}{2}))$.

We thus find that when $g \geqq 2$,

$$(106)\quad \alpha_{ng} = E(p_g(\Phi(v_{na}(y)) - \tfrac{1}{2}) - \tfrac{1}{2}(p_g(\tfrac{1}{2}) + p_g(-\tfrac{1}{2})).$$

For $g = 2$ and 3 we get:

$$(107)\quad \alpha_{n2} = \sqrt{5}\,(2\,E(\Phi(v_{na}(y)) - \tfrac{1}{2})^3,$$

$$(108)\quad \alpha_{n3} = \sqrt{7}\,(5\,E(\Phi(v_{na}(y)) - \tfrac{1}{2})^4 - \tfrac{1}{16}).$$

The corresponding variances $w^2_{na\nu}$ can be written in the form

$$(109)\quad w^2_{na\nu} = w^2_{na} - \sum_{g=0}^{\nu} \alpha^2_{ng} \geqq 0.$$

If $w^2_{na} \to 0$ *we thus must have* $\alpha^2_{ng} \to 0$ *for every* g.

As known from the theory of orthogonal polynoms $\overset{\infty}{\underset{\nu}{\uparrow}} w^2_{na\nu} \to 0$ for functions, which like $w_{na}(u_s) = W_n(a_n u_s + b_{na}) - \tfrac{1}{2} - s$ have the property $\int |d\,w_{na}(u_s)| < \text{const.}$ *If therefore* $\alpha^2_{ng} \to 0$ *for every* g, $w^2_{na} \to 0$, $w_{na}(u) \to 0$ *and* $W_n(a_n u + b_{na}) \to \Phi(u)$.

[1]. W. Feller: *Zentraler Grenzwertsatz der Wahrscheinlichkeitsrechnung* II. Mathematische Zeitschrift, 42. Band, 1937, pp. 302—303.

[2]. R. v. Mises: *Deux nouveaux théorèmes de limite de calcul des probabilités.* Revue Fac. Sc. Istanbul, 1 (1935), p. 61.

——: *Les lois de probabilité pour les fonctions statistiques.* Ann. Inst. Henri Poincaré, 6 (1936), p. 185.

[3]. H. CRAMÉR: *Mathematical Methods of Statistics*, Uppsala 1945, p. 218 and 363 foll.
[4]. ——: *Mathematical Methods of Statistics*, Uppsala 1945, p. 245.
[5]. M. MICHEL LOÈVE: Thèses presentées a la Faculté des Sciences de L'Université de Paris. 1re Thèse. — Étude asymptotique des sommes de variable aléatoires liées. 2e Thèse. — Espaces compacts et localement compacts, Paris 1946, p. 42.

(Received Oct. 1946.)

SOME NEW PRINCIPLES FOR SOLVING LINEAR PROGRAMMING PROBLEMS

By

LEO TÖRNQVIST

University of Helsinki, Finland

When linear programming problems have to be solved in practice the crucial problem is to find elements $h \in \hat{H}$, $k \in \hat{K}$ for which the optimal values $\hat{x}_k > 0$ and the optimal multipliers $\hat{\lambda}_h > 0$.

The paper gives exact criteria for $h \in \hat{H}$, $h \notin \hat{H}$, $k \in \hat{K}$, $k \notin \hat{K}$, and for the cases when no finite optimal solution exists. Because the exact rules might be difficult to apply, rules for defining good surrogate sets for $(\hat{H}, \hat{K})$ are also given.

Some numerical examples are solved by means of the new principles. The so-called transportation problem is considered in detail. Some remarks about the probability of the existence of finite optimal solutions to linear programming problems and the distribution of $(\hat{x}, \hat{\lambda})$ are made in an appendix.

We consider a linear programming problem and its dual problem. The problem is to find the optimal values

$$\hat{x} = \{\hat{x}_j;\ j \in J\}, \quad \hat{x}_j > 0 \text{ for } j \in \hat{K},\ \hat{x}_j = 0 \text{ for } j \in J - \hat{K},$$

$$\hat{\lambda} = \{\hat{\lambda}_i;\ i \in I\}, \quad \hat{\lambda}_i > 0 \text{ for } i \in \hat{H},\ \hat{\lambda}_i = 0 \text{ for } i \in J - \hat{H}$$

with the properties

$$\sum_{j \in \hat{K}} a_{0j} \hat{x}_j = \max_{x \in X_A} \sum_{j \in J} a_{0j} x_j,$$

$$\sum_{i \in \hat{H}} \hat{\lambda}_i a_{i0} = \min_{\lambda \in L_A} \sum_{i \in I} \lambda_i a_{i0},$$

where

$$x \in X_A \quad \text{if } x_i = a_{i0} + \sum_{j \in J} a_{ij} x_j \geqslant 0 \text{ for } i \in I,$$

$$x_j \geqslant 0 \text{ for } j \in J,$$

and

$$\lambda \in L_A \quad \text{if } -\lambda_j = a_{0j} + \sum_{i \in I} \lambda_i a_{ij} \leqslant 0 \text{ for } j \in J,$$

$$\lambda_i \geqslant 0 \text{ for } i \in I.$$

The crucial problem is to find the sets $\hat{H}$ and $\hat{K}$. Once these sets are known, the remaining problem is reduced to that of finding the solutions to the system of equations

$$x_h = a_{h0} + \sum_{k \in \hat{K}} a_{hk} x_k = 0 \quad \text{for } h \in \hat{H},$$

$$-\lambda_k = a_{0k} + \sum_{h \in \hat{H}} \lambda_h a_{hk} = 0 \quad \text{for } k \in \hat{K}.$$

For it can be proved that

$$\hat{x}_i \hat{\lambda}_i = 0 \quad \text{for } i \in I, \qquad \hat{\lambda}_j \hat{x}_j = 0 \quad \text{for } j \in J.$$

If we know that $h \in \hat{H}$ we can reduce the number of independent variables x_j by expressing x_k, where k preferably is an element of $\hat{K}$, by means of other independent variables according to the formula

$$x_k = -(a_{h0} + \sum_{j \in J-k} a_{hj} x_j)/a_{hk}.$$

If we know that $k \in \hat{K}$ we can express λ_h by means of other λ_i; $i \in I - h$ according to the formula

$$\lambda_h = -(a_{0k} + \sum_{i \in I-h} a_{ik} \lambda_i)/a_{hk}$$

and thus reduce the number of independent variables for which optimal values $\hat{x}_j$ and $\hat{\lambda}_i$ have to be found.

We denote $a_{\lambda j} = \sum_{i \in I} \lambda_i a_{ij}$ for $\lambda_i \geqslant 0$ and $a_{ix} = \sum_{j \in J} a_{ij} x_j$ for $x_j \geqslant 0$.

The following criteria tell us if k belongs to $\hat{K}$ or $j \in J - \hat{K}$, and if h belongs to $\hat{H}$ or $i \in I - \hat{H}$.

1. CRITERION FOR $k \in \hat{K}$.—If the set

$$L_k = \{\lambda;\ a_{\lambda 0} < 0,\ a_{\lambda k} > 0,\ a_{\lambda j} \leqslant 0 \text{ for all } j \in J - k\}$$

is not empty, then $k \in \hat{K}$, $\hat{x}_k > 0$, and $\hat{\lambda}_k = -a_{0k} - a_{\hat{\lambda} k} = 0$.

2. CRITERION FOR $j \in J - \hat{K}$.—If the set

$$X_j = \{x;\ x_j = 0,\ a_{0j} < a_{0x},\ a_{ij} \leqslant a_{ix} \text{ for all } i \in I\}$$

is not empty, then $j \in J - \hat{K}$, $\hat{x}_j = 0$, $\hat{\lambda}_j = -a_{0j} - a_{\hat{\lambda} j} \geqslant 0$.

3. CRITERION FOR $h \in \hat{H}$.—If the set

$$X_h = \{x;\ a_{0x} > 0,\ a_{hx} < 0,\ a_{ix} \geqslant 0 \text{ for all } i \in I - h\}$$

is not empty, then $h \in \hat{H}$, $\hat{\lambda}_h > 0$, $x_h = a_{h0} + a_{h\hat{x}} = 0$.

4. CRITERION FOR $i \in I - \hat{H}$.—If the set

$$L_i = \{\lambda;\ \lambda_i = 0,\ a_{i0} > a_{\lambda 0},\ a_{ij} \geqslant a_{\lambda j} \text{ for all } j \in J\}$$

is not empty, then $i \in I - \hat{H}$, $\hat{\lambda}_i = 0$, $x_i = a_{i0} + a_{i\hat{x}} \geqslant 0$.

To prove that these criteria are correct we only need to point out that for all $x_j \geqslant 0$, $\lambda_i \geqslant 0$ we have

$$\hat{x}_\lambda = a_{\lambda 0} + \sum_j a_{\lambda j} \hat{x}_j \geqslant 0 \quad \text{and} \quad \hat{x}_k > 0 \quad \text{if} \quad \lambda \in L_k,$$

$$-\hat{\lambda}_x = a_{0x} + \sum_i a_{ix} \hat{\lambda}_i \leqslant 0 \quad \text{and} \quad \hat{\lambda}_h > 0 \quad \text{if} \quad x \in X_h,$$

$$\hat{x}_i - \hat{x}_\lambda = (a_{i0} - a_{\lambda 0}) + \sum_j (a_{ij} - a_{\lambda j}) \hat{x}_j > 0 \quad \text{if} \quad \lambda \in L_i,$$

$$-\hat{\lambda}_j + \hat{\lambda}_x = (a_{0j} - a_{0x}) + \sum_i (a_{ij} - a_{ix}) \hat{\lambda}_i < 0 \quad \text{if} \quad x \in X_j,$$

while $$\hat{x}_i \hat{\lambda}_i = 0 \text{ for } i \in I, \text{ and } \hat{\lambda}_j \hat{x}_j = 0 \text{ for } j \in J.$$

Simple special cases of these criteria, where at most one $\lambda_i > 0$ for $i \in I$ or $x_j > 0$ for $j \in J$, are easily applied in practice, and may lead to the result that some of the sets mentioned can be shown to be nonempty. If L_k is nonempty we can reduce the number of independent variables x_j, and if X_h is nonempty we can reduce the number of independent λ_i. If we find that L_i is nonempty we can leave out the constraint $x_i \geqslant 0$ as an irrelevant condition. If X_j is nonempty we have $\hat{x}_j = 0$ and the terms $a_{0j}x_j$, $a_{ij}x_j$ can be left out from all the linear forms considered.

Sometimes it may be difficult to find elements in the sets mentioned. In such cases it is not worth exerting much effort for this purpose. If we find some elements $h \in H$ or $k \in K$ which seem likely to have the property $h \in \hat{H}$ or $k \in \hat{K}$, it might be preferable to transform the problem into a new form by expressing the variables x_k; $k \in K$ and λ_h; $h \in H$ as linear forms of x_j; $j \in J - K$ and λ_i; $i \in I - H$ respectively, and of the "slack" variables x_h; $h \in H$ and λ_k; $k \in K$ respectively.

We will then have identically for every square matrix $a_{HK} = \{a_{hk};\ h \in H,\ k \in K\}$

$$\det \begin{Bmatrix} x_i & a_{iK} \\ x_H & a_{HK} \end{Bmatrix} = \det \begin{Bmatrix} a_{i0} & a_{iK} \\ a_{H0} & a_{HK} \end{Bmatrix} + \sum_{j \in J-K} \det \begin{Bmatrix} a_{ij} & a_{iK} \\ a_{Hj} & a_{HK} \end{Bmatrix} x_j,$$

$$\det \begin{Bmatrix} -\lambda_j & -\lambda_K \\ a_{Hj} & a_{HK} \end{Bmatrix} = \det \begin{Bmatrix} a_{0j} & a_{0K} \\ a_{Hj} & a_{HK} \end{Bmatrix} + \sum_{i \in I-H} \lambda_i \det \begin{Bmatrix} a_{ij} & a_{iK} \\ a_{Hj} & a_{HK} \end{Bmatrix},$$

where a_{iK} denotes the row vector $\{a_{ik};\ k \in K\}$ and a_{Hj} the column vector $\{a_{hj};\ h \in H\}$, and det = "determinant of".

For $k \in K$ we have

$$(\det a_{HK})\, x_k + \sum_{h \in H} \frac{\partial (\det a_{HK})}{\partial a_{hk}} (a_{h0} - x_h + \sum_{j \in J-K} a_{hj} x_j) = 0,$$

and for $h \in H$

$$(\det a_{HK})\, \lambda_h + \sum_{k \in K} \frac{\partial (\det a_{HK})}{\partial a_{hk}} (a_{0k} + \sum_{j \in I-H} \lambda_i a_{ik} + \lambda_k) = 0$$

If we have $H = \hat{H}$, $K = \hat{K}$, we get

$$\hat{x}_k = -\sum_{h \in \hat{H}} a_{h0} \frac{\partial (\det a_{\hat{H}\hat{K}})}{\partial a_{hk}} \Big/ \det a_{\hat{H}\hat{K}}, \quad \hat{x}_h = 0,\ h \in \hat{H},\ \hat{x}_j = 0,\ j \in J - \hat{K}$$

$$\hat{\lambda}_h = -\sum_{k \in \hat{K}} a_{0k} \frac{\partial (\det a_{\hat{H}\hat{K}})}{\partial a_{hk}} \Big/ \det a_{\hat{H}\hat{K}}, \quad \hat{\lambda}_k = 0,\ k \in \hat{K},\ \hat{\lambda}_i = 0,\ i \in I - \hat{H},$$

while $$\hat{x}_i = \det \begin{Vmatrix} a_{i0} & a_{i\hat{K}} \\ a_{\hat{H}0} & a_{\hat{H}\hat{K}} \end{Vmatrix} \Big/ \det a_{\hat{H}\hat{K}} \quad \text{for } i \in I,$$

$$\hat{x}_0 = \sum_{j \in K} a_{0j} \hat{x}_j = \det \begin{Vmatrix} 0 & a_{0\hat{K}} \\ a_{\hat{H}0} & a_{\hat{H}\hat{K}} \end{Vmatrix} \Big/ \det a_{\hat{H}\hat{K}} = \hat{\lambda}_0 = \sum_{i \in H} \hat{\lambda}_i a_{i0},$$

$$\hat{\lambda}_j = -\det \begin{Vmatrix} a_{0j} & a_{0\hat{K}} \\ a_{\hat{H}j} & a_{\hat{H}\hat{K}} \end{Vmatrix} \Big/ \det a_{\hat{H}\hat{K}} \quad \text{for } j \in J.$$

The general solution to the linear programming problem and its dual problem stated can in principle be determined as follows:

$$\hat{x}_0 = x_0(\hat{H}, \hat{K}) = \sum_{j \in K} a_{0j} \hat{x}_j = \sum_{i \in H} \hat{\lambda}_i a_{i0} = \max_K \min_H x_0(H, K)$$

$$= \min_H \max_K x_0(H, K) = \min_H x_0(H, \hat{K}_H) = \max_K x_0(\hat{H}_K, K),$$

where $$x_0(H, K) = \det \begin{Vmatrix} 0 & a_{0K} \\ a_{H0} & a_{HK} \end{Vmatrix} \Big/ \det a_{HK},$$

$$\hat{x}_j = x_j(\hat{H}, \hat{K}) = \frac{\partial x_0(\hat{H}, \hat{K})}{\partial a_{0j}} \geqslant 0,$$

$$\hat{\lambda}_i = \lambda_i(\hat{H}, \hat{K}) = \frac{\partial x_0(\hat{H}, \hat{K})}{\partial a_{i0}} \geqslant 0.$$

The sets H can be restricted to those nonempty sets $H \in H_k$ for which

$$\frac{\partial x_0(H, K)}{\partial a_{h0}} > 0 \quad \text{for every } h \in H,$$

and K to the nonempty sets $K \in K_H$ for which

$$\frac{\partial x_0(H, K)}{\partial a_{0k}} > 0 \quad \text{for every } k \in K.$$

Often the sets H_K and K_H are empty sets. Then $H \neq \hat{H}$ and $K \neq \hat{K}$. If many optimal sets $\hat{K} = \hat{K}_{\hat{H}}$, $\hat{H} = \hat{H}_{\hat{K}}$ exist, then every weighted mean with non-negative weights $w(\hat{H}, \hat{K})$ of the solutions $x_j(\hat{H}, \hat{K})$ corresponding to these sets will also be an optimal solution. If many $\hat{H}$ exist, then every such mean value of the different $\lambda_i(\hat{H}, \hat{K})$

will be an optimal solution to the dual problem. It seems to be a good principle to start by assuming that both $\hat{H}$ and $\hat{K}$ have no, one, two, three, etc., elements, and then to investigate if it is possible to find nonempty sets $K = \hat{K}_{\hat{H}}$ and $H = \hat{H}_{\hat{K}}$. If we find all such sets the problem is solved also in practice.

By means of the multiplier theorem of Appendix 2 we can state the following criterion.

CRITERION 5.—We have $a_{0j}\hat{x}_j = 0$ for every $j \in J$, $\hat{\lambda}_i a_{i0} = 0$ for every $i \in I$ if, and only if, there exist such $\hat{x}_j \geqslant 0$ for $a_{0j} = 0$ and such $\hat{\lambda}_i \geqslant 0$ for $a_{i0} = 0$ that

$$\hat{x}_i = a_{i0} + \sum_{j;\, a_{0j}=0} a_{ij}\hat{x}_j \geqslant 0 \quad \text{for every } i \in I,$$

$$-\hat{\lambda}_j = a_{0j} + \sum_{i;\, a_{i0}=0} \hat{\lambda}_i a_{ij} \leqslant 0 \quad \text{for every } j \in J.$$

This criterion can be used as a test for checking if we have succeeded in transforming a linear programming problem into a form where the solution can be seen from the first row and column of the matrix of the new problem obtained.

The following criteria are obviously true.

CRITERION 6.—If there exist such $\lambda_i \geqslant 0$ that $a_{\lambda 0} < 0$, $a_{\lambda i} \leqslant 0$ for every $j \in J$, then X_A is an empty set ($a_{\hat{\lambda}0} = -\infty$).

CRITERION 7.—If there exist such $x_j \geqslant 0$ that $a_{0x} > 0$, $a_{ix} \geqslant 0$ for every $i \in I$, then the optimal value $a_{0\hat{x}} = \infty$.

CRITERION 8.—If $a_{\lambda 0} = 0$, $a_{\lambda j} \leqslant 0$ for every $j \in J$, then $a_{\lambda j}\hat{x}_j = 0$ for every $j \in J$.

CRITERION 9.—If $a_{0x} = 0$, $a_{ix} \geqslant 0$ for every $i \in I$, then $a_{ix}\hat{\lambda}_i = 0$ for every $i \in I$.

CRITERION 10.—If $a_{i0} \leqslant 0$ for every $i \in I$ and $a_{0j} \geqslant 0$ for every $j \in J$, then $a_{\hat{\lambda}0} = a_{0\hat{x}} = 0$ or no finite optimal solution exists.

In some special cases the explicit solution to a linear programming problem can be found very simply.

Example 1

If $x \in X_A(\lambda)$ when $x_\lambda = a_{\lambda 0} + \sum_{j \in J} a_{\lambda j} x_j \geqslant 0$, $x_j \geqslant 0$ for $j \in J$, $a_{\lambda j} < 0$ if $a_{0j} > 0$, and $a_{\lambda j} \leqslant 0$ if $a_{0j} = 0$, then

$$\hat{x}_0(\lambda) = \max_{x \in X_A(\lambda)} \sum_{j \in J} a_{0j} x_j = -\frac{a_{\lambda 0}\, a_{0j(\lambda)}}{a_{\lambda j(\lambda)}} = a_{0j(\lambda)}\, \hat{x}_{j(x)},$$

where the column $j = j(\lambda)$ is determined by

$$\frac{a_{0j(\lambda)}}{|a_{\lambda j(\lambda)}|} = \max_{j \in J} \frac{a_{0j}}{|a_{\lambda j}|}; \quad a_{\lambda j}\, a_{\lambda 0} < 0.$$

The optimal solution will be

$$\hat{x}_j(\lambda) = \frac{-a_{\lambda 0}}{a_{\lambda j(\lambda)}} \quad \text{for } j = j(\lambda),$$

$$\hat{x}_j(\lambda) = 0 \qquad \text{for } j \neq j(\lambda).$$

The optimal value $\hat{x}_0(\lambda)$ will have the property

$$\hat{x}(\lambda) \geqslant \hat{x}(\hat{\lambda}) = \hat{x}_0 \quad \text{for } \lambda_j \geqslant 0.$$

Example 2

If $h = \hat{H}$, $k = \hat{K}$, that is if the set $\hat{H}$ contains only the element $h \in I$ and $\hat{K}$ the element $k \in J$, then

$$\hat{x}_0 = \max_j \min_i \frac{-a_{i0}a_{0j}}{a_{ij}} = \min_i \max_j \frac{-a_{i0}a_{0j}}{a_{ij}} = \frac{-a_{h0}a_{0k}}{a_{hk}},$$

where (i, j) can be restricted to the set W for which $a_{i0}a_{ij} < 0$, $a_{0j}a_{ij} < 0$. The optimal solution will be

$$\hat{x}_k = \frac{-a_{h0}}{a_{hk}}, \quad \hat{x}_j = 0,\ j \neq k,\ j \in J,$$

$$\hat{\lambda}_h = \frac{-a_{0k}}{a_{hk}}, \quad \hat{\lambda}_i = 0,\ i \neq h,\ i \in I.$$

This example is the simplest special case of the general solution given earlier.

The necessary and sufficient conditions for the special case considered in this example are:

$h = \hat{H}$, $k = \hat{K}$ if and only if

$$\frac{a_{h0}}{|a_{hk}|} = \min_i \frac{a_{i0}}{|a_{ik}|}; \quad a_{0k}a_{ik} < 0,\ a_{i0}a_{ik} \leqslant 0,\ \frac{-a_{i0}}{a_{ik}} \leqslant \frac{-a_{h''0}}{a_{h''k}},$$

$$\frac{a_{0k}}{|a_{hk}|} = \max_j \frac{a_{0j}}{|a_{hj}|}; \quad a_{0j}a_{hj} \leqslant 0,\ a_{h0}a_{hj} < 0,\ \frac{-a_{0j}}{a_{hj}} \leqslant \frac{-a_{0k''}}{a_{hk''}},$$

where

$$\frac{-a_{h''0}}{a_{h''k}} = \min_{i:\, a_{ik}<0} \frac{-a_{i0}}{a_{ik}},$$

$$\frac{a_{0k''}}{a_{hk''}} = \min_{j:\, a_{hj}>0} \frac{-a_{0j}}{a_{hj}}.$$

Further the following conditions must be fulfilled

$$\frac{-a_{h'0}}{a_{h'k}} \leqslant \frac{-a_{h0}}{a_{hk}}, \qquad \frac{-a_{0k'}}{a_{hk'}} \leqslant \frac{-a_{0k}}{a_{hk}},$$

where

$$\frac{-a_{h'0}}{a_{h'k}} = \max_{\substack{i;\, a_{i0}<0 \\ a_{ik}>0}} \frac{-a_{i0}}{a_{ik}},$$

$$\frac{-a_{0k'}}{a_{hk'}} = \max_{\substack{j;\, a_{0j}>0 \\ a_{hj}<0}} \frac{-a_{0j}}{a_{hj}}.$$

If no point (h, k) exists fulfilling all these conditions, $(h, k) = (\hat{H}, \hat{K})$.

Example 3

If $a_{ij} = 0$ except for $j = j(i)$, and $a_{ij(i)} > 0$ if $a_{i0} < 0$, while $a_{ij} < 0$ for some $i \in I$ if $a_{0j} \geqslant 0$, then

$$\hat{x}_j = \max\left(0, \frac{-a_{i(j)0}}{a_{i(j)j}}\right),$$

where

$$\frac{a_{i(j)0}}{|a_{i(j)j}|} = \min_{i \in I_j} \frac{a_{i0}}{|a_{ij}|}$$

and $i \in I_j$ if $(a_{0j}a_{ij} < 0,\ a_{ij} > 0$ if $a_{i0} < 0)$.

If $a_{0j} = 0$ the whole interval

$$v_j = \min_{i;\, a_{ij}<0} \frac{a_{i0}}{-a_{ij}} \geqslant \hat{x}_j \geqslant \max_{i;\, a_{ij}>0} \left(\frac{-a_{i0}}{a_{ij}}, 0\right) = u_j \geqslant 0$$

is optimal. If $v_j < u_j$ for some j, the set X_A is empty, and no optimal solution exists.

The corresponding dual problem has some features in common with the so-called transportation problem.

The transportation problem is a special case of the general linear programming problem, which can be simply solved by means of methods especially developed to take account of the form in which it is mostly presented (Appendix I).

These examples, and my experience gathered during trials to solve linear programming problems numerically, suggest that we will usually be right if we guess that $i \in \hat{H}$ if $i \in H$ and $j \in \hat{K}$ if $j \in K$, when the surrogate sets H and K are defined by means of the rules to be described.

Rules for finding surrogate sets (H, K) *for* $(\hat{H}, \hat{K})$

As for every $h \in \hat{H}$ we have

$$\hat{x}_h = a_{h0} + \sum_{k \in \hat{K}} a_{hk}\hat{x}_k = 0,\ \hat{x}_{\hat{K}} > 0,$$

and for every $k \in \hat{K}$ we have

$$-\hat{\lambda}_k = a_{0k} + \sum_{h \in \hat{H}} \hat{\lambda}_h a_{hk} = 0,\ \hat{\lambda}_{\hat{H}} > 0,$$

and for every $i \in I$ we have

$$\hat{x}_i = a_{i0} + \sum_{k \in \hat{K}} a_{ik} \hat{x}_k \geqslant 0,$$

and for every $j \in J$ we have

$$-\hat{\lambda}_j = a_{0j} + \sum_{h \in \hat{H}} \hat{\lambda}_h a_{hj} \leqslant 0,$$

we can state that the sets $\hat{H}$ and $\hat{K}$ have the following easily controlled properties

(1) If $a_{0j} > 0$, there exists at least one $h \in \hat{H}$ with the property $a_{hj} < 0$.

(2) If $a_{i0} < 0$, there exists at least one $k \in \hat{K}$ with the property $a_{ik} > 0$.

(3) If $a_{0k} = 0$ and $k \in \hat{K}$, then there exists at least one pair of numbers (h, h') in the set $\hat{H}$ with the property $a_{hk} a_{h'k} < 0$.

(4) If $a_{h0} = 0$ and $h \in \hat{H}$, then there exists at least one pair of numbers (k, k') in the set $\hat{K}$ with the property $a_{hk} a_{hk'} < 0$. From the properties (1), (2), (3) and (4) it follows that the matrix

$$\begin{Bmatrix} 0 & a_{0\hat{K}} \\ a_{\hat{H}0} & a_{\hat{H}\hat{K}} \end{Bmatrix}$$

is such that in every row $h \in \hat{H}$ there exists a pair of numbers (k, k') with the property $a_{hk} a_{hk'} < 0$, and in every column $k \in \hat{K}$ there exists a pair of numbers (h, h') with the property $a_{hk} a_{h'k} < 0$. A good surrogate set (H, K) for $(\hat{H}, \hat{K})$ must have this property in common with $(\hat{H}, \hat{K})$.

The contest rule for finding (H, K)

(1) *First contest*

We say that $h \in H'$ if $h \in H'_k$ and $k \in K'$ where

$$h \in H'_k \quad \text{if } a_{0k} < 0,\ a_{hk} > 0,\ a_{h0} < 0 \text{ or if } a_{hk} < 0,\ a_{0k} \geqslant 0.$$

We say that $k \in K'$ if $k \in K'_h$ and $h \in H'$ where

$$k \in K'_h \quad \text{if } a_{h0} > 0,\ a_{hk} < 0,\ a_{0k} > 0 \text{ or if } a_{hk} > 0,\ a_{h0} \leqslant 0.$$

If to some h, $a_{h0} < 0$ correspond only one $a_{hk} > 0$, $k \in K'$ we say that $k \in K$. If to some k, $a_{0k} > 0$ correspond only one h, $a_{hk} < 0$, $h \in H'$ we say that $h \in H$. All other h and k will have competing elements h' and k' in H' and K' respectively. Only the elements in H' and K' take part in the second contest.

(2) *Second contest*

We say that

I. $h \in H$ if $h \in H_k$ and $k \in K$,

II. $k \in K$ if $k \in K_h$ and $h \in H$,

where $h \in H_k$ if

$$\frac{a_{h0}}{|a_{hk}|} = \min_i \frac{a_{i0}}{|a_{ik}|} \quad \text{for } \left(\left|\frac{a_{i0}}{a_{ik}}\right| \leqq \min_{\substack{i;a_{ik}<0 \\ a_{i0}>0}} \left|\frac{a_{i0}}{a_{ik}}\right|\right), \quad i \in H'_k$$

and $k \in K_h$ if

$$\frac{a_{0k}}{|a_{hk}|} = \max_j \frac{a_{0j}}{|a_{hj}|} \quad \text{for } \left(\left|\frac{a_{0j}}{a_{hj}}\right| \leqq \min_{\substack{j;a_{hj}>0 \\ a_{0j}<0}} \left|\frac{a_{0j}}{a_{hj}}\right|\right), \quad j \in K'_h.$$

If $a_{0j} \geqslant 0$, $a_{h'j} > 0$ for some $j \in K$, and if it is not quite clear that $a'_{0j} < 0$ in the table arrived at when x_K and x_H have exchanged their roles, we disqualify h' and say that $h' \notin H$. If $a_{i0} \leqslant 0$, $a_{ik'} < 0$ for some $i \in H$, and it is not quite clear that $a'_{i0} > 0$, we disqualify k' and say that $k' \notin K$.

If we like to get such surrogate sets (H, K) that it is not possible to see that $H \neq \hat{H}$ or $K \neq \hat{K}$ without actually calculating a'_{i0} und a'_{0j}, we can leave out of the contest the elements already found to belong to H or K and make a new contest between the elements in $H' - H$, $K' - K$, and accept new candidates until it seems possible that $H = \hat{H}$, $K = \hat{K}$.

The practical problem to find the sets (H, K) is, I think, most easily solved by first tentatively assuming that $h \in H'$ if $a_{h0} < 0$ and $k \in K'$ if $a_{0k} > 0$. Thereafter we list the most promising candidates for K respectively H. We then exclude the "weakest" candidates violating some of the beating orders by looking alternately on I and II until we get such a set (H, K) that we cannot easily see that $H \neq \hat{H}$ or $K \neq \hat{K}$.

If only a smaller number of pairs (h, k) are actually used for the exchange of the roles of x_h and x_k, we say that a pair (h, k) is better than (h', k'), if the number of competing elements in the contest (I, II) for h and k is smaller than for (h', k').

By a simple exchange the numbers a_{ij} change according to the "simplex" formulas:

$$a'_{ij} = a_{ij} - \frac{a_{ik}\, a_{hj}}{a_{hk}} \quad \text{for } i \neq h,\ j \neq k,$$

$$a'_{kj} = -\frac{a_{hj}}{a_{hk}}; \qquad a_{kh} = \frac{1}{a_{hk}}; \qquad a_{ih} = \frac{a_{ik}}{a_{hk}}.$$

Sometimes it is easy to see the signs of the numbers in the row for which $i = 0$ or $j = 0$ also after two or more pairs (x_h, x_k) have changed their roles. In this case the "best multimix" step can easily be recognized.

By trial choices of (h, k) only the new $\{a'_{i0}, a'_{0j}\}$ and a'_{ij} for $i \in H$, $j \in K$ have to be roughly estimated to make an appraisal of the goodness of competing choices of (h, k) possible.

A method for checking the calculation and for more reliable appraising of the goodness of choices made at each step is to use two different routes to the goal.

The numbers

$$x_k^{(1)} = \left|\frac{a_{h0}}{a_{hk}}\right|,\ k \in K_h \quad \text{and} \quad \lambda_h^{(1)} = \left|\frac{a_{0k}}{a_{hk}}\right|,\ h \in H_k$$

are often pretty good estimates of $\hat{x}_k$ and $\hat{\lambda}_h$. Still better estimates can be expected by means of the formulas:

$$x_k^{(2)} = \max\left(0,\ -\frac{a_{h0}^{(1)}}{a_{hk}}\right), \qquad \lambda_h^{(2)} = \max\left(0,\ -\frac{a_{0k}^{(1)}}{a_{hk}}\right),$$

where

$$a_{h0}^{(1)} = a_{h0} + \sum_{j \in K-k} a_{hj} x_j^{(1)};\quad h \in H_k,$$

$$a_{0k}^{(1)} = a_{0k} + \sum_{i \in H-h} \lambda_i^{(1)} a_{ik};\quad k \in K_h.$$

These estimates are, however, inferior to the estimates

$$x_k' = \max\,(0,\ a_{k0}'),\ \lambda_h' = \max\,(0,\ -a_{0h}')$$

arrived at if the interchange between x_K and x_H is made. If no wrong signs occur in the table $\{a_{ij}'\}$, these estimates are in fact equal to the optimal values sought.

If we interchange the roles of all x_k; $k \in K$ and all x_h; $h \in H$ found by means of the contest rule, we will often reach our goal in one or two multimix steps. This will be the case, when the number of elements in $(\hat{H}, \hat{K})$ is reasonably small and the problem is not such that the surrogate set (H, K) becomes very difficult to determine without risky choices between many competing alternatives. In such cases some preliminary simplex steps might be needed before a larger multimix step can safely be taken without resulting in work that perhaps could have been avoided by a more cautious method with appraisal of the right route to be chosen after each single step. Sometimes it might be profitable to make risky multimix steps as well because we are saved the work of deciding the right choice to be made after each small step made. Often the problem will be simpler to attack after such a risky multimix step has been taken, even if not all $h \in H$ and $k \in K$ used are such that $h \in \hat{H}$ and $k \in \hat{K}$.

In cases when many risky choices have to be made, it might be useful to make the first contest for finding H' and K' harder by requiring already at this stage that only elements $h \in H'$, $k \in K'$ for which

$$\left|\frac{a_{h0}}{a_{hk}}\right| \leqslant \min_{\substack{i;\ a_{ik}<0 \\ a_{i0}>0}} \left|\frac{a_{i0}}{a_{ik}}\right| \quad \text{and} \quad \left|\frac{a_{0k}}{a_{hk}}\right| \leqslant \min_{\substack{j;\ a_{hj}>0 \\ a_{0j}<0}} \left|\frac{a_{0j}}{a_{hj}}\right|$$

can take part in the second contest, and use only such (h, k) which are left after the second contest by the exchange of the roles of the variables x_k; $k \in K$ and x_h; $h \in H$. If we know that $h \in \hat{H}$, $k \in \hat{K}$, we first exchange x_k and x_h before any risky choices are made.

The ultimate goal is to get a table with no "wrong" signs ($a_{0j}' > 0$ or $a_{i0}' < 0$).

A rule for guessing if $h \in \hat{H}$ or not and if $k \in \hat{K}$ or not, which might be more practical than the rule defining (H, K), is the *multimix rule*:

We guess that $h \in \hat{H}$ if $h \in \overline{H}$ and $k \in \hat{K}$ if $k \in \overline{K}$ where $h \in \overline{H}$ if $k \in \overline{K}$ and $a_{0k} > 0$, $a_{h0} \leqslant 0$, $a_{hk} < 0$, or if $k \in \overline{K}$ and

$$\frac{a_{hk}}{|a_{h0}|} = \min_i \frac{a_{ik}}{|a_{i0}|} \quad \text{for } a_{i0} \text{ or } -a_{0k} < 0,\ a_{ik} \neq 0,$$

$k \in \overline{K}$ if $h \in \overline{H}$ and $a_{h0} < 0$, $a_{hk} > 0$, $a_{0k} \geqslant 0$, or if $h \in \overline{H}$ and

$$\frac{a_{hk}}{|a_{0k}|} = \max_j \frac{a_{hj}}{|a_{0j}|} \quad \text{for } a_{h0} \text{ or } -a_{0j} < 0,\ a_{hj} \neq 0.$$

The first-mentioned rule is, however, I think, better than the multimix rule in the cases when $(\overline{H}, \overline{K}) \neq (H, K)$. If both rules are used the common elements are especially well merited for being elements of $(\hat{H}, \hat{K})$.

We say that $i \in I - H$, if

$$a_{i0} \geqslant 0, \quad a_{ij} \geqslant 0 \quad \text{or} \quad a_{0j} a_{ij} > 0 \quad \text{for every } j \in \ddot{K},$$

and $j \in J - \ddot{K}$, if

$$a_{0j} \leqslant 0, \quad a_{ij} \leqslant 0 \quad \text{or} \quad a_{i0} a_{ij} > 0 \quad \text{for every } i \in \ddot{H}.$$

We call $(I - \ddot{H}, J - \ddot{K})$ *the set of presumptively superfluous conditions and variables.* We guess that $\hat{x}_i > 0$ if $i \in I - \ddot{H}$ and $\hat{x}_j = 0$ if $j \in J - \ddot{K}$.

It can often be quite clear if $h \in \hat{H}$ or not, but there may be many alternatives for the choice of elements in $\hat{K}$. This is for instance the case in the special linear programming problem called the transportation problem considered in Appendix I. In such cases it seems to be a good idea to use the following rule.

The T-rule

We first make $a_{0j} \leqslant 0$ for every $j \in J$. If $a_{h0} < 0$ we choose $k = k(h)$ so that it is possible to find non-negative numbers x_j; $j \in J_{hk}$, where $j \in J_{hk}$ if $-a_{0j}/a_{hj} \leqslant -a_{0k}/a_{hk}$, and J_{hk} contains at most one not really needed element, making $x_i = a_{i0} + \sum_{j \in J_{hk}} a_{ij} x_j > 0$ for $i = h$ and all i for which $a_{i0} \geqslant 0$. After all x_h and $x_{k(h)}$, $a_{h0} < 0$ have exchanged their roles, we eliminate in a suitable way all positive a'_{0j} arrived at, using as mix-rows (h') only such rows for which $a'_{h0} \geqslant 0$, $a'_{hj} < 0$, $h' \in I$. When this is done we eliminate alternately the remaining negative a'_{h0} and the "wrong" signed $a'_{0j} > 0$. When the last negative a'_{h0} and positive a'_{0k} have disappeared, we have reached our goal. If difficulties occur, use the contest rule.

A numerical example

TABLE 1. a_{ij}.

$i \backslash j$	0	1	2	3	4	5	6	
0	0	10	11	—8	7	12	—30	max!
7	− 70	1	10	100	− 20	− 1	10	
8	110	− 10	−100	0	0	10	− 1	
9	−280	100	0	− 10	100	0	100	
10	100	0	0	20	− 1	−100	1	
11	90	−100	20	0	0	10	− 10	
12	− 10	1	2	0	0	− 1	10	
13	− 30	− 50	50	− 1	− 20	50	1	
14	0	0	0	10	0	− 10	0	

TABLE 2. a'_{ij}.

$i \backslash j$	0	1	8	3	4	5	6
0	12.1	8.9	−0.11	− 8	7	13.1	− 30.11
7	− 59	0	−0.1	100	− 20	0	9.9
2	0.1	− 0.1	−0.01	0	0	0.1	− 0.01
9	−280	100	0	− 10	100	0	100
10	100	0	0	20	− 1	−100	1
11	112	−102	−0.2	0	0	12	− 10.2
12	− 7.8	0.8	−0.02	0	0	− 0.8	9.98
13	25	− 55	−0.5	− 1	− 20	55	0.5
14	0	0	0	10	0	− 10	0

TABLE 3. $\hat{a}_{ij}$.

$i \backslash j$	0	1	8	7	13	10	12
0	22.9072	− 5.7093	−0.3259	−0.0277	−0.3085	−0.2771	− 2.9464
3	1.4294	− 0.5844	−0.0045	0.0111	−0.0108	−0.0058	− 0.0098
2	1.2160	− 0.1079	−0.0101	0.0002	−0.0002	−0.0011	− 0.0011
9	257.2269	−109.4023	−2.4006	0.4458	−5.2961	−2.9940	10.1431
5	1.2484	− 0.0881	−0.0006	0.0022	−0.0016	−0.0109	− 0.0010
11	117.9878	−102.1676	−0.2274	0.0242	−0.0181	−0.1217	− 1.0329
6	0.8816	− 0.0872	0.0020	0.0002	−0.0001	−0.0009	0.1001
4	4.6336	− 2.9652	−0.0264	0.0054	−0.0539	−0.0297	0.0003
14	1.8107	− 4.9631	−0.0386	0.0890	−0.0915	0.0504	− 0.0888

We exchange the roles of x_8 and x_2 because our criteria tell us that $\hat{x}_8$ is surely $= 0$ and because an increase in x_2 will make the numbers a_{i0} larger in Table 2 than in Table 1.

By using the contest rule for finding (H, K) we see that the exchanges of the roles of x_7 and x_3, x_{13} and x_4, x_{10} and x_5, x_{12} and x_6 can perhaps lead to the goal of eliminating all wrong signs. When the exchanges are made, we get Table 3.

In Table 3 all $\hat{a}_{0j} < 0$ and all $\hat{a}_{i0} > 0$. The optimal solution is therefore $\hat{x}_j = 0$ for $\hat{a}_{0j} < 0$, while $\hat{x}_i = \hat{a}_{i0} > 0$. The solution to the dual problem is seen to be $\hat{\lambda}_j = -\hat{a}_{0j}$ while all $\hat{\lambda}_i = 0$ for $\hat{a}_{i0} > 0$.

APPENDIX I

The Transportation Problem

The following conditional optimalization problem is called the transportation problem:

PROBLEM T.—Find the quantities $\hat{x}_j = \hat{x}_{hk} : h = 1, \ldots, n,\ k = 1, \ldots, m,\ j = (h, k)$ that have to be transported from the place h to the place k, when the place k needs in all $x_{.k}$ units and the place h is able to deliver in all $x_{h.}$ units, and the total costs of the transportation

$$-y_0(x) = \sum_{h=1}^{n} \sum_{k=1}^{m} a_{hk} x_{hk} = \sum_{j \in J} a_j x_j; \quad J = \{(h, k)\}$$

have to be a minimum.

The admissible region A for the point $x = \{x_{hk}\}$ will then be defined by

$$y_{hk} = x_{hk} \geqslant 0, \qquad y_{h.} = -\sum_{k} x_{hk} + x_{h.} \geqslant 0, \qquad y_{.k} = \sum_{h} x_{hk} - x_{.k} \geqslant 0.$$

If we use the multiplier technique for finding the optimal transportation program, we have to solve the following system of equations (see Appendix II):

$$\frac{\partial y_0}{\partial x_j} + \sum_{h} \lambda_{h.} \frac{\partial y_{h.}}{\partial x_j} + \sum_{k} \lambda_{.k} \frac{\partial y_{.k}}{\partial x_j} + \sum_{h,k} \lambda_{hk} \frac{\partial y_{hk}}{\partial x_j} = 0; \quad j = (h', k') \in J.$$

$$\lambda_{hk} y_{hk} = 0, \qquad \lambda_{h.} y_{h.} = 0, \qquad \lambda_{.k} y_{.k} = 0.$$

We have for the above-defined functions:

$$\frac{\partial y_0}{\partial x_j} = -a_{h'k'}, \qquad \frac{\partial y_{h'.}}{\partial x_j} = -1, \qquad \frac{\partial y_{.k'}}{\partial x_j} = 1, \qquad \frac{\partial y_{h'k'}}{\partial x_j} = 1,$$

$$\frac{\partial y_{h.}}{\partial x_j} = 0;\ h \neq h', \quad \text{and} \quad \frac{\partial y_{.k}}{\partial x_j} = 0;\ k \neq k'.$$

The system of equations containing the partial derivatives can thus be written in the form:

$$-a_{hk} - \lambda_{h.} + \lambda_{.k} + \lambda_{hk} = 0; \quad (h, k) \in J,$$

where all the multipliers $\lambda'_{h.}$, $\lambda_{.k}$, λ_{hk} are non-negative.

From these results we get for $x = \hat{x}$

$$\lambda_{hk} = a_{hk} + \lambda_{h.} - \lambda_{.k} \geqslant 0, \qquad \lambda_{hk}\hat{x}_{hk} = 0, \qquad \lambda_{h.}\hat{y}_{h.} = 0, \qquad \lambda_{.k}\hat{y}_{.k} = 0.$$

Substituting $a_{hk} = \lambda_{.k} - \lambda_{h.} + \lambda_{hk}$ in $-y_0(x)$ the transportation costs can be written in the form

$$-y_0(x) = \sum_h \sum_k a_{hk} x_{hk} = \sum_k \lambda_{.k} x_{.k} - \sum_h \lambda_{h.} x_{h.} + v(\lambda, x) = c(\lambda) + v(\lambda, x),$$

where

$$v(\lambda, x) = \sum_k \lambda_{.k} y_{.k} + \sum_h \lambda_{h.} y_{h.} + \sum_h \sum_k \lambda_{hk} x_{hk}.$$

Because $x_{.k}$ and $x_{h.}$ are given constants, the optimal values x_{hk} will minimize $v(\lambda, x)$ independently of the values $\lambda_{.k}$ and $\lambda_{h.}$. When $\lambda_{h.}$, $\lambda_{.k}$, λ_{hk} and x_{hk} are optimal, we have

$$v(\hat{\lambda}, \hat{x}) = 0, \qquad -y_0(\hat{x}) = \sum_k \hat{\lambda}_{.k} x_{.k} - \sum_h \hat{\lambda}_{h.} x_{h.} = c(\hat{\lambda}) = \max_{\lambda \in L} c(\lambda),$$

where $\lambda \in L$ if $\lambda_{h.} \geqslant 0$, $\lambda_{.k} \geqslant 0$, $\lambda_{hk} \geqslant 0$.

Because $\hat{\lambda}_{hk}$ must be zero for so many different h that the required quantity $x_{.k}$ can be delivered from places h for which $\hat{\lambda}_{hk} = 0$, we can expect that $\hat{\lambda}_{hk} = 0$ for those $h \in H_k$ for which λ_{hk} are relatively smallest and $\sum_{h \in H_k} x_{h.} \geqslant x_{.k}$. For $\hat{\lambda}_{h.} > 0$ we must have $\hat{\lambda}_{hk} = 0$ for a set $k \in K_h$ such that $\sum_{k \in K_h} x_{.k} \geqslant x_{h.}$.

The first practical aim by choosing $\lambda_{h.}$ and $\lambda_{.k}$ is to get $\lambda_{hk} = 0$ for such places (h, k) for which λ_{hk} is comparatively small for given h respectively k for so many places that these requirements will be fulfilled. The following algorithm, called T-algorithm, will lead to this result.

Starting from $\lambda^{(0)}_{hk} = a_{hk}$, $\lambda^{(0)}_{h.} = \lambda^{(0)}_{.k} = 0$ we determine

$$\lambda^{(\nu+1)}_{hk} = \lambda^{(\nu)}_{hk} + \Delta\lambda^{(\nu)}_{h.} - \Delta\lambda^{(\nu)}_{.k},$$

where

(1) $$\Delta\lambda^{(\nu)}_{.k} = \min_h (\lambda^{(\nu)}_{hk} + \Delta\lambda^{(\nu)}_{h.}); \qquad x^{(\nu)}_{.k} > x_{.k}, \quad x^{(\nu)}_{.k} = \text{sum of } x_{h.}$$

for those h for which $$\lambda^{(\nu)}_{hk} + \Delta\lambda^{(\nu)}_{h.} \leqslant \Delta\lambda^{(\nu)}_{.k}$$

and

(2) $$\Delta\lambda^{(\nu)}_{h.} = \max_k (\Delta\lambda^{(\nu)}_{.k} - \lambda^{(\nu)}_{hk}).$$

These adjustments can be found in practice by first tentatively assuming that $\Delta\lambda_{h.}^{(\nu)}=0$ in (1) and correcting this assumption after $\Delta\lambda_{h.}^{(\nu)}$ has been determined by (2) and so on until such $\Delta\lambda_{h.}^{(\nu)}$ and $\Delta\lambda_{.k}^{(\nu)}$ are found which satisfy both (1) and (2).

When the T-algorithm has come to an end for $\nu=\nu'$ because all $\Delta\lambda_{.k}^{(\nu')}=0$, we will often have $\lambda_{hk}^{(\nu')}=\hat{\lambda}_{hk}=0$ and the problem can be completely solved. Sometimes we have to continue the calculations by making new moves of the same kind but now exceptionally requiring that

$$\Delta\lambda_{.k}^{(\nu')}=\min_{h}\lambda_{hk}^{(\nu')}\neq 0$$

for some k such that when the T-algorithm is continued after this exceptional move, we will get a set of multipliers $\lambda^{(\nu'')}$ such that $c(\lambda^{(\nu'')})>c(\lambda^{(\nu')})$. As long as it is possible to increase $c(\lambda)$ without making λ_{hk} negative we have not reached to such a set λ_{hk} that $\lambda_{hk}=\hat{\lambda}_{hk}=0$ for these places (h, k) for which $\hat{x}_{hk}\neq 0$. By the trials to find out if it is possible to solve the problem by assigning positive values x_{hk} to places where $\lambda_{hk}=0$, we can, in cases when we can not fulfil all the requirements for an optimal solution $\hat{x}$, usually see for which k new places with $\lambda_{hk}=0$ are needed before a solution can be found.

Many problems which do not seem to be transportation problems can be such that the solution to the problem in question is the same as that to a certain transportation problem. Such a problem is for instance Problem P: Find the permutation p which is such that $a_p=\sum_{h=1}^{n}a_{h\,p(h)}$ is a minimum over all possible permutations of $(1, \ldots, n)$ when a_{hk} are given numbers. This problem corresponds to the transportation problem with $x_{h.}=x_{.k}=1$ for $h=1, \ldots, n$ and $k=1, \ldots, n$ and the unit costs a_{hk}. This problem is of practical interest for instance when we have to assign some of n jobs to n persons in a firm when it is estimated that the person h if he gets the job k will benefit the firm by an amount measured by $-a_{hk}$.

The following examples show that the T-algorithm rapidly leads to a solution of transportation problems.

Example 1.—Dantzig's example (Koopmans: *Activity Analysis of Production and Allocation*, Wiley, New York 1951, p. 367)

In Table 1 the unit costs a_{hk} and the requirements $x_{.k}$ and the available quantities $x_{h.}$ are seen. The first row containing only zeros is reserved for the multipliers $-\lambda_{.k}$ and the first column for the multipliers $\lambda_{h.}$. The row denoted by $-\Delta\lambda_{.k}$ respectively the column $\Delta\lambda_{h.}$ are reserved for the changes to be made in these quantities, and the last column and row for the partial differences $x_{h.}\Delta\lambda_{h.}$ and $-x_{.k}\Delta\lambda_{.k}$ respectively in $-c(\lambda)$ due to these adjustments. The number (13) in the corner gives the increase $\Delta c(\lambda)$ in $c(\lambda)$. In this example the sum $(x..)$ of the requirements $x_{.k}$ is equal to the sum of available quantities $x_{h.}$. In this case we put the rest $\sum_{h}x_{h.}-\sum_{k}x_{.k}=0$ as the quantity corresponding to the first row $h=(.)$ and the first column $k=(.)$.

TABLE 1.

$x_{h.}$ \ $x_{.k}$	(0)	(3)	(3)	(3)	(2)	(2)	$\Delta\lambda_{h.}$	$x_{h.}\,\Delta\lambda_{h.}$
(0)	0	0	0	0	0	0	0	0
(1)	0	3	2	1	2	3	2	2
(5)	0	5	4	3	−1	1	0	0
(7)	0	0	2	3	4	5	0	0
$\Delta\lambda_{.k}$	0	0	−2	−3	1	−1	.	2
$-x_{.k}\,\Delta\lambda_{.h}$	0	0	−6	−9	2	−2	−15	13

TABLE 2.

$x_{h.}$ \ $x_{.k}$	(0)	(3)	(3)	(3)	(2)	(2)
(0)	(0) 0	(0) 0	(0) −2	(0) −3	(0) 1	(0) −1
(1)	(0) 2	(0) 5	(0) 2	(1) 0	(0) 5	(0) 4
(5)	(0) 0	(0) 5	(0) 2	(1) 0	(2) 0	(2) 0
(7)	(0) 0	(3) 0	(3) 0	(1) 0	(0) 5	(0) 4

TABLE 3.

$x_{h.}$ \ $x_{.h}$	(1)	(4)	(8)	(3)	(6)	$\Delta\lambda_{h.}$	$x_{h.}\,\Delta\lambda_{h.}$
(0)	0	0	0	0	0	0	0
(9)	0	5	3	1	9	−4	−36
(4)	0	15	17	3	2	−2	− 8
(9)	0	3	2	1	8	−3	−27
$-\Delta\lambda_{.k}$	4	0	1	3	0	.	−71
$-x_{.k}\Delta\lambda_{.k}$	4	0	8	9	0	21	50

Table 2 is reached from Table 1 by adding to a_{hk} the adjustments $\Delta\lambda_{h.} - \Delta\lambda_{.k}$. The T-algorithm ends after this first step. The optimal program x is seen from the numbers put in paranthesis. The smallest transportation costs are

$$\sum_k x_{.k}\,\Delta\lambda_{.k} - \sum_h x_{h.}\,\Delta\lambda_{h.} = 13.$$

The optimal $\hat{\lambda}_{h.}$ would in this case be obtained if we add one unit to all $\lambda^{(1)}_{h.}$, and $-\hat{\lambda}_{.k}$ by subtracting one unit from $-\lambda^{(1)}_{.k}$; all other $\lambda_{hk} = \lambda^{(1)}_{hk}$ in Table 2.

Example 2.—Crane's example (obtained from Jesus de Lasala: "Aportación al estudio de un método rápido de cálculo para la aplicación del método del Simplex al problema del transporte de Hitchcock, *Trabajos de Estadística*, Vol. VI, Part III, 1955, p. 214): Tables 3–5.

TABLE 4.

$x_{h.}$ \ $x_{.k}$	(1)	(4)	(8)	(3)	(6)	$\Delta\lambda_{h.}$	$x_{h.}\,\Delta\lambda_{h.}$
(0)	4	0	1	3	0	0	0
(9)	0	1	0	0	5	0	0
(4)	2	13	16	4	0	5	20
(9)	1	0	0	1	5	0	0
$-\Delta\lambda_{.k}$	0	0	0	0	− 5	.	20
$-x_{.k}\,\Delta\lambda_{.k}$	0	0	0	0	−30	−30	10

TABLE 5.

$x_{h.}$ \ $x_{k.}$	(0)	(4)	(8)	(3)	(6)
(0)	(0) 4	(0) 0	(0) 1	(0) 3	(0) −5
(9)	(1) 0	(0) 1	(5 − x) 0	(3) 0	(x) 0
(4)	(0) 7	(0) 18	(0) 21	(0) 9	(4) 0
(9)	(0) 1	(4) 0	(3 + x) 0	(0) 0	(2 − x) 0

TABLE 6.

$x_{h.}$ \ $x_{.k}$	(0)	(35)	(60)	(43)	(58)	(42)	(60)	(36)	(20)	(18)	(37)	$\Delta\lambda_{h.}$	$x_{h.}\Delta\lambda_{h.}$
(0)	0	0	0	0	0	0	0	0	0	0	0	0	0
(80)	0	15	11	30	36	27	33	9	44	33	46	−22	−1760
(120)	0	22	10	15	20	33	32	8	41	16	32	−21	−2520
(79)	0	31	9	25	34	14	13	16	23	35	41	−15	−1185
(130)	0	50	27	11	17	50	23	34	17	42	6	−18	−2340
$-\Delta\lambda_{.k}$	0	7	11	7	1	1	2	13	1	5	12	.	−7805
$-x_{.k}\Delta\lambda_{.k}$	0	245	660	301	58	42	120	468	20	90	444	2448	5357

TABLE 7.

$x_{h.}$ \ $x_{.k}$	(0)	(35)	(60)	(43)	(58)	(42)	(60)	(36)	(20)	(18)	(37)	$\Delta\lambda_{h.}$	$x_{h.}\Delta\lambda_{h.}$
(0)	0	7	11	7	1	1	2	13	1	5	12	0	0
(80)	−22	0	0	15	15	6	13	0	23	11	36	0	0
(120)	−21	8	0	1	0	13	13	0	21	0	23	0	0
(79)	−15	23	5	17	20	0	0	14	9	25	38	6	474
(130)	−18	39	20	0	0	33	7	29	0	29	0	0	0
$-\Delta\lambda_{.k}$	0	0	0	0	0	− 6	− 6	0	0	0	0	.	474
$-x_{.k}\Delta\lambda_{.k}$	0	0	0	0	0	−252	−360	0	0	0	0	−612	138

TABLE 8.

$x_{h.}$ \ $x_{.k}$	(0)	(35)	(60)	(43)	(58)	(42)	(60)	(36)	(20)	(18)	(37)
(0)	(0) 0	(0) 7	(0) 11	(0) 7	(0) 1	(0) −5	(0) −4	(0) 13	(0) 1	(0) 5	(0) 12
(80)	(0) −22	(35) 0	(22 − x) 0	(0) 15	(0) 15	(23) 0	(0) 7	(x) 0	(0) 23	(0) 16	(0) 36
(120)	(0) −21	(0) 8	(38 + x) 0	(0) 1	(28) 0	(0) 7	(0) 7	(36 − x) 0	(0) 21	(18) 0	(0) 23
(79)	(0) − 9	(0) 29	(0) 11	(0) 23	(0) 26	(19) 0	(60) 0	(0) 20	(0) 15	(0) 31	(0) 44
(130)	(0) −18	(0) 39	(0) 20	(43) 0	(30) 0	(0) 27	(0) 1	(0) 29	(20) 0	(0) 29	(37) 0

The number (1) over the first column denotes the amount which cannot be delivered at all because the sum of requirements is smaller than the sum of amounts available.

Table 4 is obtained from Table 3 by using T'-algorithm (the roles of h and k reversed).

The non-negative optimal $\hat{\lambda}_{.k}$ and $\hat{\lambda}_{h.}$ are obtained as explained in Example 1.

The solution is not uniquely determined because the number x can vary freely in the interval $0 \leqslant x \leqslant 2$. The minimal costs are $50 + 10 = 60$.

Example 3.—Restelli's example (Annamaria Restelli: La programmazione lineare applicata a problemi di distribuzione, *Bolletino del Centro per la Ricerca Operativa*, Vol. 2, 1955, p. 1): Tables 6–8.

Table 7 is obtained from Table 6 by means of the T-algorithm. Table 8 is obtained from Table 7 by means of an exceptional move resulting in a new place (h, k) with $\lambda_{hk}^{(2)} = 0$ needed because $42 + 60 > 79$. In Table 8 the optimal transportation program is seen. The program is not uniquely determined because the number x can be freely chosen in the interval $0 \leqslant x \leqslant 22$. The minimal costs are $5357 + 138 = 5495$.

APPENDIX II

The Multiplier Theorem

If the set X_A defined by $x = \{x_I, x_J\} \geqslant 0$, where

$$x \geqslant 0 \quad \text{if} \quad x_I \geqslant 0, \quad x_J \geqslant 0,$$

$$x_I = \{x_i;\ i \in I\} \geqslant 0 \quad \text{if} \quad x_i = a_{i0} + \sum_{j \in J} a_{ij} x_j \geqslant 0 \quad \text{for} \quad i \in I,$$

$$x_J = \{x_j;\ j \in J\} \geqslant 0 \quad \text{if} \quad x_j \geqslant 0 \quad \text{for} \quad j \in J$$

is not empty, and the numerical value of

$$\hat{x}_0 = \max_{x \in X_A} (x_0 = \sum_{j \in J} a_{0j} x_j) = \sum_{j \in J} a_{0j} \hat{x}_j$$

is finite, then there exists a set $\hat{L}_A \in L_A$ of non-negative multipliers

$$\hat{\lambda} = \{\hat{\lambda}_I, \hat{\lambda}_J\} \geqslant 0$$

for which

$$\hat{\lambda}_I \geqslant 0, \qquad \hat{\lambda}_J \geqslant 0, \qquad \hat{\lambda}_j + a_{0j} + \sum_{i \in I} \hat{\lambda}_i a_{ij} = 0 \quad \text{for} \quad j \in J$$

$$\hat{\lambda}_0 = \min_{\lambda \in L_A} (\lambda_0 = \sum_{i \in I} \lambda_i a_{i0}) = \sum_{i \in I} \hat{\lambda}_i a_{i0},$$

$$\hat{\lambda}_0 = \hat{x}_0, \qquad \hat{\lambda}_i \hat{x}_i = 0 \quad \text{for} \quad i \in I, \qquad \hat{\lambda}_j \hat{x}_j = 0 \quad \text{for} \quad j \in J,$$

$$\hat{\lambda}_i \geqslant 0, \qquad \hat{\lambda}_j \geqslant 0,$$

where

$$\lambda \in L_A \quad \text{if} \quad \lambda = \{\lambda_I, \lambda_J\} \geqslant 0,$$

$$\lambda_j + a_{0j} + \sum_{i \in I} \lambda_i a_{ij} = 0 \quad \text{for} \quad j \in J.$$

To the multiplier theorem corresponds a dual theorem with the roles of λ and x interchanged. In the dual theorem it is assumed that L_A is not empty and $|\hat{\lambda}_0| < \infty$ while the assumptions made in the multiplier theorem about $\hat{x}$ and $\hat{x}_0$ appear as conclusions.

Proof.—The set $X_\lambda = \{x; \sum_{i \in I} \lambda_i x_i + \sum_{j \in J} \lambda_j x_j \geqslant 0\}$ contains for every non-negative $\lambda_i \geqslant 0$, $\lambda_j \geqslant 0$ the set $X_A = \{x;\ x_I, \geqslant 0,\ x_J \geqslant 0\}$ as a subset. The set of sets

$$X_+ = \{X_\lambda;\ \lambda \geqslant 0,\ \sum_{i \in I} \lambda_i + \sum_{j \in J} \lambda_j > 0\}$$

contains all sets X_λ with the property that at least one point x in X_λ lies on the boundary of the set X_A as a subset. The set $\hat{x}_0 - \sum_{j \in J} a_{0j} x_j \geqslant 0$ is for $|\hat{x}_0 = \sum_{j \in J} a_{0j} \hat{x}_j| < \infty$ a set with this property. Because for every λ_i, λ_j we have identically

$$\sum_{i \in I} \lambda_i x_i + \sum_{j \in J} \lambda_j x_j = \sum_{i \in I} \lambda_i a_{i0} + \sum_{j \in J} (\sum_{i \in I} \lambda_i a_{ij} + \lambda_j) x_j,$$

there will thus exist such $\hat{\lambda}_i \geqslant 0$, $\hat{\lambda}_j \geqslant 0$, $\sum_{i \in I} \hat{\lambda}_i + \sum_{j \in J} \hat{\lambda}_j > 0$ that we have

$$\sum_{i \in I} \hat{\lambda}_i a_{i0} = \hat{x}_0 = \sum_{j \in J} a_{0j} \hat{x}_j,\ a_{0j} + \sum_{i \in I} \hat{\lambda}_i a_{ij} + \hat{\lambda}_j = 0.$$

Because the point $x = \hat{x}$ lies both on the boundary of $X_{\hat{\lambda}}$ and on the boundary of X_A we must have

$$x_{\hat{\lambda}} = \sum_{i \in I} \hat{\lambda}_i \hat{x}_i + \sum_{j \in J} \hat{\lambda}_j \hat{x}_j = 0,\ \hat{x}_I \geqslant 0,\ \hat{x}_J \geqslant 0,\ \hat{\lambda}_I \geqslant 0,\ \hat{\lambda}_J \geqslant 0,$$

and because a sum of non-negative terms cannot be equal to zero if every term is not equal to zero, we get $\hat{\lambda}_i \hat{x}_i = 0$ for every $i \in I$, $\hat{\lambda}_j \hat{x}_j = 0$ for every

$$j \in J,\ \sum_{i \in I} \hat{\lambda}_i + \sum_{j \in J} \hat{\lambda}_j > 0.$$

For $\lambda_i \geqslant 0$, $a_{0j} + \sum_{i \in I} \lambda_i a_{ij} = -\lambda_j$ we have identically

$$\sum_{j \in J} a_{0j} x_j = \sum_{i \in I} \lambda_i a_{i0} - \sum_{j \in J} \lambda_j x_j - \sum_{i \in I} \lambda_i x_i.$$

If $\lambda \in L_A$ we further have $\lambda_j \geqslant 0$ and for $x \in X_A$, $x_i \geqslant 0$, and thus

$$\sum_{i \in I} \lambda_i a_{i0} \geqslant \sum_{j \in J} a_{0j} \hat{x}_j \text{ if } \hat{x} \in \hat{X}_A \subset X_A \text{ and } \lambda \in L_A.$$

Because $\{\hat{\lambda}_i;\ i \in I\}$ as proved is an element of L_A for which

$$\sum_{i \in I} \hat{\lambda}_i a_{i0} = \sum_{j \in J} a_{0j} \hat{x}_j;\ \hat{x} \in \hat{X}_A,$$

we see that the point $\hat{\lambda} \in L_A$ will further have the property

$$\hat{\lambda}_0 = \sum_{i \in I} \hat{\lambda}_i a_{i0} = \min_{\lambda \in L_A} \sum_{i \in I} \lambda_i a_{i0} = \max_{x \in X_A} \sum_{j \in J} a_{0j} x_j = \sum_{j \in J} a_{0j} \hat{x}_j = \hat{x}_0.$$

We have thus shown that all statements in the multiplier theorem are correct.

APPENDIX III

Remark about the Probability of the Existence of Finite Optimal Solutions to Linear Programming Problems and the Distribution of $(\hat{x}, \hat{\lambda})$.

CASE A.—If the elements in the matrix

$$A = \begin{Bmatrix} 0 & a_{0J} \\ a_{I0} & a_{IJ} \end{Bmatrix}$$

defining a linear programming problem and its dual problem are stochastic variables, the probability that $H = \hat{H}$, $K = \hat{K}$ can be written

$$\Pr(H = \hat{H}, K = \hat{K}) = \Pr(x_K(H, K) > 0, x_{I-H}(H, K) \geqslant 0, \lambda_H(H, K) > 0, \lambda_{J-K}(H, K) \geqslant 0).$$

CASE B.—If the elements of A are independent stochastic variables with continuous symmetric distribution functions around origo, all the variables $x_i(H, K)$; $i \in I - H$, $x_j(H, K)$; $j \in K$, and $\lambda_i(H, K)$; $i \in H$, $\lambda_j(H, K)$; $j \in J - K$ will also be symmetrically distributed around origo with the same measure of probability for everyone of the 2^N orthants. We will then get

$$\Pr(H = \hat{H}, K = \hat{K}) = 2^{-N},$$

where

$$N = N(I) + N(J) = N(I) + n$$

denotes the number of non-negativity conditions.

Because the number of different (H, K) that exist is equal to $\binom{N}{n}$, the probability P_B that a finite optimal solution exists will in this case B be

$$P_B = \binom{N}{n} 2^{-N}.$$

When $N \to \infty$ this probability converges towards 0.

CASE C.—If the elements of A are independent and normally distributed, the probability P_C that a finite optimal solution exists, will converge towards P_B when the standard deviations of all elements go towards infinity while their expected values remain constant.

If the expected values of the elements in A are not equal to zero and their standard deviations converge to zero, the probability P_C that a finite optimal solution will exist converges to 1 or 0 depending on whether a finite optimal solution exists or not for $A = \bar{A}$, where $\bar{A}$ denotes the matrix corresponding to the expected values of the elements in A. When the standard deviations σ_A of the elements in A are sufficiently small compared with their expected values, the coordinates of the optimal solution will be approximately normally distributed. The coordinates are, however, mutually dependent and their expected values will differ from the corresponding values for $A = \bar{A}$; $(\sigma_A = 0)$.

REFERENCES

CHARNES, A., COOPER, W. W. & HENDERSON, A., *An Introduction to Linear Programming.* Wiley, New York, 1953.

FRISCH, R., *Principles of Linear Programming.* Memorandum fra Universitetets Socialøkonomiske Institutt. Oslo, 1954.

—— *The Multiplex-Method for Linear Programming.* Memorandum fra Universitetets Socialøkonomiske Institutt. Oslo, 1955.

VAJDA, S., *The Theory of Games and Linear Programming.* London, 1956.

Vajda's book contains many further references. During recent years many new articles have been written in mathematical and statistical journals on this subject. I have not, however, investigated them carefully enough to detect if they contain any new ideas resembling the principles used in this paper.

RÉSUMÉ

Quelques nouveaux principes pour résoudre les problèmes de programmation linéaire

Quand on tente de résoudre les problèmes de programmation linéaire en pratique, le problème essentiel est de trouver les éléments $h \in \hat{H}$ et $k \in \hat{K}$ pour lesquels les valeurs optimales $\hat{x}_k > 0$ et les multiplicateurs optimaux $\hat{\lambda}_h > 0$.

Le papier donne les critères exacts pour $h \in \hat{H}$, $h \notin \hat{H}$, $k \in \hat{K}$, $k \notin \hat{K}$ et pour les cas dans lesquels il n'y a aucune solution optimale. Parce que les règles exactes pourraient être difficile d'appliquer, on donne aussi des règles pratiques pour définir les suites (H, K) qui remplacent approximativement $(\hat{H}, \hat{K})$.

Quelques exemples numériques sont résolus au moyen de ces nouveaux principes. Le problème de transportation est considéré en détail. Quelques remarques sur la probabilité de l'existence des solutions limitées optimales pour les problèmes de programmation linéaire et sur la distribution de $(\hat{x}, \hat{\gamma})$ sont données dans l'appendice.

ON DISTRIBUTION FUNCTIONS FOR QUANTITIES RELATED TO NETWORKS

by

LEO TÖRNQVIST
University of Helsinki, Finland

In this paper is presented the solution of the problem of finding the probability distribution of the shortest distance between two points, which are distributed independently and uniformly on a network.

At the end, more general models are given, which can be used for attacking problems of traffic distributions on a network.

In connection with the planning of a sample for investigating the properties of the post parcel traffic in Finland, the question about the mean value of the shortest distance between two points chosen at random from a network was raised. After some calculations, we found out that, if the network is a ring, this mean value is a quarter of the length of the ring. When this simple result was found, a keen interest arose to find more general results. Already for the relatively simple network obtained by drawing a new connecting line between two points on a ring, the methods at hand for attacking the problem led into lengthy calculations. We thus felt a need for more powerful methods. After a lot of trials and errors, some new ideas for attacking problems related to networks began to crystallize. Our interest then turned to questions about the probability distribution $F(s)$ of the shortest distance s between two points, which are distributed independently and uniformly on a network A. When developing these ideas, we found the following symbols useful:

(1) $s_{\alpha\beta}$=the shortest distance along A between the points $\alpha\in A$, $\beta\in A$,

(2) $n_\alpha(s)$=the number of points β on A for which $s_{\alpha\beta}=s$,

(3) $n_\alpha=\lim_{s\to+0} n_\alpha(s)$=the number of directions for moving from α,

(4) $k_\alpha=n_\alpha-2$=the number of surplus directions for the point α.

It may be observed that $k_\alpha=0$ for line points of A, $k_\alpha>0$ for junction points, while $k_\alpha=-1$ for end points of A.

A point $\beta\in A$ is a *via point* of (α, γ), shorter $\beta\in\langle\alpha, \gamma\rangle$ if

(5) $s_{\alpha\gamma}=s_{\alpha\beta}+s_{\beta\gamma}$.

A points $\alpha^*\in A$ with the properties

(6) $s_{\alpha\alpha^*}>s_{\alpha\beta}$, $s_{\alpha\alpha^*}>s_{\alpha\gamma}$, whenever $\alpha^*\in\langle\beta, \gamma\rangle$, $\beta\in\langle\alpha, \alpha^*\rangle$, $\gamma\in\langle\alpha, \alpha^*)$

is called a *counter point* of α.

We make the following unessential restrictions concerning the network A:

1. The shortest way between two junction points is unique.
2. Certain critical distances, namely, the largest distances between two points on a ring or on two different rings are mutually different.

Further, we denote the largest distance between two points on A by $\hat{s}$ and the total length of A by L.

If we put

$$L_\alpha(s)=L(\{\beta \mid s_{\alpha\beta}\leq s\}), \tag{7}$$

$$f_\alpha(s)=L^{-1}n_\alpha(s) \tag{8}$$

$$F_\alpha(s)=L^{-1}L_\alpha(s), \tag{9}$$

and finally

$$\phi(C)=1, \quad \text{if the condition } C \text{ is satisfied,}$$
$$\phi(C)=0 \quad \text{otherwise,}$$

we are equipped for an attack on the problem of finding the density function $f(s)$ of $s_{\alpha\beta}$, when α and β are random points distributed independently and uniformly on A.

We first observe that $f_\alpha(s)$ and $F_\alpha(s)$, for obvious reasons, are the density function and distribution function, respectively, of $s_{\alpha\beta}$ for a given α. According to the definition of the distribution function we have

$$F(s)=\underset{\alpha}{E}\,\underset{\beta}{E}\,\phi(s_{\alpha\beta}\leq s)=\underset{\alpha}{E}\,F_\alpha(s) \tag{10}$$

and thus

$$f(s)=\underset{\alpha}{E}f_\alpha(s)=L^{-1}\underset{\alpha}{E}\,n_\alpha(s), \tag{11}$$

where E denotes "the expected value of".

To get further than to the easily found result

$$f(s)=L^{-2}\sum_{\nu}\nu L(\{a \mid n_\alpha(s)=\nu\}), \tag{12}$$

we need a new fyrmula for $n_\alpha(s)$ instead of the definition stating that $n_\alpha(s)$ is the number of points $\beta\in A$ with the property $s_{\alpha\beta}=s$.

Studying, how $n_\alpha(s)$ changes, when s increases, one realizes that it can be written in the form

$$n_\alpha(s)=2\phi(s>0)+\sum_{\beta}k_\beta\phi(s>s_{\alpha\beta})-2\sum_{\alpha^*}\phi(s>s_{\alpha u^*}). \tag{13}$$

By observing that $n_\alpha(s)=0$, when $s>\underset{\beta\in A}{\mathrm{Max}}\, s_{\alpha\beta}$, we get the following interesting formula for n^*, the number of counter points of α,

$$n^*=1+\frac{1}{2}\sum_{\beta}k_\beta. \tag{14}$$

By integrating (13), we obtain

$$L_\alpha(s)=2s\phi(s>0)+\sum_{\beta}k_\beta(s-s_{\alpha\beta})\phi(s>s_{\alpha\beta})-2\sum_{\alpha^*}(s-s_{\alpha\alpha^*})\phi(s>s_{\alpha\alpha^*}). \tag{15}$$

If we take the expected value of both sides in (13), we find

$$(16)\qquad \underset{\alpha}{E}\, n_\alpha(s)=2\phi(s>0)+\sum_\beta k_\beta \underset{\alpha}{E}\,\phi(s>s_{\alpha\beta})-2\underset{\alpha}{E}\sum_{\alpha^*}\phi(s>s_{\alpha\alpha^*})$$

$$=2\phi(s>0)+\sum_\beta k_\beta F_\beta(s)-2L^{-1}L(s).$$

The question about, how the quantity $L(s)$ is most practically calculated, is still open. The following formula seems to be the best tool for this purpose.

$$(17)\qquad L(s)=\sum_\alpha\sum_{\alpha^*}(s-s_{\alpha\alpha^*})k_\alpha\phi(s>s_{\alpha\alpha^*})-\sum_z(s-z)n_z\phi(s>z)+\sum_z a_z\phi(s>z),$$

where

$$(18)\qquad a_z=L(A_z)=L(\{\alpha\,|\,\alpha=\alpha^{**},\ s_{\alpha\alpha^*}=z\}),$$

while n_z denotes the number of end points of A_z. (A separated point is considered as having two end points.)

The distribution studied is only one of many distributions for quantities related to networks. A more general model for distributions of this kind will be

$$(19)\qquad T(A\xrightarrow{C}B)=\sum_{\alpha\in A}\sum_{\beta\in B}T(\alpha\xrightarrow{C}\beta),$$

where $T(\alpha\xrightarrow{C}\beta)$ denotes the quantity of that part of the total traffic $T(\alpha\longrightarrow\beta)$ from α to β, which fulfills the condition C.

If we put

$$T(A\xrightarrow{C}B)=T(A\longrightarrow B)F_{A\to B}(C),$$

$$T(\alpha\xrightarrow{C}\beta)=T(\alpha\longrightarrow\beta)F_{\alpha\to\beta}(C),$$

$$p_{\alpha\beta}=p_{\alpha\to\beta}=T(\alpha\longrightarrow\beta)/T(A\longrightarrow B)$$

this model can also be written in the from

$$(20)\qquad F_{A\to B}(C)=\sum_{\alpha\in A}\sum_{\beta\in B}p_{\alpha\beta}F_{\alpha\to\beta}(C)=\underset{\alpha\in A}{E}\{\underset{\beta\in B}{E}F_{\alpha\to\beta}(C)\}.$$

Using such tools the traffic distribution problems can be attacked on a broader front.

It may be, for instance, interesting to observe that, if $p_{\alpha\beta}=g(s)$ for $s=s_{\alpha\beta}$, we obtain the density function of the corresponding traffic according to the distance between the different points α and β by means of the formula

$$(21)\qquad f_{A\to A}(s)=\frac{f(s)g(s)}{\int f(s)g(s)ds},$$

where $f(s)$ is the density function studied.

The formulas and concepts mentioned in this paper may give an impulse towards the developing of a sound theory for distribution functions related to networks. Such a theory would, for instance, be of value for scientists interested in traffic problems. Because in these problems the interesting objects of study are changes in the statistical attributes of some entity, such a theory may also have something to contribute to the statistical methodology for analyzing other phenomena in our world of time and space.

One of my pupils, Seppo Mustonen in Finland has shown a keen interest in

these problems. I would like to thank him for his very valuable help in reaching the sesults presented, and I hope he will be able to develop a more complete theory and draw out in more details the consequences of ideas mentioned in this contribution to the ISI meeting in Tokyo 1960.

RÉSUMÉ

Dans cette étude on présente la solution du problème de trouver la distribution de probabilité de la distance la plus courte entre deux points distribues indépendamment et uniformément sur un réseau.

A la fin on donne des modèles plus généraux qui peuvent être employés en traiant des problèmes de la distribution de circulation sur un réseau.

REVIEW OF THE INTERNATIONAL
STATISTICAL INSTITUTE
Volume 31 : 1, 1963

THE THEORY OF REPLICATED SYSTEMATIC CLUSTER SAMPLING WITH RANDOM START [1]

by

L. Törnqvist

Institute of Statistics, University of Helsinki, Finland

We consider the problem of estimating, for instance, the value V_I of the agricultural production in the country I. The estimating procedure could be based on a list showing the area, a_i, $(i \in I)$, of every farm in I. The total number of farms may be N.

The areas of the farms may be listed in a certain order j, $j \in J = \{1, \ldots, N!\}$. The area of the farm i listed in the order j may be denoted $a_{i,j}$. Let in this order, j, $A_{i-1,j}$ denote the total area of all farms preceding the farm no. i. We will then have

$$A_{ij} = A_{i-1,j} + a_{ij} \tag{1}$$

A_{ij} thus denote the area of all farms preceding the farm $i + 1$ in the list j. Specially we have

$$A_{oj} = 0 \qquad \text{for every } j, \text{ and} \tag{2}$$

$$A_{Nj} = A \qquad \text{for every } j, \tag{3}$$

where A is the total area of the country.

Let v_{ij} denote the value of the product of the farm i in the list j. Let accordingly in this order, j, $V_{i-1,j}$ denote the total value of the product of the farms preceding the farm no. i. We then have

$$V_{ij} = V_{i-1,j} + v_{ij} \tag{4}$$

$$V_{oj} = 0 \text{ and } V_{Nj} = V \qquad \text{for every } j, \tag{5}$$

where V denotes the total value of the farm production. The problem is to estimate V or $\frac{V}{A} = z_A$ by means of an appropriate sampling procedure.

Let

$$z_{ij} = \frac{v_{ij}}{a_{ij}} \tag{6}$$

denote the value of the production per unit of area for a certain farm (i) in a certain list (j).

If we now write

$$p_{ij} = \frac{a_{ij}}{A}, \qquad \text{then } \sum_i p_{ij} = 1 \text{ and} \tag{7}$$

$$z_A = \sum_i p_{ij} z_{ij} \tag{8}$$

[1] When working as UN-expert in Indonesia in 1960-61 the author became interested in the theory of systematic sampling. The first version of this paper was presented in January 1962 in a statistical seminarium at the University of Copenhagen. The author wishes to express his thanks to professor A. Hald for valuable comments made in this meeting. The author is indebted to Mr. S-E. Bergström and Mr. H. Järvinen who have checked the formulas and performed the numerical calculations.

For the cumulative sums of p_{ij} we have

$$P_{hj} = \sum_{i=1}^{h} p_{ij} = \frac{A_{hj}}{A}. \tag{9}$$

We define for every integer $n \geqslant 1$ the step function $z_j(t, n)$ by means of the rule

$$z_j(t, n) = z_{ij} \text{ if } nP_{i-1,j} < t \leqslant nP_{ij} \tag{10}$$

extended by means of the circular rule $z_j(t + n, n) = z_j(t, n)$. We then have

$$z_A = \frac{1}{n}\int_0^n z_j(t, n)\, dt = \frac{1}{n}\sum_{g=0}^{n-1} \bar{z}_j(\tau + g, n) \tag{11}$$

where

$$\bar{z}_j(\tau + g, n) = \int_0^1 z_j(((\lambda - \frac{1}{2}) + g + \tau), n)\, d\lambda.$$

Theorem 1: If τ is a random equidistributed variable in the interval (0,1) then the estimator

$$z_j(\tau \mid n) = \frac{1}{n}\sum_{g=0}^{n-1} z_j(\tau + g, n) \tag{12}$$

gives an unbiased estimate of z_A.

Proof. The random variable τ has, according to our assumption, the cumulative distribution function $P(\tau \leqslant x) = x$ for $x \in [0,1]$.

The theorem states, that

$$E(z_j(\tau \mid n)) = \int_0^1 z_j(\tau \mid n)\, d\tau = z_A. \tag{13}$$

By making use of the formula (12) we obtain after some simple calculations

$$\begin{aligned} E(z_j(\tau \mid n)) &= \frac{1}{n}\sum_{g=0}^{n-1}\int_0^1 z_j(\tau + g, n)\, d\tau = \\ &= \frac{1}{n}\sum_{g=0}^{n-1} \bar{z}_j((g + \frac{1}{2}), n) = \frac{1}{n}\int_0^n z_j(t, n)\, dt = z_A. \end{aligned} \tag{14}$$

The question about how $z_j(\tau \mid n)$ is distributed is the pertinent problem of the theory of systematic sampling. The solution to this problem is stated in the following theorem.

Theorem 2: The cumulative distribution function of $z_j(\tau \mid n)$ is equal to

$$\begin{aligned} &P\{z_j(\tau \mid n) \leqslant z\} = F_{jn}(z) \\ &= \sum_{\nu=1}^{N} (\alpha_{\nu j}^{(n)} - \alpha_{\nu-1,j}^{(n)})\, \varnothing(z_j(\alpha_{\nu j}^{(n)} \mid n) \leqslant z) = \\ &= \sum_{\nu=1}^{N} \beta_{\nu j}^{(n)}\, \varnothing(z_j(\alpha_{\nu j}^{(n)} \mid n) \leqslant z), \end{aligned} \tag{15}$$

where $\alpha_{0j}^{(n)} = 0 \leqslant \alpha_{1j}^{(n)} \leqslant \alpha_{2j}^{(n)} \leqslant \ldots \leqslant \alpha_{Nj}^{(n)}$ denote the sequence of numbers obtained, when the numbers $(nP_{ij}) = nP_{ij} - [nP_{ij}]$ ($[nP_{ij}]$ = the largest integer $< nP_{ij}$) are ordered according to size, and $\varnothing(c)$ is a step function with the property $\varnothing(c) = 1$ if the conditions represented by the symbol c are fulfilled while $\varnothing(c)$ otherwise is equal to zero.

Proof: According to formulas (12) and (10) the function $z_j(\tau \mid n)$ will change only if τ surpasses some of the values (nP_{ij}) and will thus remain constant in the interval $\alpha_{\nu-1,j}^{(n)} < \tau \leqslant \alpha_{\nu j}^{(n)}$. The length, $\overset{+}{p}_{\nu j}^{(n)} = \alpha_{\nu j}^{(n)} - \alpha_{\nu-1,j}^{(n)}$, of this interval thus measures the probability that $z_j(\tau \mid n) = z_j(\alpha_{\nu j}^{(n)} \mid n)$. Accumulating the lengths $[\alpha_{\nu j}^{(n)} - \alpha_{\nu-1,j}^{(n)}]$ of all intervals for which $z_j(\alpha_{\nu j}^{(n)} \mid n) \leqslant z$ we get the cumulative distribution function $F_{jn}(z)$ as stated in the theorem.

Theorem 3: The variance of the estimates $z_j(\tau \mid n)$ of z_A can be written in the following forms:

$$\sigma^2(z_j(\tau \mid n)) = \int_0^1 [z_j(\tau \mid n) - z_A]^2 \, d\tau \tag{16}$$

$$\sigma^2(z_j(\tau \mid n)) = \int_0^1 \frac{1}{n} \sum_{g=0}^{n-1} (z_j(\tau + g, n) - \bar{z}_j(\tau + g, n))^2 \, d\tau \tag{17}$$

$$= \frac{1}{n^2} \int_0^1 \sum_{g=0}^{n-1} \sum_{g'=0}^{n-1} \varepsilon_j(\tau + g, n) \, \varepsilon_j(\tau + g', n) \, d\tau$$

where $\varepsilon_j(\tau + g, n) = z_j(\tau + g, n) - \bar{z}_j(\tau + g, n)$ denote the deviation between the observations $z_j(\tau + g, n)$ and the mean $\bar{z}_j(\tau + g, n)$ of $z_j(t, n)$ over the interval $[(\tau + g - \frac{1}{2}) < t \leqslant (\tau + g + \frac{1}{2})]$. Expression (17) can also be written in the form

$$\sigma^2(z_j(\tau \mid n)) = \frac{1}{n^2} \sum_{\nu=0}^{n-1} \int_0^n \varepsilon_j(t, n) \, \varepsilon_j(t + \nu, n) \, dt = \tag{18}$$

$$= \frac{1}{n} \bar{\sigma}_{jn}^2 \left(1 + \sum_{\nu=1}^{n-1} \rho_{\nu jn}\right) = \frac{1}{n} \bar{\sigma}_{jn}^2 (1 + (n-1) \, \bar{\rho}_{jn})$$

where $$\bar{\sigma}_{jn}^2 = \frac{1}{n} \int_0^n (\varepsilon_j(t, n))^2 \, dt \text{ and}$$

$$\rho_{\nu jn} = \frac{1}{n} \int_0^n \varepsilon_j(t, n) \, \varepsilon_j(t + \nu, n) \, dt / \bar{\sigma}_{jn}^2$$

denotes the autocorrelation coefficient of $z_j(t, n)$ with the lag ν and $\bar{\rho}_{jn}$ denotes the mean of the n-1 values $\rho_{\nu jn}$.

Proof: The form (16) follows from the general definition of variance when applied to the case where τ is equidistributed over the interval [0,1]. The form (17) follows from (16) by making use of the forms (12) and (11). The form (18) follows from (17) by changing the order of integration and summations and observing that

$$\int_0^n z_j(t + g, n) \, z_j(t + g', n) \, dt = \sum_{g=0}^{n-1} \int_0^1 z_j(t + g, n) \, z_j(t + g', n) \, dt$$

because $z_j(t + g, n)$ is a periodic function with the period n. The rest of the formulas in the theorem is obtained by making use of shorter notations.

Remark: If the mean of the autocorrelation coefficient is zero the variance is equal to $\frac{\bar{\sigma}_{jn}^2}{n}$, which is an analogue to the variance of the mean obtained by random sampling. However the variance $\bar{\sigma}_{jn}^2$ denotes the variance of the deviations ε from the trend curve $\bar{z}_j(\tau + g, n)$ and will thus generally be smaller than the variance $\sigma^2(z(t)) = \sigma^2$ to be used in the corresponding variance formula $\frac{\sigma^2}{n}$, when random sampling is used. The mean value $\bar{\rho}_{jn}$ belongs to the interval $[-\frac{1}{n-1}, 1]$.

If we are interested to get numerical values for $\sigma^2(z_j(\tau \mid n))$ the formulas of theorem 3 are only of limited value. By making use of the distribution function $F_{jn}(z)$ defined in theorem 2 we get explicit formulas for this variance as stated in theorem 4.

Theorem 4: The variance $\varepsilon_{nj}^2 = \sigma^2(z_j(\tau \mid n))$ is equal to

$$\varepsilon_{nj}^2 = \sum_{\nu=1}^{N} \beta_{\nu j}^{(n)} (z_j(\alpha_{\nu j}^{(n)} \mid n) - z_A)^2 = \tag{19}$$

$$= \sum_{\nu=1}^{N} \beta_{\nu j}^{(n)} (z_j(\alpha_{\nu j}^{(n)} \mid n) - z_j(\alpha_{1j}^{(n)} \mid n))^2 \tag{20}$$

$$- \left[\sum_{\nu=1}^{N} \beta_{\nu j}^{(n)} (z_j(\alpha_{\nu j}^{(n)} \mid n) - z_j(\alpha_{1j}^{(n)} \mid n))\right]^2 =$$

$$= \sum_{\nu=1}^{N} \beta_{\nu j}^{(n)} \left(\frac{\sum_h \Delta z_{hj} \, \underset{+}{\varnothing}((nP_{hj}) < \alpha_{\nu j}^{(n)})}{n}\right)^2 \tag{21}$$

$$- \left[\frac{\sum_{\nu=1}^{N} \beta_{\nu j}^{(n)} \sum_h \Delta z_{hj} \, \underset{+}{\varnothing}((nP_{hj}) < \alpha_{\nu j}^{(n)})}{n}\right]^2$$

$$= \sum_{\nu=1}^{N} \beta_{\nu j}^{(n)} \left(\frac{u_{\nu jn}}{n}\right)^2 - \left(\sum_{\nu=1}^{N} \beta_{\nu j}^{(n)} \frac{u_{\nu jn}}{n}\right)^2 \tag{22}$$

$$= \frac{1}{n^2} \left(\sum \beta_{\nu j}^{(n)} u_{\nu jn}^2 - \left(\sum \beta_{\nu j}^{(n)} u_{\nu jn}\right)^2\right)$$

The formula (22) can also be written in the forms

$$n^2 \varepsilon_{nj}^2 = \sum_i \sum_h (1\text{-Max}(\underset{+}{(nP_{ij})}, \underset{+}{(nP_{hj})})) \Delta z_{ij} \Delta z_{hj} - \left(\sum_i \underset{+}{(nP_{ij})} \Delta z_{ij}\right)^2 \tag{23}$$

$$= \sum_i \sum_{h<i} - |\underset{+}{(nP_{ij})} - \underset{+}{(nP_{hj})}| \Delta z_{ij} \Delta z_{hj} - \left(\sum_i \underset{+}{(nP_{ij})} \Delta z_{ij}\right)^2 \tag{24}$$

$$= \sum_i \sum_h \text{Min}(\underset{+}{(nP_{ij})}, \underset{+}{(nP_{hj})}) \Delta z_{ij} \Delta z_{hj} - \left(\sum_i \underset{+}{(nP_{ij})} \Delta z_{ij}\right)^2 \tag{25}$$

$$= \sum_i \sum_h \left(\frac{1}{12} - \frac{1}{2} \underset{+}{(n(P_{ij} - P_{hj}))} \underset{+}{(n(P_{hj} - P_{ij}))}\right) \Delta z_{ij} \Delta z_{hj} \tag{26}$$

Proof: The probability that $z_j(\tau \mid n) = z_j(\alpha_{\nu j}^{(n)} \mid n)$ is according to theorem 2 equal to $(\alpha_{\nu j}^{(n)} - \alpha_{\nu-1,j}^{(n)})$ and we thus obtain (19). The formula (20) is obtained from (19) by changing the origo to $z_j(\alpha_{1j}^{(n)} \mid n)$. Next we have to show that

$$z_j(\alpha_{\nu j}^{(n)} \mid n) - z_j(\alpha_{1j}^{(n)} \mid n) = u_{\nu jn}/n \, .$$

To see this we note that when τ surpasses the point $\alpha_{\nu j}^{(n)}$ every term of the form z_{ij}/n included in the mean $z_j(\alpha_{\nu j}^{(n)} \mid n)$ will be exchanged to $z_{i+1,j}/n$ if $\underset{+}{(nP_{ij})} = \alpha_{\nu j}^{(n)}$ and therefore $z_j(\alpha_{\nu j}^{(n)} \mid n) - z_j(\alpha_{1j}^{(n)} \mid n) = \Sigma \Delta z_{hj} \, \underset{+}{\varnothing}((nP_{hj}) < \alpha_{\nu j}^{(n)})/n = u_{\nu jn}/n$.

The formulas (23) – (26) are obtained from (21) by expanding the squaresum in the first term of (21) and performing the first summation over ν and making use of the following relations between the quantities introduced:

$$\begin{cases} \sum_\nu \beta_{\nu j}^{(n)} \, \underset{+}{\varnothing}((nP_{hj}) < \alpha_{\nu j}^{(n)}) = 1 - \underset{+}{(nP_{hj})} = \underset{+}{(-nP_{hj})} = \underset{-}{(nP_{hj})} \\ \sum_\nu \beta_{\nu j}^{(n)} \, \underset{+}{\varnothing}((nP_{hj}) < \alpha_{\nu j}^{(n)}) \, \underset{+}{\varnothing}((nP_{ij}) < \alpha_{\nu j}^{(n)}) = 1\text{-Max}(\underset{+}{(nP_{hj})}, \underset{+}{(nP_{ij})}) \\ \text{Max}(x, y) = \dfrac{x+y}{2} + \dfrac{|x-y|}{2}, \quad \text{Min}(x, y) = \dfrac{x+y}{2} - \dfrac{|x-y|}{2} \\ \sum_i \Delta z_{ij} = 0, \qquad \sum_i \sum_h x_h \Delta z_{ij} \Delta z_{hj} = 0 \qquad \text{for every } x_h \, . \end{cases}$$

The main formula (26) we write also in the following forms

$$(26') \qquad n^2 \varepsilon_{nj}^2 = \sum_i \sum_h \left(\frac{1}{12} - \frac{1}{2} \underset{+}{(}n (P_{ij} - P_{hj})\underset{+}{)} \underset{+}{(}n (P_{hj} - P_{ij})\underset{+}{)}\right) \Delta z_{ij} \Delta z_{hj}$$

$$= \frac{1}{2} \sum_i \sum_h \left(\frac{1}{6} - \underset{+}{(}n (P_{ij} - P_{hj})\underset{+}{)}\right) \Delta z_{ij} \Delta z_{hj}$$

$$= \sum_i \sum_h \frac{1}{2} \underset{+}{(}n (P_{ij} - P_{hj})\underset{-}{)} (z_{i+1,j} - z_{ij}) (z_{hj} - z_{h+1,j})$$

where $\underset{+}{(}x\underset{-}{)} = \underset{+}{(}x\underset{+}{)} \underset{+}{(}-x\underset{+}{)}$.

Remark: The variance ε_{nj}^2 can also be written in the form

$$(26'') \qquad \varepsilon_{nj}^2 = 2 \sum_{g=1}^{\infty} | A_j (gn) |^2$$

where

$$A_j (gn) = -\frac{1}{2\pi\sqrt{-1}\, g} \sum_i z_{ij} \left(e^{-2\pi\sqrt{-1}\, g\, nP_{ij}} - e^{-2\pi\sqrt{-1}\, g\, nP_{i-1,j}}\right)$$

is the Fourier-coefficient corresponding to frequence $\nu = gn$ for $z_j (u) = z_j (nu, n)$. The formula (26″) gives some idea about how the variance diminishes with growing n, because the formula for ε_{nj}^2 contains starting with the term $2 | A_j (n) |^2$ only every n : *th* term of the terms, which are contained in ε_{1j}^2.

How to calculate the mean $n^2\varepsilon_n^2$ over j of $n^2\varepsilon_{nj}^2$ will be investigated in the following theorem.

Theorem 5: The mean value of $n^2\varepsilon_{nj}^2$ over different possible orderings $j = 1, \ldots, N!$ of the frame is equal to

$$(27) \qquad n^2 \varepsilon_n^2 = \frac{1}{N!} \sum_{j=1}^{N!} n^2 \varepsilon_{nj}^2 = \sum_{g=1}^{N-1} \sum_{\nu=1}^{\binom{N}{g}} \underset{+}{(}nL_{\nu g}\underset{-}{)}\, w_{\nu g}$$

when

$$L_{\nu g} = \sum_{i \in C_{\nu g}} p_i, \text{ where } p_i = p_{i1} \text{ and } C_{\nu g} \text{ is a sample of } g \text{ elements from the}$$

set $C_{1N} = (1, \ldots, N)$ and

$$(28) \qquad w_{\nu g} = \frac{N}{2\binom{N}{g}} \left[\frac{\left(\sum_{i \in C_{\nu g}} z_i\right)^2 - \binom{g'}{2} \sum_{i \in C_{\nu g}} z_i^2}{g^{[g']}} \right.$$

$$\left. - 2 \frac{\sum_{i \in C_{\nu g}} z_i \cdot \sum_{h \notin C_{\nu g}} z_h}{g(N-g)} + \frac{\left(\sum_{h \notin C_{\nu g}} z_h\right)^2 - \binom{g''}{2} \sum_{h \notin C_{\nu g}} z_h^2}{(N-g)^{[g'']}}\right]$$

where $z_i = z_{i1}$, $g' = \text{Min}\,\{g, 2\}$, $g'' = \text{Min}\,\{(N-g), 2\}$.

The quantity $w_{\nu g}$ can also be written in the following forms.

$$(29) \qquad w_{\nu g} = \frac{N}{2\binom{N}{g}} \left[\sum_{i \in C_{\nu g}} \frac{z_i\, 1_i^{[]}}{g^{[]}} - \sum_{h \notin C_{\nu g}} \frac{z_h\, 1_h^{[]}}{(N-g)^{[]}}\right]^2$$

$$(30) \qquad = \frac{N}{2\binom{N}{g}} \left[(z_{\nu g} - z_{\nu, N-g})^2 - s_{\nu g}^2 - s_{\nu, N-g}^2\right]$$

where $z_{\nu g} = \frac{1}{g} \sum_{i \in C_{\nu g}} z_i$, $z_{\nu, N-g} = \frac{1}{N-g} \sum_{h \notin C_{\nu g}} z_h$,

$$s^2_{\nu g} = \frac{1}{g^{[g']}} \sum_{i \in C_{\nu g}} (z_i - z_{\nu g})^2, \quad s^2_{\nu, N-g} = \frac{1}{(N-g)^{[g^*]}} \sum_{h \notin C_{\nu g}} (z_h - z_{\nu, N-g})^2$$

and the factorial operator symbols $1_i^{[\,]}$, $g^{[\,]}$ and $(N-g)^{[\,]}$
have the properties $\sum_{i \in C_{\nu g}} 1_i^{[\,]} = g^{[\,]}$, $\sum_{h \notin C_{\nu g}} 1_h^{[\,]} = (N-g)^{[\,]}$,

$$\left[\frac{\sum_{i \in C_{\nu g}} z_i \, 1_i^{[\,]}}{g^{[\,]}}\right]^2 = \frac{\sum_{i \in C_{\nu g}} \sum_{\lambda \in C_{\nu g - i}} z_i \, z_\lambda}{g^{[2]}} = \frac{(\sum_{i \in C_{\nu g}} z_i)^2 - \sum_{i \in C_{\nu g}} z_i^2}{g^{[2]}}$$

when $g \geqslant 2$.

For $g = 1$ this expression is equal to z_i^2; $i \in C_{\nu 1}$,
$g^{[2]} = g\,(g-1)$, $[1_i^{[\,]}\, 1_h^{[\,]}] = 1$ if $i \neq h$, $[1_i^{[\,]}\, 1_i^{[\,]}] = 0$.

Remark: By introducing the symbols

$$\overset{+}{z}_i = z_i - \frac{1}{N} \sum_{i=1}^{N} z_i = z_i - z_{1N}, \tag{31}$$

$$\sigma^2_{1N} = \frac{1}{N-1} \sum_{i=1}^{N} \overset{+}{z}_i^2 = \frac{1}{N-1} \left(\sum_{i=1}^{N} z_i^2 - N z^2_{1N}\right) \tag{32}$$

$$\sigma^2_{\nu g} = \frac{1}{g-1} \left(\sum_{i \in C_{\nu g}} z_i^2 - \frac{1}{g} \left(\sum_{i \in C_{\nu g}} z_i\right)^2\right) \tag{33}$$

$$z_{\nu g} = \frac{1}{g} \sum_{i \in C_{\nu g}} z_i \tag{34}$$

we can for $N > 2$ modify the expression for $n^2 \varepsilon_n^2$ as follows

$$n^2 \varepsilon_n^2 = \sum_{i=1}^{N} \underset{+}{\underset{-}{(np_i)}} \frac{N \overset{+}{z}_i{}^2 - \sigma^2_{1N}}{(N-2)} + \left[\varphi\,(N > 3) + \varphi\,(N > 4)\right] R_{N2} + R_{N3} \tag{35}$$

where

$$R_{N2} = \sum_{i=1}^{N} \sum_{h<i} \underset{+}{\underset{-}{(np_i + np_h)}} \left[\frac{N \overset{+}{z}_i \overset{+}{z}_h + (z_i - z_h)^2 - \sigma^2_{1N}}{(N-2)\,(N-3)}\right]$$

$$R_{N3} = \sum_{g=3}^{N-3} \sum_{\nu=1}^{\binom{N}{g}} \frac{(g-2)!}{2\,(N-2)^{[g]}} \underset{+}{\underset{-}{(nL_{\nu g})}} \left[N\left((z_{\nu g} - z_{1N})^2 - \frac{1}{g}\sigma^2_{\nu g}\right) + 2\,\sigma^2_{\nu g} - \sigma^2_{1N}\right]$$

For $N = 2$ we have $n^2 \varepsilon_n^2 = \underset{+}{\underset{-}{(np_1)}}\,(z_1 - z_2)^2$.

The formula (35) is recommended for practical use. The formula (27) combined with (30) shows best the character of $n^2 \varepsilon_n^2$. From (30) it is easy to prove that $\sum_{\nu} w_{\nu g} = 0$ for $g > 1$, $N-g > 1$. When $g = 1$ or $N-1$, $s^2_{\nu g} = 0$. The sum $s^2_{\nu g} + s^2_{\nu, N-g}$ is for $g > 1$, $N-g > 1$ the usual unbiased estimate of the variance of $z_{\nu g} - z_{\nu, N-g}$ (when random sampling is used).

If $p_i = \frac{1}{N}$ the formula for ε_n^2 can be simplified to

$$\varepsilon_n^2 = \frac{(N-n)}{Nn} \sigma^2_{1N}$$

In this case ε_n^2 corresponds to the variance of the mean obtained by means of random sampling without replacement. This was first proved by Madow (1944).

Proof: According to the theorem 4 formula (26′) we can write

$$n^2 \varepsilon_n^2 = \frac{1}{2\,N!} \sum_{j=1}^{N!} \sum_{i=1}^{N} \sum_{h=1}^{N} \underset{+}{(n\,(P_{ij} - P_{hj}))}\,(z_{i+1,j} - z_{ij})\,(z_{hj} - z_{h+\ ,j})$$

We have had great help of a graph in our attempts to transform this formula into new formulas useful for practical purposes. In this graph the frame is represented by means of a probability circle C_n. This circle is divided in arcs c_k, $k = 1, \ldots, N$, of length np_k.

On the circle C_n we consider two clockwise endpoints of arcs c_{hj} and c_{ij} which divide the circle C_n into two parts $C_{\nu\,g}\,(h, i)$ and $C_{\nu,\,N-g}\,(i, h)$ consisting of g and respectively $N-g$ arcs $c_{\mu j}$. The quantity $nL_{\nu\,g}$ is the length of $C_{\nu\,g}\,(h, i)$ and $n\,(1-L_{\nu\,g})$ is the length of $C_{\nu,\,N-g}\,(i, h)$.

The quantity $\underset{+}{(nL_{\nu\,g})} = \underset{+}{(n\,(P_{ij} - P_{hj}))}$ is then equal to the product of the rests (modulo 1) of the lengths of the arcs $C_{\nu\,g}\,(h, i)$ and $C_{\nu,\,N-g}\,(i, h)$. These two rests are $\underset{+}{(nL_{\nu\,g})}$ and $\underset{+}{(n\,(1-L_{\nu\,g}))} = \underset{+}{(-nL_{\nu\,g})} = \underset{-}{(nL_{\nu\,g})}$.

The quantity $\underset{+}{(nL_{\nu\,g})}$ is invariant for a subgroup $J_{\nu\,g}$ of $Ng!\,(N-g)!$ permutations j. The permutations belonging to this group $J_{\nu\,g}$ have the property that the set $C_{\nu\,g}$ of g arcs $c_k \in C_{\nu\,g}\,(h, i)$ is invariant and the total length of arcs c_k belonging to $C_{\nu\,g}$, $\sum_{k \in C_{\nu\,g}} np_k = nL_{\nu\,g}$ is constant. We can thus write

$$n^2 \varepsilon_n^2 = \sum_{g=1}^{N-1} \binom{N}{g} \sum_{\nu=1} \frac{N}{2\binom{N}{g}} \underset{+}{(nL_{\nu\,g})} \left[\sum_{i \in C_{\nu\,g}} \sum_{\lambda \in C_{\nu,\,N-g}} \sum_{\rho \in C_{\nu\,g} - i \varnothing (g'=2)} \right.$$

$$\left. \sum_{h \in C_{\nu,\,N-g} - \lambda \varnothing (g''=2)} (z_\lambda - z_i)\,(z_h - z_\rho) \right] \frac{1}{N_g}$$

where $N_g = g^{[g']}\,(N-g)^{[g'']}$ denotes the number of terms in the brackets.

The results stated in this theorem are obtainable from this basic formula most elegantly by making effective use of the factorial operators $1_l^{[]}$, $g^{[]}$, $(N-g)^{[]}$ etc.

The results have been checked by direct algebraic calculation transforming the sums in the bracket and finally also by means of numerical examples[1] for $N \leqq 5$.

The next theorem solves the problem about the mean of $n^2\,\varepsilon_{nj}^2$ over different $n = 1, \ldots, \infty$.

Theorem 6: If all $(P_{ij} - P_{hj})$ are irrational numbers then the mean value of $n^2\,\varepsilon_{nj}^2$ over $n = 1, \ldots,$ is for a given permutation j equal to

$$E_{Nj}^2 = \frac{1}{12} \sum_{i=1}^{N} (\Delta\, z_{ij})^2 \tag{36}$$

and the mean of $n^2\,\varepsilon_n^2$ over all $n = 1, 2, \ldots$ is equal to

$$E_N^2 = \frac{1}{6}\,N\,\sigma_{1N}^2\,. \tag{37}$$

Proof: We assume tentatively that $P_{ij} - P_{hj}$ is a rational number of the form K_{hij}/M_{hij}, where K_{hij} and M_{hij} have no common factor. If we let n vary from $n = 1$ to $n = M_{hij}$ the set of numbers $\{\underset{+}{(n\,(P_{ij} - P_{hj}))},\ n = 1, \ldots, M_{hij}\}$ is equal to the set of numbers $\{\nu/M_{hij},\ \nu = 1, \ldots, M_{hij}\}$.

[1] For details see appendix 1.

If our last statement were not fulfilled we must for some $n' \neq n$ have $(\overset{+}{n} K_{hij}/M_{hij}) = (\overset{+}{n'} K_{hij}/M_{hij})$. We could then conclude that $(\overset{+}{n-n'})\, K_{hij}/M_{hij}$ must be an integer. This is not true, because according to our assumptions $(n - n') < M_{hij}$ and K_{hij}, M_{hij} are integers without common factor.

The mean value of $(\overset{+}{n}(P_{ij}-P_{hj}))$ over all $n = 1, \ldots, M_{hij}$ will thus be equal to

$$\sum_{n=1}^{M_{hij}} \frac{1}{M_{hij}} (\overset{+}{n}(P_{ij} - P_{hj})) = \sum_{\nu=1}^{M_{hij}} \frac{1}{M_{hij}} \left[\frac{\nu}{M_{hij}} \left(1 - \frac{\nu}{M_{hij}}\right) \right] = \frac{1}{6} \left\{ 1 - \left(\frac{1}{M_{hij}} \right)^2 \right\}.$$

If $P_{ij} - P_{hj}$ are irrational numbers, they can be considered as limit values for sequences of rational numbers of type $K_{hij}^{(k)}/M_{hij}^{(k)}$, where $M_{hij}^{(k)} \to \infty$, for $k \to \infty$. We conclude that

$$\operatorname{limes}_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} \left[\frac{1}{6} - (\overset{+}{n}\,(P_{ij} - P_{hj})) \right] = 0$$

The first statement of theorem 6 is obtained if we write in accordance with formula (26')

$$\operatorname{limes}_{M \to \infty} \sum_{n=1}^{M} \frac{1}{M} n^2 \varepsilon_{nj}^2 = \operatorname{limes}_{M \to \infty} \left\{ \frac{1}{M} \sum_{n=1}^{M} \sum_{i=1}^{N} \frac{1}{12} (\Delta z_{ij})^2 + \right.$$

$$\left. \frac{1}{2} \sum_{i=1}^{N} \sum_{h \neq i} \Delta z_{ij}\, \Delta z_{hj} \frac{1}{M} \sum_{n=1}^{M} \left[\frac{1}{6} - (\overset{+}{n}\,(P_{ij} - P_{hj})) \right] \right\} = \sum_{i=1}^{N} \frac{1}{12} (\Delta z_{ij})^2.$$

The mean E_N^2 is derived from formula (35) after using formula (30) to find out that the mean values of R_{N3} and R_{N2} over n are equal to zero. Taking the mean value over n of the first term in formula (35) we get

$$E_N^2 = \operatorname{limes}_{M \to \infty} \sum_{i=1}^{N} \sum_{n=1}^{M} \frac{1}{M} (\overset{+}{n p_i}) \cdot \frac{N \overset{+}{z_i^2} - \sigma_{1N}^2}{N-2} = \frac{1}{6} \sum_{i=1}^{N} \frac{N \overset{+}{z_i^2} - \sigma_{1N}^2}{N-2} = \frac{1}{6} N \sigma_{1N}^2.$$

Sampling variance estimates. When we have to estimate the variance of the mean in systematic sampling, the crux of the problem is that only one sample point $z_j(\tau \mid n)$ is available. One idea to circumvent this difficulty is to somehow replicate the sampling.

If we choose at random m different values $(\tau_1, \tau_2, \ldots, \tau_m) = T_m$ we will get an unbiased estimate $s^2(z_j(\tau \mid n))$ for $\sigma^2(z_j(\tau \mid n)) = \varepsilon_{nj}^2$ by means of the classical formula:

$$s_m^2(z_j(\tau \mid n)) = \frac{1}{m-1} \sum_{\nu=1}^{m} [z_j(\tau_\nu \mid n) - \bar{z}_j(T_m \mid n)]^2 \tag{38}$$

where

$$\bar{z}_j(T_m \mid n) = \frac{1}{m} \sum_{\nu=1}^{m} z_j(\tau_\nu \mid n).$$

An unbiased estimate of the variance of the over all mean $\bar{z}_j(T_m \mid n)$ is obtained by means of the usual formula for the variance of a mean.

$$s^2(\bar{z}_j(T_m \mid n)) = \frac{1}{m} s_m^2(z_j(\tau \mid n)) \tag{39}$$

Remark: If $n < N$ and n is large we may sometimes with good reason assume that the mean value $\bar{\rho}_{jn}$ of the autocorrelation coefficients in the "time series" $\varepsilon_j(t, n)$ is approximately zero. If this is true we would get useful approximations for the variance ε_{nj}^2 by making use of the variate difference method. We will then have:

$$\varepsilon_{nj}^2 \approx \frac{1}{2n^2} \sum_{g=0}^{n-1} (z_j(g+\tau, n) - z_j(g+1+\tau, n))^2 \tag{40}$$

$$= \frac{1}{2n^2} \Sigma (\Delta_g z_j(g+\tau, n))^2$$

$$\approx \frac{1}{6n^2} \Sigma (\Delta_g^2 z_j(g+\tau, n))^2 \tag{41}$$

$$\approx \frac{1}{20n^2} \sum_{g=0}^{n-1} (\Delta_g^3 z_j(g+\tau, n))^2 . \tag{42}$$

and more generally:

$$n^2 \varepsilon_{nj}^2 \approx n^2 \varepsilon_{njk\tau}^2 = \sum_{g=0}^{n-1} (\Delta_g^k z_j(\tau+g, n))^2 / \binom{2k}{k};\ 0 < 2k < n \tag{43}$$

A slight modification of these formulas developed in the years 1923–24 by J. W. Lindeberg have been used in Finland in determining the accuracy of line estimates of the amount of timber in forests.

The expected value of $n^2 \varepsilon_{njk\tau}^2$ is equal to

$$n^2 \varepsilon_{njk}^2 = \sum_i \sum_h \Delta z_{ij} \Delta z_{hj} d_{ih,njk} \tag{44}$$

where

$$d_{ih,njk} = \frac{(-1)^{k-1}}{\binom{2k}{k}} \delta_u^{2k} \frac{|u|}{2} = \frac{1}{2} \left[\frac{(k-1)^{[g]}}{(-(k+1))^{[g]}} \left(\frac{k+g}{2k-1} \pm (|u|) \right) \right]$$

for $u = n(P_{ij} - P_{hj})$ and $|u| - 1 < g \leqslant |u|$,

$$\delta_u^2 \frac{|u|}{2} = \frac{1}{2}(|u+1| - 2|u| + |u-1|) = (1-|u|)\, \text{ø}\, (|u| < 1),$$

$$(-1)^{k-1} \delta_u^{2k} \frac{|u|}{2} = (-1)^g \left(\binom{2k-2}{k-1-g} - \binom{2k-1}{k-g} (|u| - g) \right);\ 0 \leqslant |u| - g < 1$$

If we observe also $z_{i+1,j}$ and $z_{i-1,j}$ whenever z_{ij} belongs to our sample we could get a good approximation to ε_{nj}^2 by means of the formula

$$n^2 * \varepsilon_{nj}^2 = \sum_{i=1}^{N} \frac{\nu_{ij}}{n p_{ij}} \left[\frac{(\Delta z_{ij})^2}{12} + \left(\frac{1}{6} \pm (n p_{ij}) \right) \Delta z_{ij} \Delta z_{i-1,j} \right] \tag{45}$$

where ν_{ij} denotes the number of integers g for which $z_{ij} = z_j(\tau+g, n)$; $g = 0, \ldots, n-1$. The expected value of $*\varepsilon_{nj}^2$ is equal to the approximation to ε_{nj}^2 obtained from (26′) when terms for which $|h-i| > 1$ are neglected. Often we get a useful approximation to ε_{nj}^2 if we instead of $z_{i+1,j}$, $z_{i-1,j}$ use as substitutes the neighbouring really observed values $z_j(\tau+g-1, n)$ and $z_j(\tau+g+1, n)$ to $z_j(\tau+g, n)$ in the sample.

It can be shown by means of Fourier analysis that:

$$\varepsilon_{n_o j}^2 - \varepsilon_{nj}^2 = \underset{\tau}{E}\, Q_{jn_o}^2(\tau, h) \text{ for } n = n(h) = n_o 2^h . \tag{46}$$

where

$$Q_{jn_o}^2(\tau, h) = \sum_{\nu=1}^{h} (\Delta_\nu z_j(\tfrac{\tau n(\nu-1)}{n} \mid n(\nu-1)))^2 \tag{47}$$

The function $Q_{jn_o}^2(\tau, h)$ gives thus valuable information about the increase in accuracy of the estimates $z_j(\tau \mid n(h))$ by successive doubling the amount of data used. Curves of $Q_{jn_o}^2(\tau, h)$ can also be used for predicting the value of $\varepsilon_{nj}^2 = \varepsilon_{nj}^2 - \varepsilon_{\infty j}^2$. A formula which usually gives a slight overestimate of ε_{nj}^2 is for instance

$$\varepsilon_{nj}^2 \approx \frac{(Q_{jn}^2(\tau, h) - Q_{jn_o}^2(\tau, h-q))\,(Q_{jn_o}^2(\tau, h) - Q_{jn_o}^2(\tau, h'))}{Q_{jn_o}^2(\tau, h') - Q_{jn_o}^2(\tau, h'-q) - Q_{jn_o}^2(\tau, h) + Q_{jn_o}^2(\tau, h-q)}.$$

Here the positive integers $q < h' <$ h can be varied to give a whole set of different estimates for ε_{nj}^2.

In the literature of the theory of systematic sampling (see list of references) also cases when the points to be observed are defined by means of regular systems of lattices of points in a multidimensional space are treated. These studies generally make use of assumptions about the correlation coefficients between observations located by points at a certain distance from each other in this space. Our theorem 3 is a contribution along this line of approach. The theory of systematic sampling is also nearly related to the theory of the analysis of timeseries.

Remark about the use of clustering in systematic sampling

The theory developed can as such be applied also to cluster sampling, if we let z_i and p_i denote the means and weights for clusters of final observational units. It seems to be profitable to use clustering in cases when otherwise the weights would become relatively very small or if there are parts in the list where the intracluster variance of z would be very large if these clusters were partioned into smaller units. Whenever some units in such a cluster have to be observed according to the plan for systematic sampling, then the whole cluster of observational units should be observed.

The optimal use of parameters of action in systematic sampling

The parameters of action in systematic replicated cluster sampling are:

(1) the choice of the weights $a_{i\mu}$ to which the final observations $v_{i\mu}$ are related and the construction of clusters with weights $a_i = \sum_\mu a_{i\mu}$ and cluster means $z_i = \sum_\mu v_{i\mu} / a_i = v_i / a_i$.

(2) the ordering (j) of the list of clusters.

(3) the choice of the number (n) of clusters to be observed.

(4) the number of replications (m) to be made.

1. If there exists a number of competing systems of weights $\{p_i^{(\nu)}\}$ the theory developed seems to support such a choice which will minimize the overall variance σ_{1N}^2, that is we have to choose a_i so that v_i is approximately proportional to a_i.

2. The ordering (j) is to be chosen so that $\sum_i (\Delta z_{ij})^2$ will become relatively small. This will be the case if $z_j(t, n)$ is as smooth as possible with only one maximum and minimum.

3. If $\sum_i (np_i)$ has a relative minimum for varying $n > 1$, then the corresponding choice $\overset{\vee}{n}$ of n will usually give an exceptionally high precision for the mean obtained.

4. The number of replicates m has to be chosen just large enough to give the required precision for the estimates to be made.

5. Systematic (equidistant) replicated cluster sampling can be combined with stratification and complete enumeration of those clusters for which $[np_i] \geqslant 1$. This idea was utilized in construction of a master sample of local offices to be investigated by means of sampling in studies concerning the economy of Post and Telegraph Office in Finland.[1]

[1] For details see appendix 2.

Considering the merits of systematic replicated cluster sampling we have come to the conclusion that if this flexible tool is really effectively used one can obtain unbiased estimates with a higher precision than by other sampling procedures requiring the same amount of labour to be done. It would therefore be worthwhile to study the properties of this tool when applied to examples based on empirical data in cases when N and n are large enough to be of practical interest. Such a study would necessarily require the use of electronic computers.

REFERENCES

Madow, W. G., Madow. L. H. On the theory of systematic sampling, I. Annals of mathematics, 15, 1944, p 1–24.

Lindeberg, J. W. Todennäköisyyslasku ja sen käytäntö tilastotieteessä. 1927, p. 104–120.

Textbooks dealing with systematic sampling:

Cochran, W. G. Sampling techniques.

Hansen, M. H., Hurwitz, W. N., Madow, W. G. Sample survey methods and theory, I and II.

Literature concerning systematic sampling and related topics:

papers before 1950 are cited in the article of:

Buckland, W. R. A review of the literature of systematic sampling. Journal of the Royal statistical society (B), 13, 1951, p 208–215.

Dalenius, T., Hajek, J., Zubrzycki, S. On plane sampling and related geometrical problems. Proceedings of the fourth Berkeley symposium, Volume, I, 1961, p 125–150.

Das, A. C. Two dimensional systematic sampling and the associated stratified and random sampling. Sankyā, 10, 1950, p 95–108.

Eckler, A. R. Rotation sampling. Annals of mathematical statistics, 26, 1955, p 664–686.

Gautschi, W. Some remarks on systematic sampling. Annals of mathematical statistics, 28, 1957, p 385–394.

Hald, A. The decomposition of a series of observations composed of a trend, a periodic movement, and a stochastic variable. Copenhagen, 1948.

Madow, W. G. On the theory of systematic sampling, III. Comparison of centered and random start systematic sampling. Annals of mathematical statistics, 24, 1953, p 101–106.

Mattila, S. Some tests based on moving average operations on time series. Annales adademiae scientiarum fennicae, Series A 226, 1956.

Stange, K. Probenahme vom Band. Metrika, 1, 1958, p 177–222.

Williams, R. M. The variance of the mean of systematic samples. Biometrika, 43, 1956, p 137–148.

Appendix 1. *A numerical example*

We have calculated a hypothetical example with $N = 5$, $n = 1, \ldots, 10$

i	p_{i1}	z_{i1}	P_{i1}	$(10P_{i1})$	ν	ν	$\alpha_{\nu 1}^{(10)}$	$\beta_{\nu 1}^{(10)}$	$z_1(\alpha_{\nu 1}^{(10)} \mid 10)$
1	0.11	5	0.11	0.10	1	1	0.10	0.10	5.4
2	0.16	3	0.27	0.70	4	2	0.20	0.10	5.2
3	0.19	6	0.46	0.60	3	3	0.60	0.40	4.9
4	0.26	7	0.72	0.20	2	4	0.70	0.10	5.0
5	0.28	4	1.00	1.00	5	5	1.00	0.30	5.3
$\bar{z}$ A		5.11							5.11
ε_{n1}^2		2.1379							0.0389

We first calculate $z_j(\alpha_{\nu j}^{(n)} \mid n)$ for $\nu = 1$ which is done as follows:

$$z_j(\alpha_{\nu j}^{(n)} \mid n) = \left(\sum_i [np_{ij}]\, z_{ij} + \sum_i \varnothing\left(\left(\underset{+}{nP_{ij}}\right) < \left(\underset{+}{nP_{i-1,j}}\right)\right) z_{ij}\right)/n;$$

$5.4 = 0.1\,((5 + 3 + 6 + 7 + 7 + 4 + 4) + (5 + 6 + 7))$. The others are obtained by means of formula $dz_j(\tau \mid n) = \left(\sum_i (z_{i+1,j} - z_{ij})\, \varnothing\left(\left(\underset{+}{nP_{ij}}\right) = \tau\right)\right)/n$ as follows:

$5.2 - 5.4 = (3 - 5)/10 = -0.2$, $4.9 - 5.2 = (4 - 7)/10 = -0.3$, etc.

In our example there are in all $(N-1)!/2 = 12$ possible permutations j with different distributions of z_j $(\tau \mid n)$. The permutations $j = 1, \ldots, 12$ used are as follows:

$j =$	1	2	3	4	5	6	7	8	9	10	11	12
z_{1j}	5	5	5	5	5	5	5	5	5	5	5	5
z_{2j}	3	3	3	6	6	6	7	7	7	4	4	4
z_{3j}	6	4	7	7	3	4	4	4	3	3	6	3
z_{4j}	7	6	4	4	7	3	3	6	6	6	7	7
z_{5j}	4	7	6	3	4	7	6	3	4	7	3	6
$\sum_i (\Delta z_{ij})^2$	24	14	34	16	36	26	24	30	34	16	26	20

For each permutation j the quantity $n^2\varepsilon^2_{nj}$ was calculated by means of formula (24) (table 1). We have calculated from these results the empirical means over the 12 permutations j and compared them with the means $n^2\varepsilon^2_n$ calculated directly by means of formula (35). The two means were naturally found to be equal in every case (for 10 different n). The variation among variances $n^2\varepsilon^2_{nj}$ over j was also calculated.

The empirical means $^*E^2_{Nj}$ over the ten different n used for each permutation j as well as the variance of $n^2\varepsilon^2_{nj}$ about these means were also calculated. Finally an analysis of variance was made.

Table 1. An example about the quantities $n^2\varepsilon^2_{nj}$ in systematic sampling when $N = 5$

$n^2\varepsilon^2_{nj}$ $j =$	1	2	3	4	5	6	7	8	9	10	11	12	$n^2\varepsilon^2_n$	$s^2(n^2\varepsilon^2_{nj})$
$n =$ 1	2.14	2.14	2.14	2.14	2.14	2.14	2.14	2.14	2.14	2.14	2.14	2.14	2.138	0.000
2	1.81	0.89	6.73	0.61	5.77	2.09	1.33	6.01	4.97	0.77	1.93	1.45	2.865	5.275
3	2.72	1.24	2.12	2.12	3.80	3.88	4.10	2.12	4.36	1.66	3.40	2.44	2.831	1.081
4	1.85	0.89	1.05	1.05	1.61	1.37	1.61	1.85	2.33	1.13	2.01	1.29	1.500	0.199
5	3.55	1.05	3.85	1.15	3.55	2.25	1.45	1.45	1.85	1.85	3.45	1.95	2.281	1.064
6	1.54	1.42	2.54	1.70	2.02	2.54	2.62	2.86	2.06	1.34	1.42	1.70	1.984	0.291
7	0.88	0.56	1.36	0.88	2.32	1.68	1.60	0.88	2.32	0.64	1.36	1.04	1.290	0.355
8	2.67	1.39	1.71	1.71	1.95	2.35	2.43	1.71	2.03	1.79	2.67	1.95	2.026	0.170
9	0.63	1.09	5.85	0.51	5.85	1.13	0.51	5.73	5.77	0.55	1.21	1.13	2.496	6.018
10	3.89	2.69	3.09	3.09	3.29	3.89	3.89	2.69	3.09	3.09	3.89	3.29	3.323	0.210
$^*E^2_{Nj}$	2.17	1.34	3.04	1.50	3.23	2.33	2.17	2.74	3.09	1.50	2.35	1.84	2.274	0.428
$s^2(n^2\varepsilon^2_{nj})$	1.12	0.41	3.63	0.65	2.39	0.86	1.29	3.05	2.02	0.61	0.92	0.47	0.393	.
$\sum_i (\Delta z_{ij})^2/12$	2.00	1.17	2.83	1.33	3.00	2.17	2.00	2.50	2.83	1.33	2.17	1.67	2.083	0.391

Table 2. Analysis of variance of the quantities $n^2\varepsilon^2_{nj}$

Cause of variation	dgr of fr	Q^2	s^2	F
Between the permutations (j)	11	47.087	4.28	3.71^{xxx}
Between the number of observations (n)	9	42.494	4.72	4.09^{xxx}
Rest variance	99	114.209	1.15	

Appendix 2. *The construction of a master sample of local post and telegraph offices in Finland 1950*

Population. The population was composed of 577 clusters of post and telegraph offices. The local principal office and its suboffices formed a cluster. The clusters varied very much in both economic and geographic size. For instance, the expenditures of the clusters ranged from 0.2 millions to 513.3 millions of marks. The distribution of expenditures, as well as many other economic quantities relating to the offices, was found to be nearly lognormal, with standard deviation approximately equal to the mean. Equidistant sampling was performed on the cumulative sums of the expenditures of the offices belonging to clusters.

Sampling. Among the largest offices there were many specialized ones, which only offered either postal or telephone services. This was one reason to stratify the population into two strata. The size of sample was then planned so that the 95%-intervals would be at most $\pm 2\%$ of the estimates to be made. In the following table we give some knowledge describing the population sampled:

Stratum	Population (clusters)	Expenditures % of total expenditures	Sample size
I	42	69%	42
II	535	31%	59
Total	577	100%	101

The first stratum was enumerated completely. The expenditures of the clusters in the first stratum were more than 18 millions of marks. In the second stratum the mean of expenditures was 2.2 millions. Clusters were listed according to the size of expenditures. Equidistant sampling was then performed at the intervals of 20 millions of marks.

Results. The sample was used for economic investigations. In this appendix we only refer to calculations, which were performed to be sure that the risk level stated was correct. We calculated for three different years the estimates for quantities such as: revenues of stamps and franks, total postal revenues, revenues of telephone calls, total revenues, expenditures, etc. The error of estimation was found to be more than 2% in only one case of the 20 cases actually investigated. The estimates concerning only the second stratum, however, were not so exact. The greatest error of estimation observed was 20%. In this stratum the mean value of quotients between the square of the actual error of estimate and the corresponding variance according to the classical formula σ_z^2/n was 0.5. When we in the same way instead of the variance σ_z^2/n used the formulas (40)-(42), we obtained the mean 0.7. If we would have used a simple random sample from a list of the clusters in the second stratum we would have obtained an error variance about 40 times as large as that obtained by using systematic sampling as described.

RESUME

La théorie de l'échantillonnage systématique en grappes, avec répétition, les origines étant choisies au hasard. L'article traite de l'échantillonnage systématique en grappes, pour estimer, par exemple une moyenne pondérée inconnue z_A avec les poids connus $p_{ij} = P_{ij} - P_{i-1,j}$, $\Sigma p_{ij} = 1$.

En définissant, pour chaque nombre entier $n \geqslant 1$ la fonction en escalier $z_j(t, n)$ par:

$$z_j(t, n) = z_{ij} \text{ (où } n P_{i-1,j} < t \leqslant n P_{ij}),$$

le théorème 1 montre que l'estimateur $z_j(\tau \mid n)$, défini par

$$z_j(\tau \mid n) = \frac{1}{n} \sum_{g=0}^{n-1} z_j(\tau + g, n),$$

donne une estimation sans biais de z_A. Le théorème 2 donne la fonction de répartition cumulative de cet estimateur, et le théorème 3 établit diverses formules pour la variance ε_{nj}^2 des estimations $z_j(\tau \mid n)$ de z_A. Ces formules n'ont qu'un intérêt limité si l'on désire obtenir des valeurs numériques pour la variance ε_{nj}^2 ; c'est pourquoi on donne, dans le théorème 4 quelques formules explicites pour cette variance.

Le théorème 5 définit ensuite la valeur moyenne de $n^2\varepsilon_{nj}^2$ pour les différentes séries ordonnées $j = 1, \ldots . N!$ de la base, tandis que le théorème 6 donne une formule pour le valeur moyenne de $n^2\varepsilon_{nj}^2$ relative à la série ordonnée $n = 1, \ldots, \infty$. Enfin des formules permettant l'estimation de la variance d'échantillonnage sont établies. La première annexe donne une description complète de ε_{nj}^2 pour un exemple hypothétique avec $N = 5$ et $n = 1, \ldots 10$. La deuxième annexe concerne la construction d'un échantillon principal des bureaux locaux de poste et de télégraphe en Finlande en 1950.

CHAPTER 9

Minimaxing and Optimal Programming

By LEO TÖRNQUIST*

We shall study the following problems:

I. The minimaxing problem.

Find

$$\tilde{x} = (\hat{x}_J, \check{x}_I) = \{\hat{x}_1, \ldots, \hat{x}_n, \check{x}_{n+1}, \ldots, \check{x}_N\}$$

when

$$W(\tilde{x}) = \max_{x_J} (\min_{x_I} W(x_J, x_I)).$$

II. The optimal programming problem.

Find $\hat{x}_J$ when

$$U(\hat{x}_J) = \max_{x_J} U(x_J); \quad a_i^{(-)} \leqq v_i(x_J) \leqq a_i^{(+)} \qquad \text{for all } i \in I$$

where (;) denotes the expression "under the conditions . . .". The upper bounds $a_i^{(+)}$ and the lower bounds $a_i^{(-)}$ are independent of x_J. The functions v_i are continuous for all x_J.

III. The minimax programming problem.

Find $\tilde{x}$ when

$$W(x) = U(x_J) + \sum_{i \in I} x_i(a_i^{(\operatorname{sign} x_i)} - v_i(x_J)).$$

If $\tilde{x} = (\hat{x}_J, \check{x}_I)$ is a solution to the minimax programming problem (III), $\hat{x}_J$ is a solution to the optimal programming problem (II). When the conditions of the optimal programming problem are fulfilled we have

$$x_i(a_i^{(\operatorname{sign} x_i)} - v_i(x_J)) \geq 0$$

for all $i \in I$ and thus $\min_{x_I} W(x) = U(x_J)$, while $\min_{x_I} W(x) = -\infty$ if some of the conditions

$$x_i(a_i^{(\operatorname{sign} x_i)} - v_i(x_J)) \geq 0$$

are not fulfilled. If $W(\tilde{x}) \neq -\infty$ we have $W(\tilde{x}) = U(\hat{x}_J)$, where

$$U(\hat{x}_J) = \max_{x_J} U(x_J); \quad a_i^{(-)} \leq v_i(x_J) \leq a_i^{(+)}$$

for all $i \in I$.

The minimax programming problem is a special case of the minimaxing problem (I). If we can find a method for solving the general minimaxing problem (I), this method will also give a solution to the optimal programming problem (II).

* University of Helsinki.

Definitions and assumptions.

DEFINITION 1. We call the point $\tilde{x} = (\hat{x}_J, \check{x}_I)$ a minimax point of $W(x)$. The value of $W(x)$ for $x = \tilde{x}$ we denote $\tilde{W}$.

DEFINITION 2. The function $J(h) = 1$ for $h \in J$, $J(h) = -1$ for $h \in I$. We define $I(h) = -J(h)$. If $h \notin I + J$ we put $J(h) = 0$, $I(h) = 0$. We call the functions $J(h)$ and $I(h)$ the partition functions of $I + J$.

DEFINITION 3. The function

$$\check{U}(x_J) = \min_{x_I} W(x_J, x_I) = W(x_J, \check{x}_I(x_J))$$

we call the preference function of the minimaxing problem (I).

The definition 1 can be stated more precisely as follows:

DEFINITION 1′. The point $x = \tilde{x} = (\hat{x}_J, \check{x}_I)$ is a minimax point of $W(x)$ if and only if

1) $x_I = \check{x}_I(x_J)$, that is if $\Delta_{x_I} W(x) \geq 0$ for every Δx_I where

$$\Delta_{x_I} W(x) = W(x_J, x_I + \Delta x_I) - W(x)$$

and if

2) $x_J = \hat{x}_J$, that is if $\Delta_{x_J} \check{U}(x_J) \leq 0$ for every Δx_J where

$$\begin{aligned} \Delta_{x_J} \check{U}(x_J) &= \check{U}(x_J + \Delta x_J) - \check{U}(x) \\ &= W(x + \Delta x) - W(x), \\ \Delta x &= (\Delta x_J, \Delta_{x_J} \check{x}_I(x_J)). \end{aligned}$$

We make the following regularity assumptions about the function $W(x)$:

ASSUMPTION 1. The function $W(x)$ is a continuous real function of x uniquely defined for all real x.

ASSUMPTION 2. The partial derivatives

$$\frac{\partial W(x)}{\partial x_h} = a_h^{(\operatorname{sign} x_h)} - v_h(x), \quad h \in I + J$$

are uniquely determined and continuous for $x_h \neq 0$. The real numbers $a_i^{(+)}$ and $a_i^{(-)}$ can be equal and some of them may be $\pm\infty$.

We make the convention:

$$x_h \frac{\partial W(x)}{\partial x_h} = 0 \qquad \text{for } x_h = 0.$$

Under the regularity assumptions made we have

$$\Delta_x W(x) = \sum_h \frac{\partial W(\dot{x})}{\partial \dot{x}_h} \Delta x_h \qquad \text{for some } \dot{x}$$

fulfilling $0 < |\dot{x}_h - x_h| < |\Delta x_h|$ whenever $x_h(x_h + \Delta x_h) \geq 0$. This formula remains valid for all Δx_h if we in the case when $x_h(x_h + \Delta x_h) < 0$, let

$\frac{\partial W(\overset{1}{x})}{\partial \overset{1}{x}_h}$ for $\overset{1}{x}_h = 0$ denote some suitably chosen mean value of the limes values of this partial derivative for $\overset{1}{x}_h \to 0$, $\overset{1}{x}_h > 0$ and for $\overset{1}{x}_h \to 0$, $\overset{1}{x}_h < 0$.

In the following we restrict if not otherwise stated our studies to functions $W(x)$ fulfilling the regularity assumptions made above.

Theorems.

THEOREM 1. *If the point x has the property $J(h)\frac{\partial W(\overset{1}{x})}{\partial \overset{1}{x}_h}\Delta x_h \leq 0$ for every h and Δx_h and some $\overset{1}{x} = \overset{1}{x}(x, \Delta x)$, $|\overset{1}{x}_h - x_h| < |\Delta x_h|$ such that $\Delta_x W(x) = \sum_h \frac{\partial W(\overset{1}{x})}{\partial \overset{1}{x}_h}\Delta x_h$, then x is a minimax point of $W(x)$.*

PROOF. To prove this theorem it is sufficient to show that $x = \overset{1}{x} + \Delta\overset{1}{x}$, $|\Delta\overset{1}{x}_h| < |\Delta x_h|$ fulfills the conditions of Definition 1′. We consider separately the cases $J(h) = -1$ and $J(h) = 1$. According to Definition 2 $J(h) = -1$ if $h \in I$ and $J(h) = 1$ if $h \in J$. We have according to our assumptions

$$\sum_{h\in I} J(h)\frac{\partial W(\overset{1}{x})}{\partial \overset{1}{x}_h}\Delta x_h = -\sum_{h\in I}\frac{\partial W(\overset{1}{x})}{\partial \overset{1}{x}_h}\Delta x_h = -\Delta_{x_I} W(x) \leq 0$$

for some $\overset{1}{x}$ and all Δx_I and thus $\Delta_{x_I} W(x) \geq 0$, that is the condition 1) of Definition 1′ is fulfilled for x. For $h \in J$ we have $J(h) = 1$ and

$$\begin{aligned}\Delta_{x_J}\check{U}(x_J) &= W(x + \Delta x) - W(x)\\ &= \Delta_{x_J} W(x) + \min_{\overset{1}{x}_I}(W(x_J + \Delta x_J, \overset{1}{x}_I) - W(x_J + \Delta x_J, x_I)).\end{aligned}$$

The first term $\Delta_{x_J} W(x) = \sum_{h\in J} J(h)\frac{\partial W(\overset{1}{x})}{\partial \overset{1}{x}_h}\Delta x_h \leq 0$ according to the assumption of Theorem 1, while the second term $\leq W(x_J + \Delta x_J, x_I) - W(x_J + \Delta x_J, x_I) = 0$. Thus the condition (2) of Definition 1′ is fulfilled.

REMARK. A minimax point fulfilling the conditions of Theorem 1 we call a strong minimax point. A minimax point which is not strong we call a weak minimax point.

We can generalize the concept of a strong minimax point to all real functions $W_R(x)$ by defining a strong minimax point as a point $x = \tilde{x}$ for which $\Delta_{x_J} W_R(x) \leq 0$ and $\Delta_{x_I} W_R(x) \geq 0$ for all $\Delta x = (\Delta x_J, \Delta x_I)$.

If a strong minimax point exists, such a point is thus a common solution to the problem to find $\tilde{x}$ for which

$$W(\tilde{x}) = \max_{x_J}\min_{x_I} W(x_J, x_I)$$

and the problem to find $\bar{\bar{x}}$ for which

$$W(\bar{\bar{x}}) = \min_{x_I}\max_{x_J} W(x_J, x_I).$$

When $\tilde{x} = \bar{\bar{x}}$ and $W(x)$ is uniquely defined, we have $W(\tilde{x}) = W(\bar{\bar{x}})$.

THEOREM 2. *If $\tilde{x}$ is a strong minimax point, we have for $x = \tilde{x}$, either* $x_h = 0$ *or* $\dfrac{\partial W(x)}{\partial x_h} = 0$, *and thus* $x_h \dfrac{\partial W(x)}{\partial x_h} = 0$, *whenever Assumption 2 is valid.*

PROOF. According to the conditions of Theorem 1 $\dfrac{\partial W(\dot{x})}{\partial \dot{x}_h}$ and Δx_h change their signs simultaneously. If $\dfrac{\partial W(\dot{x})}{\partial \dot{x}_h}$ is continuous for $\dot{x} = \tilde{x}$, we must have $\dfrac{\partial W(\dot{x})}{\partial \dot{x}_h} = 0$ for $\dot{x} = \tilde{x}$, while a stepwise change in $\dfrac{\partial W(\dot{x})}{\partial \dot{x}_h}$ can occur only for $\dot{x}_h = \tilde{x}_h = 0$. Theorem 2 is thus proved.

THEOREM 3. *For a strong minimax point we have when Assumption 2 is valid*

$$a_h^{(J(h))} \leq v_h(\tilde{x}) \leq a_h^{(I(h))}$$

and

$$\tilde{x}_h(a_h^{(\operatorname{sign} \tilde{x}_h)} - v_h(\tilde{x})) = 0.$$

PROOF. Substituting in $J(h) \dfrac{\partial W(\dot{x})}{\partial \dot{x}_h} \Delta x_h \leq 0$ the expression of $\dfrac{\partial W(x)}{\partial x_h}$ given in Assumption 2 and using the assumptions about the regularity of $v_h(x)$ we obtain after some calculation the inequalities of Theorem 3.

THEOREM 4. *If the function $W(x)$ fulfills the assumptions required for the mean value theorem, that is if*

$$\Delta_{x_k} W(x) = \frac{\partial W(\dot{x})}{\partial \dot{x}_k} \Delta x_k \qquad \text{where } \dot{x} = x + \delta_{x_k} x,\ |\delta x_k| < |\Delta x_h|,$$

$\Delta x_k \delta x_k \geq 0$
for every Δx_k, $k \in J + I$, we have

$$J(k)\, \Delta_{x_k} W(x) = J(k) \frac{\partial W(\dot{x})}{\partial \dot{x}_k} \Delta x_k \leq 0$$

for every $k \in I$ if $x_I = \check{x}_I(x_J)$
and for every $k \in J$ if $x_J = \hat{x}_J$ and $x_I = \check{x}_I(\hat{x}_J + \Delta_{x_k} x_J)$.

REMARK. The minimax points

$$\tilde{x}^{(s_k,k)} = \{\hat{x}_J, \check{x}_I^{(s_k,k)}\}$$

where $s_k = \operatorname{sign} \Delta x_k$; $|\Delta x_k| \to 0$ and

$$\check{x}_I^{(s_k,k)} = \operatorname*{limes}_{|\Delta_k x| \to 0} (\check{x}_I(x_J + \Delta_{x_k} x_J)); \quad s_k = \operatorname{sign} \Delta x_k$$

will not necessarily be equal for $s_k = +$ and $s_k = -$. For all minimax points $\tilde{x}^{(s_k,k)}$ we have

$$W(\tilde{x}^{(s_k,k)}) = \tilde{W} = \max_{x_J} \min_{x_I} W(x_J, x_I) = U(\hat{x}_J)$$

and

$$J(k) s_k \left[\frac{\partial W(x)}{\partial x_k}\right] \leqq 0 \quad \text{for } x = \tilde{x}^{(s_k,k)}.$$

If $\tilde{x}^{(+,k)} = \tilde{x}^{(-,k)}$: we have

$$\tilde{x}_k^{(s_k,k)}\left[\frac{\partial W(x)}{\partial x_k}\right] = 0 \quad \text{for } x = \tilde{x}^{(s_k,k)}.$$

For $\tilde{x}^{(+,k)} \neq \tilde{x}^{(-,k)}$ this equation need not be fulfilled.

For every $h \in I$ and every minimax point $\tilde{x}$ we must as proved in connection with Theorem 2 have

$$\check{x}_h\left[\frac{\partial W(x)}{\partial x_h}\right]_{x=\tilde{x}} = 0.$$

PROOF. For a minimax point (x_J, x_I) we have according to Definition 1′ and Definition 3 $\Delta x_k W(x) \geq 0$ for $k \in I$, $J(k) = -1$, $x_I = \check{x}_I(x_J)$ and

$$\Delta_{x_k}\check{U}(x_J) = \Delta_{x_k} \min_{x_I} W(x_J, x_I) \leq 0.$$

We always have for $x = \tilde{x}$, $\Delta_{x_k} x_I = \check{x}_I(x_J + \Delta_{x_k} x_J) - \check{x}_I(x_J)$, $J(k) \neq +1$

$$\begin{aligned}[\Delta_{x_k}\check{U}(x_J)]_{x_J=\hat{x}_J} &= [W(\hat{x}_J + \Delta_k x_J, \check{x}_I + \Delta_{x_k} x_I) \\ &\quad - W(\hat{x}_J, \check{x}_I + \Delta_{x_k} x_I)] \\ &\quad + [W(\hat{x}_J, \check{x}_I + \Delta_{x_k} x_I) - W(\hat{x}_J, \check{x}_I)] \leq 0,\end{aligned}$$

where

$$W(\hat{x}_J + \Delta_k x_J, \check{x}_I + \Delta_{x_k} x_I) = \check{U}(\hat{x}_J + \Delta_k x_J) = \min_{x_I} W(\hat{x}_J + \Delta_k x_J, x_I),$$

and $W(\hat{x}_J, \check{x}_I + \Delta_{x_k} x_I) - W(\hat{x}_J, \check{x}_I) \geq 0$ since $W(\hat{x}_J, \check{x}_I) = \check{U}(\hat{x}_J) = \min_{x_I} W(\hat{x}_J, x_I)$. We thus have

$$W(\hat{x}_J + \Delta_k x_J, \check{x}_I + \Delta_{x_k} x_I) - W(\hat{x}_J, \check{x}_I + \Delta_{x_k} x_I) \leq 0.$$

Applying the mean value theorem to this result we obtain Theorem 4.

THEOREM 5. *The point $\tilde{x}$ is a minimax point of $W(x)$ if and only if*

$$\tilde{W} = W(\tilde{x}) \geq W(x)$$

whenever $W(x_J, x_{\dot{I}} + \Delta x_I) \geq W(x)$ for every Δx_I.

This theorem is only a reformulation of Definition 1′.

THEOREM 6. *If $\underline{W}(W(x_J(y_j), x_I(y_j, y_{\dot{I}}))) = \overline{W}(y_j, y_{\dot{I}})$ where $\underline{W}(W)$ is a strictly increasing function of W and $y_j, y_{\dot{I}}$ are such new independent variables that to every x_J corresponds some y_j for which $x_J = x_J(y_j)$ and to every x_I and y_j corresponds some $y_{\dot{I}}$ for which $x_I = x_I(y_j, y_{\dot{I}})$, and $\{\hat{y}_j, \check{y}_{\dot{I}}\}$ is a minimax point of $\overline{W}(y_j, y_{\dot{I}})$, then $\{\tilde{x}_J = x_J(\hat{y}_j), \check{x}_I = x_I(\hat{y}_j, \check{y}_I)$ is a minimax point of $W(x_J, x_I)$.*

PROOF. Since $\underline{W}$ is a strictly increasing function of W, both $\underline{W}$ and W must have the same minimax points. Thus if $\tilde{x}$ is a minimax point of W it is also a minimax point of $\underline{W}$. Since the dependent variables $\{x_J, x_I\}$ take

all values when the new independent variables vary, we have

$$\begin{aligned}\overline{W}(\hat{y}_J, \check{y}_I) &= \max_{y_J} \min_{y_I} \overline{W}(y_J, y_I) \\ &= \max_{y_J} \min_{y_I} \underline{W}(W(x_J(y_J), x_I(y_J, y_I))) \\ &= \underline{W}(W(x_J(\hat{y}_J), x_I(\hat{y}_J, \check{y}_I))) \\ &= \underline{W}(W(\hat{x}_J, \check{x}_I)) = \max_{x_J} \min_{x_I} \underline{W}(W(x_J, x_I)).\end{aligned}$$

REMARK. By a suitable transformation ($x_J = x_J(y_I)$, $x_I = x_I(y_J, y_I)$) it is possible to simplify the problem of finding minimax points and avoid difficulties connected with minimax points for which some coordinates of x or $W(x)$ are infinite.

General discussion of possibilities of finding minimax points.

In principle we could try to find the solutions of the system of equations and inequalities mentioned in the Theorems 1–4. Every such solution fulfilling the conditions of Definition 1′ is then a solution to our minimax problem. This method for finding the solution of our minimaxing problem we call the direct method.

We could try to solve our problem also by making use of such transformations of the minimaxing problem to another minimaxing problem that the new minimaxing problem is easier to solve by means of the direct method. If the transformation used fulfills the conditions of Theorem 6, we can from the solution of the transformed problem obtain the solution of the original problem. Methods using a finite number of transformations we call transformation methods. If the method needs in principle an infinite number of transformations for obtaining the solution we call it a sequential approximation method. The rules required for choosing the successive transformations are called transition rules. A method with given transition rules usually leading to a minimax point we call a minimax algorithm. In the following we suggest some minimax algorithms which may be useful in practice.

Linear minimax algorithms.

Assuming that strong minimax points exist we can try to use linear minimax algorithms for finding these points.

In such an algorithm the transform rules are of the following form

$$x_h = b_{h0}^{(\nu)} + \sum_{p \in P_{h\nu}} b_{hp}^{(\nu)} y_p^{(\nu)}$$

where ν is the ordering number of the last transition made and $h \in J + I$.

The function to be minimaxed is denoted by $W(x_J, x_I)$ and the transformed minimax function by

$$W_\nu(y_{J_\nu}^{(\nu)}, y_{I_\nu}^{(\nu)}) = W(x_J, x_I).$$

The transition $\nu \to \nu + 1$ is defined by the formulas

$$y^{(0)} = x.$$
$$y_p^{(\nu)} = a_{p0}^{(\nu)} + \sum_{q \in Q_p} a_{pq}^{(\nu)} y_q^{(\nu+1)}$$

where $a_{pq}^{(\nu)} = 0$ if $q \in J_{\nu+1}$ and $p \in I_{\nu+1}$ and $a_{pp}^{(\nu)} \neq 0$, $p \in Q_{p\nu}$.

These substitutions fulfill the conditions of Theorem 6. We obtain the following transition rule for the coefficients $b_{ho}^{(\nu)}$ and $b_{hp}^{(\nu)}$

$$b_{h0}^{(\nu+1)} = b_{h0}^{(\nu)} + \sum_{q \in P_{h\nu}} b_{hq}^{(\nu)} a_{q0}^{(\nu)}$$
$$b_{hp}^{(\nu+1)} = \sum_{q \in P_{h\nu}} b_{hq}^{(\nu)} a_{qp}^{(\nu)}, \; P_{h,\nu+1} = Q_{h\nu}$$

To obtain a completely specified minimax algorithm we have to define the numbers $a_{po}^{(\nu)}$, $a_{pq}^{(\nu)}$. The rules for determining these numbers must be somehow related to the partial derivatives of first and second order of $W_\nu(y_{J_\nu}, y_{I_\nu})$ for $y_{J_\nu} = y_{I_\nu} = 0$. By constructing these rules the following guiding principles may be useful.

PRINCIPLE 1. If we can prove that in a minimax point of $W_\nu(y_{J_\nu}, y_{I_\nu})$, $\tilde{y}_h = 0$ for some h, we put $y_h = 0$ in the following steps thus reducing the number of independent variables.

PRINCIPLE 2. If it seems to be likely that $\tilde{y}_h^{(\nu)} = 0$ for some h, we use in the following step the transform $y_h^{(\nu)} = y_h^{(\nu+1)}$ for these h.

PRINCIPLE 3. If it seems to be likely or if it can be proved that $\tilde{y}_h^{(\nu)} \neq 0$, $\tilde{x}_h \neq 0$ for $h \in H_\nu$ and consequently according to the remark of Theorem 4 we expect or otherwise can prove that $\dfrac{\partial W_\nu(y^{(\nu)})}{\partial y_h^{(\nu)}} = 0$ for $h \in H_\nu$, $y^{(\nu)} = \tilde{y}$, we have to use such a transform for $\nu \to \nu + 1$ that $\dfrac{\partial W_{\nu+1}(y^{(\nu+1)})}{\partial y_h^{(\nu+1)}} \approx 0$ for $h \in H_\nu$ when $y_h^{(\nu+1)} \approx 0$.

PRINCIPLE 4. Taking into account Principles 1–3 we try to minimize the work to be done in finding sufficiently good approximations $b_{ho}^{(\nu)}$ for the minimax values $\tilde{x}_h$ sought.

The crucial problem is to determine the sets H_ν mentioned in Principle 3. Algorithms for finding such sets for linear programming problems are presented in my paper "Some New Principles for Solving Linear Programming Problems" published in *Bulletin de L'Institut International de Statistique*, Tome 36-3. Livraison (pp. 335–356).

AN ATTEMPT TO ANALYZE THE PROBLEM OF AN ECONOMICAL PRODUCTION OF STATISTICAL DATA

By *Leo Törnqvist**).

The problem of rationalizing the production of statistical data is one of the most difficult of all the rationalization problems which anyone who sets out to apply the basic principles of economics to various undertakings can encounter. In such work it has to be determined what investigations may be regarded as profitable, what degree of reliability should be aimed at in the investigations decided upon, and in what manner they are to be carried out in order that a reasonably estimated net yield of the work of statistical production may be as large as possible. If we consider that the magnitudes of the figures required to solve this problem satisfactorily are unknown except as a more or less subjectively determined, approximate value, the difficulty is fully realized. But although in practice any altogether satisfactory solution cannot be hoped for, it appears to me to be worth the trouble of trying to analyze the problem, to the extent that this is possible.

Let us first take the simple case in which it is desired, by means of random samples, to find an approximate value x_k for a figure x which in some way characterizes a definite statistical parent population. The utility value y_k of this approximate value x_k depends on the form of the distribution function which indicates the probability $P(u_k \leq u)$ that the error in the rough value x_k, *i. e.*

$$u_k = x_k - x, \tag{1}$$

will be less than or equal to u. This probability distribution function we will call

$$F_k(u) = P(u_k \leq u). \tag{2}$$

The suffix k in the formulæ (1) and (2) we will interpret in the following as an indication of the growth of the costs K of the statistical data production

$$k = \underset{x_k}{\Delta} K \tag{3}$$

*) University of Helsinki, Finland.

caused and/or expected to be caused by the determination of x_k. The figure k may be assessed with relative exactness at the time the rough value x_k is made known. For an exact assessment of k there should also be included an item which measures the value of the services rendered without cash remuneration by the persons providing the information, also the interest on disbursements made prior to the date of the evaluation and the discounted value of the future disbursements caused by the determination of the rough value.

The utility value y_k of the investigation result x_k under consideration denotes the increase $\underset{x_k}{\Delta} Y$ in the value Y of the available statistical data as a result of the ascertaining of the figure x_k. This figure should be evalued at the same time and in the same measuring unit as the corresponding costs k.

$$y_k = \underset{x_k}{\Delta} Y\,(dF_k\,(u)) = \int_{-\infty}^{\infty} y\,(u)\,dF_k\,(u) + \cdots \tag{4}$$

The figure y_k, as we have tried to show in formula (4), is a function of the form of the distribution function $F_k(u)$ for the error $u_k = x_k - x$. By developing y_k serially according to the power of the growths $dF_k(u)$ we find that y_k in the first approximation may be calculated as the average value as measured by the standards $dF_k(u)$ of a function

$$y\,(u) = \left(\frac{\delta y_k}{\delta\,(dF_k\,(u))} + \tilde{y}\right). \tag{5}$$

The figure $y\,(u)$ may be interpreted as the utility value which may be ascribed to an approximate value x_k with the property $x_k - x = u$. The evaluation function $y\,(u)$ attains its maximum value when $u = 0$, *i. e.* when $x_k = x$.

The gain function whose maximization interests us we define as the difference

$$v_k = y_k - k. \tag{6}$$

In order to obtain an idea of the extent to which v_k depends on k we must study a few special assumptions concerning the function $y\,(u)$.

Ex. 1. Assume that $y\,(u) = y\,(0)\,(1 - u^2/s^2)$.

By calculating the mathematical expectancy of $y\,(u)$, we find

$$y_k = y\,(0)\left(1 - \frac{\sigma_k^2}{s^2}\right), \tag{1,1}$$

where

$$\sigma_k^2 = \int_{u=-\infty}^{\infty} u^2 \, dF_k(u) \tag{1,2}$$

denotes the variance of the error $u_k = x_k - x$.

For further progress we must make the most plausible assumption possible with regard to the manner in which the variance σ_k^2 may be expected to depend on the costs k. In doing so we might rely on the law of large figures, according to which the variance in an approximate value which is calculated on the basis of random samples declines in almost inverse proportion to the size of random sample n_k. The size of the material depends in turn almost linearly on the costs $k = k_0 + h \cdot n_k$. It thus seems plausible that the variance σ_k^2 should be produced with relative exactness as a broken linear function of k of the formula:

$$\sigma_k^2 = \frac{\sigma^2}{n_k} + \epsilon\,\sigma^2 = \frac{\sigma^2 \cdot h}{k - k_0} + \epsilon\,\sigma^2; \quad \sigma^2 = \sigma_{k_0+h}^2(1-\epsilon); \quad \sigma_\infty^2 = \epsilon\,\sigma^2. \tag{1.3}$$

In this formula k_0 denotes the so-called fixed costs, which do not depend on the bulk of the primary material, and h the supplementary costs per studied statistical unit.

In theory, for random samples from an unlimited parent population, σ^2 is interpreted as the variance of the distribution function which denotes how the units of the parent population are distributed according to the variable x under consideration. The figure ϵ is mostly, in theory, set $= 0$. If the parent population consists of a limited number N of units, $\epsilon = -1/N$. In practical statistics it is simpler to regard the formula (1,3) as an equalizing formula for the series of variances which would be obtained if we calculated the variance in the figure $x_{k\lambda}$ for a series of approximate values $x_{k\lambda}$, calculated on the basis of various primary material whose volume is determined by the fact that the determination of the rough values $x_{k\lambda}$ all call for equally large costs k. Thus it would usually prove that the figure ϵ is a *positive* magnitude that is not altogether negligible. For in practice, in the study of statistical units and the working up of the primary material, it is not altogether possible to avoid a number of errors creeping in, with the effect that not even an investigation which in principle sets out to examine every unit in the parent population gives a wholly reliable result. If we try to observe statistical populations which change with time, the remainder term $\epsilon\,\sigma^2$ (especially if the attribute x in question may vary continually) may be a relatively large figure. If a so-called complete investigation is carried out with the aid of inadequately trained personnel, the variance in the final result may even be greater than with a thoroughly executed examination of random samples on a substantially smaller material.

When we introduce the expression (1,3) into the formula (1,1), we obtain for the net gain v_k the expression

$$v_k = y(0)\left(1 - \frac{\sigma^2}{s^2}\left(\frac{h}{k - k_0} + \epsilon\right)\right) - k. \tag{1,4}$$

This expression attains its maximum value with variations of k when

$$\frac{\partial v_x}{\partial k} = y(0)\frac{\sigma^2}{s^2}\frac{h}{(k - k_0)^2} - 1 = 0. \tag{1,5}$$

When we solve the equation (1,5) for k we obtain the optimal value $\hat{k}$ for the costs k.

$$\hat{k} = k_0 + \frac{\sigma}{s}\sqrt{hy(0)}. \tag{1,6}$$

The extent of the primary material when $k = \hat{k}$ is obtained by the formula

$$n_{\hat{k}} = \frac{\sigma}{s}\sqrt{\frac{y(0)}{h}}. \tag{1,7}$$

The variance $\sigma_{\hat{k}}^2$ is then obtained according to (1,3), and is thus

$$\sigma_{\hat{k}}^2 = \epsilon\,\sigma^2 + \sigma\,s\sqrt{\frac{y(0)}{h}}. \tag{1,8}$$

The maximum attainable gain is thus, in this example,

$$v_{\hat{k}} = y(0) - k_0 - \epsilon\frac{\sigma^2}{s^2}y(0) - 2\frac{\sigma}{s}\sqrt{h \cdot y(0)}. \tag{1,9}$$

The more rapidly the evaluation function $y(u)$ approaches zero with increasing errors u, the smaller is the figure s. The figure s denotes the margin of error beginning with which the evaluation function $y(u)$ becomes negative; $y(u) < 0$ for $u > s$. As s diminishes, the optimal costs increase. The optimal variance diminishes linearly with s. The maximum attainable net gain declines with diminishing s, *i. e.* when the demand for exactness rises.

Ex. 2. We assume in this example that $y(u) = e^{-u^2/s^2}y(0)$. To obtain a simple expression for y_k we study only the special case where the distribution function of the error is normal $F_k(u) = \Phi\left(\frac{u}{\sigma_k}\right)$. In this case we get

$$y_k = y(0) \int_{-\infty}^{\infty} e^{-\frac{u^2}{s^2}} \, d\Phi\left(\frac{u}{\sigma_k}\right) = \frac{1}{\sqrt{2\pi}\,\sigma_k} \int e^{-\frac{u^2}{2}\left(\frac{2}{s^2}+\frac{1}{\sigma_k^2}\right)} du = \frac{s}{\sqrt{s^2+2\sigma_k^2}}, \quad (2,1)$$

whereby the net gain

$$v_k = y(0)\frac{s}{\sqrt{s^2+2\sigma_k^2}} - k = y(0)\frac{1}{\sqrt{1+2\kappa_k}} - h; \quad \kappa_k = \frac{\sigma_k^2}{s^2} = \frac{\sigma^2}{s^2}\left(\frac{h}{k-k_0}+\epsilon\right). \quad (2,2)$$

By determining the value of k which renders this expression a maximum, we find that:

$$\hat{k} = k_0 + \Theta_{\hat{k}}\frac{\sigma}{s}\sqrt{h\,y(0)}; \quad \Theta_{\hat{k}} = \frac{1}{(1+2\kappa_{\hat{k}})^{\frac{3}{4}}} \approx 1 - \frac{3}{2}\frac{\sigma^2}{s^2}\left(\frac{h}{\hat{k}-k_0}+\epsilon\right) < 1. \quad (2,3)$$

i. e. the optimal costs in this case are less than in Example 1. The corresponding variance

$$\sigma_{\hat{k}}^2 = \epsilon\sigma^2 + \frac{1}{\Theta_{\hat{k}}}\,\sigma\cdot s\sqrt{\frac{h}{y(0)}} \quad (2,4)$$

is greater than in Ex. 1.

The maximum attainable net gain is also greater than in the example first studied. The net gain

$$v_{\hat{k}} = y(0) - k_0 - \epsilon\frac{\sigma^2}{s^2} - 2\,\Theta_k\frac{\sigma}{s}\sqrt{h\,y(0)} \quad (2,5)$$

exceeds the corresponding gain in Ex. 1, on the assumption, of course, that the figures $y(0)$, k_0, σ, h and s have the same figure value in both the cases under consideration. The result rests on the fact that in this case the evaluation function $y(u)$ diminishes to zero at a slower rate than in Ex. 1.

We will now study an extreme, special case in regard to the form of the evaluation function.

Ex. 3. We consider the case where the utility value $y(u)$, of the rough value x_k is equal to $y(0)$, if this approximate value x_k is on the same side of a critical value x_0 as the figure x. If the figure x_0 lies between the figure x_k and x, we assume that the utility value of the approximate value obtained is negative $= -\beta\,y(0)$ or that $F_k(u) = \Phi\left(\frac{u}{\sigma_k}\right)$.

We introduce the term

$$y = y(0)(1+\beta) \quad (3,1)$$

for the value by which a "correct answer" exceeds the value of an

"incorrect answer", to the question whether the figure x looked for exceeds or falls short of the critical figure x_0.

The figure value of the difference $x_k - x_0$ we will denote s_k.

$$s_k = | x_k - x_0 |. \tag{3,2}$$

The evaluation function $y(u)$ in this case may be expressed by the formula

$$y(u) = y\,\Phi\left(\frac{(x_k - x_0)\ \ (x_k - x_0 - u)}{0}\right) - \beta\, y(0). \tag{3,3}$$

For the mathematical expectancy of $y(u)$ we find the express:

$$y_k = y \cdot \Phi\left(\frac{s_k}{\sigma_k}\right) - \beta\, y(0), \tag{3,4}$$

and for the net gain v_k the formula:

$$v_k = y_k - k = y \cdot \Phi\left(\frac{s_k}{\sigma_k}\right) - \beta\, y(0) - k. \tag{3,5}$$

In this expression, s_k is a statistical variable and not a constant, for which reason an exact determination of the optimal costs is not possible before the completion of the examination. The figure s_k may, however, be understood as being independent of the costs k in the sense that there is no systematic fluctuation of s_k with increasing k. We could therefore maximize v_k under the assumption that $\frac{\partial s_k}{\partial k} = 0$. By partial derivation in respect of the costs k we find as a condition of k being at least approximately optimal:

$$\frac{\partial v_k}{\partial k} = -y\,\phi\left(\frac{s_k}{\sigma_k}\right)\frac{s_k}{\sigma_k^3}\,\frac{1}{2}\frac{\partial \sigma_k^2}{\partial k} - 1 = 0; \quad \phi\left(\frac{s_k}{\sigma_k}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{s_k^2}{\sigma_k^2}}. \tag{3,6}$$

Calculating with the aid of formula (1,3) $\frac{\partial \sigma_k^2}{\partial k}$, we obtain the formula

$$\frac{\partial v_k}{\partial k} = \frac{y}{2} \cdot \left(\frac{s_k}{\sigma_k}\right)^3 \phi\left(\frac{s_k}{\sigma_k}\right) \cdot \frac{\sigma^2}{s_k^2} \cdot \frac{h}{(k - k_0)^2} - 1 = 0. \tag{3,7}$$

From (3,7) follows

$$\hat{k} = k_0 + \Theta_{\hat{k}} \cdot \frac{\sigma}{s_{\hat{k}}} \sqrt{hy}, \tag{3,8}$$

where the figure

$$\Theta_{\hat{k}} = \left(\frac{s_{\hat{k}}}{\sigma_{\hat{k}}}\right)^{\frac{3}{2}} \frac{e^{-\frac{1}{4}\frac{s_{\hat{k}}^2}{\sigma_{\hat{k}}^2}}}{\sqrt[4]{8\pi}} \leqq \sqrt[4]{\frac{27}{8\pi e^3}} \approx 0.48. \tag{3,9}$$

The figure Θ_k reaches its maximum value when $\sigma_k^2 = \frac{1}{3} s_k^2$. From (3,8) and (3,9) follows

$$s_{\hat{k}} (\hat{k} - k_0) \leq 0.48 \cdot \sigma \sqrt{h y}. \tag{3,10}$$

This formula may be employed to advantage to determine the point at which the collection of new material may no longer be regarded as profitable. In the event, however, of the figure $s_k = x_k - x_0$ being very small relative to σ, another formula is of greater value. If we set down the conditions for the determination of the optimal costs (3,6) in the form

$$\tfrac{1}{2} y \frac{s_k}{\sigma_k} \phi\left(\frac{s_k}{\sigma_k}\right) \cdot \frac{\frac{\delta \sigma_k^2}{\delta k}}{\sigma_k^2} - 1 = 0, \tag{3,11}$$

and introduce $\sigma_k^2 = \dfrac{\sigma^2 h}{k - k_0} + \epsilon, \quad \dfrac{\delta \sigma_k^2}{\delta k} = -\dfrac{\sigma^2 h}{(k - k_0)^2},$

we find that the optimal costs $\hat{k}$ also satisfies the equation

$$(k - k_0) + \epsilon \frac{(k - k_0)^2}{h} = \tfrac{1}{2} \frac{s_k}{\sigma_k} \phi\left(\frac{s_k}{\sigma_k}\right) \cdot y. \tag{3,12}$$

If we assume that $\epsilon \geq 0$, it follows from (3,12) that

$$k = k_0 + \grave{\Theta}_{\hat{k}}\, y - \epsilon \frac{(k - k_0)^2}{h} < k_0 + \grave{\Theta}_{\hat{k}}\, y, \tag{3,13}$$

where the figure

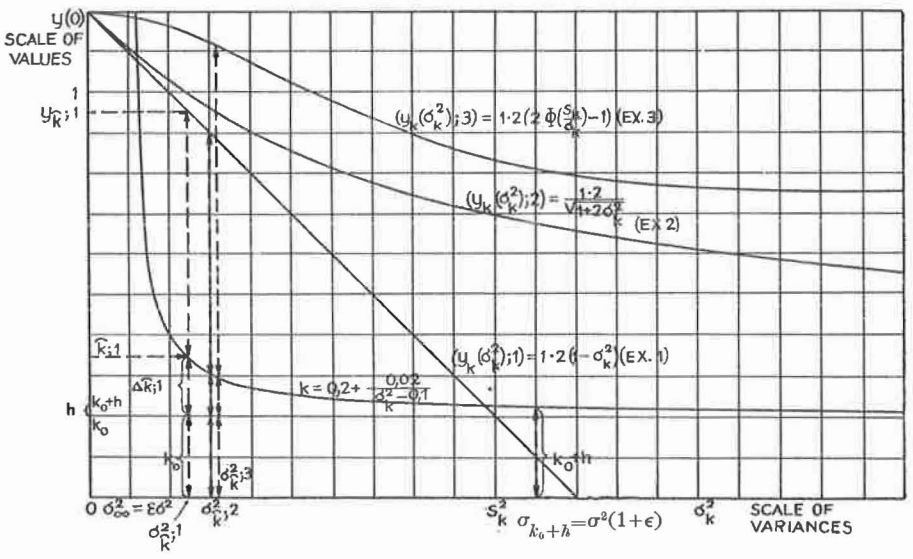
$$\grave{\Theta}_{\hat{k}} = \tfrac{1}{2} \frac{s_{\hat{k}}}{\sigma_{\hat{k}}} \phi\left(\frac{s_{\hat{k}}}{\sigma_{\hat{k}}}\right) \leq \tfrac{1}{2} \phi(1) = \frac{1}{2\sqrt{2\pi e}} \approx 0.121. \tag{3,14}$$

If $\sigma_{\hat{k}}^2 \approx s_{\hat{k}}^2$, we obtain from the formula

$$\hat{k} \approx k_0 + 0.1\, y - 0.01 \frac{\epsilon y^2}{h} < \hat{k}_0 + 0.121\, y; \quad \epsilon > 0, \tag{3,15}$$

a good approximate value for the optimal costs. By utilising the formula (3,10) and (3,15) it is possible to work out a relatively convenient and economically rational method (a form of sequential analysis) answering the question whether the unknown figure x is greater or less than a critical figure x_0.

In order to obtain a better general view of the contents of the figure magnitudes and functions discussed here we have set out three special cases of the discussed examples in diagram form.



Space will not permit of a more detailed analysis of the case where there is a simultaneous interest in obtaining a whole table of approximate values. We merely point out that the formulæ here might easily be varied so as to indicate how the problem can be solved even in more complicated cases than the ones considered here. If the maximum net gain $v_{\hat{k}} = \underset{x\hat{k}}{\Delta V} = \underset{x\hat{k}}{\Delta (Y-K)}$, which usually depends on what other figures it is decided to include in the table, is negative, it is advantageous to abstain from any idea of determining an approximate value x_k for the figure x.

To give some indication of the manner in which the formulæ given here, when the posing of the problem is generalized, yield to new, similar formulæ, it may be said that the most important formula in the theory of stratified sampling from the point of view of profitability (viz. the formula which gives the amount of the primary material which

should be collected from the various strata of the parent population), represents a special case of the formula (1,7). If $x = \Sigma\sigma_\lambda z_\lambda$, and the only motive for calculating the approximate value of the figure z_λ is to facilitate a calculation of the figure x, we should choose the number of statistical units $n_{\lambda\hat{k}}$, taken at random from the strata number λ, so that:

$$n_{\lambda\hat{k}} = \frac{\sigma(c_\lambda z_\lambda)}{\Theta_{\hat{k}} \cdot s} \sqrt{\frac{y(0)}{h_\lambda}} = \frac{c_\lambda \cdot \sigma(z_\lambda)}{\Theta_{\hat{k}}\, s} \sqrt{\frac{y(0)}{h_\lambda}}; \quad h_\lambda = \frac{\partial k}{\partial n_\lambda}, \qquad (7)$$

i. e. proportionally with the product of the standard deviation for the statistical variable z_λ and this figure's measurement coefficient in sum expression which determines the figure x wanted — and *vice versa* proportionally with the square root of the supplementary costs per statistical unit, if the number of studied units from the strata is increased. The total number of units $n_{\hat{k}} = \Sigma n_{\hat{k}\lambda}$ increases proportionally with the square root of the value $y(0)$ attributable to an exact determination of the figure x.

It may be worth pointing out that the utility value $y(0)$ of an exact determination of a statistical magnitude varies with the time of the result. This fact greatly complicates the problem of finding in practice an optimal solution to the problem of rationalizing the production of statistical data. For it has to be borne in mind that the figure $y(0)$ may very quickly diminish if the time of announcing the result is advanced. If, on the other hand, an attempt is made to press forward with the work, then the supplementary costs per studied statistical unit begin to increase rapidly as we approach the shortest limit of time in which it is at all possible to answer the statistical problems posed.

As it is a question of solving the method-technical problem of selecting from several available methods the one which gives the maximum net gain, it is clear that we can disqualify such methods as, with an equally great or greater ϵ, have both higher fixed costs k_0 and higher supplementary costs h per studied unit than any other known method. In order to make the right choice among the methods that are characterized by having relatively greater fixed costs k_0 counteracted by relatively lower supplementary costs h (or vice versa), it is necessary to estimate the corresponding net gains $v_{\hat{k}}$ before it is possible to form any well-founded idea as to which method may be regarded as the best for the purpose in question.

The question of minimizing the variance with a constant n_k and/or k, that has been rather thoroughly treated in mathematical-statistical literature (Cochran, Mahalanobis, Neyman, Wilks, Wald, Yates, et al.), and the corresponding inverse question of minimizing k with a previously

fixed variance, might be regarded as extreme, special cases of the above-mentioned problem of maximizing $V = Y - K$. The solution of these problems might be utilized as a means to finding the solution of the problem of maximizing V.

To carry out in practice an analysis of the profitability of the production of statistical data, according to the principles sketched above, is of course very difficult. It is not possible to avoid rather rough estimates and subjective evaluations. But if the views set forward here can contribute, at least in some cases, to increasing the probability that the practical solutions to the problems obtained become approximately optimal, then the object of this article will have been achieved.

HOW TO COPE WITH THE URGENT NEED FOR STATISTICAL SERVICES IN DEVELOPING COUNTRIES.

by

Leo Törnqvist, Professor,

University of Helsinki, Finland.

1. The importance of statistical services for developing countries.

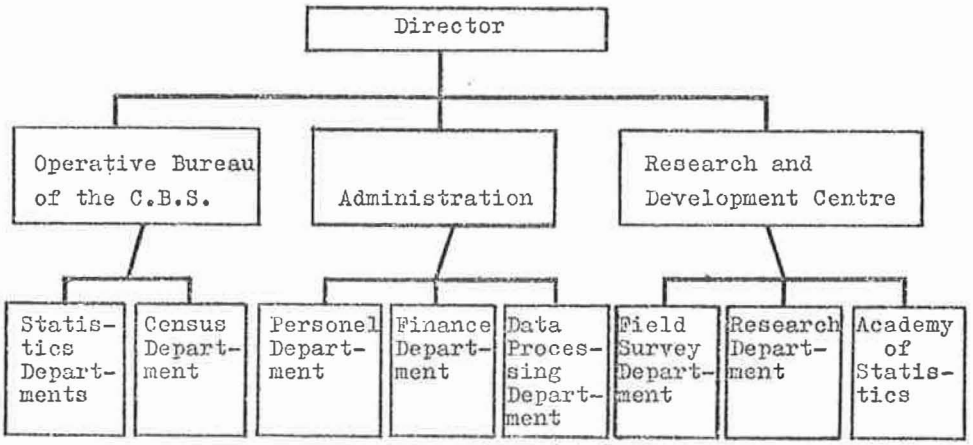
When working as a UN-expert in Indonesia in the years 1960-61, I saw that in developing countries, there really is an urgent need for establishing and improving many kind of economic, cultural and social statistics. Relevant and reliable quantitative knowledge about the development of the economic situation and the social and cultural structure of the country are necessary, both for the government and for the public and private enterprises, in their efforts to make more efficient plans for the future, and for observing if their policies really have influenced the development in the desired direction and for anticipating the consequencies of different competing plans.

To meet this urgent need for statistical information, we must first educate a sufficiently large number of statisticians who can produce the information needed and can correctly interprete the statistical information available. Further, the trained statisticians must be effectively organised to get the new data needed as rapidly as possible. This will require that a statistical research and development centre is built up, in which the best trained statisticians work as a team aiming to function as a dynamic force, doing the hard work to get the new activities started on a sound base and analyzing the results obtained. This centre must also be

responsible for the training of newcomers in the statistical field. During the first five to ten years it will be necessary to get experts from abroad to build up such a centre because the few outstanding statisticians available in these countries are already overloaded with administrative work. These statisticians will need the centre as a brainstrust for working out the "blue prints" of the new activities.

2. The organization plan for the Central Bureau of Statistics in Indonesia.

The following organization chart shows how the Research and Development Centre according to the request approved by U.N. Special Fund has been related to the Central Bureau of Statistics in Indonesia



This type of organization is, I think, a good model also for other countries. It shows the main tasks of a statistical organization for a country. It was developed by a comittee planning how the organization of the Bureau could be strengthened so that it would become an effective tool for producing the statistical information needed.

It can be mentioned that the Data Processing Department has to assist the departments of the Operative Bureau and the Centre. The experts in data processing will also have the duty to teach in the

Academy.

The building of the Centre will require international assistance. The need of foreign experts during the building of the Centre is seen in the following scheme.

Area of specialization, duration and timing of expert assistance.

Statistical Activity	first year	second year	third year	fourth year	fifth year
1. Academy of Statistics	x	x	x	x	x
2. National Income	x	x	x	x	
3. Trade and Industry	x	x	x	x	
4. Agriculture	x	x			
5. Vital Statistics, Labour and Manpower	x	x			
6. Transportation and Communication	x	x			
7. Prices and Finance	x	x	x	x	
8. Field Organization	x	x	x		
9. Data processing	x	x	x	x	
Total number of experts engaged	9	9	6	5	1

The qualification required from the experts is a thorough knowledge in the field of specialization. The senior advisors in statistical studies and research and the expert in national income statistics should have doctorates according to the job descriptions. All the experts will have the duty of contributing to the training programme of the Academy of Statistics.

3. The tasks of the Statistical Research and Development Centre in Indonesia.

The Centre consists of three departments called the Field Sur-

vey Department, the Research Department and the Academy of Statistics. These names refer to the three main tasks of the Centre: 1) To plan how to obtain new data from fields not earlier covered by statistical information. 2) To do research work to find out the useful facts obtainable from existing data and 3) To train new workers in the field who can really do the practical work needed before valuable statistical information can be obtained. It should be kept in mind that a member of the Centre has to make his contribution to solving these three tasks according to his abilities. The big planning projects will require the united efforts of all members. The Centre as a whole will assist the director of the Bureau as a staff for reaching the overall goal of developing the statistical services of the country.

The Field Survey Department has to strengthen the existing field organization for obtaining new statistical data by means of National Sampling Surveys of a type well known in India. The main task of the Research Department will be to develop a National Income Accounting system for the country. Before this task can be done succesfully, many different branches of statistics must be improved. A Census has already been taken in Indonesia in 1961. The results of this Census will form the starting point for demographic research to be done in this department.

The Akademi Ilmu Statistik has been working since 1 October 1958. The aim of the Academy is to provide a three year course in theoretical and practical statistics with such supporting subjects as English, economics, sociology and mathematics, necessary in order to train technical and semi-technical personnel for the development of public administration. The Academy is the only educational institution in the whole of Indonesia giving professional education to statisticians.

4. How to train the statisticians needed.

During my stay in Indonesia, one of my main tasks besides teaching statistics was to establish standards for the requirements for the certificate of merits in statistics and the new academic degrees, Bachelor of Statistics and Bachelor of Science with Honours in Statistics.

Every student of the third year class obtained a certificate

of merits at the end of the obligatory courses. This certificate showed how many lectures were given in the courses attended and the relative standing of the student in the tests of these courses. At most, one third of the students in the class got the highest mark 3, at most one third the weakest mark 1, and the others the medium mark 2. No student was failed because every pupil trained was considered valuable for some kind of statistical work. For the choice of suitable positions for these students such a certificate was considered to be very useful. Most students continued the studies for the Bachelor of Statistics degree after the obligatory courses. For this new degree, it was required that the student prepared and defended a seminar paper and read thoroughly about 1500 pages of statistical textbooks. A standard choice was as follows:
Hald: Statistical Theory with Engineering Applications.
Snedecor: Statistical Methods.
Cochran: Sampling Techniques.
Many variants of this reading course were however possible, depending on the books available in the library, and the interest of the student. Variety was considered to be useful for obtaining statisticians with different special knowledges. A good library in statistics must be built up to make it possible for the statisticians to obtain such further information which might be necessary in different applications.

The requirements for the Bachelor of Statistics degree corresponds to at least 2/3 of the efforts required in the University of Helsinki for a Master of Social Sciences majoring in statistics (laudatur). The full course for a Master degree would still require in Helsinki a master thesis (about 50-80 pages) and further reading of advanced or special books of about 1000 pages.

The students were given an opportunity every month to sit for B.St. examination. To pass the examination 50% of the maximum credits obtainable in the tests was required. After the written examination an oral examination was held. Sometimes students on the boarder line of failing were given six exercises to be solved as home work before they were accepted for the oral examination.

Students with more than 75% of the maximum credits in the written test and who had written good seminar papers and performed well in the oral examination were given the honour remark: Magna Cum laude. The "laudatur in statistics" is reserved for bachelors

of statistics who continue their studies for Bachelor of Science, with honours in Statistics or a master degree.

My experience in attemping to produce good statisticians speaks for a system of education, where the time is about evenly divided into three parts: Hearing lectures, Reading books and Doing exercises or performing research tasks. These means of learning should be used at every stage of the development towards a degree. At the beginning, the lectures play a relatively larger role than in the later stages when reading literature and research tasks become the main tools. The lectures and seminars should put their main efforts into making the students interested in the subject, researchminded, and aware of the high quality requirements on reliable statistical information to be produced.

The students should become selfcritical and obtain the habit of trying to estimate the probability that a statement made is wrong. They have to learn to think systematically and constructively. Without being succesful in developing their minds in this direction, the pupils will not master statistics and become really useful for improving and enlarging the production of statistical information needed. We have thus to remember that it is not enough that the new statisticians have a sufficiently wide knowledge of statistics. They must also become quality minded and accept the ethos of science before good results of the training process can be expected in the long run. The students should be encouraged to teach their assistants the basic elements of statistical thinking and learn how to train others who have not been taught statistics in the Academy.

The weaker the intellectual capabilities of the students are, the more time they need for exercises and reading. To fulfill satisfactorily the requirements for a "Bachelor of Statistics" degree, the weaker students will need almost one year longer than the brightest students who usually have already at the start a better command of English and Mathematics. Some of the students will allways be too weak to be able to reach a degree in a reasonable period of time, but, if they have got a sufficient amount of training in basic statistical work, they may, nevertheless, be very useful in a statistical organization and deserve to get a certificate telling the employer how well they have performed. The certificates of merits should show their relative standing in the class.

The brightest students must be considered as the most valuable

assets for building up an effective statistical service organization for the country. It will pay off to encourage them to continue their studies up to the upper limit of their potentialities. If the teaching staff of the Academy had one fully competent professor in charge of the further education of the Bachelors of Statistics, it would be possible to prepare them well before they go abroad for postgraduate work on Ph.D. level. It must be considered as a goal for all statistically less developed countries to have competent professors in statistics of their own, who can take over the main responsibility for making their countries well developed in the field of statistics. A country well developed in the statistical field has, in my opinion, good chances of making rapid progress also in other fields needing planned efforts.

About Different Kinds of Research Projects

illustrated by a search process for finding new approximation formulas to the normal distribution function and its inverse

by Leo Törnqvist

Dept of Statistics, University of Helsinki

When I was asked to write an article in honour of my old friend Herman Wold about how scientific research is really done a sequence of odd feelings went through me. Mostly these feelings were repulsive to the idea of exposing myself and my seemingly irrational working habits, which I know are in many ways different to my good friend's well organized methods. Nevertheless I accepted, in the hope that before the time for preparing the manuscript had elapsed I would be able to say something of interest about the subject.

When this happened about a year ago my mind was occupied with the problem of finishing my book about decision making. It has now come out in a first Finnish version: *Päätäntätieteen keskeisiä ongelmia*; that is: "The Central Problems of the Science of Decision Making". This book is a result of a project based on my deep interest in the wellbeing of mankind as a whole. The preparation has been going on for about fifteen years. There has been no hurry to get it finished. Parts of the results have reached the articulated stage during lecturing on the subject. It has developed slowly, often seeming to stand still for a year or two. This kind of synthetic scientific work uses the deep "cellars of my brain". The field of interest is of the broad type. In working with synthesis the brain is satisfied with slow progress and most of the thinking is done unconsciously. The suggestions for explicite formulations of opinions about the problems and how to attack them may rise from the cellar of my brain when no other more precisely definable problem occupies the main floor of my mind. On this floor my brain works when producing scientific articles and attacking well defined problems. When the problems are hard enough it may take about half a year of very intensive work before I am satis-

fied with the results or in some few cases I am prepared to accept the fact that I am not able to produce any new interesting results valuable enough to make it worth while to take the trouble connected with publishing them or even to prepare a manuscript. Rarely, however, can my brain be persuaded to give up a project; more often it happens that the projects require more and more thinking energy until some progress is done or the interest takes a more fruitful direction. After a project requiring a full mobilization of my brain power a long rest seems to be needed. The "main floor" of my brain is then almost closed down for a time comparable with the time it has worked hard. During the resting time of the creative parts of my brain more energy is available for actual living, for making observations about the outside world, paying attention to other people's problems, lecturing, directing other researchers, and for answering questions rapidly but often superficially. During such periods it will however happen that an answer given does not satisfy a deeper more critical part of my brain. Then my brain will tell me after some hours or at least the next morning that the answer was wrong. Sometimes it is then still possible to correct the answers already given. But enough of preliminaries and general remarks about myself as a "persistent regula di falsi" type researcher. It is, I think, more useful to present the results, with comments about how they were obtained, of an actual research project still going on.

The first interest in this specific search process awakened for about 25 years ago when I gave a lecture about the normal distribution function and found that the quotient approximation

$$\Phi(-x)=\varphi(x)/(x+v)$$

where

$$v=\sqrt{\frac{2}{\pi}} \quad \text{and} \quad \varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

was a slight improvement (giving the correct value $\Phi(o)=1/2$) to a well-known classical formula (correxponding to the case when we put $v=0$) applicable for large values of x. Thereafter the scientific interest in this function has been dormant until one of my assistants, working with a new kind of nonlinear smoothing formulas useful in cost analysis, formulas, which could be investigated by means of studying more precisely the properties of truncated normal distributions, suggested that we should make a program for calculating this function on a new electronic desk top computer aquired to the Econo-

metric Research Institute of the Post and Telegraph Administration of Finland. This institute, which I am directing is organized in an unconventional free form. It has slowly developed from a "gentlemen's agreement" between the late director general of the Post and Telegraph Administration S. J. Ahola and me, made 20 years ago. I promised to help him with problems requiring research work. From a small beginning with consultations once or twice every month this cooperation has slowly grown into a research institute giving full time or part time work to about ten assistants most of them working towards a master of Political Science or a Doctor's Degree specilizing in statistics. The new assistants are normally chosen among my most promising and brilliant students at the University of Helsinki.

The question of how to enlarge the entrances to those parts of the human brain where new ideas in science are produced and thus to make excellent researchers out of almost perfect research assistants interests me very much. A perfect research assistant can do all the checkings, calculations and other time and patience consuming things required to obtain final results out of a sample of promising ideas to attack a scientific problem. How to find a promising sample of ideas for attacking a new problem seems to be difficult to teach and learn.

But now back to the question of finding approximation formulas for the function $\Phi(-x)$. The first idea was to make use of the Taylor expansion which can be written in the following recursive form.

$$\Phi_{g+1}(-x)=\Phi_g(-x)+\Delta_g(x),$$

$$\Phi_0(-x)=\tfrac{1}{2},\ \Delta_0(x)=-\frac{v}{2}\cdot x$$

$$\Delta_{g+1}(x)=-\frac{x^2}{2g}\cdot\frac{2g-1}{2g+1}\cdot\Delta_g x$$

$$\Phi(-x)=\operatorname*{limes}_{g\to\infty}\Phi_g(-x).$$

The formulas suit the desk top computer and work well for $x^2<2$, but require a rapidly increasing number of steps g before $|\Phi(-x)-\Phi_g(x)|$ becomes sufficiently small for $x^2>2$. This idea was thus not good enough. The logically seen next step would have been to investigate formulas giving good approximations for $x\to\infty$ such as the formula $\Phi(-x)=\varphi(x)/(x+v)$ already mentioned. However research is rarely done along the logically most promising

lines for reaching the final goals. After a long excursion in trials to use instead of polynomials quotients between polynomials to get good enough interpolation formulas to the tables of Hald and of Owen we came to the conclusion that the formulas obtainable in this way were not good enough when we had to take into acccount the stubborn fact that our desk-top computer was not really prepared to handle programs with more than four given input constants and pretty small number of orders in the same run. When this became clear we were ready to investigate formulas of the type

$$\Phi(-x)=\varphi(x)\left(\frac{1}{x}-\frac{1}{x^3}+\frac{3}{x^5}-\frac{15}{x^7}+\ldots\right)$$

which we knew where good for sufficiently large values of x but had some well-known drawbacks of beeing only restrictively convergent and clearly bad for $x<1$. At this point I remembered the almost forgotten results from 25 years ago, to use $1/x+\sqrt{2/\pi}$ instead of $1/x$. An excursion in this direction lead to promising formulas of the type

$$\Phi(-x)=\varphi(x)\left(\frac{\alpha_1}{x+a}+\frac{\alpha_2}{(x+a)^2}+\frac{\alpha_3}{(x+a)^3}+\frac{\alpha_4}{(x+a)^4}+\ldots\right)$$

but they were not really tested very carefully with respect to their usefulness before I got the idea to try if it would be possible to express $\Phi(-x)/\varphi(x)$ by means of the method of continued fraction in a form useful for practical calculations. Because I had almost forgotten how to change over from polynomials to continued fraction a lot of time was spent in attempts to remember how to perform the trick—already the tempation was large to go to a large library and try to find some book telling us how to do it—when my old memories were revitalized. However we made a fantastic number of erroneous calculations before it became clear that the result could be written in the elegant form.

$$\frac{\Phi(-x)}{\varphi(x)}=1/x+2/x+3/x+\ldots+g/(x+P_{g+1}(x))$$

where every new fraction $P_g(x)$ is connected with the next one by means of the simple rules

$$P_g(x)=\frac{g}{x+P_{g+1}(x)},\quad g=1,2,3\ldots$$

$$P_{g+1}(x)=\frac{g}{P_g(x)}-x$$

$$\Phi(x)=\varphi(x)/(x+(P_1(x)).$$

The discovery seemed too good to be true, but was repeatedly checked and found to be correct. This type of recursion formulas fitted our desk-top computer perfectly if only the continued fractions could be stopped for instance for some given $g\leqslant 20$ by making use of an approximately correct non-recursive step giving a sufficiently good approximation to $\Phi(-x)/\varphi(x)$ for $0<x<1$.

How such a good final step was found is too long a story to be told completely. It can be mentioned that quite a number of more or less complex but surprisingly accurate formulas have participated in the contest. The pertinent difficulty was to keep the number of coefficients to be used as computer input below 5. The highlights on the way towards our target were the following findings:

If we develop $P_g(0)/P_g(x)$ in a Taylor expansion (1.0)

$$(\theta_g(x))^{-1}=\frac{P_g(0)}{P_g(x)}=b_g(x)+a_g x=1+a_g x+b_g x^2+c_g x^3+d_g x^4+\ldots \tag{1.1}$$

then the coefficients a_g, b_g etc. can be determined recursively by means of the nice formulas:

$$P_{g+1}(0)=\frac{g}{P_g(0)}\quad P_1(0)=v=\sqrt{\frac{2}{\pi}} \tag{1.2}$$

$$a_{g+1}=\frac{P_g(0)}{g}-a_g\quad a_1=v^{-1}-v \tag{1.3}$$

$$b_{g+1}=a_{g+1}^2-b_g\quad b_1=v^{-2}-\tfrac{3}{2} \tag{1.4}$$

$$c_{g+1}=a_{g+1}(b_{g+1}-b_g)-c_g\quad c_1=v(v^{-4}-2v^{-2}+\tfrac{2}{3}) \tag{1.5}$$

$$d_{g+1}=-d_g+a_{g+1}(c_{g+1}-c_g)-b_g b_{g+1}\quad d_1=v^{-4}-\tfrac{5}{2}v^{-2}+\tfrac{35}{24} \tag{1.6}$$

These formulas can easily be derived from the recursion formulas for $P_g(x)$. Starting values are directly obtained by expanding

$$\left(\frac{P_1(x)}{v}\right)^{-1}=\left(\frac{\varphi(x)}{v\Phi(-x)}-\frac{x}{v}\right)^{-1}=\frac{2v\Phi(-x)}{2\varphi(x)-2\Phi(-x)}$$

The numerical values of the above-mentioned coefficients are presented in Table 1.

2. The functions $P_g(x)$ and $b_g(x)$ can be written in many different forms which might be useful in searching for nonrecursive approximation formulas.

$$P_g(x) = P_g(0)\,\theta_g(x) \tag{2.11}$$

$$= \frac{P_g(0)}{a_g x + b_g(x)} \tag{2.12}$$

$$= P_g(0)\,(b_{g-1}(x) - a_g(x)) \tag{2.13}$$

$$= \frac{2}{x + \sqrt{x^2 + 4(P_{g+1}(0))^2 \dfrac{a_g x + b_g(x)}{a_{g+1} x + b_{g+1}(x)}}} \tag{2.14}$$

$$= \frac{2g}{x + \sqrt{x^2 + 4g \dfrac{P_{g+1}(x)}{P_g(x)}}} \tag{2.15}$$

$$= \frac{2g}{x + \sqrt{x^2 + 4(g+1) - 4P_{g+1}(x)\,(P_{g+2}(x) - P_{g+1}(x))}} \tag{2.16}$$

Similar identities are obtained for $b_g(x)$

$$b_g(x) = a_{g+1} x + \theta_{g+1}(x) \tag{2.21}$$

$$= a_{g+1} x + \frac{1}{a_{g+1} x + b_{g+1}(x)} \tag{2.22}$$

$$= -a_g x + (b_{g-1}(x) - a_g x)^{-1} \tag{2.23}$$

$$= \sqrt{a_{g+1}^2 x^2 = \frac{a_{g+1} x + b_g(x)}{a_{g+1} x + b_{g+1}(x)}} \tag{2.24}$$

$$= \sqrt{a_{g+1}^2 x^2 + 1 + \frac{b_g(x) - b_{g+1}(x)}{a_{g+1} x + b_{g+1}(x)}} \tag{2.25}$$

$$= \sqrt{1 + 2b_g x^2 + 2c_g x^3 + (2d_g + b_g^2)\, x^4 + \ldots} \tag{2.26}$$

These formulas can easily be derived from the basic recursion formula for

$$P_g(x) = \frac{g}{x + P_{g+1}(x)} \tag{2.0}$$

by solving equations of at most second degree.

3. In actual search for useful approximation formulas, we have tried to combine information in the Table 1 and the above-mentioned recursion formulas in order to make use of information both from origo and infinity and thus to get approximation formulas behaving asymptotically correctly and having correct derivates of at most fourth order in origo. It might be mentioned that rough approximations will be obtained by substituting $P_{g+1}(x)=P_g(x)$ on the right side of 2.15 or $b_{g+1}(x)=b_g(x)$ on the right side of 2.25 Approximation formulas thus obtained can again and again be substituted in the corresponding formulas. If in 2.16 we put $P_{g+2}(x)=P_{g+1}(x)$ we will get the upper limit of $P_g(x)$ and if in 2.14 we put $a_{g+1}x+b_{g+1}(x)=a_gx+b_g(x)$ we will get the lower limit of $P_g(x)$. Utilizing more efficiently the information in the Table 1 we get more accurate approximation formulas such as the following ones:

$$b_g(x) \approx \sqrt{\frac{1}{1-2c_gx^3}+2b_gx^2} = \hat{b}_g(x) \tag{3.1}$$

$$\approx \sqrt{a_{g+1}^2x^2+\frac{a_{g+1}x+\hat{b}_g(x)}{a_{g+1}x+\hat{b}_{g+1}(x)}} \tag{3.2}$$

$$\approx \sqrt{(\hat{b}_g(x))^2+\varepsilon_g(\hat{b}_g(x))} \tag{3.3}$$

where

$$\varepsilon_g(\hat{b}_g(x)) = \left(1-\frac{g^3}{g^3+\lambda_gx^4}\right)\left(\frac{2a_{g+1}}{a_{g+2}+a_{g+1}}+a_{g+1}^2-2b_g\right)x^2 \tag{3.4}$$

where

$$\lambda_g = \frac{(a_{g+1}+a_{g+2})}{2a_{g+1}}(2d_g+b_g^2)\cdot g^3 \approx 0.004 \tag{3.5}$$

Formulas 3.2 and 3.3 are asymptotically correct but they are already too complicated for the memory of our desk-top computer. Quite a number of formulas has participated in the contest of choosing the best approximation formula but they are not valuable enought to be mentioned. Without further comments we will also give a very efficient method for computation of the inverse function to the normal distribution function. The derivation of these formulas will be left to the final report of the project. Before the final report is published we will investigate the very accurate formulas (3.2) and (3.3) by using a big memory computer and study the literature on the subject.[1]

[1] Note added while the manuscript was in print:
The basic continued fraction form of $\Phi(-x)$ is according to Shenton (Biometrica 41, 1954) derivable from a formula due to Laplace (1805). The formula (3.2) has been found to be very accurate giving (when $g=49$) at least 15 significant figures compared with 13 significant figures given by (3.1.).

The research process illustrated above is in many respects a typical example of a search process for obtaining an efficient work routine. Some difficulties may arise in judging when the search has reached the point where the probability of finding any new significant improvements is sufficiently low. It might be a wise policy to establish well defined goals and means in the beginning of such a project. To judge the efficiency of the competing suggestions a set of practical trials might be needed. The heavy burden of the practical trials makes it desireable to make a preliminary elimination of final loosers by means of whatever scant information is available ex ante actual trials. Usually it is easy to decide which of many alternatives of a search project is the best if only a sufficient number of trials is made.

In research projects of the analytic type no practical trials are needed before the results can be classified as correct or wrong. Thus the quality of analytic results can be determined by pure logic and critical thinking. Most analytic problems can be solved by relying on wellknown basic mathematical facts but nevertheless there exist scientifically intersting cases which require creative thinking. Challenging problems of this kind may require a formidable amount of thinking energy before a solution to the problem is available or the problem is given up. In order to obtain a solution to such a problem it is usually necessary to build a new language of mathematical symbols and concepts. Such an enlargement of our scientific terminology can lead to new progresses in that domain to which the solved problem belongs.

According to my experience of scientific work all kinds of research projects require a deep interest in the subject, scientific imagination and hard work. I like to underline the importance of the art of asking such a sequence of relatively simple questions that the answers to them lead towards the solution of the main problem. During the idea producing phase of the process it might be dangerous to let critical thinking play such an important role that it blocks the flow of promising ideas, but some control is needed for avoiding the possibility of being caught too early by one idea when another maybe clearly better one still waits for being detected. Quality control of the final results should be carefully done by checking the chain of thoughts and the formulations of the findings over and over again.

Finally I like to emphasize the importance of having as myself an experienced clerical staff and a team of efficient research assistants available for reaching the targets of started research projects.

Table 1. Input coefficients for program 1.

g	$P_g(0)$	a_g	b_g	c_g	d_g
1	0.7978845606	0.4554295770	0.0707963275	−0.0060039903	−0.0012563831
2	1.2533141376	0.3424543836	0.0464790882	−0.0023235694	−0.0007737872
3	1.5957691212	0.2842020852	0.0342917370	−0.0011401012	−0.000483717.
4	1.8799712064	0.2477209552	0.0270739346	−0.0006478997	−0.000322766.
5	2.1276921616	0.2222718464	0.0223308391	−0.0004063568	−0.000228129.
6	2.3499640080	0.2032665859	0.0189864658	−0.0002734425	−0.000168837.
7	2.5532305939	0.1883940821	0.0165058643	−0.0001938881	−0.000129563.
8	2.7416246760	0.1763531456	0.0145945676	−0.0001431750	−0.000102389.
9	2.9179778217	0.1663499389	0.0130777345	−0.0001091500	−0.000082814.
10	3.0843277605	0.1578698190	0.0118451452	−0.0000854386	−0.000068349.
11	3.2421975796	0.1505629570	0.0108240588	−0.000068299.	−0.000057282.
12	3.3927605366	0.1441822775	0.0099644703	−0.000055638.	−0.000048748.
13	3.5369428141	0.1385477672	0.0092310134	−0.000045980.	−0.000041896.
14	3.6754905813	0.1335247569	0.0085978473	−0.000038562.	−0.000036480.
15	3.8090153383	0.1290102846	0.0080458062	−0.000032656.	−0.000031934.
16	3.9380256228	0.1249240712	0.0075602173	−0.000028005.	−0.000028312.
17	4.0629496942	0.1212025302	0.0071298360	−0.000024157.	−0.000025124.
18	4.1841522242	0.1177945106	0.0067457107	−0.000021090.	−0.000022610.
19	4.3019467350	0.1146583907	0.0064008358	−0.000018452.	−0.00002026..
20	4.4166051256	0.1117598585	0.0060894301	−0.000016349.	−0.00001847..
21	4.5283649842	0.1090703977	0.0058069215	−0.000014463.	−0.00001667..
22	4.6374353819	0.1065660301	0.0055493972	−0.000012979.	−0.00001538..
23	4.7440014120	0.1042264872	0.0053137634	−0.000011579.	−0.00001395..
24	4.8482278993	0.1020344437	0.005097264.	−0.000010511.	−0.00001302..
25	4.9502623429	0.0999750521	0.004897746.	−0.000009435.	−0.00001183..
26	5.0502373951	0.0980354416	0.004713201.	−0.000008656.	−0.00001117..
27	5.1482728366	0.0962044582	0.004542096.	−0.000007804.	−0.00001015..
28	5.2444772949	0.0944723135	0.004382921.	−0.000007232.	−0.00000970..
29	5.3389496084	0.0928304470	0.004234570.	−0.000006538.	−0.00000879..
30	5.4317800554	0.0912712636	0.004095873.	−0.000006120.	−0.00000851..

Program 1

1. Task Computation of

$$\Phi(-x)=\int_{-\infty}^{-x}\varphi(x)\,dx=\frac{ve^{-x^2/2}}{2(P_1(x)+x)}$$

Given:

$$x\geqslant 0,\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}=\frac{ve^{-x^2/2}}{2},\quad \text{Table 1}$$

Table 2. Values of $\hat{P}_{g\to 1}(x)$, $|\hat{P}_{7\to 1}(x) - P_1(x)| < 5 \cdot 10^{-7}$

x	$g=2$	$g=3$	$g=4$	$g=5$	$g=6$	$g=7$
0.3	0.698165	0.698166	0.698166	0.698166	0.698166	0.698166
0.5	0.641070	0.64108.	0.641077	0.641078	0.641078	0.641078
1.0	0.52508.	0.52515.	0.525132	0.525136	0.525135	0.525135
1.6	0.4239..	0.42415.	0.424123	0.424130	0.424129	0.424129
2.0	0.3730..	0.37325.	0.37321.	0.373217	0.373215	0.373216
2.6	0.3138..	0.31409.	0.314053	0.314058	0.314057	0.314057
3.0	0.2828..	0.2831..	0.283095	0.283099	0.283099	0.283099
3.6	0.2456..	0.24583.	0.245811	0.245814	0.245813	0.245813
4.0	0.22559	0.22562.	0.225605	0.225607	0.225607	0.225607
4.4	0.20810.	0.20827	0.208270	0.208272	0.208271	0.208271

2. *Choice of* g for instance $g = 10$

3. *Choice of approximation for* $P_g(x)$

$$P_g(x) \approx P_g(0)/(a_g x + \hat{b}(x)) = \hat{P}_{g\to g}(x)$$

where

$$\hat{b}_g(x) = \sqrt{2b_g x^2 + 1/(1 - 2c_g x^3)}$$

4. *Backward step*

$$\hat{P}_{g\to h}(x) = h/(\hat{P}_{g\to h+1}(x) + x)$$

$$h = g-1, g-2, \ldots, 1$$

5. *Solution*

$$\Phi(-x) \approx \Phi_{g\to 0}(-x) = \frac{ve^{-x^2/2}}{2(\hat{P}_{g\to 1}(x) + x)}$$

6. *Evalution of error*

$$\varepsilon_g(x) = |\hat{P}_{g\to 1}(x) - P_1(x)| < 10^{-10} \quad \text{for} \quad g \geqslant 25$$

7. *Values of* $\hat{P}_{g\to 1}(x)$ see Table 2

Findings on the way to the forward steps chosen in Program II

If $\quad 2\int_0^{x_L} Q_1(x)\,dx = L,\ x_L \approx x_k \quad$ and $\quad y_k = L - x_k^2 - 2\log(\theta_1(x_k)/v)$

then

$$x_L - x_k \approx y_k/Q_1(x_k) + \sqrt{Q_1^2(x_k) + [Q_1'(x_k) + Q_1''(x_k)(x_L - x_k)/3]y_k}$$

and if

$$Q_1(x) = P_1(x) + x$$

then

$$2\int_0^x Q_1(x)\,dx = x^2 + 2\log(Q_1(x)/v)$$

and $Q_1'(x) = P_1'(x) + 1 = P_1(x)\,Q_1(x)$

$$Q_1''(x) = -Q_1(x) + P_1(x)\,Q_1(x)\,(P_1(x) + Q_1(x))$$

Program II

1. Task Computation of

$$x = x_\phi \quad \text{when} \quad \Phi(-x_\phi) = \Phi < \tfrac{1}{2}$$

Given:

$$L = -2\log 2\Phi$$

$$x_0 = L/(0.21 + \sqrt{L+2})$$

2. Forward step $k \to k+\dot{1}$

$$x_{k+1} = x_k + y_k/\left[q_k + \sqrt{q_k^2 + y_k\left(1 - \frac{(1-p_k q_k)}{1 + y_k(p_k^2 + p_k q_k - 1)/6(1-p_k)\,q_k}\right)}\right]$$

$$q_k = x_k + p_k$$

$$p_k = \hat{P}_{25\to 1}(x_k)$$

$$y_k = L - L_k$$

$$L_k = x_k^2 + 2\log(q_k/v)$$

3. Solution

$$x_\phi = x_1$$

4. Evalution of errorr

$$|x_0 - x_\phi| < 0.015$$

$$|x_1 - x_\phi| < 10^{-10}.$$

BIBLIOGRAPHY

(1933) Arbetslösheten och den höga räntefoten (Unemployment and the high rate of interest; in Swedish), *Köpmannen*, No. 10, 134-138.

(1934a) Likviditetsproblemet (The liquidity problem; in Swedish), *Affärsekonomi*, Nov., 136-137.

(1934b) Den värdestabila penningenheten (The stable monetary unit; in Swedish), in *Hälsning till Svensk allmoge i Finland* (edited by Åbo Akademi studentkår) 18-20.

(1935a) Kostnadsjämförelse, ett sätt att hjälpa detaljhandlaren till större vinst (Cost comparison, a way of helping the retailer to gain greater profit; in Swedish), *Köpmannen*, No. 3, 34-36.

(1935b) Om grundandet av ett affärsekonomiskt forskningsinstitut i Finland (On founding a research institute for business economics in Finland; in Swedish), *Köpmannen*, No. 7-8, 109-110, 123 ja No. 9, 127-128, 136.

(1935c) Ekonomiska institutet för medelståendet i Haag, Holland (The institute of economics in Haag, Holland; in Swedish), *Köpmannen*, No. 11, 172.

(1936a) Levnadskostnadsindexerna i Finland och Sverige, deras tillförlitlighet och jämförbarhet (The Finnish and Swedish cost-of-living indices, their reliability and comparability; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 37, 1-35.

(1936b) Ålderdoms- och invaliditetsförsäkringen (Old-age and invalidity insurance; in Swedish), *Köpmannen*, No. 1, 4-5 och *Kommunaltidningen*, No. 1, 1-3.

(1936c) The Bank of Finland's Consumption Price Index, *Bank of Finland, Monthly Bulletin*, No. 10, 27-34.

(1936d) *Kriterien für die reellen algebraischen Zahlen, aritmetische Ketten und diophantische Approximationen*, Acta Academiae Aboensis, Mathematica et Physica X.7.

(1937a) Finlands Banks konsumptionsprisindex (The Bank of Finland's consumption price index; in Swedish), *Nordisk Tidskrift for Teknisk Økonomi*, No. 2, 73-95.

(1937b) Något om ändamålet med företagsräkningar (On the objective of business accounting; in Swedish), *Köpmannen*, No. 9, 130-132 ja No. 10, 147.

(1937c) Vilket är förhållandet mellan finska markens och svenska kronans köpkraft (together with Erik Törnqvist) (What is the relationship between the purchasing powers of the Finnish markka and the Swedish krona?; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 39, 1-39.

(1939a) Säsongfluktuationerna i Finlands Banks sedel- och myntcirkulation åren 1870-1937 (Seasonal fluctuations in the Bank of Finland's notes and coins in circulation in 1870-1937; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 44, 1-36.

(1939b) Diskussionsinlägg om tariffpolitiken vid Finlands järnvägar (A contribution to the discussion on the tariff policy of the Finnish railways; in Swedish), *Nordisk Järnbanetidskrift*, No. 7, 8, 209-210, 235-236.

(1939c) Handeln i Norge i företagsräkningens belysning (Trade in Norway in the light of business accounting; in Swedish), *Köpmannen*, No. 9, 140-142.

(1939d) Prissättning, reglementering och penningvärdeförsämring under krigstid (Price fixation, regulation and the deterioration in the value of money in war time; in Swedish), *Köpmannen*, No. 11-12, 166-168.

(1939e) Erling Petersen: Macrodynamic Aspects of the Equation of Exchange (a book review in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 45, 46-47.

(1940a) Tulojen vaikutus kulutukseen (The influence of income on consumption; in Finnish), *Mercantile*, No. 6, 102-106.

(1940b) Prispolitiken och folkförsörjningen under krigstid (Price policy and the war-time supply situation; in Swedish), *Köpmannen*, No. 1-3, 1-3.

(1941) Herman Wold: Efterfrågan på jordbruksprodukter och dess känslighet för pris- och inkomstförändringar (a book review in Swedish), *Ekonomiska Tidskrift*, No. 43, 216-225.

(1942a) Ekonometrisen suhdannetutkimuksen saavutuksia (Results achieved by econometric business cycle research; in Finnish), *Kansantaloudellinen aikakauskirja*, No. 3, 1-16.

(1942b) Om beräkning av skenvinster och verkliga vinster (On the calculation of apparent profits and actual profits; in Swedish), *Köpmannen*, No. 5-6, 36-39.

(1942c) Voidaanko luottaa viralliseen elinkustannusindeksiin? (Can the official cost-of-living index be relied upn?; in Finnish), *Talouselämä ja Suomen ulkomaankauppa*, No. 3-4, 35-36.

(1942d) Handelns kostnader och lönsamhet i statistisk belysning (The costs and profitability of internal trade in the light of statistics; in Swedish), *Köpmannen*, 7-8, 52-53.

(1943a) Tarkastus- ja tilastotyön koneellistuttaminen (Mechanization of inspection and statistical work; in Finnish), *Rautatieliikenne*, No. 2-3, 3-4.

(1943b) Hurudan är den ideala löneformeln? (What is the ideal wage system like? in Swedish), *Affärsekonomi*, No. 13, 1-3.

(1943c) Palander, Tord och Lindahl, Erik: Kostnadsberäkningar och taxepolitik vid statens järnvägar (a book review, in Swedish), *Weltwirtschaftliches Archiv*, No. 47, 26-28.

(1944a) Om tariffpolitiken vid statsjärnvägarna i Finland åren 1938-43 (On the tariff policy of the Finnish State railways in the years 1938-43; in Swedish) in *Vårt näringsliv och kriget*, festskrift utgiven av Ekonomiska Samfundet i Finland, 179-184.

(1944b) Vad förstås med ekonometri? (What is meant by econometrics? in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 60, 85-90.

(1945a) Prisbildningsproblemet betraktat ur sannolikhetsteoretisk dynamisk synvinkel (The problem of price formation looked at from a probability-theoretical dynamic point of view; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 65, 47-65.

(1945b) On the Economic Theory of Lottery-Gambles, *Skandinavisk Aktuarietidskrift*, 228-246.

(1946a) De finska järnvägstarifferna ha fått en ny gestalt, (The Finnish railway tariffs have been given a new shape; in Swedish), *Nordisk Järnbanetidskrift*, No. 5, 146-153.

(1946b) On the Distribution Function for a Function of n Statistic Variables and the Central Limit Theorem in the Mathematical Theory of Probability, *Skandinavisk Aktuarietidskrift*, 206-229.

(1947a) Tåghastighetens inverkan på ekonomin (The impact of the speed of trains on profitability; in Swedish), *Nordisk Järnbanetidskrift*, No. 5, 1-3.

(1947b) Sequential Analysis (Discussion), *Proceedings of the International Statistical Conferences*, Vol. III, 476.

(1948a) Rautatietaloudellisen tutkimuslaitoksen tarpeellisuus, (The need for a research institute for railway economy; in Finnish), *Rautatieliikenne*, No. 2, 3-4.

(1948b) Junanopeuden vaikutus rautateiden talouteen (The impact of the speed of trains on the finances of the railways; in Finnish), *Rautatieliikenne*, No. 6, 7-11.

(1948c) Tutkimustyöstä ja sen tuloksista Amerikan Yhdysvaltojen rautateillä (On research work and its results in the railways of the United States of America; in Finnish), *Rautatieliikenne*, No. 4, 5-7.

(1948d) Rautatietaloudellisesta tutkimustoiminnasta Amerikassa (On the research activities concerning railway economy in America; in Finnish), *Rautatieliikenne*, No. 8, 5-6.

(1948e) Valtionrautateiden taloudellinen kehitys 1911-1947 (The economic development of the (Finnish) State Railways in 1911-1947; in Finnish), *Rautatieliikenne*, No. 9, 5-18.

(1948f) En befolkningsstatistisk jämförelse mellan Finland och Sverige (A population-statistical comparative analysis of Finland and Sweden; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 3, 139-171.

(1948g) An Attempt to Analyze the Problem of an Economical Production of Statistical Data, *Nordisk Tidskrift for Teknisk Økonomi*, 265-274.

(1949a) *Calculations concernant la population de la Finlande, sa reproduction et son development futur* (together with Jorma Hyppölä and Aarre Tunkelo), Central Statistical Office.

(1949b) Valuuttakurssien muutoksen vaikutus hintaindekseihin (The effect of changes in the rates of exchange on price indices; in Finnish), *Kansallis-Osake-Pankki, Taloudellinen Katsaus*, No. 1, 9-14.

(1950a) LT-painoilmaisijoiden kokeilusta valtionrautateillä (On experimenting with LT weight indicators in the State Railways; in Finnish), *Rautatieliikenne*, No. 6, 5-6.

(1950b) Rautateiden kappaletavaraliikenteen lisäkustannuksista, (On the additional costs of the case-goods transport of railways; in Finnish), *Rautatieliikenne*, No. 12, 4-6.

(1951a) Tilastotieteen asema tieteiden piirissä (The place of statistics among the sciences; in Finnish), *Kansantaloudellinen aikakauskirja*, No. 1, 43-52.

(1951b) Nyblén, Göran: The Problem of Summation in Economic Science (a book review; in Swedish), *Ekonomiska Samfundets Tidskrift*, 251-254.

(1952a) Posti- ja lennätinlaitoksen tariffit vuosina 1926-1950 indeksilaskelmien valossa (The tariffs of the post and telegraph services in the years 1926-1950, in the light of index calculations; in Finnish), Appendix to *Posti- ja Lennätinlaitoksen vuosikertomukseen vuodelta 1950*, 1-53.

(1952b) Fougstedt, Gunnar: Finlands svenska befolkning åren 1936-1945 (a book review; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 1, 37-39.

(1953a) A General and Realistic Approach to the Decision Problem (An abstract), *Econometrica*, No. 2, 346.

(1953b) Some Game-theoretic Points of View on Scientific Research (An abstract), *Econometrica*, No. 3, 473.

(1953c) Wold, Herman and Juréen, Lars: Demand Analysis. A Study in Econometrics (a book review; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 2, 209-214.

(1953d) Discussion of Controlled Stochastic Processes (An abstract), *Econometrica*, No. 3, 476.

(1955a) Om olika typer av beslut (On various types of decision; in Swedish), in E. Appelberg, N. Meinander, J. Roos and G. Stjernschantz (eds.) *Ekonomi och kultur, Festskrift tillägnad Hugo E. Pipping*, Söderström, 244-248.

(1955b) Long-term Trends in Food Consumption. A Multi-Country Study, by Lars Juréen (Discussion), *Econometrica*, No. 2, 214.

(1955c) Wald Abraham: Selected Papers in Statistics and Probability (a book review; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 3, 196-197.

(1956a) Formation and Evaluation of Anticipations in Decision Making (An abstract), *Econometrica*, No. 3, 335.

(1956b) Pöyhönen, Pentti: Ekonometrinen tutkimus tonttien hinnoista (a book review; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 4, 200-201.

(1956c) Eino Haikala: Maatalouden ominaissuhdanteet ja Cobweb-teoria (a book review; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 4, 202.

(1957a) Beslut och osäkerhet (Decisions and uncertainty; in Swedish), *Affärsekonomisk Revy*, No. 3, 65-70.

(1957b) A Method for Calculating Changes in Regression Coefficients and Inverse Matrices Corresponding to Changes in the Set of Available Data, *Skandinavisk Aktuarietidskrift*, 219-226.

(1957c) Posti- ja lennätinlaitoksen tariffit ja taloudellinen kehitys vuosina 1931-1955 indeksilaskelmien valossa (The tarifs and economic development of the post and telegraph services in the years 1931-1955, in the light of index calculations; In Finnish), Appendix to *Posti- ja lennätinlaitoksen kertomukseen vuodelta 1955*, 1-87.

(1958a) Industrieproduktionens utveckling i Finland åren 1925-1956 (The course of industrial production in Finland in 1925-1956; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 3, 150-163.

(1958b) Some New Principles for Solving Linear Programming Problems, *Bulletin de L'Institut International de Statistique, 30^e Session*, 335-356.

(1958c) Statistical Methods in Finnish Industries, *Bulletin de L'Institut International de Statistique, 30^e Session*, 523-524.

(1959a) Larna, Kaarlo: The Money Supply, Money Flows and Domestic Product in Finland 1910-1956 (a book review; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 4, 303-308.

(1959b) Regressiokertoimien estimaattien variansseista ja kovariansseista (On the variances and covariances of the estimates of regression coefficients; in Finnish), *Arkhimedes*, No. 1, 4-8.

(1960a) The Interplay between Decision Makers, Management Sciences, Models and Techniques, Vol I, *Proceedings of the 6th International Meeting of the Institute of Management Science*, 323-327.

(1960b) On Distribution Functions for Quantities Related to Networks, *Bulletin de L'Institut International de Statistique, 32^e Session*, 609-612.

(1960c) Undersökning av Finska Post- och Telegrafverkets ekonomiska utveckling (together with Pertti Kukkonen) (A study of the economic development of the Finnish post and telegraph Services; in Finnish), *Nordisk Post Tidskrift*, No. 2, 26-33.

(1962) Posti- ja lennätinlaitoksen tariffit ja taloudellinen kehitys vuosina 1936-1969 indeksilaskelmien valossa (The tariffs and economic development of the Finnish post and telegraph services in the years 1936-1969, in the light of index calculations; in Finnish), Appendix to *Posti- ja lennätinlaitoksen kertomukseen vuodelta 1960*), 1-100.

(1963) The Theory of Replicated Systematic Cluster Sampling with Random Start, *Review of International Statistical Institute*, No. 1, 11-23.

(1964) Tilastotiede (The science of statistics, in Finnish), in *Yhteiskuntatieteiden käsikirja II*, Otava, 797-799.

(1965a) Sattuman leikki päätöksentekoprosessissa (The play of change in the decision-making process; in Finnish), *Taloustieteellisen seuran vuosikirja*, 8-11.

(1965b) Resultat från undersökningar i Finland (Kapitel 12) (Results of investigations conducted in Finland; in Swedish), Chapter 12 in G. Andersson, E. Johansson, T. Dalenius et al. (eds.) *Efterfrågeanalys av postverkets trafikvolym*, Elander, Göteborg, 170-172.

(1966) Posti- ja lennätinlaitoksen taloudellinen kehitys vuosina 1936-1965 (The economic development of the Finnish post and telegraph services in the years 1936-1965; in Finnish), Appendix to *Posti- ja lennätinlaitoksen kertomukseen vuodelta 1965*), 1-98.

(1967a) Minimaxing and Optimal Programming, in *Essays in Mathematical Economics* (In Honor of Oskar Morgenstern) Princeton University Press, 168-182.

(1967b) The Post-War Population Development of Finland Compared with Predictions Made after the War, in *World Population Conference 1965, Vol. III, Projections, Measurement of Population Trends*, UN, Department of Economic and Social Affairs, 44-46.

(1968a) En analys av tariffproblemet för transportföretag (An analysis of the tariff problem of the transport business; in Swedish), *Ekonomiska Samfundets Tidskrift*, No. 2, 168-182.

(1968b) *Päätäntätieteen keskeisiä ongelmia* (together with Leif Nordberg) (Central problems in the decision-making science; in Finnish), WSOY.

(1970a) Beslutvetenskapen och människans strävan att styra skeendet (The decision-making science and man's endeavour to steer the course of events; in Swedish), Societas Scientarium Fennica, XVLIII, No. 6, 3-5.

(1970b) About Different Kinds of Research Projects, in *Scientists at Work* (Festschrift in Honour of Herman Wold edited by T. Dalenius et al), Almqvist & Wicksell, 236-246.

(1971a) *Models for Non-Exosive World Population Development*, in Helsingin yliopiston Valtiotieteellisen tiedekunnan promootiojulkaisu, 11-17.

(1971b) Posti- ja lennätinlaitoksen taloudellinen kehitys vuoteen 1970 (The economic development of the Finnish post and telegraph services up to 1970; in Finnish), Appendix to *Posti- ja lennätinlaitoksen kertomukseen vuodelta 1970*), 1-151.

(1974a) *Aikasarjojen analyysi ja ennustaminen* (The analysis of time series and forecasting; in Finnish), edited by Pekka Tavaila, Gaudeamus.

(1974b) *Aikasarjojen konstruointi* (The constructing of time series; in Finnish), edited by Anitta Kautio-Toikka, Gaudeamus.

(1977) Posti- ja lennätinlaitoksen taloudellinen kehitys vuoteen 1975 (The economic development of the Finnish post and telegraph services up to 1975; in Finnish), Appendix to *Posti- ja lennätinhallituksen kertomukseen vuodelta 1975*), 1-159.